The devil is in the detail: A mixed-methods exploration of the utility of a clinical prediction tool to identify abusive head trauma in children less than three years of age with intracranial injury

Laura Elizabeth Cowley

Thesis submitted for the degree of Doctor of Philosophy

School of Medicine, Division of Population Medicine

December 2018

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ……………………………………….. (candidate) Date ……………………….

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ……………………………………….. (candidate) Date ……………………….

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ……………………………………….. (candidate) Date ……………………….

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ……………………………………….. (candidate) Date ……………………….

Acknowledgements

> "Acquire the art of detachment, the virtue of method, and the quality of thoroughness, but above all the grace of humility." Sir William Osler

First and foremost I would like to thank my fantastic supervisors Professor Alison Kemp, Dr Daniel Farewell, and Dr Sabine Maguire (prior to her retirement) for their unwavering support and encouragement throughout this PhD. Thank you for your belief and faith in a student endowed with a (probably unhealthy) dose of self-doubt. Each of you have known when I needed additional encouragement, and when you were able to leave me to my own devices. You have not only shared your skills and expertise but have been wonderful mentors who are never short of a kind word. You have truly been the best supervisors I could have hoped for, and I feel tremendously fortunate and proud to have been your student.

Special thanks go to Dr Giles Haythornthwaite, Mrs Aimee White, Ms Jenni Fryer and Mrs Heather Julier at Bristol Royal Hospital for Children and to Dr Malcolm Gajraj and Mrs Donna Newell at the University Hospital of Wales for their invaluable help and involvement in the feasibility study. Thanks also go to all of the professionals who took the time to participate in the studies presented in this thesis.

I would like to express my heartfelt gratitude to Professor Franz Babl and Ms Helena Pfeiffer for inviting me to collaborate on some exciting projects related to this PhD. I have really enjoyed working with you, it has been a fun, fruitful and intellectually stimulating experience and has greatly contributed to my knowledge of and input to the field and to some of the work presented in this thesis.

Thank you to Health and Care Research Wales for funding this PhD. I am extremely grateful that this funding gave me the opportunity to undertake this fascinating research, and to travel to conferences in some amazing places to present my work.

I am eternally grateful to my lovely friends and colleagues at the Division of Population Medicine for their friendship and support over the years. In particular I would like to thank Annette Evans, Dr Becky Richards, Bethan Carter, Diane Nuttall, Dr Cheryl McQuire, Dr Lauren Copeland, Harriet Quinn-Scoggins, Laura Harding, Laura Merrett, and Sophie Wood. A special mention must go to Dr Verity Bennett for the copious amounts of coffee, walks in the park, chats, emotional support and general kindness over the past few months. Thank you to my lovely house-mates Neil and Jess for the many cups of tea and outings to the local, and for supporting me and putting up with me during a challenging time. To my best friends Sammy and Milly, thank you for understanding, for the silliness, fun and laughter, for your belief in me, and for just being you.

This thesis is the product of enduring love, support and encouragement from my family, who have never once doubted me or my determination. Special thanks to Mum, who I am lucky enough to call not just my mum but my best friend, Mike, Mart, Dad, Penny, Jazz, Andy, Sarah, and my beloved Grandma.

Last but not least, thank you Steve, from the bottom of my heart, for the many sacrifices you have made in order that I could undertake this PhD, and for the many ways in which you have supported and motivated me throughout. You have been my rock, I could not have done this without you, and I don't have the words to express how much it means to me.

Finally, I am very grateful for the opportunity to submit this thesis and sincerely appreciate the time and effort that will go into the examination procedure.

Summary

Abusive head trauma (AHT) is the leading cause of traumatic death in infants. The Predicting Abusive Head Trauma (PredAHT) clinical prediction rule (CPR) was developed to assist clinicians in deciding which children less than three years of age with intracranial injury require additional specialist clinical or multidisciplinary investigations for possible AHT. The impact of PredAHT has not been tested in clinical practice. However, first there is a need to understand if this is feasible, and whether PredAHT is acceptable to child protection professionals. To address this gap, a computerised version of PredAHT was developed, and a series of novel empirical studies were conducted exploring the acceptability and potential impact of PredAHT, and the feasibility of evaluating its impact in clinical practice.

A literature review highlighted the many difficulties faced by clinicians in identifying AHT, suggesting that evidence-based CPRs would be of value to clinicians working in this field. A scoping review of clinical decision-making theories and the logic underpinning CPRs underscored the need for and relevance of decision support in suspected AHT cases, and informed the development of the computerised PredAHT. The computerised PredAHT provides predicted probabilities and likelihood ratios of AHT for 729 possible combinations of six clinical features (retinal haemorrhages, head/neck bruising, apnoea, seizures, rib fractures and long-bone fractures), depending on whether each is present, absent, or unknown.

The results show that the computerised PredAHT is acceptable to a range of child protection professionals, and has the potential to standardise the clinical investigation of AHT and provide clinicians with confidence and reassurance in their diagnostic decisions. Further feasibility and/or development work is recommended before the impact of PredAHT can be tested in a clinical trial. The studies presented in this thesis make important contributions to knowledge in the field of AHT diagnosis.

Abbreviations

AAP	American Academy of Pediatrics
AHT	Abusive head trauma
ALTE	Apparent life-threatening event
BESS	Benign enlargement of the subarachnoid space
BRHC	Bristol Royal Hospital for Children
CDC	Centers for Disease Control and Prevention
CPR	Clinical prediction rule
CP	Child protection
CPSW	Child protection social worker
CT	Computed tomography
DAI	Diffuse axonal injury
EBM	Evidence-based medicine
ED	Emergency department
EDH	Epidural haemorrhages
ICH	Intracranial haemorrhage
ICI	Intracranial injury
ICP	Intracranial pressure
IPH	Intraparenchymal haemorrhages
LR	Likelihood ratio
MICE	Multiple imputation by chained equations
MRC	Medical Research Council
MRI	Magnetic resonance imaging
nAHT	Non-abusive head trauma
NPV	Negative predictive value
PICU	Paediatric intensive care unit
PPV	Positive predictive value
PredAHT	Predicting Abusive Head Trauma clinical prediction tool
RCPCH	Royal College of Paediatrics and Child Health
RH	Retinal haemorrhages
SAH	Subarachnoid haemorrhages

SBS	Shaken baby syndrome
SDH	Subdural haemorrhages
SES	Socioeconomic status
UHW	University Hospital of Wales

Peer reviewed journal articles

Published peer reviewed journal articles arising from this thesis

1. **Cowley LE**, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO, Kemp AM. (2018). Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals. *Child Abuse & Neglect,* 81: 192-205.
2. **Cowley LE**, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO, Kemp AM. (2018). Factors influencing child protection professionals' decision-making and multidisciplinary collaboration in suspected abusive head trauma cases: A qualitative study. *Child Abuse & Neglect,* 82: 178-191.
3. **Cowley LE**, Farewell DM, Kemp AM. (2018). Potential impact of the validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A clinical vignette study. *Child Abuse & Neglect*, 86: 184-196.

Manuscripts accepted for publication

1. **Cowley LE**, Farewell DM, Maguire S, & Kemp AM (2019). Methodological standards for the development and evaluation of clinical prediction rules: A review of the literature. *BMC Diagnostic & Prognostic Research*, In Press.

Manuscripts submitted and under review

1. Pfeiffer H, **Cowley LE**, Kemp AM, Dalziel SR, Smith A, Cheek JA, Borland ML, et al. (2019). Validation of the PredAHT prediction tool for abusive head trauma. Manuscript submitted to *Archives of Disease in Childhood*.

Published peer reviewed journal articles that relate to this thesis

1. Pfeiffer H, Smith A, Kemp AM, **Cowley LE**, Cheek JA, Dalziel SR, Borland ML, O'Brien S, Bonisch M, Neutze J, Oakley E, Crowe L, Hearps SJC, Lyttle MD, Bressan S, Babl FE; Paediatric Research in Emergency Department International Collaborative (PREDICT). (2018). External validation of the PediBIRN clinical prediction rule for abusive head trauma. *Pediatrics,* 141(5): e20173674.
2. Pfeiffer H, Crowe L, Kemp AM, **Cowley LE**, Smith AS, Babl FE; Paediatric Research in Emergency Departments International Collaborative (PREDICT). (2018). Clinical

prediction rules for abusive head trauma: A systematic review. *Archives of Disease in Childhood,* 103(8):776-783.

3. **Laura E Cowley**, Sabine A Maguire, Daniel M Farewell, Alison M Kemp. Letter to Editor. *Law, Probability and Risk*, 14(3):275-277.
4. Maguire S, **Cowley L**, Farewell D, Kemp A. (2016). Theoretical re-analysis of two previously published datasets. *The Journal of Pediatrics,* 171: 321.
5. **Cowley LE**, Morris CB, Maguire SA, Farewell DM, Kemp AM. (2015). Validation of a prediction tool for abusive head trauma. *Pediatrics,* 136(2): 290-298.
6. Kemp A, **Cowley L**, Maguire S. (2014). Spinal injuries in abusive head trauma: Patterns and recommendations. *Pediatric Radiology,* 44(Suppl 4): S604-S612.

"Proof," I said, "is always a relative thing. It's an overwhelming balance of probabilities. And that's a matter of how they strike you."

- Raymond Chandler, Farewell, My Lovely, 1940

Table of Contents

1 Introduction

1.1 Chapter overview

This PhD thesis aims to describe the development of a computerised version of the Predicting Abusive Head Trauma (PredAHT) clinical prediction rule (CPR), and to present the findings of three empirical studies that were conducted to determine the utility of PredAHT in assisting in the identification of abusive head trauma (AHT) in children. This introductory chapter will provide context to the thesis by defining the main concepts, providing a brief historical overview of research into AHT, and describing the epidemiology and outcomes of AHT and the mechanisms by which it is proposed to occur as documented in the existing literature. It will explain the rationale for the research, and identify the gaps in the evidence-base that will be addressed. The aims, objectives and structure of the thesis are presented at the end of this chapter.

1.2 Abusive head trauma

In simple terms, AHT is defined as "child physical abuse that results in injury to the head or brain".[1] It is colloquially known as shaken baby syndrome (SBS), however broader proposed mechanisms of injury are recognized. Evidence from perpetrator confessions suggests that AHT sometimes occurs because adult care-givers become angry or frustrated by the child's crying.[2] It is proposed that the perpetrator grips the child by the torso or extremities and violently shakes them, subjecting the head to acceleration-deceleration and rotational forces as it moves back and forth.[3] Impact forces to the head may occur as a result of a direct blow to the head, or if the child is thrown onto a nearby surface, against a wall or to the ground after a shaking episode.

AHT is a leading cause of traumatic death in infants.[4] It is the primary cause of fatal child abuse[5]; the majority of fatal head injuries in children aged less than two years are due to physical abuse.[6, 7] Studies have found that between 21%–33% of children less than two years old admitted to hospital with head injury had suffered AHT.[8-10] The prevalence of AHT is higher again in children admitted to a paediatric intensive care unit (PICU).[11] AHT is mostly restricted to children less than three years of age, with the majority of cases occurring in children less than one year of age.[11-17] Mortality in children with AHT ranges from 6%[18]–36%.[19] In addition, morbidity for children who survive AHT is significant; the majority suffer from some degree of impairment in their motor and cognitive abilities, language, vision, and behaviour.[20] A recent extended follow-up study found serious neurological impairment in 40% of children who had suffered severe AHT.[21]

The cardinal clinical features of AHT include subdural haemorrhages (SDH; bleeding between the skull and the surface of the brain), retinal haemorrhages (RH; bleeding in the eyes) and encephalopathy (widespread brain injury), and associated injuries include bruising, rib fractures, long-bone fractures, skull fractures, visceral injury and spinal injury.[17, 22-29] However, children can present with no external injuries, non-specific clinical features and no history of trauma, making identification difficult.[9, 12, 30, 31]

1.3 Historical overview of abusive head trauma

The first classical description of child abuse is attributed to Auguste Ambroise Tardieu, a French pathologist who in the mid-19th century published a forensic study on the cruelty and ill treatment of 32 children.[32-34] The study was one of the first to describe the association between child physical abuse and SDH.[32] This was followed by an epidemiological analysis by Tardieu of infanticide cases,[35] where he attributed the presence of SDH, in the absence of signs of external injury, to probable inflicted head trauma.[33, 36] The pathologist distinguished the features seen in abuse cases from those seen in accidental falls, and urged that such cases "must not catch off guard the physician, often the only one capable of denouncing the crime to the legal authorities".[32(p.327)] The recognition of child abuse was likely facilitated by the explicit documentation of radiographic findings by the paediatric radiologist John Caffey.[37] In 1946, Caffey described six children who presented with chronic SDH and long-bone fractures with no history of trauma.[37, 38] At the time, Caffey tentatively considered the possibility of "intentional ill-treatment" in one of the infants, but ultimately concluded that "the traumatic episodes and causal mechanism remain obscure".[37, 38(p.758)] It was not until 1962 that the medical community broadly recognized and accepted the existence of child abuse, following the seminal publication of "The Battered-Child Syndrome" by Kempe and colleagues.[39] These authors reported that fractures, SDH, soft tissue swellings and skin bruising were clinical indicators of abuse; they called attention to the significance of healing injuries, and recommended that abuse be considered "where the degree and type of injury is at variance with the history given".[39(p.17)] Caffey eventually postulated that "battered baby syndrome", or "parent-infant stress syndrome" was the cause of injury in all six children in his 1946 study.[40]

SDHs were associated with a shaking mechanism in 1971 by Norman Guthkelch, a British neurosurgeon, who again observed their presence, in addition to RH, in infants without external evidence of injury.[41] Guthkelch proposed that repeated acceleration/deceleration forces rather than blunt trauma may be the cause of the SDH and RH and that infants may be susceptible to shaking injuries due to their large heads and weak necks.[41] This theory was supported by Caffey,[42, 43] who detailed the role of "whiplash-shaking" and jerking in the causation of brain, retinal and skeletal lesions in the absence of external signs of trauma, and associated a shaking mechanism with the bony and cerebral injuries observed in his 1946 case series.[43] Caffey coined the term "whiplash shaken infant syndrome", suggesting that in some cases the label of "battered infant" was misleading. Research into the phenomenon of infant shaking intensified in the 1980's, and the term "shaken baby syndrome" (SBS) was devised to describe the constellation of injuries apparent in such cases.[14] In 1993, the American Academy of Pediatrics (AAP) Committee on Child Abuse and Neglect released a statement affirming in a somewhat strong summary that SBS is a "clearly definable medical condition",[44(p.874)] and in 2001 they published a technical report emphasising the dangers of shaking, stating that "The act of shaking leading to shaken baby syndrome is so violent that individuals observing it would recognize it as dangerous and likely to kill the child".[45(p.206)] However, over the decades studies had begun to highlight the presence of impact injuries in many cases, including scalp haematomas, skull fractures, and brain contusions,[13, 46-48] leading to the adoption of the term "shaken-impact" syndrome by many researchers. The definitions in use today and the definition adopted in this thesis are outlined below.

1.4 Nomenclature in abusive head trauma

It is now widely accepted in the medical community that AHT includes inflicted cranial, cerebral and spinal injuries following blunt force trauma, shaking, or a combination of forces.[49-54] Resulting injuries may be primary (occurring at the moment of trauma) or secondary (occurring as a result of physiological processes set in motion by the primary injury). The classic injury pattern that is associated with abuse includes SDH, RH, and encephalopathy. These three injuries have been traditionally referred to as "the triad" of injuries, however this term is not used in clinical practice, but is used in the legal arena to incorrectly suggest that the diagnosis of AHT is based on the presence of these three features alone.[55, 56]

There have been many terms used to describe AHT in the literature; in addition to the terms noted above, AHT has been referred to as "inflicted traumatic brain injury", "inflicted head injury", and "intentional (or inflicted) childhood neurotrauma". The term "non-accidental head injury" has been suggested by the UK courts as it does not assume a particular mechanism of injury and is less emotive than other terms.[57] In 2009, the AAP Committee on Child Abuse and Neglect published a policy statement recommending adoption of the term "abusive head trauma" to describe an inflicted injury to the head and its contents.[50] This broad term was recommended as it is less mechanistic than SBS, and inclusive of all potential mechanisms of injury, and because it accounts for the numerous primary and secondary brain injuries that can arise from AHT. The statement acknowledged that SBS is a subset of AHT, and that injuries resulting from both shaking and impact can cause death or permanent neurologic disability.[50] The authors recognized that the term SBS has become familiar to the public and the media and thus is useful for facilitating prevention efforts and is a terminology that remains in relatively wide use in the literature. The official definition of AHT published by the Centers for Disease Control and Prevention (CDC) is similar to the AAP definition; AHT is defined as "an injury to the skull or intracranial contents of an infant or young child (less than five years of age) due to inflicted blunt impact and/or violent shaking".[58(p.10)]

1.4.1 Definition of abusive head trauma used in this thesis

Following recommendation from the AAP Committee on Child Abuse and Neglect,[50] the terminology AHT will be used throughout this thesis to describe infants and children who have sustained head injuries as a result of physical abuse. The term non-abusive head trauma (nAHT) will be used to describe children who have sustained head injuries as a result of accidents or medical causes. However, it should be noted that PredAHT was derived for children < 3 years of age admitted to hospital with intracranial injury (ICI) confirmed on neuroimaging.[59, 60] Unlike the AAP[50] and CDC[58] definitions, this includes children with or without skull fractures, but excludes children with skull fractures and no accompanying ICI. The systematic review[24] and derivation study[59] on which PredAHT is based included primary studies that were conducted in the late 1990s and early 2000s and therefore pre-dated the 2009 AAP[50] and 2011 CDC[58] definitions of AHT that include ICI and/or injuries to the skull. Whilst all of the primary studies included children with ICI with or without skull fractures, they did not all include children with skull fractures alone, because, as described above, children with AHT classically present with ICI on neuroimaging, with or without concomitant skull fracture.[61]

1.5 Mechanisms of injury and pathophysiology

As described above, mechanisms involving both shaking and/or impact have been proposed to explain the constellation of cranial, intracranial and associated injuries observed in AHT.[2, 13, 15, 43, 46, 47] Population studies show that patients with AHT typically have severe brain injury with few or no signs of external injury and no history of significant trauma, whereas patients with nAHT typically present with an explanation of trauma and external signs of impact injury to the head.[9, 30, 62] Although scalp swelling and isolated skull fractures are seen in both AHT and nAHT, epidemiological studies and systematic reviews have found that they are more strongly associated with nAHT.[27, 63, 64] However, children with nAHT occasionally present with ICI without skull fractures or cranial soft tissue injury.[65] In children with AHT, impact injuries appear to be more common in older children than younger children.[62] It is not possible to differentiate between impact and non-impact head trauma based on neuroradiological or skeletal findings[66] and neither is it possible to predict the intent behind a care-givers actions from the mechanism or severity of injury.

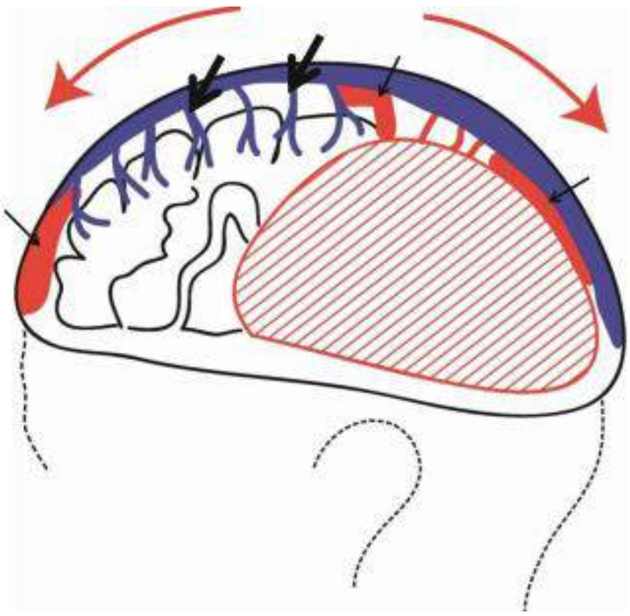
1.5.1 Primary brain injury

The proposed mechanism that produces the injuries seen in AHT is rotational movement of the brain within the cranial cavity.[67-69] Rotational forces are generated by either direct impact or shaking, which produce abrupt acceleration or deceleration of the head.[67] These forces cause the brain to turn suddenly on its axis, relative to the more stationary skull and dura.[67, 70] Evidence from animal studies suggests that infant brains are more vulnerable to injury resulting from rotational acceleration-deceleration forces than adult brains, and have highlighted the particular vulnerability of the immature brain to repeated, mild rotational forces in comparison to a single isolated incident.[71, 72]

It is proposed that forces generated by rotational acceleration and deceleration of the cranial vault result in rupture of the bridging veins that run from the cerebral cortex to the dural venous sinuses, causing SDH.[41-43, 67, 68] The finding of bridging vein rupture on neuroimaging or at autopsy at least confirms the traumatic nature of the SDH.[73-75] Figure 1.1 shows a schematic representation of the acceleration-deceleration forces generated by the movement of the infant head during a shaking incident and the resulting SDH. Tearing of bridging veins also causes subarachnoid haemorrhage (SAH; bleeding into the subarachnoid space), which is present in virtually all fatal cases of AHT but may be small and thus difficult to identify.[67] Clinical and experimental data have shown that rotational movements of the brain result in shearing of the brain parenchyma, which can cause focal or diffuse traumatic axonal injury, brain swelling, gliding contusions and focal parenchymal tears.[67, 70, 76-80] Anatomical and developmental features of the infant brain and skull make infants particularly susceptible to diffuse or shearing brain injuries.[13, 43, 68] The pliability of the skull and soft consistency of the brain facilitate brain deformation following impact, and the incomplete myelination and small axonal size of infant brains leaves them prone to shearing injury.[68] In addition, the relatively large infant head and weak neck musculature enables greater movements of the head and brain when the head is subjected to rotational acceleration-deceleration forces.[67]

In contrast, the mechanism that produces the injuries seen in nAHT, such as those resulting from a short fall, is primarily translational deceleration (movement of the head in a straight line) and cranial impact.[77] In these cases, injuries are typically focal, although exceptions have been noted.[81] Impact forces may result in head bruising, skull fractures, epidural haemorrhage (EDH), focal SDH, or focal haemorrhagic contusions.[68, 82] Therefore, the evidence suggests that it may be possible to discriminate between AHT and nAHT on the basis of injury patterns.

Figure 1.1 Schematic representation of the acceleration-deceleration forces generated by the "to and fro" motion of the head during a shaking episode



The curved red arrows represent the "to and fro" motion of the head during a shaking episode. It is proposed that this action causes disruption of the bridging veins that run from the cortex to the dural venous sinus, indicated by the thick black arrows, resulting in subdural haemorrhages, indicated by the thin black arrows. The striated red lines over the lateral convexity represent another older/more chronic subdural haemorrhage. Reprinted with permission from Springer Nature: Springer India, The Indian Journal of Pediatrics, Imaging of Abusive Head Trauma, Shekdar, K, 2016.

1.5.2 Secondary brain injury

In addition to primary traumatic brain injury, AHT can also induce secondary brain injury as a result of a cascade of biochemical, cellular and metabolic responses to the traumatic injury.[83] Neuronal damage is caused by the release of excitatory neurotransmitters, such as acetylcholine, glutamate, and aspartate.[84] Compromised cerebral perfusion may lead to hypoxia and/or ischaemia,[85] which has been correlated with poor outcomes.[86, 87]

Clinical and neuropathology studies have highlighted the significance of secondary hypoxic-ischaemia in the pathogenesis of cerebral injury in children who have suffered AHT.[26, 86-93] Neuropathology studies have found cervical EDHs and focal traumatic axonal damage to the corticospinal tracts in the brainstem and the spinal nerve roots in children with AHT, in conjunction with widespread hypoxic-ischaemia.[88-90] The authors of these studies suggest that local traumatic damage to the brainstem, including the respiratory centres, results from stretch injury from a cervical hyperextension/flexion mechanism.[88-90] It is proposed that such damage may be responsible for apnoea and subsequent hypoxia, leading to brain swelling and ensuing death, and that the diffuse axonal injury (DAI) in the majority of cases is therefore hypoxic rather than traumatic.[88-90] A recent comparative study of infants with AHT and nAHT found that in those with AHT, cervical ligamentous injury was positively correlated with hypoxic-ischaemic injury,[26] while in a retrospective review of children evaluated for suspected AHT, 83% of children with both cervical spine injuries and diffuse hypoxic brain injury were diagnosed with AHT,[94] lending support to this interpretation. Further studies have similarly documented an association between cervical spine injuries and hypoxic-ischaemia.[95, 96] However, since hypoxic brain injury is present both with and without cervical spinal injury, it is doubtful that cervical spinal injury is the sole contributor to the pathogenesis of hypoxic-ischaemia in AHT.[26, 89, 94, 97] Researchers have suggested that hypoxemia from respiratory insufficiency results from delayed medical attention and/or repeated traumatic events, or loss of airway protective reflexes.[86, 98] Some have proposed that seizure activity in AHT patients may be related to hypoxic-ischaemia, and that seizures exacerbate brain injury via excitotoxic mechanisms, or by inciting further respiratory insufficiency.[86]

1.5.3 Associated injuries

It is generally accepted that the major mechanism by which RH occurs in AHT is vitreoretinal traction caused by transmission of force through the soft tissue connections between the eye and the brain (lens, vitreous and retina) during repetitive acceleration-deceleration.[99-102] Impact forces to the head can cause deformation of bone beyond its failure strength, resulting in skull fracture.[82] Rib or long-bone fractures are caused by compression of the child's thorax, arms or legs, or jerking of the limbs during a shaking incident.[42] Visceral injuries including both hollow and solid organ injuries typically result from direct impact from blows or kicks.[29]

Concerning spinal injury, several mechanisms have been proposed.[25] As noted above, cervical damage is thought to arise from stretching injury caused by hyperextension/flexion trauma.[88-90] Certain anatomical characteristics of the infant neck render infants more vulnerable to cervical injury, including low muscle tone, horizontally oriented facet joints, underdeveloped intervertebral joints, and laxity of spinous ligaments; in addition, the relatively large head to body ratio of an infant makes them more susceptible to cervical injury.[103-105] These features may explain the presence of injury to the cervical spinal cord, ligaments and extra-axial structures in children with AHT with few associated spinal skeletal fractures.[15, 26, 88-90, 97, 106-108] Thoracolumbar spinal SDHs have also been reported in children with AHT.[26, 106, 108, 109] Gruber and Rozzelle[110] purport that the subdural bleeding is caused by ruptured blood vessels around the spinal cord resulting from hyperflexion if the infant is gripped and shaken by the thorax. Other authors have hypothesized that tracking of intracranial SDH into the spinal compartment can explain these findings.[106, 109] Further studies are needed to explore the proposed mechanisms of spinal injury in AHT.[25]

1.5.4 Can shaking alone cause intracranial injury?

Several authors have questioned the assertion that shaking alone is sufficient to cause ICI,[13, 111-113] leading to intense courtroom debate and even overturned convictions in some cases.[53, 114-116] In a landmark biomechanical study using doll models, Duhaime et al.[13] assessed the forces generated by both shaking and impact and concluded, based on established injury thresholds, that the forces subjected to an infant's head during shaking in the absence of impact were insufficient to cause concussion, SDH or axonal injury. However, the biofidelity of the model used in this study was criticized, based on observations in subsequent studies that changes in various parameters produced substantially different findings.[117, 118] In Cory and Jones[117] study, angular head accelerations with shaking were produced that exceeded those demonstrated in the Duhaime study[13] and also exceeded concussion thresholds. The sensitivity of the findings to surrogate design strongly underscores the need for accuracy in the biofidelity of surrogates used to represent infants in studies investigating injury mechanisms in abusive or accidental trauma.[119] In addition, the established injury thresholds used in the Duhaime study[13] were derived from adult primates, and are therefore not necessarily applicable to human infant brains.[120] As described above, the physical features of the infant brain are significantly different from the adult brain, and there is increasing evidence that the biochemical and metabolic responses to head injury, and AHT in particular, in infants are fundamentally different to the responses of older children or adults.[121-126] In short, injury thresholds for the infant head and brain are unknown.[127] Finite element models have demonstrated that shaking alone could induce rupture of the bridging veins and therefore cause SDH[128, 129] however such models are limited by their ability to approximate a real-world scenario, and the requirement for validation against experimental data.[119] Further research is required to establish more appropriate injury thresholds and to develop biofidelic surrogates in addition to more accurate computer models.[119, 130] Therefore, firm conclusions regarding whether or not shaking alone can cause the injuries seen in AHT cannot be drawn based on the current biomechanical evidence-base, and it is impossible to state with any degree of accuracy the amount of force necessary to cause ICI in children.[91, 119] Human experimental studies are clearly out of the question for ethical reasons.

Despite this, there is a consensus among experts that the constellation of serious injuries associated with AHT do not result from normal handling of the child, rough play, or minor trauma, and that violent forces are implicated.[45, 51, 131-133] In addition, many high quality comparative studies have demonstrated that the nature and pattern of ICI following AHT are clinically, radiographically and pathologically distinct from ICI resulting from accidental falls.[8, 10, 19, 70, 98, 134-137] Evidence of impact is also significantly less common in children with AHT than in children with accidental injuries.[98] Finally, many researchers have documented perpetrator confessions of shaking.[2, 137-144] In these studies, perpetrators have admitted to shaking alone, impact alone, or shaking with impact. In one study, the frequency of SDHs was similar in all three scenarios, while RH were more frequent in the shaking only cases, leading the authors to conclude that shaking alone is sufficient to cause both SDH and RH.[138] Dias[120] emphasized that the consistent confessions by perpetrators of shaking their child overwhelmingly suggests that shaking is an important element of AHT whether or not associated impact injury is found in the child. When all of the considerations are taken together, there is ample evidence that violent shaking of infants and children can cause serious head injury.[120]

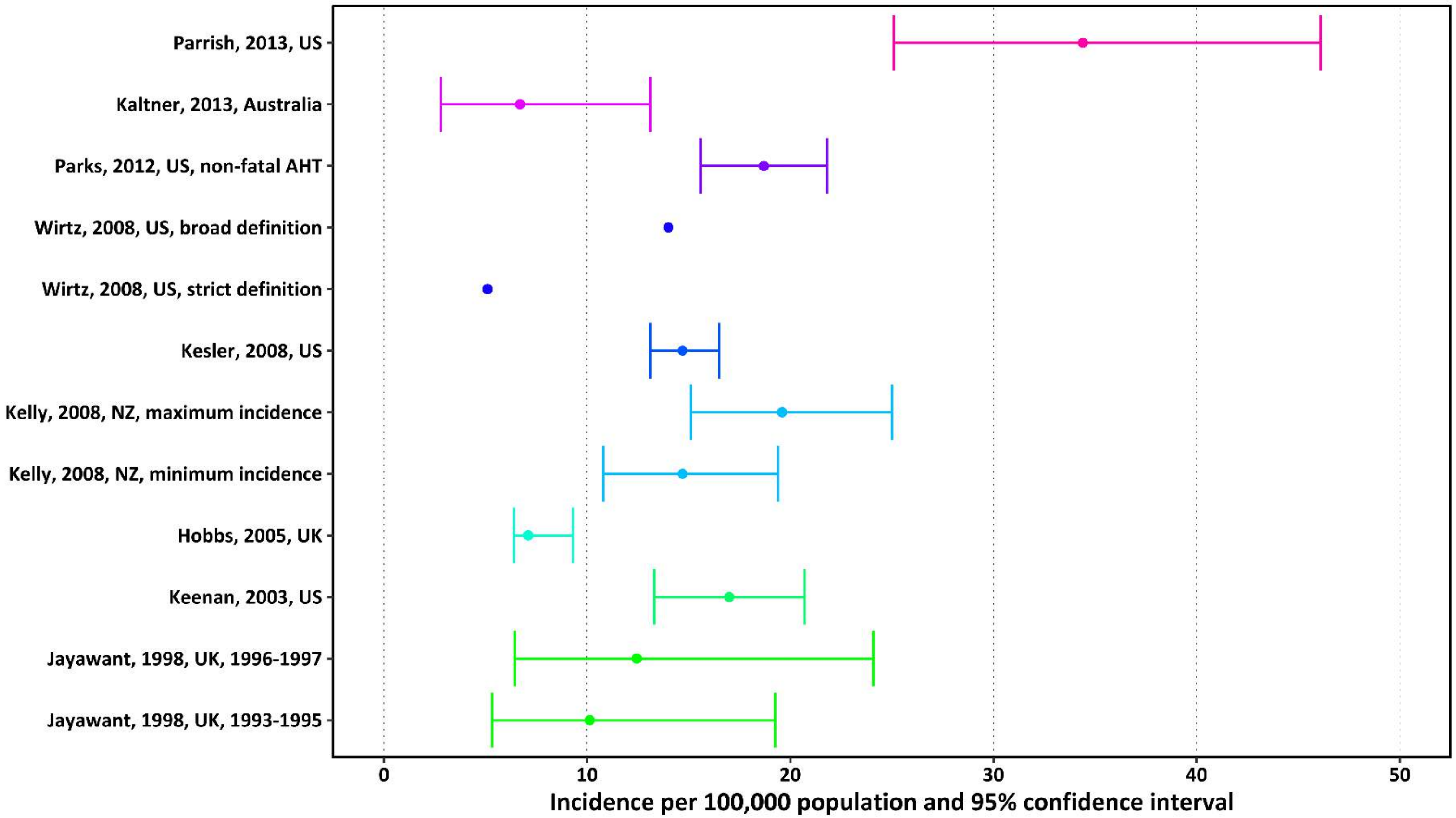
1.6 Epidemiology of abusive head trauma

1.6.1 Incidence of abusive head trauma

Worldwide estimates of the incidence of AHT are remarkably similar.[145, 146] Population-based studies of AHT estimate an incidence ranging from 5.1–34.4/100,000 population for children less than two years of age,[11, 17, 147-153] shown in Figure 1.2, and 8.5–55.9/100,000 population for children less than one year of age,[11, 18, 147-161] shown in Figure 1.3. Studies assessing the effect of prevention programs on the incidence of AHT have estimated similar rates in these age groups.[162-164] In infants aged less than six months old the incidence rises to 36/100,000 population.[17] Although the majority of AHT occurs in children less than one year of age,[165] AHT has also been identified in older children.[166] Incidence estimates of AHT in children less than three, four, five and six years old are comparable to the rates estimated for younger children.[167-174] The incidence of fatal AHT has been estimated at 0.20 per 100,000 population for children less than four years old[175] and 2.3–4.4 for infants less than one year of age.[161, 169, 175, 176]

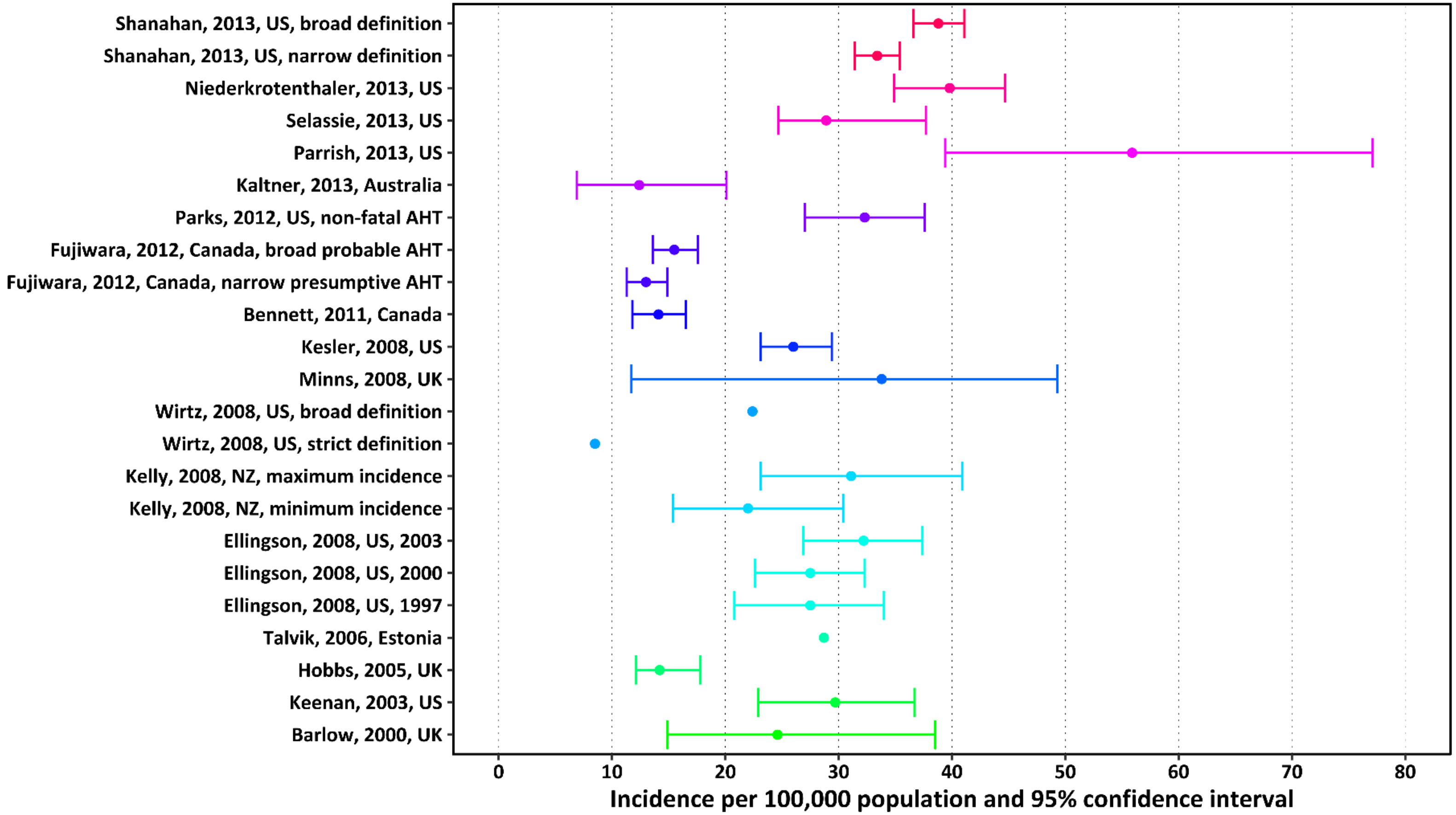
However, the true incidence of AHT is unknown as it is difficult to establish, and incidence rates are likely to be grossly underestimated.[177-179] This is because AHT can be missed by medical professionals due to the challenges in identifying AHT,[31, 180-182] or AHT can be recognized but not reported to child protection (CP) authorities.[183] Some children with mild symptoms or "subclinical" injury resulting from AHT may not present to hospital and thus will not be seen by medical professionals at all.[184, 185] Anonymous parental self-reports of physical abuse highlight that the incidence of AHT is likely much higher than the rate suggested by hospital presentations.[184-186] One survey of 1,435 mothers in the Carolinas demonstrated that 2.6% reported shaking their children less than two years of age as a means of discipline.[184, 185] In addition, incidence studies use different populations, data sources, definitions of AHT, age groups and case ascertainment strategies.[177-179]

Figure 1.2 Estimated annual incidence of abusive head trauma in children less than two years of age



Studies are listed by first author, date of publication, country, years the incidence estimates relate to, if applicable, and the definition of abusive head trauma or incidence used, if applicable. One study (Wirtz, 2008) did not give 95% confidence intervals for their incidence estimates.

Figure 1.3 Estimated annual incidence of abusive head trauma in children less than one year of age



Studies are listed by first author, date of publication, country, year the incidence estimates relate to, if applicable, and the definition of abusive head trauma or incidence used, if applicable. Two studies (Talvik, 2006 and Wirtz, 2008) did not give 95% confidence intervals for their incidence estimates.

1.6.2 Risk factors associated with abusive head trauma

Risk factors associated with AHT include child characteristics, perpetrator/family characteristics and societal risk factors. Children with AHT can also present with no identifiable risk factors; these cases may be particularly challenging for decision-makers.[187] Risk factors associated with AHT should not be used to create a "patient profile" of abuse; to do so would generate heuristics that would give rise to cognitive diagnostic errors.[188] Rather, risk factors should be considered as additional background information that may support a diagnosis of AHT or alternatively inform a CP plan or future provision of services to the family, when considered in the context of the entire case.[188, 189] The role of bias and heuristics in decision-making in possible AHT cases is discussed further in Chapters 2 and 3.

1.6.2.1 Child characteristics

Child risk factors associated with AHT include young age, with infants less than six months old being at particular risk.[10, 16] Male child gender has been reported as an associated risk factor in some studies,[11, 147, 190, 191] however gender was found to be non-discriminatory in a systematic review of features associated with AHT and nAHT.[24] Relentless crying has been consistently associated with AHT, and therefore children who cry a lot or who have care-givers who are unable to tolerate persistent crying may be at increased risk for AHT.[2, 186, 192, 193] Perpetrator admissions and anonymous surveys show that care-givers report shaking, slapping or smothering infants in order to stop them crying.[2, 186] Studies have correlated the peak age-related incidence of AHT with normal patterns of infant crying, suggesting that infant crying may act as a trigger for AHT.[192-194] Other child characteristics associated with AHT include perinatal illness, prematurity and birth defects.[191, 195, 196]

1.6.2.2 Perpetrator/family characteristics

Perpetrators in AHT cases where the child is hospitalised are most often male.[12, 190] In one study, it was found that 70% of the perpetrators were the father, stepfather or the male partner of the mother.[12] In several retrospective case series, evidence of repeated abuse was noted in over 60% of victims of AHT, highlighting that children who have been abused in the past are at increased risk for AHT.[12, 180, 197] Family risk factors associated with AHT include young maternal age,[11, 196] familial dysfunction including domestic violence, psychiatric illness and drug or alcohol abuse,[195, 198] and lower levels of parental education.[19, 149] Unrealistic care-giver expectations of an infant's ability to regulate their behaviour have also been linked to

AHT.[199] The literature on the relationship between ethnicity/race and AHT is conflicting.[11, 62, 148, 149, 153, 172, 175] Sinal et al.[172] found no significant difference in AHT incidence between Caucasian and non-Caucasian infant populations in a US sample, whilst other studies suggest that minority group ethnicity may be an associated risk factor for AHT.[148, 149, 153] Compared to other ethnic groups, fatal AHT incidence estimates are higher for non-Hispanic African-Americans.[175] Researchers have suggested that the apparent association between AHT and ethnicity may be confounded by socioeconomic status (SES) or social adversity.[11, 62, 148, 200] In addition, ethnic and racial bias in the classification and referral of AHT has been reported, where medical professionals are more likely to classify a head injury case as AHT when the infant belongs to a minority ethnic group.[31, 201, 202]

1.6.2.3 Societal risk factors

Societal risk factors related to AHT include low family SES and social deprivation.[147, 154, 167, 203, 204] An AHT incidence study in South-East Scotland found that 64% of victims of AHT were categorised in the lowest quintile of social deprivation at the time of the incident, based on the Scottish Index of Multiple Deprivation.[154] It is hypothesized that low SES is associated with high levels of parental stress and psychopathology, which in turn increase the likelihood of AHT occurring.[205-207] However, a recent comparative study of AHT and nAHT cases referred to a hospital CP team found that the probability of AHT was similar regardless of either ethnicity or SES.[62] The authors cautioned that the decision to refer a head-injured child for assessment for possible AHT should not be influenced by these demographic variables and that screening for AHT on this basis would be unsafe.[62] Finally, evidence suggests that the incidence of AHT may be higher for some children of military families,[191] during and after an economic recession,[170, 208, 209] and in the months following a natural disaster.[210]

1.7 Outcomes of abusive head trauma

Numerous studies and reviews have demonstrated that outcomes for victims of AHT are poor.[20, 21, 211-224] Most studies report mortality rates of around 20–25%,[12, 20, 30, 45, 190, 212, 216, 225-227] while mortality resulting from accidental head injury is estimated at 2% for children less than three years of age.[224] Clinical predictors of mortality in children with AHT include an initial Glasgow Coma Scale score of 3 or 4-5, RH, intraparenchymal haemorrhage (IPH), and cerebral oedema.[228]

Short-term and long-term outcomes for children who survive AHT are grave. Impairments include neurologic manifestations such as hemiplegia, quadriplegia, sight, hearing and speech problems, cerebral palsy, and epilepsy,[212, 213, 215-217] in addition to pervasive and enduring cognitive, behavioural and developmental deficits.[21, 221-223] Even those who appear to have few or no impairments at short-term follow-up have been found to exhibit poor neurocognitive functioning some year's later.[215, 216, 229]

The clinical presentation and injuries are much more severe, and outcomes much worse, in AHT than nAHT.[10, 16, 98, 135, 195, 230, 231] Around two thirds of children with AHT have moderate or severe neurodisability.[12, 17, 169, 216, 224, 225] Two studies found that 53%–54% of children with AHT had serious neurological sequelae compared to 7%–11% with nAHT.[9, 214] Approximately half of severely injured survivors of AHT die before the age of 21 years and it is estimated that severely injured patients have a reduction in their health-related quality of life of 55.5%.[218] Even those with minor injuries have an estimated reduction of 15.5% in their quality of life.[218]

A variety of clinical and radiographic factors are associated with poorer outcomes in AHT.[221] As expected, poorer outcomes are related to the severity of the clinical presentation and injuries and the extent of the brain lesions observed on neuroimaging.[98, 216, 221, 232] Cerebral oedema, seizures, apnoea, and hypoxic-ischaemia are also associated with worse neurologic and developmental outcome.[91, 233-237] Overall, the severe outcomes observed in children with AHT affect their everyday functioning and ability to interact with the environment, leading to future problems with education, community integration and social attainment.[20, 221]

1.7.1 Costs of abusive head trauma

The poor outcomes for victims of AHT result in significant societal costs, including direct costs related to medical care and rehabilitation, and indirect costs associated with reduced productivity, lost earnings or legal proceedings.[167, 238-242] Many children who survive AHT require long-term and even life-long multidisciplinary medical care, rehabilitation and therapy, and specialized educational support.[12, 216] A recent study conducted in the US estimated lifetime costs to be 2.6 million US dollars per surviving victim of AHT,[242] while a similar study conducted in New Zealand estimated direct lifetime costs of over 1 million NZ dollars per child with AHT.[239] Compared with children with nAHT, children with AHT are hospitalized for longer and have higher inpatient medical costs, even after controlling for injury severity.[238] Their use of ancillary medical resources such as home health services and occupational, physical and speech therapy is also higher.[230] In addition, a study comparing medical care between patients with and without AHT diagnoses found that a diagnosis of AHT is associated with increased medical use and costs for many years following discharge.[240]

1.8 Legal issues in abusive head trauma cases

Although AHT is accepted within the medical community as a legitimate medical diagnosis,[49-54] its validity has been questioned in the legal literature,[243, 244] in the media,[245-247] in judicial decisions,[248, 249] and in some medical literature.[250-252] Critics often perpetuate the view that, contrary to medical opinion, there is widespread “controversy” or “debate” regarding the scientific basis for AHT within the medical community.[245, 253] This has caused confusion in the courtroom, has impacted upon clinician’s confidence in making a diagnosis of abuse and testifying in court,[254] and in some cases has had a devastating impact on family and criminal proceedings and the future protection of children.[255-257]

Convictions based on the triad of SDH, RH, and encephalopathy alone, without evidence of extracranial injuries such as bruising and fractures, have been appealed and quashed.[116] This was largely due to the emergence of a “unified hypothesis” in the medical literature proposing that the triad of injuries could be caused by hypoxia, leading to brain swelling, which, combined with raised intracranial pressure (ICP), could cause both SDH and RH.[258] This hypothesis was quickly dismissed, including by the author who proposed it,[116] but it continued to be propagated by others, in various forms, in subsequent cases.[259] A recent report out of Sweden queries the existence of AHT based on the fact that there are alternative explanations for each component of the triad, and the incorrect and ignorant assumption that clinicians diagnose AHT based solely on the presence of the triad.[251, 252] Although they are *not* diagnostic or pathognomonic for AHT, the features of the triad, whether combined or in isolation, are highly suggestive of AHT *in the absence of an adequate accidental explanation or differential diagnosis*, and warrant a full child abuse work-up *alongside a comprehensive work-up for other potential diagnoses.*[52] The Swedish report has been broadly condemned by numerous experts in the field,[56, 260-265] and the Royal College of Paediatrics and Child Health (RCPCH) has called for its withdrawal and retraction “for the sake of the unbiased protection of children who may have been physically assaulted and suffer AHT”.[266(p.609)] Similarly, an earlier piece of medical literature oft-cited by opponents of AHT and described as an evidence-based systematic review,[250] suffers flaws in its search strategy, review question and application of criteria for assessing bias and study quality, and has been widely discredited by expert clinicians, lawyers and researchers alike.[51, 53, 267]

Numerous other scientifically unsupported theories have been proposed in court to explain the findings resulting from AHT,[268-271] exacerbating the confusion within the legal arena, the media and the public. Many such theories, and general approaches in the courtroom, focus on individual injuries in isolation, conveniently ignoring other injuries and failing to take into account the full clinical picture and the pattern of the presenting injuries as a whole.[51, 272] Tragically, in one recent UK case in which a father had a conviction for shaking his five-week old daughter quashed, he subsequently went on to kill the child within a year following her re-release into his care, inflicting "catastrophic skull and brain injuries from which she very quickly died".[256(p.1)] In this case, the original shaking conviction was overturned as the RH "resolved without evidence of residual damage",[255(p.24)] leading to the conclusion that "there was no rational basis on which a jury, in the light of the ophthalmological evidence, could reject an unknown cause".[255(p.27)] This was in spite of an acknowledgement that "there is no doubt that there was an encephalopathy and that there were subdural haemorrhages".[255(p.25)] An alternative theory of causation put forward in this case was that the SDH was a re-bleed of a birth-related chronic SDH, a theory for which there is no sound scientific basis.[273-275]

1.9 The importance of timely and accurate identification of abusive head trauma

Timely and accurate identification of AHT is crucial for the protection of the child and any siblings or other children who may be exposed to the perpetrators of the abuse. Identification of AHT can be lifesaving. In a seminal study describing missed cases of AHT, Jenny et al.[31] found that four of five deaths in children with a missed diagnosis of AHT may have been prevented if the abusive mechanisms had been recognized during prior medical evaluations. Altogether, 54/173 children (31.2%) with AHT had been seen previously by clinicians for symptoms associated with head injury, but the diagnosis went unrecognized. Of these, 22 (40.7%) suffered medical complications associated with the missed diagnosis, and 15 (27.8%) were subsequently re-injured following discharge. In a similar, multicentre study investigating prior opportunities to identify abuse in children with AHT almost two decades later, Letson et al.[182] demonstrated that 59/232 children (25%) had at least one prior opportunity to identify abuse in a medical setting, highlighting that abuse continues to go unrecognized by clinicians, with devastating consequences for children. As in the study conducted by Jenny et al.,[31] some deaths due to AHT in children in whom abuse was initially missed were potentially preventable.[182]

Numerous other studies have shown that physical abuse is rarely a one-off event, and that children are subjected to repeated and sustained episodes of abuse, many of which frequently go undiagnosed.[2, 180, 276-284] King et al.[281] showed that seven of 37 children (19%) who had died from abuse were evaluated by a clinician during the month preceding their fatal injury. Repeated and violent episodes of shaking were reported in 55% of cases in a study of judicial admissions of AHT, with some perpetrators admitting to violently shaking their child in up to 30 separate incidents.[2] Sheets et al.[277] conducted a retrospective case-control study investigating the occurrence of sentinel injuries (previous injuries suspicious for abuse) including bruising and intraoral injuries, in infants evaluated for physical abuse in a hospital setting. They found that 27.5% (55) of 200 infants diagnosed with definite abuse had previous sentinel injuries, compared with none of the 101 children diagnosed with accidental injury or medical causes.[277] Further evidence for missed AHT is offered by studies reporting that children admitted to hospital with a normal neurological examination and injuries suspicious for abuse including rib fracture, healing fractures, facial injury or apparently isolated bruising, exhibit high rates of occult head injury including ICI, when they are subsequently screened for AHT.[181, 285-288]

However, on the other hand, an incorrect diagnosis of abuse has profound psychological, emotional, societal and legal consequences for the families involved, including disruption of the family if the child is removed from the home and placed in foster care, job loss, unnecessary social services investigation and legal proceedings, and wrongful prosecution and incarceration.[289] Children may undergo unnecessary tests and be needlessly exposed to radiation from radiological investigations.[290, 291] Given the clinical dilemma and the high stakes associated with an incorrect diagnosis, an intervention to assist in the accurate identification of AHT is clearly warranted.

1.10 Clinical prediction rules for assisting in diagnosis

CPRs are decision-making tools for clinicians that provide a probability of a disease or outcome or suggest a diagnostic or therapeutic course of action, based on three or more variables from a patient’s history, physical examination or diagnostic tests.[292] They are designed to improve the accuracy of clinical decision-making, and are of most value in situations where decision-making is difficult, the clinical stakes are high, and a clinician’s intuition may be misleading.[293] The current introductory review has demonstrated that AHT is a serious and potentially debilitating or fatal condition with a complex and variable clinical picture, and that it may be missed in the clinical setting. Epidemiological studies highlight that AHT is rare enough that some clinicians may only see a handful of cases in their career. There is doubt seeded around the validity of the diagnosis of AHT in the legal arena and in the media, which has impacted upon clinician's confidence to diagnose AHT. The evaluation and interpretation of clinical features forms a crucial part of the assessment process in children with suspected AHT, within the context of identified risk and predisposing factors, and while the mechanisms of AHT and nAHT may share some similarities, there are differences that result in differing injury patterns and pathology. Therefore, it is feasible that clinical features differ between children with AHT and children with nAHT, and that certain clinical features may be predictive of AHT. Taken together, the literature thus far suggests that the diagnosis of AHT is complex and of great consequence, and that evidence-based clinical decision-making tools may aid clinicians in discriminating between AHT and nAHT. It is for this reason that the PredAHT CPR was created. Chapter 2 provides further evidence that the diagnosis of AHT is challenging for clinicians, and Chapter 3 further describes the logic underlying CPRs.

1.11 The Predicting Abusive Head Trauma clinical prediction rule

The PredAHT CPR was developed to assist clinicians in deciding which children < 3 years of age with ICI identified on neuroimaging, require additional specialist clinical, multidisciplinary or multiagency investigations for possible AHT.[59, 60] PredAHT was derived following a systematic review of the clinical features associated with AHT and nAHT[24] and a pooled analysis of individual patient data.[59] PredAHT estimates the probability of AHT in children < 3 years of age with ICI, based on combinations of six clinical features (Table 1.1). It is intended for use by any clinician involved in the evaluation of children where AHT may be considered within the differential diagnosis, alongside their clinical judgment and in combination with all other information about each case. In an external validation study, PredAHT performed with a sensitivity of 72.3% and a specificity of 85.7% using a 50% probability cut-off.[60] The derivation and validation of PredAHT, and its development into a computerised CPR, is described in detail in Chapter 4.

Table 1.1 The six clinical features included in the Predicting Abusive Head Trauma clinical prediction rule

Feature	Description
Head or neck bruising	Any documented bruising to head or neck
Seizures	Any documented seizures from a single seizure to status epilepticus
Apnoea	Any apnoea documented in the initial history or during inpatient stay
Rib fracture	Any rib fracture documented after appropriate radiologic imaging
Long-bone fracture	Any long-bone fracture documented after appropriate radiologic imaging
Retinal haemorrhage	Any retinal haemorrhage documented after indirect ophthalmologic examination by a paediatric ophthalmologist

Reproduced with permission from Cowley et al., Pediatrics, Vol. 136, Page 292, Copyright © 2015 by the American Academy of Pediatrics

1.12 Gaps in the evidence-base and rationale for the research

It is widely acknowledged in the literature that there are three main stages involved in the development of CPRs; derivation; validation and refinement; and impact analysis to determine their impact on clinician behaviour and patient care.[293-300] The three main stages of development correspond to increasing hierarchies of evidence.[293, 299, 300] Stiell and Wells[301] consider a further stage to be important prior to the derivation of a CPR, namely, identifying the need for a CPR. The current introductory review and the literature review reported in Chapter 2 address the issue of whether a CPR for AHT is needed.

Evaluating the impact of a CPR has been described as “the next painful step” in the development process.[302] Unfortunately, compared to the number of new CPRs that are derived, few are externally validated and even fewer have undergone impact analysis to establish their actual effect on process outcomes or relevant patient outcomes.[303-307] Many CPRs are developed for the same purpose,[308-314] and there is an urgent need to change the focus from the derivation of new CPRs to the validation and impact analysis of existing ones.[300] In addition, it is recommended that extensive exploratory and preparatory work is undertaken prior to a formal experimental impact analysis study, to assess the acceptability of the CPR and the feasibility of conducting such a study in clinical practice.[300] Recent systematic reviews of impact studies of CPRs have found that the risk of bias in many studies is either high or unclear,[306, 307] and that few CPRs are adequately prepared for experimental impact analysis.[306] PredAHT has been derived and externally validated; the next stage is to test its impact in clinical practice. However, first there is a need to understand if this is feasible, and whether PredAHT is acceptable to CP professionals. This thesis therefore presents a series of empirical studies exploring the acceptability and potential impact of PredAHT, and the feasibility of evaluating its actual impact in clinical practice.

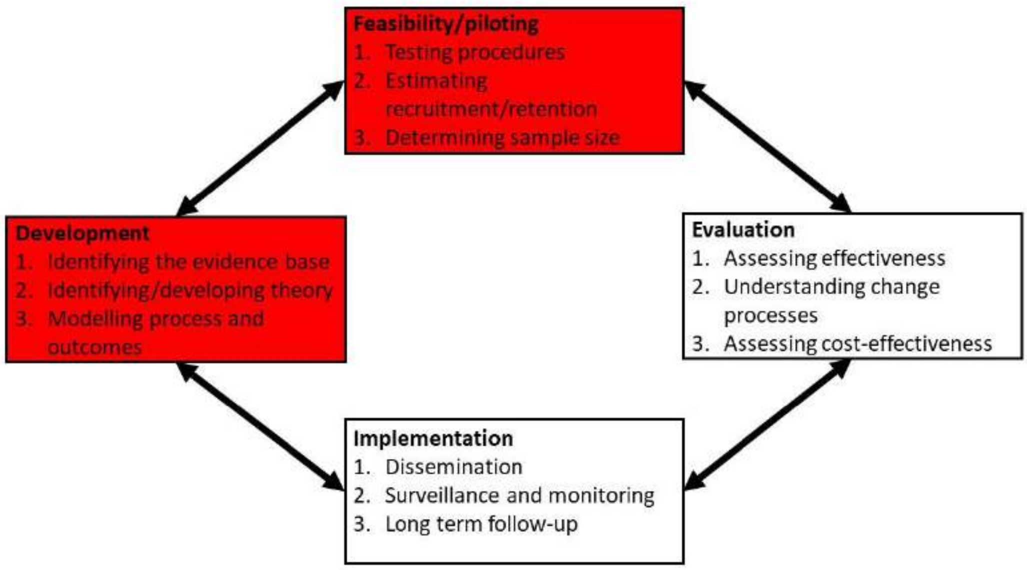
1.13 Frameworks to guide the development and evaluation of a clinical prediction rule

1.13.1 The Medical Research Council framework for the development and evaluation of complex interventions

Since CPRs aim to influence clinician behaviour and decision-making, they should be regarded as complex interventions.[315-321] By convention, a complex intervention is defined as an intervention comprised of several interacting components.[322] The impact of a CPR on clinical practice will depend on multiple components that interact, including the accuracy of the rule, clinicians’ interpretation of probabilities, and clinicians’ adherence to the rule.[317]

Thus, the Medical Research Council (MRC) framework was used to guide the development and evaluation of PredAHT, and the planning and design of the empirical studies presented in this thesis.[322, 323] The MRC framework provides guidance for researchers on the development, evaluation and implementation of complex interventions to improve health, and advocates a systematic, phased approach to intervention development and evaluation.[322, 323] Four phases of development and evaluation are outlined, which need not follow a linear or cyclical sequence (Figure 1.4).

Figure 1.4 The four phases of the development and evaluation of complex interventions as outlined in the Medical Research Council guidance



Adapted with permission from BMJ Publishing Group Limited. [Developing and evaluating complex interventions: The new Medical Research Council guidance, Craig, P., 337; a1655, 2008]. The work undertaken in this thesis relates to the "development" and "feasibility/piloting" phases, shaded in red.

During the development phase, researchers should identify the evidence-base, identify relevant theories to develop a theoretical understanding of the likely process of change, and model processes and outcomes of the intervention. Studies relating to the development of PredAHT are reported in Chapters 2, 3 and 4. Chapter 2 reviews the literature on the challenges associated with identifying AHT, to demonstrate that a CPR that provides an evidence-based estimate of the probability of AHT is likely to be of value to clinicians. Chapter 3 reviews theories of clinical diagnostic decision-making and details the mechanisms by which

CPRs may improve clinician's decision-making. Chapter 4 describes the previous derivation of the PredAHT regression model based on a systematic review of the clinical features associated with AHT and nAHT, and the subsequent validation of PredAHT on novel data. The findings from Chapter 3 were used to inform the development of a computerised version of PredAHT based on the regression model, detailed in Chapter 4.

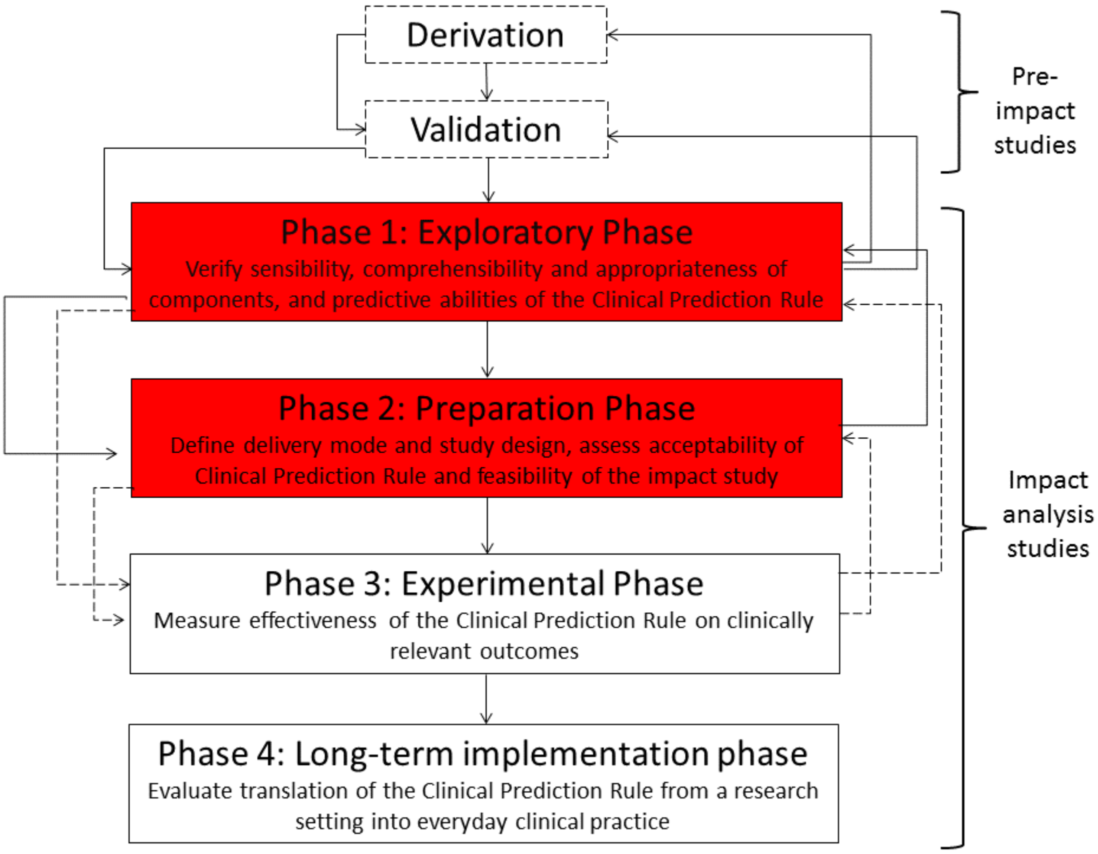
The "feasibility and piloting" phase involves initial testing of the acceptability and the feasibility of the intervention, prior to a full-scale evaluation study. The acceptability of PredAHT to CP professionals is investigated in Chapter 5, and the potential impact of PredAHT on clinical decision-making is tested in a clinical vignette study reported in Chapter 6. Finally, a multisite feasibility study was conducted to assess whether the impact of PredAHT can be evaluated in clinical practice; this is reported in Chapter 7. The "evaluation" phase is used to assess the effectiveness and cost-effectiveness of the intervention and to develop an understanding of change processes. In the final phase, if the intervention is acceptable, feasible, and demonstrates evidence of effectiveness, it can be implemented, which requires dissemination, surveillance and monitoring of the intervention, and long-term follow-up.

1.13.2 Frameworks for the development and evaluation of clinical prediction rules

Although there is an abundance of methodological guidelines for the derivation and validation of CPRs,[324] in comparison there is a lack of clear guidance for the design, conduct and reporting of impact analysis studies of CPRs. To this end, Wallace and colleagues[300] formulated an iterative four-phased framework for the impact analysis of CPRs, specifying the importance of substantial preparatory and feasibility work prior to the conduct of a full-scale formal experimental study (Figure 1.5). This framework was used to guide the work presented in this thesis.

Phase 1 involves determining whether the CPR is ready for impact analysis i.e. whether it has been rigorously derived and broadly validated according to pre-defined methodological standards. Phase 2 includes assessing the acceptability of the CPR and identifying potential barriers to its uptake and implementation, as well as assessing the feasibility of conducting an impact study. Evaluating the feasibility of carrying out an impact study involves consideration of multiple factors including the most appropriate study design for measuring relevant outcomes, and how the CPR will be delivered at the point of care or integrated into the clinical workflow. Phase 3 involves formally testing the impact of the CPR using a comparative study design. Phase 4 involves long-term dissemination and implementation of the CPR. The body of work undertaken in this PhD thesis relates to phases 1 and 2. In Chapter 4, consideration is given as to whether PredAHT is ready for impact analysis, and the acceptability, potential impact and feasibility of PredAHT are investigated in Chapters 5, 6 and 7, respectively.

Figure 1.5 The four phases of impact analysis for a clinical prediction rule



Reproduced with permission from Wallace et al. (2011). The work undertaken in this thesis relates to Phases 1 and 2, shaded in red.

1.14 Aims and objectives of this PhD thesis

This PhD thesis aims to build on the knowledge gained from the derivation[24, 59] and validation[60] of PredAHT. The primary aims of this thesis were to 1) develop a computerised version of PredAHT for use in clinical practice, and 2) to determine the utility of PredAHT in assisting in the identification of AHT, using mixed methods. Three empirical studies were conducted based on existing frameworks and guidance for the development and evaluation of CPRs.[300, 301, 322, 323]

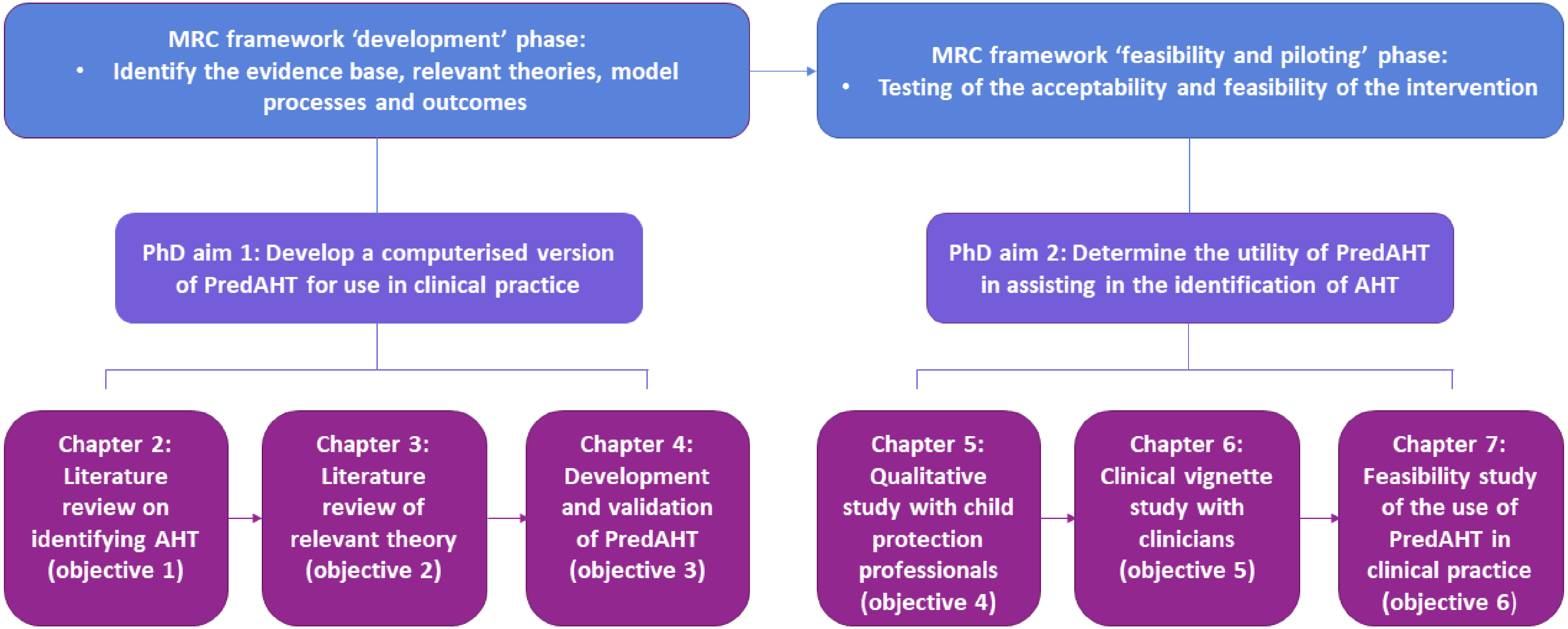
There were six objectives of this PhD thesis:

1. To undertake a review of the literature on the challenges associated with the identification of AHT using systematic search methods.

2. To review relevant theories of clinical decision-making and the logic of CPRs using systematic search methods.
3. To describe the previous derivation and validation of PredAHT, present a critical appraisal of PredAHT and other CPRs for AHT, develop a computerised version of PredAHT using Shiny, a Web application framework for the R language and environment for statistical computing, and to validate the computerised PredAHT in an Australian/New Zealand population.
4. To assess the acceptability of PredAHT with a range of CP professionals, using qualitative methods.
5. To explore the potential impact of PredAHT on clinicians' judgments and decision-making, using clinical vignettes.
6. To assess the feasibility of evaluating the impact of PredAHT in clinical practice, using mixed methods.

Each phase of work presented in this thesis was guided by two phases of the MRC framework.[322, 323] Phase one, which involved the development of a computerised version of PredAHT for use in clinical practice, included a literature review of the challenges involved in identifying AHT, a review of potentially relevant theories related to clinical diagnostic decision-making in suspected AHT cases, and a review of the derivation and validation of PredAHT. The computerised PredAHT was then created based on the findings of this work. Phase two consisted of evaluating the utility of PredAHT in assisting in the identification of AHT and involved a qualitative study with a range of CP professionals, a clinical vignette study with clinicians, and a multisite feasibility study of the use of PredAHT in clinical practice. Figure 1.6 presents a schematic of the work conducted for this thesis according to the relevant phases of the MRC framework and the aims and objectives of the PhD.

Figure 1.6 A schematic of the work presented in this thesis, according to the aims and objectives of the PhD and relevant phases of the Medical Research Council framework



1.15 Thesis structure

This introductory chapter has described the context, main concepts, rationale, gaps in the evidence-base, and aims and objectives for this thesis. This thesis contains 7 further chapters.

Chapter 2. Identifying Abusive Head Trauma: The challenge for clinicians

Chapter 2 presents a review of the literature of the challenges associated with identifying AHT in clinical practice. This review was conducted to demonstrate why a CPR in this field is required and why it is likely to be beneficial to clinicians in assisting in the identification of AHT.

Chapter 3. Clinical decision-making and the logic of clinical prediction rules

Chapter 3 presents a review and critical appraisal of a selection of clinical decision-making models, theories and approaches that are relevant to the diagnosis of AHT, and a review of the logic underpinning CPRs. This review was conducted to gain a theoretical understanding of clinical decision-making in suspected AHT cases and the mechanisms by which CPRs may improve clinician's decision-making. The findings informed the development of the computerised PredAHT tool reported in Chapter 4.

Chapter 4. Development of the Predicting Abusive Head Trauma clinical prediction tool

Chapter 4 reports the systematic process used to create the novel computerised PredAHT clinical prediction tool, and a critical appraisal of PredAHT against other CPRs for the identification of AHT. The chapter describes the previous derivation and validation of PredAHT, a systematic review, comparison and critical appraisal of validated CPRs for AHT, the development of PredAHT into a computerised tool, and the external validation of the computerised PredAHT on an Australian/New Zealand dataset.

Chapter 5. Acceptability of the Predicting Abusive Head Trauma clinical prediction tool: A qualitative study with child protection professionals

Chapter 5 presents findings from a novel qualitative interview study with clinicians, child protection social workers (CPSWs), police officers, pathologists and lawyers. Interviews explored factors influencing decision-making and multidisciplinary collaboration in suspected AHT cases, and attitudes towards PredAHT.

Chapter 6. Potential impact of the validated Predicting Abusive Head Trauma clinical prediction tool: A clinical vignette study

Chapter 6 presents findings from a novel vignette-based cross-sectional survey study with clinicians involved in suspected AHT cases. Using six clinical vignettes, this study explored the impact of PredAHT on clinicians' probability estimates of AHT, and their proposed CP actions, and assessed the degree of agreement between clinicians' opinions both before and after they used PredAHT.

Chapter 7. Evaluating the impact of the Predicting Abusive Head Trauma clinical prediction tool in clinical practice: A feasibility study

Chapter 7 presents findings from a novel feasibility study of the evaluation of PredAHT in clinical practice. A multisite, non-randomised, before—after study within the same clinicians was conducted, with a substantial qualitative element. Clinicians applied PredAHT to all consecutive children less than three years of age admitted to two UK teaching hospitals with ICI, and participated in an interview where they discussed PredAHT in relation to the case.

Chapter 8. General discussion

Chapter 8 summarises the main findings of this thesis, highlights the novel contributions of this thesis to the evidence-base, and discusses the findings in relation to theory. The strengths and limitations of the frameworks used to guide the research presented in this thesis are discussed and the advantages of the mixed-methods approach are highlighted. Finally, the implications of the thesis for research and practice are outlined.

2 Identifying Abusive Head Trauma: The challenges for clinicians

2.1 Chapter overview

This chapter presents a review of the literature of the challenges faced by clinicians in identifying AHT. The introduction to this chapter will provide the background and rationale for this literature review. Then, the aim of the chapter will be presented, followed by the literature search strategy and a description of the challenges associated with the identification of AHT in clinical practice. Challenges identified relate to the accuracy of the history provided by the care-giver, variability in the clinical manifestations of AHT, potential differential diagnoses, forensic considerations, clinician bias, alternative theories of causation and the evaluation of suspected AHT. The findings will be discussed, and the implications of this literature review for this thesis will be described.

2.2 Introduction

As described in Chapter 1, the MRC framework for developing and evaluating complex interventions emphasizes the importance of a systematic approach to intervention development based on the best available evidence.[322, 323] It is imperative to identify the relevant, existing evidence-base in order to gain an understanding of whether an intervention is likely to be effective.[322, 323] The very first stage in the development of CPRs is to identify if there is a need for the CPR.[301] Investigators rarely justify why a CPR is needed,[325, 326] or why a CPR for a particular condition is likely to be of value to clinicians.[301] The literature reviewed in Chapter 1 provided preliminary evidence that the diagnosis of AHT is complex, that clinicians are failing to accurately diagnose AHT, and that there are serious consequences associated with an incorrect diagnosis. Cases of AHT are frequently missed, and many children who are returned to an abusive environment are subsequently reinjured or killed, emphasizing the challenges in identifying AHT and protecting children from repeated abuse or subsequent fatal AHT.[31, 182, 277, 282]

Although a great deal has been written about the diagnosis and identification of AHT with regard to injury mechanisms, outcomes, clinical features, differential diagnoses, and the evaluation of suspected AHT, in comparison there is little information in the literature about the challenges faced by clinicians in identifying AHT.[289] There are some useful commentaries,[289, 327] but few literature reviews that specifically focus on the difficulties associated with the identification of AHT in clinical practice. A review of the identification of

AHT was therefore conducted, from the perspective of the challenges involved, to help further elucidate whether a CPR in this field will be of value and be likely to assist clinicians in identifying this devastating condition.

2.2.1 Aim of this chapter

The aim of this chapter is to review the literature on the challenges confronting clinicians in the identification of AHT.

2.3 Methods

This is a literature review and was not intended to meet the Cochrane Collaboration definition of a systematic review.[328] Nevertheless, a systematic search of key databases, texts, and supplementary sources was conducted to identify relevant literature.

2.3.1 Literature search strategy

An existing comprehensive search strategy was used to retrieve literature regarding the identification of AHT. This search strategy was developed by the Cardiff Child Protection Systematic Reviews (Core Info) team with the assistance of an information specialist, to identify literature on the clinical and neuroradiological features indicative of AHT, and was used to locate studies included in three systematic reviews conducted by the team.[24, 136, 329] The full list of search terms used and databases searched are detailed online,[330] and in Appendix 1. As a former member of the Core Info team, the researcher had access to the full database of articles retrieved from these reviews. Weekly email auto-alerts were then set up using this search strategy in order to keep up-to-date with the literature; these were restricted to the following databases: MEDLINE, EMBASE, PsycINFO, Scopus, and Web of Science.

Supplementary sources included electronic alerts from the New England Journal of Medicine Journal Watch for “Hospital Medicine”, “Emergency Medicine”, “Pediatric and Adolescent Medicine” and “Neurology”, electronic table of contents alerts from *Pediatrics*, *Pediatric Radiology*, and *The Lancet*, and electronic newsletters from the Child and Maternal Health Knowledge Update by Public Health England, the Royal College of Paediatrics & Child Health, and the North South Child Protection Hub. Literature was sourced from the Wild Iris Continuing Education Pediatric Abusive Head Trauma accredited online course (https://wildirismedicaleducation.com/courses/pediatric-abusive-head-trauma-training-ceu), and the proceedings for the 15th and 16th International Conference on Shaken Baby Syndrome/Abusive Head Trauma and the 6th Penn State Health International Conference on Pediatric Abusive Head trauma, which the researcher attended. Other resources searched included the Quarterly Update, a journal that reviews peer-reviewed articles in the medical literature on the diagnosis, prevention and treatment of child abuse and neglect, *UpToDate*, an evidence-based resource for clinicians, Safety Lit, a bibliographic database of published research relevant to injury prevention and safety promotion, the RCPCH Child Protection Companion child abuse handbook, and the BMJ best practice guide to AHT in infants. Systematic reviews and review articles were used as an efficient way to gain a summary of evidence. Finally, literature was sourced from the reference lists of identified articles, and from key textbooks in the field, including: Child Abuse and Neglect: Diagnosis, Treatment, and Evidence[331] and Shaking and Other Non-accidental Head Injuries in Children.[332]

2.4 Challenges in identifying abusive head trauma

Identifying suspected abuse and reporting reasonable suspicions to social services has been described as one of the most challenging, difficult and unsettling responsibilities for paediatricians.[333] As described in Chapter 1, children with AHT typically present with ICI and few or no signs of external injury, with no explanation of significant trauma, or an explanation that is inconsistent with the severity of the injury or the developmental stage of the child.[9, 12, 30, 59] In common with many areas of medicine, there is no gold-standard diagnostic test for AHT. A clinical diagnosis of AHT is made following a thorough clinical and forensic assessment that excludes other potential causes of the clinical presentation including accidental trauma and medical conditions, and confirms abuse from the constellation of findings.[333] The evaluation and investigation of suspected AHT requires a multidisciplinary team approach whereby paediatricians collaborate with clinicians from other specialities (e.g. neurosurgeons, radiologists, ophthalmologists, pathologists), CPSWs, and police officers.[334] The multidisciplinary team must piece together all available information to jointly determine the likelihood of AHT. The investigative process may include deliberation in the family or criminal courts, and the identification of children requiring future protection and the legal determination of the child's injuries as a criminal offence, are understood to be the responsibility of the justice system.[51, 334] The current review uncovered a number of reasons why the identification of AHT can be challenging for clinicians. These are described below.

2.4.1 Inaccurate history

Many of the challenges associated with understanding how a child was injured are related to the history of the injury.[54, 289, 327, 333] Firstly, there are rarely any independent witnesses to AHT, and abused children may be very young and nonverbal or severely injured, and therefore unable to explain what happened, or they may be unwilling to disclose abuse out of fear.[289, 327, 333] Secondly, perpetrators of AHT often lie about what happened because of the severity of the potential consequences, and therefore the history is frequently inaccurate or incomplete.[9, 31, 54, 138, 140, 231, 289, 327, 333] Many children who have sustained AHT present to medical facilities without a history of preceding trauma, or with a history that is incompatible with their physical examination findings or developmental capabilities.[9, 30, 333] This compounds the difficulties in identifying AHT because clinicians are taught to rely on an accurate history to guide their diagnostic decision-making, and inaccurate or incomplete histories may misdirect the evaluation process and impede a clinician's ability to come to the correct diagnosis.[54, 289, 327, 333] A recent study investigated the accuracy of parental reports of infant trauma history, classifying infants into low-risk, middle-to-high-risk and unknown-risk based on the mechanism of injury put forward by the parents.[335] The study found that infants in the unknown-risk group, where no mechanism of injury was given, had the highest injury severity, worst neurological outcomes and a higher frequency of injuries suggestive of AHT, while the same pattern of findings was observed for infants in the low-risk group compared to those in the middle-to-high risk group. The authors concluded that the history of injury provided in cases of AHT is likely inaccurate and that clinicians should investigate all cases with inconsistencies between the history offered and the severity of injury observed.[335]

The absence of a history of trauma in children with ICI has been found to be associated with AHT in many studies.[8-10, 16, 27, 30, 62, 70, 135, 231, 336] In contrast, among children with nAHT, care-givers have consistently reported a history of a traumatic incident.[8, 30, 231] In several retrospective case series, a history of trauma was absent in 64%–97% of children with AHT.[8, 9, 16, 30] In one study, the absence of a history of trauma had a specificity of 97% and a positive predictive value of 92% for AHT.[9] In those children with persistent neurological impairment, having a history of a low impact fall (defined as a fall from less than three feet) had perfect specificity and PPV for AHT. Histories of home resuscitative efforts, and histories that changed over time were strongly associated with AHT.[9]

2.4.2 Variability in the clinical manifestations of abusive head trauma

Difficulties in the diagnosis of AHT can be partly attributed to the fact that infants and children with AHT present with a variety of signs, symptoms and physical findings, that range from subtle and nonspecific to severe and life-threatening.[30, 31, 147, 181, 285, 337] The clinical presentation depends on the type and severity of AHT sustained and therefore the severity of the subsequent brain injury. When considered in isolation, each of the clinical manifestations of traumatic brain injury is nonspecific for AHT, as each can result from either AHT, accidental injury or medical causes.[54, 338] Given the extraordinary variability in presentation, a high index of suspicion for AHT is warranted.

2.4.2.1 Clinical presentation

Children who have sustained AHT frequently present with nonspecific signs and symptoms, such as irritability, abnormal mental status, poor feeding, vomiting, and lethargy, which can hinder diagnosis.[31, 138, 147, 214] Such nonspecific signs and symptoms may be mistakenly attributed to common childhood conditions such as otitis media, gastroenteritis, influenza, or upper respiratory tract infections.[31, 339] In the aforementioned seminal study describing missed cases of AHT conducted by Jenny et al.,[31] the most common misdiagnoses made for children with missed AHT were viral gastroenteritis or influenza, accidental head injury and "rule out sepsis". Children with AHT can also have a completely normal neurologic examination,[181, 285] and many have no external signs of injury.[12, 30, 214, 340] In two studies,[12, 30] 35% and 40% of children with AHT had no external evidence of trauma on presentation, respectively. However, many children with AHT present with apnoea or other breathing difficulties, coma, and seizures and the majority initially have an abnormal neurologic examination.[9, 12, 30] Minns and Busuttil[341] distinguish between four different patterns of clinical presentation of AHT. The first is hyperacute encephalopathy, where the majority of patients are dead on arrival or die shortly afterwards; this type of presentation is associated with axonal damage at the craniocervical junction, acute respiratory failure, cerebral oedema and hypoxic injury. The second is acute encephalopathy, which is the most common presentation and is characterised by altered mental status, apnoea, bilateral SDH, widespread RH and associated skeletal or other injuries. The third pattern is termed subacute non-encephalopathic presentation, where the brain injury is less intense and the outcome typically better. The final type of presentation is chronic extracerebral presentation, where infants present late with an isolated SDH, expanding head circumference, raised ICP, irritability and vomiting, but little encephalopathy. This type of presentation is the most difficult to relate to AHT, and AHT may not be considered in this group of patients.[341]

Children who have suffered AHT may present with a history of an apparent life-threatening event (ALTE).[337, 342-346] These children pose a unique diagnostic dilemma for clinicians, because an ALTE is characterized by nonspecific symptoms such as apnoea, colour change, alterations in muscle tone and choking or gagging.[347] Children may appear well and present without any associated discernible injuries. An ALTE can have a number of aetiologies including gastro-oesophageal reflux, seizure, and respiratory infections, as well as AHT.[348] In one study almost half of the children with AHT presenting with an ALTE were initially missed in the emergency department.[344] All of the children with missed AHT had nonspecific symptoms and unremarkable physical examination findings, compared to the children in whom AHT was recognized, who had associated pertinent physical findings including bruising and RH.[344] In a recent retrospective observational study, children presenting with ALTE-associated SDH were more likely to exhibit extracranial injuries suspicious for abuse including retinoschisis, high-specificity bruising and internal abdominal injury, than children presenting with non-ALTE SDH.[349] This supports the interpretation that an ALTE, in and of itself, does not cause the clinical findings seen in AHT, as has been theorized in the literature,[271, 350] although a major limitation is that definitive outcome data regarding abuse and non-abuse were not collected in this study.[349] In summary, it is clear that in the absence of obvious injuries or neurologic signs on presentation, clinicians may not consider the possibility of AHT.

2.4.2.2 Clinical and radiological features associated with abusive head trauma

AHT has a broad clinical spectrum. As outlined in Chapter 1, the characteristic features of AHT include SDH, RH and encephalopathy, and AHT can also result in a range of additional associated physical findings such as bruising, fractures, spinal injury and visceral injury. Some or all of these injuries may be found in any given child with AHT. Knowledge about the expected range and patterns of injury produced following AHT and nAHT, and knowledge regarding the specificity of individual clinical and radiological features for AHT is essential for differentiating between AHT and nAHT. The clinical and radiological features associated with AHT are discussed below and summarised in Table 2.1.

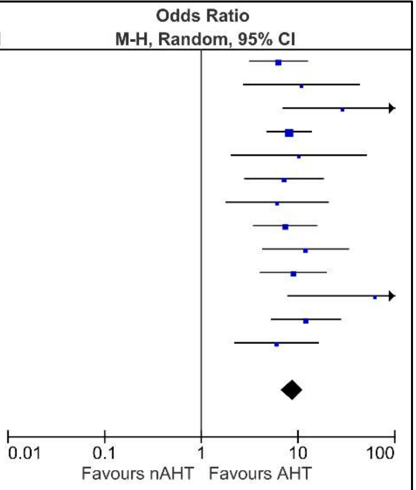
2.4.2.2.1 Intracranial injury

Intracranial haemorrhage (ICH) is a distinctive feature of AHT and may include SDH, IPH, SAH, EDH, or a combination of these.[136] The most common neuroradiological finding in children with AHT is SDH. In one systematic review of 21 studies examining the neuroradiological features associated with AHT and nAHT, SDH was present in 68% of children with AHT compared to 23% of those with nAHT.[136] In this review, SDH was strongly associated with AHT and had an odds ratio for AHT of 8.2 (95% CI 6.1–11; p<0.0001).[136] This association remained highly statistically significant following an update of the review in 2014,[330] in which four additional studies were identified and included.[143, 351-353] Data regarding SDH were available in three of these studies,[143, 351, 352] and inclusion of these data into the meta-analysis resulted in a similar odds ratio of 8.7, as depicted in the forest plot in Figure 2.1. Updated forest plots of the additional neuroradiological features examined in Kemp et al.[136] are displayed in Appendix 2.

Figure 2.1 Forest plot depicting the association between subdural haemorrhage and abusive head trauma in children

	AHT		nAHT			Odds Ratio
Study or Subgroup	Events	Total	Events	Total	Weight	M-H, Random, 95% CI
Adamo 2009	28	54	24	164	13.1%	6.28 [3.16, 12.49]
Bechtel 2004	12	15	18	67	3.3%	10.89 [2.75, 43.09]
Duhaime 1992	13	24	3	76	3.1%	28.76 [7.05, 117.36]
Ettaro 2004	58	89	54	288	22.4%	8.11 [4.79, 13.73]
Goldstein 1993	8	14	3	26	2.4%	10.22 [2.06, 50.76]
Hettler 2003	43	49	57	114	7.2%	7.17 [2.83, 18.16]
Hymel 1997	16	39	4	39	4.2%	6.09 [1.81, 20.52]
Hymel 2013	85	95	61	114	11.0%	7.39 [3.48, 15.66]
Keenan 2004	75	80	40	72	6.0%	12.00 [4.34, 33.20]
Reece 2000	25	51	14	144	10.3%	8.93 [4.10, 19.44]
Thalayasingam 2012	38	39	17	45	1.4%	62.59 [7.86, 498.50]
Vinchon 2005	47	57	26	93	9.3%	12.11 [5.34, 27.48]
Vinchon 2010	37	45	17	39	6.3%	5.99 [2.22, 16.14]
Total (95% CI)		**651**		**1281**	**100.0%**	**8.71 [6.79, 11.17]**
Total events	485		338			

Heterogeneity: Tau² = 0.00; Chi² = 9.63, df = 12 (P = 0.65); I² = 0%
Test for overall effect: Z = 17.02 (P < 0.00001)



M-H = Mantel-Haenszel statistic, AHT = abusive head trauma, nAHT = non-abusive head trauma. Adapted from Kemp et al.,[136] with the addition of three new studies identified in an update of the systematic review conducted in 2014.[330] Data were pooled using a random effects model to allow for both within-study and between-study variance. Subdural haemorrhage is the most common neuroradiological finding in abusive head trauma and was significantly associated with abusive head trauma in all 13 studies.

In the Kemp et al.[136] review, particular characteristics of SDHs demonstrated a greater association with AHT than others. Specifically, multiple SDHs, bilateral SDHs, interhemispheric and infra-tentorial/posterior fossa SDHs, and SDHs over the convexities were significantly associated with AHT[136] (see Appendix 2). In addition, a descriptive analysis revealed an association between multiple SDHs of different density and low density SDHs, and AHT,[136] findings which have previously been construed as evidence for prior or repeated episodes of AHT.[134, 354] However, these findings should be interpreted with caution, as there are many other factors that may affect the appearance of SDH on neuroimaging,[355] and low density SDH may arise sooner than originally thought.[227] This issue is discussed further in section 2.4.4.1.

SAH was not significantly associated with AHT or nAHT, while EDH (bleeding into the space between the dura and the inner surface of the skull) was significantly associated with nAHT[136] (see Appendix 2). EDH is a contact injury, requiring impact to the head, and in young children results most commonly from falls.[356] EDH is associated with skull fracture in approximately 85% of cases,[357] and in rare cases can cause death due to increased intracranial pressure.[358] Of note, RHs have been found in infants with EDH resulting from accidental trauma.[359] As described in Chapter 1, parenchymal injuries can occur as the result of impact forces, rotational acceleration-deceleration forces, and/or hypoxic-ischemia. In the original meta-analysis, focal parenchymal injury was not significantly associated with AHT or nAHT.[136] In a refined analysis, including only those with IPH specifically, there was again no significant association between IPH and AHT or nAHT (Appendix 2). In contrast, a recent study identified parenchymal brain lacerations in 18 patients with AHT and no patients with accidental trauma.[360] Cerebral oedema, hypoxic-ischemia, DAI/shear injury, and closed head injury (ICI without skull fracture) were all significantly associated with AHT[136] (see Appendix 2).

A separate systematic review utilising a different search strategy and inclusion criteria yielded similar results.[27] In this review SDH, cerebral ischaemia, cerebral oedema, and ICI co-occurring with skull fracture were significantly associated with AHT, however cerebral oedema was no longer associated with AHT (or nAHT) when including only high quality studies.[27] EDH was significantly associated with nAHT, while SAH and DAI were not significantly associated with either AHT or nAHT.[27]

2.4.2.2.2 Retinal haemorrhages

RHs are commonly but not universally noted in victims of AHT. They are seen in around 75% of AHT cases.[23, 361] They are most likely to occur in children who have died and least likely to occur in children who are neurologically unimpaired,[181, 285, 362] and the severity of RH correlates with the severity of brain injury,[363, 364] as well as the likelihood of abuse.[365] RHs, although not pathognomonic for AHT, are strongly associated with AHT, and occur much less commonly among children with nAHT.[23, 24, 27, 143, 365] In one systematic review of the literature, RHs were found in 78% of children with AHT compared to 5% of children with nAHT, and in a child with head trauma and RH, the probability of abuse was estimated to be 91%.[23] It is crucial to note that not all RHs are the same, and the specificity of RH for AHT depends on the number, type, and distribution pattern of the RH.[366] Bilateral, multi-layered, too-numerous-to-count RHs that extend into the periphery are highly specific for AHT.[23, 361, 365, 367]

In one large prospective study, severe RHs were 100% specific for AHT.[195] However there is no one pattern of RH that is unique to AHT.[23]

2.4.2.2.3 Additional retinal findings

Additional retinal findings that have been identified in children with AHT include perimacular retinal folds, traumatic retinoschisis, optic nerve sheath haemorrhage, vitreous haemorrhage, papilledema and anterior segment injuries.[23, 361, 368, 369] Retinal folds and traumatic retinoschisis are found more commonly in fatal AHT cases than non-fatal AHT cases.[361, 370] In two systematic reviews of the literature, there were no comparative studies found that reported retinal folds or traumatic retinoschisis in nAHT, highlighting the specificity of these features for AHT.[23, 361] Optic nerve sheath haemorrhages are also frequently found in fatal AHT cases and were significantly more common in AHT than nAHT in one post-mortem study.[371] Their specificity for AHT has been estimated as 71%.[361] Vitreous haemorrhage, papilledema and anterior segment injuries are relatively uncommon in children with AHT but are associated with poor prognosis.[364, 366, 368, 372, 373]

2.4.2.2.4 Rib and long-bone fractures

Rib fractures and long-bone fractures have consistently been found to be associated with AHT in infants.[27, 59] Rib fractures in particular are highly specific for abuse in the absence of obvious accidental trauma or an organic cause.[374, 375] In children with AHT, rib fractures result from squeezing forces on the immature skeleton, generated by a tight grip around the infants chest; this mechanism means that posterior rib fractures in particular are highly specific for abuse.[376] Younger children, particularly those less than 12 months of age, are at greater risk for abusive rib fractures.[377] Rib fractures in children are a recognized marker of severe trauma and are associated with a high risk of death, but acute rib fractures may be difficult to detect clinically and radiographically.[376, 378] In one study, 80% of acute rib fractures found at post-mortem were undetectable on chest radiographs.[379] Healing rib fractures were strongly associated with abuse in one study, and were only seen in abused children.[380] Rib fractures resulting from cardiopulmonary resuscitation in children are rare, but when reported, they are anterior and may be multiple; there are no reports in the literature of a child with a posterior rib fracture due to cardiopulmonary resuscitation.[381, 382] Classic metaphyseal lesions of the long-bones were one of the earliest injuries to be associated with AHT and result from compression and jerking of the extremities during a shaking incident.[42, 43]

Metaphyseal fractures have a high specificity for abuse,[383] particularly in infants less than 12 months of age,[384] and in infants who are not yet walking.[385] They are the most common long-bone fracture identified in infants who die with evidence of abusive injury.[386] Long-bone fractures caused by accidental mechanisms occur more commonly with increasing age.[387-390] In one systematic review, neither rib or long-bone fractures significantly discriminated between AHT and nAHT,[24] but both helped to discriminate between AHT and nAHT when entered into a subsequent multivariable analysis with other clinical features.[59] In another systematic review, rib fractures, long-bone fractures and metaphyseal fractures were all significantly associated with AHT.[27]

2.4.2.2.5 Skull fractures

Skull fractures are seen frequently in children with AHT, but are also common in those with accidental head trauma.[391-393] In one systematic review, skull fractures were more commonly reported after nAHT than after AHT and the probability of abuse in a child with a skull fracture was estimated to be 30%.[374] Similarly in another review, skull fractures were more strongly associated with nAHT than AHT; the probability of AHT given a skull fracture in this review was calculated as 44%.[24] However, when included in a subsequent multivariable analysis, given all other information, the presence or absence of skull fractures did not help to discriminate between AHT and nAHT.[59] Linear, parietal skull fractures are the most common type of skull fracture in both abused and non-abused children.[10, 391, 394, 395] Complicated skull fractures, such as those that are complex, multiple, bilateral, depressed, diastatic, or crossing suture lines, have been reported to be associated with AHT, however the specificity of complicated skull fractures for AHT varies.[10, 391, 394, 396] In one study, multiple, diastatic, growing, depressed, complex, or bilateral skull fractures, or those that crossed suture lines, were much more frequent in children with AHT than children with nAHT.[396] In another study abused children had more multiple or bilateral fractures or fractures that crossed suture lines, than non-abused children.[394] However neither Leventhal et al.[391] or Reece and Sege[10] found any significant difference in the prevalence of complicated skull fractures between abused and non-abused children. In a recent systematic review of clinical and radiographic features associated with AHT and nAHT, skull fracture co-occurring with ICI was significantly associated with AHT, while isolated skull fractures were significantly associated with nAHT.[27] Children who present with isolated skull fractures, a clear history of trauma, no extracranial injuries and no social concerns have most likely sustained accidental trauma.[397]

2.4.2.2.6 Bruising

Head and neck bruising was found to be non-discriminatory for AHT in one systematic review,[24] and significantly associated with nAHT in another,[27] however this association disappeared when the analysis was limited to high quality studies.[27] In a multivariable analysis, head and neck bruising assisted in discriminating between AHT and nAHT.[59] In one retrospective study, 54% of children with AHT had no bruising recorded at initial presentation.[12] Similarly, Ingham et al.[398] found that only 16% of infants with fatal AHT had bruising. Bruising associated with rib or extremity fractures is uncommon and the presence of bruising therefore cannot be used to differentiate between abusive and non-abusive fractures.[399-401]

2.4.2.2.7 Spinal injury

There is increasing recognition in the literature of the risk of spinal injuries in children with AHT.[402] Spinal injuries seen in AHT include extra-axial haemorrhage, ligamentous and soft-tissue abnormalities, and spinal fractures.[25] Spinal injury has been documented in both clinical and autopsy case series.[26, 97, 106] Many injuries are found in the cervical region,[402] however thoracic and lumbar injuries are also reported.[106, 109, 403, 404] Spinal injuries have been infrequently recorded in association with AHT in the scientific literature, and it is likely that they are underreported as they may not be recognized by clinicians at initial presentation.[103, 106, 403] In one study 8/18 children with AHT (44%) had spinal SDHs noted on magnetic resonance imaging (MRI) however all of these were clinically occult and were likely masked by the altered mental status of the children.[109] In a retrospective study of children aged less than two years of age, cervical spine ligamentous injury was discovered in 78% of children with AHT compared to 46% of children with accidental trauma and just 1% of children with non-traumatic conditions.[26] The percentage of cervical spinal soft tissue injuries found in children less than three years of age undergoing evaluation for suspected AHT was 36% in one study.[94] However, prevalence estimates of cervical spine injury in children with AHT vary due to the differences in patient populations included in different studies.[402] Cervical spine injury may be more common in fatal AHT.[15, 97] Spinal fractures in AHT are relatively uncommon, with an estimated prevalence of 0.3%-2.7%.[25]

2.4.2.2.8 *Other associated injuries*

Children with AHT have been found to have other associated injuries including visceral injuries,[28, 405] burns,[406] and oral injuries.[9]

2.4.2.2.9 *Combinations of clinical features*

Although systematic reviews and meta-analyses of individual clinical features are helpful for discriminating between AHT and nAHT, it is clear that no single physical finding is specific for AHT, and in practice children with head trauma typically present with different combinations of multiple clinical features.[407] With the increasing wealth of published literature on the clinical features associated with AHT and nAHT, clearly a critical challenge for clinicians is the ability to synthesize and apply this data to a decision about the likelihood of AHT in a child with a given set of features.[289] There is increasing evidence that combinations of clinical features can identify children with ICI and a high likelihood of AHT, who thus require a thorough evaluation for abuse.[59] As summarised in Chapter 1, in a multivariable analysis of individual patient data, the following six features were significantly associated with AHT: rib fracture, RH, long-bone fracture, seizure, apnoea and head/neck bruising.[59] The Maguire et al.[59] study describes the derivation of the PredAHT CPR that is the focus of this thesis, and will therefore be discussed in further detail in Chapter 4.

Table 2.1 Clinical and radiological features associated with AHT

	AHT	**Non-discriminatory**	**nAHT**
Neuroradiology Extra-axial haemorrhage	SDH[27, 136] Multiple[136] Bilateral[330] Interhemispheric[136] Convexity[136] Posterior fossa[136]	SAH[27, 136]	EDH[27, 136]
Intracerebral features	Hypoxic ischaemic injury[27, 136] Cerebral oedema[27, 136] DAI/shear injury[330]	Focal parenchymal injury[136] IPH[330] DAI[27] Cerebral oedema (high quality studies only)[27]	
Other neuroradiological features	Closed head injury (ICI without skull fracture)[136] ICI co-occurring with skull fracture[27]		

History and clinical features[9]	No history of trauma Low impact fall with persistent neurological impairment Out-of-hospital cardiopulmonary resuscitation Initial history changes Other trauma explanations	Low impact trauma without neurological impairment	High impact trauma
Clinical features	Rib fractures[27, 59] Long-bone fractures[27, 59] Metaphyseal fractures[27] Retinal haemorrhage[23, 24, 27, 59] Apnoea[24, 27, 59] Seizures[27, 59] Head and neck bruising[59]	Skull fractures[24, 59] Long-bone fractures (univariable analysis)[24] Head and neck bruising (univariable analysis)[24] Head and neck bruising (high quality studies only)[27] Seizures (univariable analysis)[24] Rib fractures (univariable analysis)[24]	Isolated skull fractures[27] Head and neck bruising[27]
Retinal haemorrhage[23, 361]	Bilateral Multi-layered Extend to periphery Numerous	Other retinal features	Rare but when occur: Unilateral Posterior pole Scattered Few in number

SDH = subdural haemorrhage, SAH = subarachnoid haemorrhage, EDH = epidural haemorrhage, IPH = intraparenchymal haemorrhage, DAI = diffuse axonal injury, ICI = intracranial injury. Adapted from Kemp, 2011[407]

2.4.3 Differential diagnosis

Review articles have documented a long list of medical conditions that are proposed to "mimic" AHT, and both ICH and RH have an extensive differential diagnosis,[274, 368, 408-414] further complicating the identification of AHT. However, many of the conditions cited in the literature are not true mimics of AHT but are conditions in which ICH or RH can occur as part of their clinical spectrum.[408] Most of these can be differentiated from AHT by careful consideration of the history, physical examination findings and radiological and laboratory studies.[415] Conditions associated with ICH include, but are not limited to, birth and other accidental trauma, congenital malformations, genetic and metabolic conditions, coagulation and haematological disorders, oncologic disease, infections, neurosurgical complications, poisoning, nutritional deficiencies and benign enlargement of the subarachnoid space (BESS)[407, 416] (Table 2.2). Most genetic and metabolic diseases are less common than AHT, and the discovery of a rare disease does not necessarily rule out child abuse.[415] A detailed discussion of each of these conditions is beyond the scope of this thesis, however the conditions with clinical findings that are more likely to be mistaken for those seen in AHT will be briefly reviewed. These include birth trauma, some coagulation and metabolic disorders, and accidental trauma.

Table 2.2 Differential diagnosis of intracranial haemorrhage in children

Condition	Characteristics	Comment
Accidental trauma	SDHs are more likely to result from high impact falls or motor vehicle crashes. Low impact falls (< 3 feet) rarely cause SDH	RH is rarely associated with accidental trauma[23]
Neurosurgical complications	SDH is commonly reported as a postoperative complication of neurosurgery	Neurosurgery will be apparent from the medical record
Birth trauma	Asymptomatic neonatal SDHs have been reported in any type of delivery. They are characteristically small and most resolve by 1 month, all by 3 months[417-419]. SDHs in asymptomatic neonates are of a different pattern to those reported in AHT[419, 420]	Birth-related RH are more common after instrumental deliveries. They are commonly bilateral, and predominantly intraretinal and posterior. They resolve rapidly, and rarely persist beyond 6 weeks[421]
Congenital malformations	Spontaneous bleeding from vascular malformations – for example aneurysms. Less serious trauma can result in SDH when arachnoid cyst is present[422, 423]	Congenital malformations are relatively rare in the paediatric population[424]
Cerebral infections	Meningitis: post-infective subdural effusions are reported[425]	Meningococcal septicaemia can be associated with RH[426]
Coagulation and haematological disorders	Leukaemia[427] Sickle cell anaemia[428] Disseminated intravascular coagulation Haemophilia Von Willebrand disease Haemorrhagic disease of the newborn (Vitamin K deficiency)[429] Idiopathic thrombocytopaenia purpura[430]	These disorders will also predispose to RH[431] and bruising In a recent study of congenital bleeding disorders, nontraumatic ICH occurred most commonly in severe haemophilia. In this study Von Willebrand disease was not supported as a "mimic" of AHT[432]
Metabolic disorders	Glutaric aciduria type 1 is associated with fronto-temporal atrophy that can predispose to SDH Galactosaemia Menkes kinky hair syndrome	Case reports describe associated RH in glutaric aciduria[433, 434] A systematic review of SDHs in children with glutaric aciduria found that the SDHs are accompanied by other brain abnormalities specific for the disorder[435] Vitreous haemorrhage has been reported in galactosaemia Levy 1996

		Menkes syndrome has associated femoral spurs that can be confused with fractures[436]
Genetic disorders	Osteogenesis imperfecta (OI) Ehlers–Danlos syndrome (EDS)	Fractures and RH have been reported in OI[437] Although ICH have rarely been described in EDS[438], SDH and RH were found in a 3 month old child who was placed in foster care but subsequently found to have EDS VI. The child was returned home following family court proceedings[439] Children with EDS are predisposed to bruising but there is no evidence that they are predisposed to fractures[440-442] Recently a mixed EDS and OI phenotype has been identified in which children have an increased susceptibility to fractures[443, 444]
Hypernatraemia	SDH is described in association with salt poisoning, hypernatraemic dehydration	Hypernatraemia may also be a complication of the intracranial trauma[445]
BESS	Benign extra-axial fluid collections of infancy must be differentiated from low attenuation SDH. If this coexists with SDH the cause must be investigated. There is debate in the literature as to whether benign extra-axial fluid of infancy predisposes an infant to SDH[446, 447]	Literature suggests that the rate of haemorrhagic SDH is low (1.7%) in children with BESS[448-453]. AHT or accidental trauma was reported in 41.6% of cases. A recent study reported that concomitant injuries were found in 50% of children with BESS and SDH[454]

SDH = subdural haemorrhage, RH = retinal haemorrhage, ICH = intracranial haemorrhage, OI = osteogenesis imperfecta, EDS = Ehlers-Danlos syndrome, BESS = benign enlargement of the subarachnoid space. Adapted from Kemp, 2011[407]

2.4.3.1 *Birth trauma*

Asymptomatic ICH has been described in healthy new-borns after birth, and in any type of delivery.[417-420] In one prospective study in which MRI was performed on 111 infants within 48 hours of birth, SDH was found in nine infants (8.1%), after 3/49 (6.1%) vaginal deliveries and 6/37 (16.2%) forceps or vacuum assisted deliveries.[419] There were no SDHs recorded in infants following caesarean delivery.[419] In another study, the incidence of asymptomatic ICH following vaginal delivery was 26%, and 0% following delivery by caesarean section.[420] Similarly, Rooks et al.[418] found SDH in 46/101 (46%) infants, including 32/63 (51%) infants born by spontaneous vaginal delivery, 10/16 (63%) infants born by assisted delivery and 4/11 (18%) infants born by caesarean section. Assisted deliveries may increase the risk of injury compared to normal deliveries.[419] Birth-related SDHs are characteristically small, most of them resolve by one month of age, and all of them resolve by three months of age.[418, 419] They are typically of a different pattern to SDHs reported in AHT.[419, 420]

2.4.3.2 *Accidental trauma*

Household accidents or short falls are a common history given for both children with AHT and children with nAHT, and therefore these histories can present a diagnostic dilemma for clinicians attempting to distinguish between AHT and nAHT.[356] However, while short falls in children can occasionally cause skull fractures and associated underlying focal ICI, they rarely cause severe or fatal brain injury.[455-464] One study showed a lower rate of skull fractures or ICI after a fall from standing or from low height furniture than after a fall from a carer's arms, windows, or other building components.[456] The risk of death resulting from short falls < 1.5 metres in infants and young children has been estimated as 0.48 per 1 million per year.[465] Accidents involving severe forces e.g. falls from a significant height or motor-vehicle accidents can cause SDHs,[17, 231, 466, 467] however AHT is the commonest cause of SDH in infants and children < 3 years of age,[17, 22, 147, 231, 466] occurring three times more frequently in children with AHT than in children with nAHT.[136]

2.4.3.3 *Coagulation and metabolic disorders*

Spontaneous or traumatic ICH can occur in children with severe bleeding disorders such as haemophilia,[432, 468] representing a challenge in the evaluation for possible abuse.[469] The frequency of ICH associated with congenital and acquired disorders of coagulation varies by condition and severity of disease.[432, 468] In one recent study the highest prevalence of ICH was in severe haemophilia A (9.1%) and B (10.7%); in this study only one child with type 1 von Willebrand disease had SDH, the findings of which were not congruent with AHT.[432] The AAP has published a clinical report outlining the probability that a child would have a specific coagulopathy causing ICH by condition, to assist in minimizing the possibility of misdiagnosis.[468] It is important to distinguish between coagulopathy due to a primary bleeding disorder and that due to secondary parenchymal brain injury.[470] Glutaric aciduria type 1 (GA1) is a metabolic disorder that shares some clinical features with AHT and can predispose infants to SDH.[434] Children with GA1 have been misdiagnosed with child abuse,[434] however a recent systematic review has highlighted additional characteristic abnormalities associated with the disease, thus it should be easily distinguishable from AHT on neuroimaging.[435]

2.4.3.4 Differential diagnoses for retinal haemorrhages and related findings

RHs have been documented in asymptomatic new-borns.[421] These are more frequent after instrumental deliveries. They are commonly bilateral, and predominantly intraretinal and posterior. They resolve rapidly, and rarely persist beyond six weeks of age.[421] RHs have been reported in association with severe accidental trauma such as fatal motor vehicle crashes[471] or crush injuries,[472] however they are rare, and typically occur in a different pattern from RHs associated with AHT; RHs from accidental trauma are predominantly unilateral, few in number and confined to the posterior pole.[23] However, in one systematic review,[23] one fifth of children who had sustained an accidental crush injury had extensive, multi-layered RH more commonly seen in AHT, with associated traumatic retinoschisis and perimacular retinal folds.[473-475] Another systematic review of the differential diagnosis of retinal haemorrhages in children with clinical features associated with abuse identified nine conditions that have considerable clinical/retinal overlap with child abuse, and therefore pose the greatest diagnostic dilemma, including metabolic diseases, bony dysplasias, bleeding disorders and vascular malformations.[409] This review found that reports of RH in children with organic diseases are rare, but nevertheless these conditions should be actively considered when investigating a child with RH.[409] Where described, RHs were bilateral and present only in the posterior pole. There was no evidence that cardiopulmonary resuscitation, cough, or ALTE, causes RH.[409] Recent studies examining patterns of RH associated with non-traumatic increased ICP have determined that raised ICP alone is unlikely to cause numerous, multi-layered RH.[476, 477]

2.4.4 Forensic considerations

The difficulties associated with the identification of AHT originate in large part from the necessity to interpret the child's clinical, investigation and historical findings within a *forensic* context.[338] Unlike in other areas of medicine, the evaluation of suspected AHT requires clinicians to operate outside of their normal environment and engage with children's social services and law enforcement.[478] In addition to defining and communicating the likelihood of AHT, clinicians must contribute to decisions regarding the future welfare and risk of harm to the child and any siblings.[407] They may be asked to present evidence in the family or criminal court as an expert medical witness. Many clinicians are not comfortable with this responsibility.[479, 480] The diagnosis of AHT is typically an evolutionary process; a clinician's level of certainty regarding whether or not AHT has occurred may fluctuate as the investigation unfolds and further information is acquired[289]. Uncertainty plays an important role in the forensic evaluation of possible child abuse and has been categorised into four main types: technical uncertainty, personal uncertainty, conceptual uncertainty and uncertainty beyond the medical diagnosis[189, 481] (Table 2.3).

2.4.4.1 Technical uncertainty

In addition to the clinical management and treatment of the child, if AHT is part of the differential, the clinician must first decide whether the findings are due to trauma or a medical cause, and, if they are due to trauma, whether they are due to accidental trauma, birth trauma, or AHT.[189, 289] The clinician must carefully consider whether the care-giver's account of the incident is a plausible explanation for the child's findings, and specifically whether the history explains the mechanism, severity and timing of the injuries.[189, 289, 338] In order to do this, detailed information regarding the dynamics of any proposed traumatic event is needed, including the height and forces of an alleged fall or impact, the type of surface the child is reported to have struck, and the child's immediate response and symptoms following the injury.[416] Clinicians may be inexperienced in conducting such "forensic style" interviews,[416] however they will most likely face forensic questions from law enforcement and children's social services regarding the most likely perpetrator of the abuse, the timing of the injury, or the biomechanics of the purported injury mechanism, which carry profound medico-legal implications.[137, 289] Establishing the onset of the child's symptoms can aid in determining when they were injured. Studies demonstrate that children with AHT become symptomatic immediately after they are injured,[137-139, 482, 483] and there is no evidence of a so-called protracted, asymptomatic "lucid interval" prior to catastrophic neurologic collapse in children with fatal head trauma[482] or children with serious confirmed AHT.[483] Whilst neuroimaging may be helpful in providing *general* age ranges of ICH,[355, 484] a recent systematic review concluded that the time intervals of the different appearances of SDHs are broad and overlapping, and thus imaging cannot be used to accurately date SDHs.[485] In a study by the same research group, a survey of Dutch radiologists highlighted considerable practice variation regarding the age determination of SDHs, implying that precise dating of SDHs should be avoided.[486] Adamsbaum et al.[484] propose that the finding of different density in two distant SDHs can be considered the only reliable indicator of age-different injuries and may suggest possible prior abuse or repeated violence. A recent study examined the evolution of RH and identified informative patterns to help clinicians establish the timing of injury.[487] Although the sample size is small and further research is needed in this area, this study is the first of its kind and can assist clinicians in determining injury timing with greater precision than was previously possible.[488] Aging of bruises is not possible, and while fracture dating is challenging and has been described as an "inexact science", there are recognised stages of healing that can facilitate the dating of fractures in children with unknown injury timing.[382, 489]

2.4.4.2 Personal uncertainty

Clinicians may be uncertain about whether AHT has occurred due to their personal involvement with the family or particular characteristics of the family.[189] A diagnosis of AHT places strain on the patient-doctor relationship, and clinicians may find it difficult to balance their responsibilities as family and child advocates.[490, 491] Regarding the characteristics of the family, as described in Chapter 1, although evidence-based risk factors should be considered as part of a thorough evaluation for AHT, the presence of risk factors should not automatically lead to a conclusion of abuse, and vice versa.[188, 189] Clinician personal bias is discussed in further detail below and in Chapter 3.

2.4.4.3 Conceptual uncertainty

Conceptual uncertainty may arise if clinicians have difficulty applying the vast amount of literature and evidence-based research to their individual patients.[189] This type of uncertainty may result from either a lack of experience evaluating children with suspected abuse, a lack of familiarity with or understanding of the literature, or a lack of definitive, evidence-based data regarding a particular finding or theory to which to turn.[189] In addition, clinical colleagues or colleagues from social services and law enforcement may have differing opinions on the likelihood of abuse, and each may have their own uncertainties. Moles and Asnes[189] point out that this "additive nature of uncertainty results in multiple opportunities for a correct diagnosis to be missed or clouded" (p. 1025).

2.4.4.4 Uncertainty beyond the medical diagnosis of abusive head trauma

An additional difficulty in identifying suspected AHT relates to the level of certainty in the diagnosis required for the future protection of the child and for a criminal prosecution of a possible perpetrator.[189] The civil standard of proof in the UK family courts is proof on "a balance of probabilities", meaning that AHT is more likely to have occurred than not, whereas the criminal standard of proof is proof "beyond a reasonable doubt". Suspected AHT cases often rise to the level of certainty at which social services can intervene to protect a child from future harm, but may not rise to the level at which an alleged perpetrator faces trial or is criminally convicted of AHT.[189]

Table 2.3 Types and sources of uncertainty in the evaluation of suspected physical abuse

Type of uncertainty	Definition	Sources of uncertainty
Technical uncertainty	Inadequate scientific knowledge exists to predict disease processes or outcomes	Unclear diagnosis of abuse Possible accident Possible medical condition Uncertain timing of injury
Personal uncertainty	Personal involvement or connection to the patient may affect clinical decisions	Care-giver characteristics that influence level of concern (e.g. parents sought care appropriately, intact family, educated parents)
Conceptual uncertainty	Difficulty applying abstract information to a concrete scenario	Clinician's inexperience with possible abuse Differing opinions of subspecialty clinicians on likelihood of abuse
Uncertainty beyond the medical diagnosis	Level of certainty in the diagnosis of AHT required by the multidisciplinary team	Meaning and interpretation of civil and criminal standards of proof

Adapted from Moles and Asnes, 2014[189]

2.4.5 Clinician bias

Studies have demonstrated that clinicians exhibit biases in their evaluations and reporting of child physical abuse and AHT related to the family's SES, marital status, race and ethnicity.[31, 201, 202, 490, 492-499] In the study on missed AHT conducted by Jenny et al.,[31] AHT was more likely to be missed in white children whose parents were married. In another study, publicly insured or uninsured infants with head trauma were more likely to have a skeletal survey than privately insured infants, and this effect was modified by race; the difference in the number of skeletal surveys performed in patients with public or no insurance versus private insurance was greater among white infants than among black or Hispanic infants.[201] Although there were fewer skeletal surveys among white infants than black or Hispanic infants, white infants were more likely to be diagnosed with AHT, indicating over-evaluation for AHT in black or publicly insured/uninsured infants, and under-evaluation in white or privately insured infants.[201] Similarly, a recent study found that minority race/ethnicity children were more often evaluated and reported for suspected AHT than white/non-Hispanic children, and that such disparities occurred almost exclusively in lower-risk children.[202] Clinicians may find it difficult to believe that parents are capable of hurting their children, and as such they may not consider AHT as part of the differential diagnosis, or they may perform an abuse evaluation but remain conflicted about the cause of the injury even when the findings are highly suggestive of AHT, particularly if they have an established relationship with the family.[189, 289] Clinician bias is discussed further in Chapter 3.

2.4.6 Alternative theories of causation for findings associated with abusive head trauma

As described in Chapter 1, a variety of scientifically unsupported theories of causation have been proffered in the courtroom to explain the findings associated with AHT.[269] Such theories have been advanced in the literature,[350] generating confusion among the medical community regarding the diagnosis of AHT and thus potentially impacting upon medical and legal decision-making.[269] Articles contesting the validity of AHT consist of deliberately misleading case reports that exclude important facts,[350] unproven hypotheses,[258] descriptions of fabricated conditions such as “temporary brittle bone disease”, and commentaries or letters with no accompanying scientific evidence.[271, 500] Nevertheless, high profile legal cases have highlighted areas of uncertainty surrounding the diagnosis of AHT, and while the evidence base for many of the alternative theories of causation for AHT is weak, in some cases there may be a lack of evidence-based data toward which to turn.[189, 268, 501] A critical interpretation of the available evidence is necessitated in order to make informed decisions.[268] Theories thus far alluded to in the current chapter and in Chapter 1 propose that shaking alone cannot cause ICI,[13] and that the signs and symptoms of AHT are caused by hypoxia,[258] raised ICP,[502] an

ALTE,[350] short falls,[503] a rebleed of a chronic birth-related SDH,[504] or BESS.[446] Additional theories purport that SDH is caused by immunizations[505] or primary venous sinus thrombosis.[506] An in-depth discussion of the evidence-base for the many alternative theories of causation proposed is beyond the scope of this thesis, however the interested reader is referred to three excellent reviews,[268, 501, 507], a comprehensive legal treatise[508] and the recent consensus statement on AHT.[51]

2.5 Evaluation of suspected abusive head trauma

If the clinical and radiological features associated with AHT are to be identified or ruled out, and alternative diagnoses excluded, there are a series of investigations and assessments that must be performed in every child where AHT is suspected.[407] These are summarised in Table 2.4. A thorough and well-documented history is an essential element of the evaluation of suspected AHT.[333, 509, 510] Clinicians should obtain a detailed description of the events surrounding the child's injuries or proposed traumatic incident, including information regarding the mechanics of the event, when the child was last asymptomatic and the child's developmental capabilities, among other details.[333, 509, 510] A comprehensive multidisciplinary evaluation should be carried out in collaboration with a CP paediatrician, social services and law enforcement, and a psychosocial history obtained, which may identify additional evidence to support or refute a diagnosis of AHT.[333, 509-511]

The initial preferred imaging modality for the evaluation of children with symptomatic suspected AHT is unenhanced computed tomography (CT), as it is readily available and reliably identifies injuries requiring urgent intervention.[329, 512-515] UK guidelines also recommend a CT scan for any child aged less than one year with suspected physical abuse, regardless of whether they are neurologically symptomatic.[512] However, early indications of cerebral oedema or DAI may not be evident on CT.[410, 516, 517] A recent study suggests that MRI may be effective as a screening protocol in well-appearing, asymptomatic children.[518] All patients with abnormal CT findings should undergo follow-up MRI,[329, 512, 519] as MRI is superior to CT in identifying the pattern, extent and timing of head injuries and may identify additional injuries missed on the initial CT.[329, 410, 513, 516, 520] With the increasing recognition of the association between spinal injuries and AHT, spinal MRI is also recommended.[512] All children less than two years of age with suspected AHT should have a skeletal survey to identify any skeletal injuries, and a repeat skeletal survey should be performed around two weeks later to detect possible additional healing fractures and assist with dating the injuries.[512, 514] Skeletal survey's in older children may be appropriate and should be considered on a case-by-case basis.[512, 514]

Guidelines for the eye examination in the evaluation of child abuse recommend that children with suspected AHT should undergo indirect ophthalmoscopy conducted by an ophthalmologist.[521, 522] The UK guidelines include a standardised pro forma for documenting retinal findings.[521] Additional investigations that should be performed in children with suspected AHT include a complete blood count, platelet count and coagulation screen to exclude underlying bleeding disorders.[468] The evaluation of suspected AHT thus involves numerous elements and investigations and presents a unique challenge to clinicians, necessitating consideration of a wide range of differential diagnoses, acquisition of an extremely detailed injury history, testing for occult injury, and difficult questions relating to the psychosocial history.[509] One Delphi study found that experts highly recommended 37 separate critical elements for the evaluation of children with ICH.[509] In addition, clinicians must balance the risk of harm caused by missing an abusive injury with the harm caused by unnecessary testing, abuse evaluation and increased radiation exposure in those children ultimately deemed as non-abused,[290, 291] although it is generally agreed that the harm caused by missed AHT is greater than the harm caused by a false accusation of abuse.[523]

Table 2.4 Evaluation of children with suspected abusive head trauma

History and psychosocial history[333, 509-511]	Detail proposed traumatic explanations—for example ▸ Height of fall ▸ Nature of surface of impact ▸ Timing of trauma ▸ When child was last asymptomatic Obtain psychosocial history
Examination and observation (and history)[333, 509]	▸ Associated injuries ▸ Apnoea ▸ Seizures
Neuroimaging[329, 512, 514]	▸ Initial CT ▸ Early MRI with diffusion weighted imaging, consider follow-up MRI if neurological symptoms persist ▸ Consider extending MRI to include spinal cord
Axial skeletal imaging[512, 514]	▸ According to UK standards[512] ▸ Include oblique views thorax ▸ Consider repeat skeletal survey at 10–14 days or early bone scan if concerns persist or equivocal findings
Ophthalmology[521, 522]	▸ By ophthalmologist ▸ Indirect ophthalmoscopy through dilated pupils ▸ Standardised record of findings (laterality, extent, layer involved, associated features)
Additional investigations[468]	▸ Full blood count ▸ Coagulation screen to exclude coagulopathy ▸ Blood cultures to exclude sepsis

CT = computed tomography, MRI = magnetic resonance imaging. Adapted from Kemp, 2011[407]

2.5.1 The child protection process in the United Kingdom

When a clinician suspects that a child has suffered AHT, they have a duty to refer the child to social services, who will then lead a multi-agency assessment to decide whether there is reasonable cause to suspect that the child is suffering or is likely to suffer significant harm.[524] A strategy discussion is usually convened, involving clinicians including CP paediatricians and hospital CP nurses, police officers, and social workers, to share available information, decide whether to pursue a criminal investigation, and to decide whether to initiate an enquiry under section 47 of the Children Act 1989 to determine whether any action should be taken to safeguard the child.[524] If concerns of significant harm are substantiated, a CP conference may be organised and a CP plan put in place for the child.[524] In fatal cases, a multi-agency investigation into unexpected death in infancy is conducted.[525] Cases may be tried in the family or criminal courts. Clinicians are expected to communicate a clear opinion of the likelihood of AHT to their multidisciplinary colleagues and in their court reports.[333] Studies have reported barriers to multidisciplinary working, highlighting challenges relating to conflicting agendas and assumptions regarding the roles of each agency.[526-528]

2.6 Discussion

This chapter aimed to review the literature on the challenges faced by clinicians in identifying AHT. The purpose of this exercise was to determine the need for a CPR that predicts the probability of AHT in children with ICI, and to gain an understanding of the likely value of introducing a CPR in this field. The review highlighted that clinicians face many challenges in the identification of AHT related to the history provided by the care-giver, variability in the clinical manifestations of AHT, potential differential diagnoses, forensic considerations, personal bias, alternative theories of causation, and the evaluation of suspected AHT.

Unfortunately, due to the complexities in defining and diagnosing AHT, numerous studies have established that clinicians struggle to define a “reasonable suspicion” or “reasonable medical certainty” of abuse,[529, 530] lack the confidence to identify abuse,[531] exhibit broad variability in their perceptions of the likelihood of abuse,[532, 533] and demonstrate inconsistencies in their investigations and diagnoses of suspected abuse.[534-537] In addition, there is evidence that clinicians are uncomfortable rendering a definitive opinion of the likelihood of abuse in both the clinical and legal arena,[479, 538, 539] a task which is critical to enable CPSWs and law enforcement to safeguard children from future harm.[333, 334, 479] Lane & Dubowitz emphasize that “*protective service workers may not know how to respond unless* *they are provided a clear statement that the injuries are abusive, non-inflicted, or indeterminate*".[479(p.82)] There is also a lack of agreement between medical experts providing opinions to the courts in suspected child abuse cases,[540] and a requirement that clinicians testifying in court can support their opinions with scientific evidence.[525, 541]

A recent proposed research agenda regarding the forensic medical evaluation of child maltreatment emphasized that research is required to assist clinicians in evaluating suspected AHT more objectively, to minimize inconsistencies and variability in abuse evaluations.[542] Specifically, the agenda recommends that clinicians embrace and implement more evidence-based, probabilistic and/or Bayesian approaches to the screening and diagnosis of AHT, to aid in determining which clinical, radiological and laboratory findings should prompt a thorough abuse evaluation in children with head trauma, and which clinical, radiological and laboratory findings carry a high specificity for AHT in children with head trauma.[542] The authors argue that an evidence-based, probabilistic and/or Bayesian approach can support clinicians in responding to and overcoming the challenges associated with identifying AHT, for example in refuting the claims of those who dispute the validity of AHT as a medical diagnosis.[542] More recently, in a review of the medical progress made over the last decade in minimizing cases of missed physical abuse, clinicians explicitly recommended the use of CPRs for assisting in the early recognition and improved identification of physical abuse.[290]

2.7 Conclusions and implications for this thesis

This chapter reviewed the literature on the challenges confronted by clinicians in the identification of AHT. The findings from this review highlight that the identification of AHT is difficult for a myriad of different reasons. Despite this, there is evidence that certain combinations of clinical features may help discriminate between AHT and nAHT. Taken together, the review of the literature presented in the current chapter and in Chapter 1 demonstrate that a CPR for AHT that integrates a child's clinical data and calculates an evidence-based patient-specific probability of AHT is urgently needed and would be of value to clinicians working in this field. This literature review was conducted in line with phase one of the MRC framework for the development and evaluation of complex interventions, in addition to guidelines for the development of CPRs,[301, 322, 323] as it has reviewed the evidence-base to establish that there is a need for the PredAHT CPR. Chapter 3 therefore reviews clinical decision-making theories, and the logic underpinning CPRs, in order to illustrate the mechanisms by which CPRs may improve clinician's decision-making and the identification of AHT.

3 Clinical decision-making theories and the logic of clinical prediction rules

"Errors in judgment must occur in the practice of an art which consists largely of balancing probabilities." Sir William Osler[543]

3.1 Chapter overview

This chapter presents a scoping review of literature relating to clinical diagnostic decision-making theories and the logic of CPRs. The introduction to this chapter will provide the background and rationale for this scoping review. Then, the aims of the chapter will be presented, followed by a description and critical evaluation of selective theories and models of clinical diagnostic decision-making and an overview of the current prevailing model of clinical diagnostic decision-making, the dual process model proposed by Croskerry.[544] CPRs will then be introduced as a strategy for improving clinical diagnostic decision-making, and the stages in their development and evaluation described. Next, generic tools and CPRs for identifying child physical abuse will be briefly reviewed, and existing CPRs for the identification of AHT will be identified. The findings will be discussed, and the implications of clinical decision-making theories and the logic of CPRs for this thesis will be described. This chapter will help to inform the development of the computerised PredAHT, and the interpretation of qualitative interviews with clinicians, reported in subsequent chapters.

3.2 Introduction

As described in Chapter 1, the MRC framework for developing and evaluating complex interventions emphasizes the importance of identifying relevant theories to explain the rationale for an intervention, and to gain a deeper understanding of the processes underlying the behaviour to be targeted by an intervention, the changes that may be expected and how change is to be achieved.[322, 323] This should be done whether the researcher is developing a new intervention or evaluating an existing intervention.[322, 323] Previous studies, as described in Chapters 1 and 2, have provided evidence that the identification of AHT is challenging, but that combinations of clinical features assist in discriminating between AHT and nAHT. Thus, it has been established that a CPR for AHT would be of value to clinicians who come into contact with children with possible AHT. If PredAHT is to be introduced into clinical practice, it is important to understand how decisions regarding the diagnosis of AHT are usually made, and the mechanisms by which PredAHT may improve clinical decision-making and the identification of AHT. Therefore, a scoping review was conducted to explore relevant theories and models of clinical diagnostic decision-making and literature regarding the logic of CPRs, to develop an understanding of decision-making processes in suspected AHT cases and how a CPR may conceivably act to alter decision-making, and to identify potential features of a computerised tool that might best support clinicians in their decision-making.

3.2.1 Aims of this chapter

1. To describe and critically evaluate a selection of clinical diagnostic decision-making theories relevant to the diagnosis of AHT and provide an overview of the current prevailing model of clinical diagnostic decision-making, namely the dual process model[544]
2. To review the literature on the logic of CPRs and the mechanisms by which they may improve clinical decision-making and the identification of AHT
3. To describe the methodological stages involved in the development and evaluation of CPRs
4. To provide a brief overview of existing tools for the detection of child physical abuse, and to identify CPRs for assisting in the diagnosis of AHT

3.3 Methods

This is a literature review and was not intended to meet the Cochrane Collaboration definition of a systematic review.[328] Nevertheless, a systematic search of key databases, texts, and supplementary sources was conducted to identify relevant literature.

3.3.1 Literature search strategy

Theories and models of clinical diagnostic decision-making were identified by consulting key literature including three key textbooks: Learning Clinical Reasoning,[545] The Evidence Base of Clinical Diagnosis,[546] and Medical Decision Making[547]; a systematic review[548]; a selective literature review[549]; and two comprehensive reviews of the issues surrounding the medical diagnosis of AHT.[53, 508] Rather than being exhaustive, this approach was selected to include only theories and models contributing to our current understanding of clinical diagnostic decision-making when AHT is suspected. To identify literature regarding the development of CPRs, the electronic databases MEDLINE, EMBASE and PsycINFO were searched from inception using the Ovid platform. Search terms used were developed with the assistance of a medical librarian and information specialist, and are detailed in Table 3.1. Duplicates were removed using the Ovid de-duplicate function. Weekly email auto-alerts were set up in order to keep up-to-date with the literature. Supplementary sources included electronic alerts from the New England Journal of Medicine Journal Watch for Hospital Medicine, Emergency Medicine, Pediatric and Adolescent Medicine and Neurology. Literature was sourced from the reference lists of identified articles, and from a key textbook in the field: Clinical Prediction Models: A Practical Approach to Development, Validation and Updating.[324] Generic screening tools and injury-specific CPRs for identifying physical abuse were identified from the search strategy detailed in Chapter 2. Finally, CPRs for the identification of AHT were retrieved by combining the search strategy detailed in Chapter 2 with validated search filters for locating prediction studies.[550] These filters are detailed in Table 3.2 and have a high sensitivity for finding studies on CPRs when combined with outcome-specific search terms.

Table 3.1 Search terms used to identify literature on the development of clinical prediction rules

Clinical predict* rule Clinical decision rule Clinical predict* tool Clinical decision tool Clinical predict* model* Clinical decision model*	AND	Method* Standard* Deriv* Validat* Impact Implement* Evaluat* Guid* Barrier* Acceptab* Attitud* Opinion* Aware* Priorit* Framework* Survey Valu*

Table 3.2 Validated search filters for locating clinical prediction rule studies

(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
OR
‘‘Stratification’’ OR ‘‘ROC Curve’’[Mesh] OR ‘‘Discrimination’’ OR ‘‘Discriminate’’ OR ‘‘c-statistic’’ OR ‘‘c statistic’’ OR ‘‘Area under the curve’’ OR ‘‘AUC’’ OR ‘‘Calibration’’ OR ‘‘Indices’’ OR ‘‘Algorithm’’ OR ‘‘Multivariable’’

Reproduced from Geersing et al., 2012[550]

3.4 Clinical decision-making when abusive head trauma is suspected

The process of forming a medical diagnosis requires clinicians to identify probable internal and external/environmental causes of a patient's signs and symptoms.[551] To do this, they must use clinical reasoning, defined as "the range of strategies that clinicians use to generate, test and verify diagnoses".[552(p.1118)] Clinical diagnostic reasoning is the most important of a clinician's skills.[544] Keen diagnostic acumen is critical in the formulation of an actionable diagnosis and is essential for the safe, timely and effective management and treatment of patients.[545] As described in Chapter 2, diagnosing AHT is challenging as there is no gold-standard diagnostic test, and a requirement to interpret the child's clinical and historical information within a forensic context. Nevertheless, the approach to the medical diagnosis of AHT is the same as the approach to any other medical diagnosis.[53, 508]

When arriving at a diagnosis, clinicians must critically assess the trade-offs between the harms and benefits of tests and treatments.[545] If an incorrect diagnosis is given, patients may not receive appropriate treatment or may undergo unnecessary treatment, investigations or tests.[545] When faced with a child with an ICI evident on neuroimaging, clinicians must decide whether to pursue an abuse evaluation and undertake additional investigations such as an ophthalmology exam, skeletal survey, and coagulation screen, subjecting the child to additional radiation exposure and possible unnecessary testing, and causing distress for their family. Conversely, if an abuse evaluation is not launched, and investigations not undertaken, AHT may be missed and the child returned to an abusive environment. An incorrect diagnosis of nAHT could endanger the child's life or future wellbeing, whereas an incorrect diagnosis of AHT could result in unjustified removal of the child from their family and a delay in the correct diagnosis and treatment of any unrecognized medical cause for their symptoms. Clinicians must be able to convey information about the likelihood of a diagnosis to other interested persons[551]; in cases of suspected AHT this includes professionals from social services, law enforcement and sometimes the judiciary.[333]

In practice, clinicians must integrate several patient characteristics and symptoms to estimate the likelihood of a particular diagnosis or sometimes several competing diagnoses at once.[553] The diagnostic work-up involves a series of sequential steps in which an array of clinical information is obtained from the history, physical examination and various diagnostic tests.[553] When AHT is suspected, clinicians conduct a thorough physical examination, obtain a complete medical and psychosocial history, cooperate with multiple agencies and medical disciplines to gather additional historical and clinical information, and examine radiologic and laboratory data to determine the need for further testing.[1] In theory, following each step, the clinician (implicitly or explicitly) integrates the information into a judgment regarding the probability of the condition of interest being present.[553] The information provided at each step is combined with the previous information, and the estimated probability of the diagnosis is refined.[553] Clinical diagnostic decision-making is therefore inherently multivariable.[553-556]

3.4.1 Medicine as a science of uncertainty and an art of probability

It is widely recognized that medicine is not an exact science and that clinicians regularly encounter diagnostic uncertainty in clinical practice.[557] The significance of probability in clinical decision-making was recognized as long ago as 1922: "good medicine does not consist in the indiscriminate application of laboratory examinations to a patient, but rather in having so clear a comprehension of the probabilities and possibilities of a case as to know what tests may be expected to give information of value".[558] Although diagnostic reasoning is grounded in the roots of the scientific method,[53] "all diagnostic hypotheses represent probabilistic judgments that are based on observed medical facts that have variable probabilities of being correct. Each fact (symptom, sign, or test abnormality) also has only a variable probability of being found in a given condition that is typically characterized by its presence".[551(p.465)] Clinicians make probabilistic judgments every day, both prior to and after carrying out laboratory tests or other investigations that may confirm or refute a suspected diagnosis.[551] In addition, clinicians' judgments and actions are influenced by their personal behavioural, social, economic and cultural norms, cognitive biases, preconceptions, and emotional temperature.[559-561] In the 1980's, the evidence-based medicine (EBM) movement was pioneered by Dr David Sackett, in an attempt to minimise subjectivity in clinical decision-making and provide a framework to aid clinicians in navigating uncertainty in clinical practice.[562] The central notion of EBM is the integration of the best available research evidence with individual clinical expertise.[563] The "art of medicine" has been defined as the ability to apply EBM to individual patients.[564] Patel et al.[565] describe the science of medicine as "correlating or applying principles in an axiomatic or deductive fashion to a patient's symptoms, yielding a precise diagnostic solution", and the art of medicine as "the use of intuition, experience and holistic perceptions in making clinical judgments and in the delivery of humane care" (p. 75). Making decisions under conditions of uncertainty is fundamental to the definition of a professional, and it has been argued that for those working in CP, every evidence-based decision is essentially a risk decision.[566] Ultimately, clinicians must astutely balance their scientific knowledge of the pathophysiological processes of the human body and the statistical and epidemiologic data in the medical literature, their emotions and personal biases, and their clinical experience when considering a diagnosis of AHT for an individual patient.[53, 508]

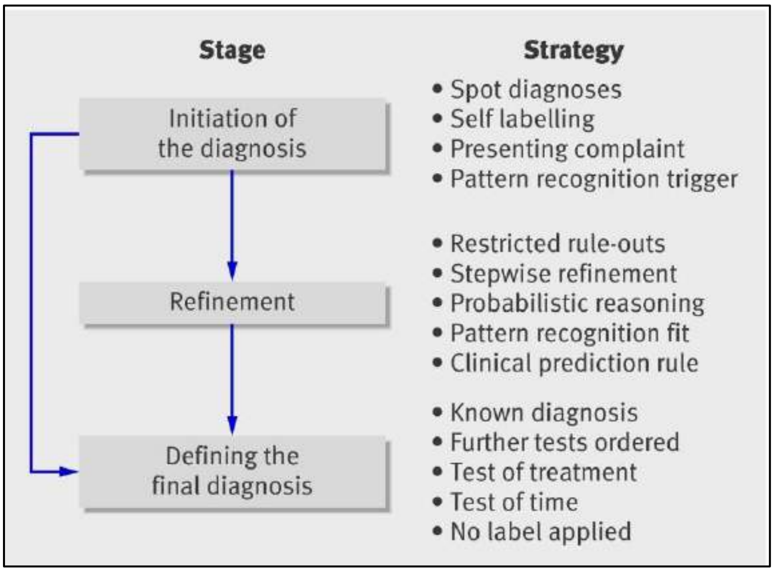
3.4.2 The hypothetico-deductive model of clinical decision-making

Whereas clinical diagnostic reasoning was once thought to proceed in a discretely linear, sequential manner, influential research conducted in the 1970's demonstrated that diagnosis follows a dynamic, non-linear, unstructured process involving both inferential and deductive reasoning.[567, 568] Possible diagnoses are generated early in the patient encounter, which guide subsequent investigations and history taking.[567, 568] This method of decision-making is known as the hypothetico-deductive model of reasoning,[567] or the "differential diagnosis methodology" in the AHT literature.[53, 508] The hypothetico-deductive model is a proposed description of scientific method; according to the model, scientific inquiry proceeds by formulating hypotheses in a form that could conceivably be falsified by a test on observable data.

Heneghan et al.[569] devised a three-stage model of hypothetico-deductive reasoning in clinical decision-making: generation of initial diagnostic hypotheses; refinement of the diagnostic hypotheses; and definition of the final diagnosis (Figure 3.1). At the hypothesis generation stage, information about the chief complaint, the history of the circumstances surrounding the presenting symptoms(s), a comprehensive medical and psychosocial history, and the physical examination findings is assembled, and a list of differential diagnoses formulated; this list then informs the investigations needed to confirm or exclude these potential diagnoses.[545] Either the most likely or the most severe diagnosis is tested first.[553] As additional clinical information is obtained, diagnostic hypotheses are progressively refined; some are excluded, some become more or less likely, and new hypotheses may be triggered. This process continues until a "working diagnosis" is reached.[545] This working diagnosis must be verified, by evaluating its coherency (do the patient's symptoms and signs match those expected for the diagnosis?), its adequacy (does the diagnosis explain enough of the patient's findings?), and its parsimony (is the diagnosis a simple explanation for the patient's findings?), and rejecting competing diagnoses (can any other diagnosis explain the patient's findings better than the working diagnosis?). The verified working diagnosis informs the next step in patient management and the definition of the final diagnosis.[545]

In the context of a child with ICI where AHT is among the differential, the chief complaint usually comprises the initial presenting symptom(s) such as apnoea, irritability, altered mental status, seizures, lethargy, vomiting, or others, as detailed in Chapter 2. Based on this, the clinician performs a physical examination, orders diagnostic tests, beginning with a head CT scan, and obtains a comprehensive medical history, to identify injuries and rule out other causes for the findings. This includes a total body examination to identify signs of trauma, a detailed history of the events surrounding the presenting symptom(s), a trauma history, a medical history, and a developmental history.[1] The clinician then considers the potential diagnoses that might explain the clinical presentation and initial radiologic findings, known as "differentials".[53] Differentials are formulated for all presenting symptoms/injuries (e.g. bruises, fractures, ICI), and the clinician then embarks on the "complex inferential and deductive process of differential refinement".[53(p.572)] Additional tests may be ordered and additional clinical information obtained from medical specialists including neurosurgeons, radiologists and ophthalmologists.[1, 570] If AHT is suspected, a multidisciplinary assessment is conducted and a comprehensive psychosocial history is obtained from social workers, including identification of psychosocial stressors, mental health disorders, substance abuse, domestic violence, and previous child maltreatment concerns.[1, 570] Investigative details will be obtained from police officers regarding e.g. parental interviews and scene examination.[334] The findings must be analysed to reach a unifying diagnosis that meets the criteria of "adequacy", "parsimony" and "coherency".[508] Heneghan et al.[569] found that multiple different strategies are used to combine, integrate, and interpret data at each of the three stages of reasoning when making diagnostic inferences (Figure 3.1).

Figure 3.1 Stages and strategies in reaching a clinical diagnosis



Reproduced from [Diagnostic strategies used in primary care, Heneghan et al., 338, b946, 2009][569] with permission from BMJ Publishing Group Ltd.

3.4.3 Hypothesis generation: pattern recognition

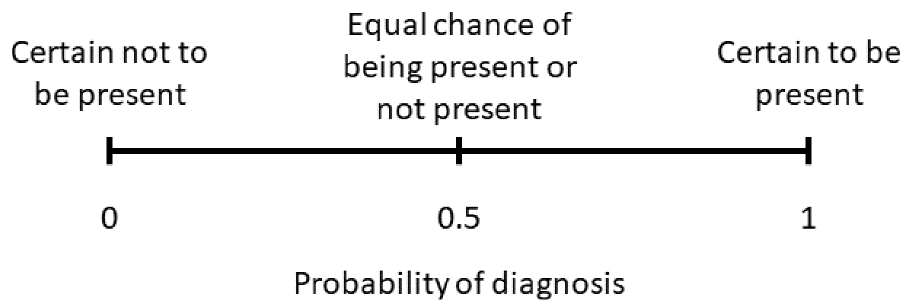
One such strategy used at the hypothesis generation stage is the direct automatic retrieval of relevant knowledge via the recognition of specific patterns of signs or symptoms.[549, 569] A number of different theories attempt to explain how new patients with specific clinical patterns are compared to a clinician's existing diagnostic knowledge, and how such knowledge is stored and retrieved.[545, 549] Exemplar-based models posit that new cases are categorised by their similarity to previously encountered cases.[571] Prototype-based models propose that new cases are compared to abstract mental models or prototypes developed through clinical experience.[572] Another theory suggests that new cases are interpreted with reference to case-specific knowledge stored in symbolic structures defined as "illness scripts".[573, 574] Illness scripts contain information about the context, manifestations and consequences of conditions, drawn together with causal links and arranged in a temporal sequence. Scripts can be both general prototypes of conditions, or specific descriptions of individual patients (exemplars). Expert diagnosticians are thought to have access to multiple stored exemplars or "instance scripts" due to exposure to many different cases.[573] This concept is particularly attractive as it allows for the possibility that different clinicians may develop different scripts for the same condition based on their individual experiences, and may explain why more experienced clinicians are better diagnosticians.[545, 573]

Indeed, expertise in diagnostic reasoning varies greatly between clinicians and depends upon their proficiency in a particular field. Some researchers have questioned the relevance of the hypothetico-deductive model of clinical reasoning described above, since diagnostic accuracy appears to rely at least in part on clinical knowledge rather than clinical reasoning strategies *per se.*[549, 571] It is probable that clinicians lacking knowledge of a particular condition or clinical area and novice clinicians adopt a hypothetico-deductive strategy to generate diagnostic hypotheses, while experienced clinicians use pattern recognition strategies, and reserve the hypothetico-deductive approach for especially difficult cases.[549]

3.4.4 Hypothesis refinement: probabilistic diagnostic reasoning

Within the hypothetico-deductive framework of clinical decision-making described above, a diagnostic hypothesis is progressively refined as new information successively increases or decreases a clinician's belief in that hypothesis. Probabilistic reasoning is one strategy used at the hypothesis refinement stage of clinical decision-making to assist clinicians in their interpretation of new diagnostic information, and has been described as an approach that can help clinicians to deal with the uncertainty inherent in clinical decision-making.[575] Probability is a quantitative means of expressing uncertainty and is defined as a number between 0 and 1 that expresses a clinician's opinion about the likelihood of a condition being present or of an event occurring in the future.[547] Any level of uncertainty can be expressed on this scale. The probability of a condition or future event that a clinician believes is certain to occur is equal to 1. The probability of a condition or future event that a clinician believes is certain not to occur is equal to 0 (Figure 3.2).

Figure 3.2 A quantitative means for expressing uncertainty: the probability scale



Crucially, a probabilistic reasoning approach is based on the principle that the interpretation of new diagnostic information depends on what you believed beforehand, and that reaching a diagnosis involves updating an initial opinion with imperfect information, i.e.

the clinical evidence.[576, 577] The formal mathematical rule for this task is Bayes' theorem.[578, 579] With this approach, a clinician identifies a condition that could plausibly explain the given set of clinical findings, and subsequently quantifies the diagnostic uncertainty via three cognitive steps.[547, 580] The first step involves making an initial judgment about the likelihood of the patient having the condition by estimating the pre-test probability of the condition. The pre-test probability is defined as the probability of the condition before acquiring new information and is also known as the prior probability. The pre-test probability can simply be the known prevalence of the condition or the clinician's subjective impression of the probability of the condition based on their personal experience and knowledge of the clinical literature. The second step involves gathering additional information, from the history, physical examination or diagnostic tests. Clinicians must know how to measure the accuracy of diagnostic tests, and possess knowledge of the discriminatory power of clinical and historical findings, in order to interpret additional information in light of their prior beliefs.[581, 582] Test performance can be conveniently summarized in a 2x2 contingency table showing the numbers of patients with and without the condition who had a positive or negative test result.[583] Commonly used measures of test accuracy can then be calculated from a contingency table (Table 3.3). The third step involves updating the initial probability estimate, that is, the pre-test or prior probability of the condition, to the post-test probability of the condition. The post-test probability is defined as the probability of the condition after acquiring new information and is also known as the posterior probability. The post-test probability is a function of the pre-test probability and the discriminatory power of a test or clinical finding. Bayes' theorem is a normative rule which explains how clinicians should reason, but it does not claim to describe how diagnostic hypotheses are actually refined in practice.[549] It has been demonstrated that clinicians with training in evidence based medicine are more likely than untrained clinicians to adopt a Bayesian approach to the interpretation of diagnostic information.[563] The role of Bayes' theorem in probabilistic reasoning is depicted in Figure 3.3. Bayes' theorem is discussed in further detail in Chapter 4.

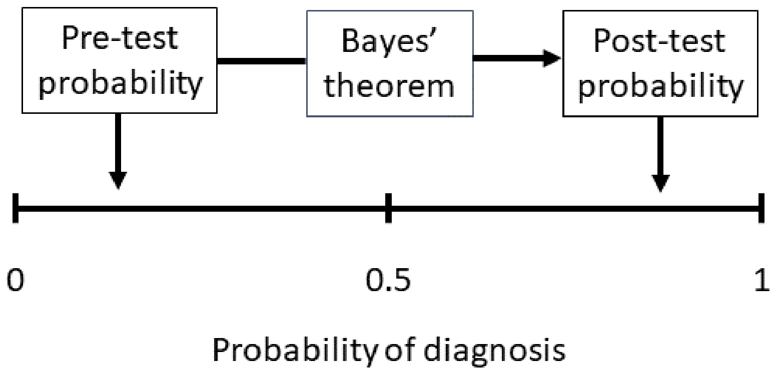
Table 3.3 Test performance parameters

	Condition Present	Condition Absent	Totals
Test positive	TP	FP	TP + FP
Test negative	FN	TN	FN + TN
Totals	TP + FN	FP + TN	TP + FP + FN + TN

Parameter	Definition	Calculation
True positives	The number of patients with the target condition who have a positive test result	TP
True negatives	The number of patients without the target condition who have a negative test result	TN
False positives	The number of patients without the target condition who have a positive test result	FP
False negatives	The number of patients with the target condition who have a negative test result	FN
Prevalence (pre-test probability)	The number of patients in the study population who have the target condition	(TP + FN) / (TP + FP + FN + TN)
Sensitivity (Sens)	The number of patients with the target condition who have a positive test result	TP / (TP + FN)
Specificity (Spec)	The number of patients without the target condition who have a negative test result	TN / (FP + TN)
Positive predictive value	The probability that a patient with a positive test result has the target condition	TP / (TP + FP)
Negative predictive value	The probability that a patient with a negative test result does not have the target condition	TN / (FN + TN)
Positive likelihood ratio: LR(+)	The number of times more likely a patient with the target condition is to have a positive test result compared with a patient without the target condition	Sens / (1 – Spec)
Negative likelihood ratio: LR(-)	The number of times more likely a person with the target condition is to have a negative test result compared with a person without the target condition	(1 – Sens) / Spec
Overall accuracy	The probability that a patient will be accurately classified by a test	(TP + TN) / (TP + FP + FN + TN)
Pre-test odds (PTO)	The odds that a patient will have the target condition before the test is carried out	Prevalence / (1 – Prevalence)
Positive post-test odds (PPTO)	The odds that a patient will have the target condition given a positive test result	PTO x LR(+)
Negative post-test odds (NPTO)	The odds that a patient will have the target condition given a negative test result	PTO x LR(-)
Positive post-test probability	The probability that a patient will have the target condition given a positive test result	PPTO / (PPTO + 1)
Negative post-test probability	The probability that a patient will have the target condition given a negative test result	NPTO / (NPTO + 1)

Adapted from Whiting et al., 2013[583]

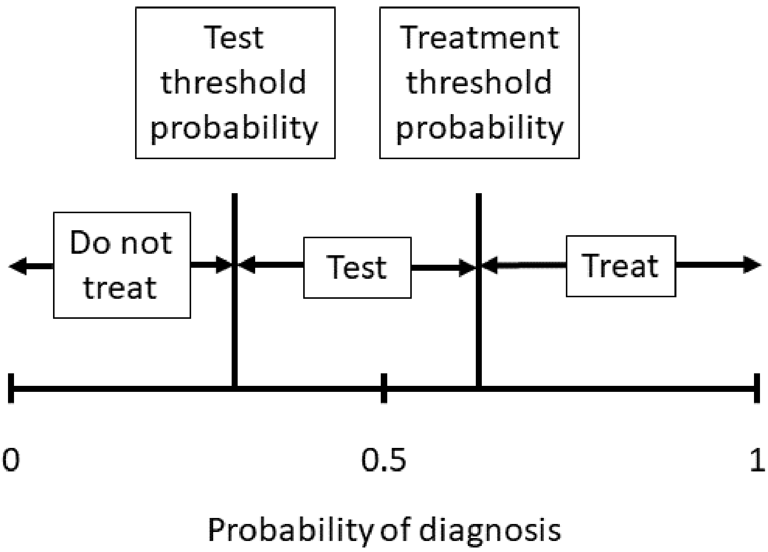
Figure 3.3 The role of Bayes' theorem in the probabilistic reasoning approach to clinical diagnosis



3.4.4.1 The threshold approach to diagnosis

In order to appreciate how a diagnostic test, clinical finding, or indeed a CPR may conceivably act to alter testing or management decisions in a case of suspected AHT, it is important to understand the threshold approach to diagnosis.[580, 584, 585] Within the framework of probabilistic reasoning, the selection and/or interpretation of a diagnostic test depends on the level of certainty at which an individual clinician is willing to order a test or initiate treatment. This is known as the threshold concept, which explicitly takes into account both the probability of a condition and the trade-offs between the harms and benefits of tests and treatments.[545, 580, 585] The test threshold is the probability at which a clinician should be indifferent between not treating the patient and carrying out further tests. The treatment threshold is the probability at which a clinician should be indifferent between carrying out further tests and treating the patient (Figure 3.4). Once these thresholds have been calculated by clinicians, whether implicitly or explicitly, when the probability of a condition falls below the test threshold, the diagnosis under consideration is effectively excluded and the optimal choice is to withhold treatment or further testing; when the probability of a condition falls above the treatment threshold, the diagnosis is effectively confirmed and the optimal choice is to treat the patient; and when the probability of a condition falls in the middle of the two thresholds, the diagnosis remains unconfirmed and the optimal choice is to conduct further tests and to either treat or not treat the patient depending on the results.[545, 580] A test should only be ordered if it would cause the probability of a condition to cross the treatment threshold and thus alter the management of the patient.[547] In the context of identifying AHT, the test threshold can be thought of as the threshold above which a diagnosis of AHT is *considered*, and an abuse evaluation and additional tests ordered e.g. an ophthalmology exam and skeletal survey, while the treatment threshold can be thought of as the threshold above which a diagnosis of AHT is *suspected*, and a referral to social services warranted.[586, 587] It is well known that individual clinicians' probability thresholds for considering, suspecting and referring AHT are somewhat varied.[490, 529, 530, 532, 533]

Figure 3.4 Test and treatment threshold probabilities



3.4.5 Causal reasoning

To determine a medical diagnosis for a patient, and therefore the most appropriate next step in management, clinicians seek to identify the cause of a patient's signs and symptoms.[588] Throughout the diagnostic reasoning process, clinicians use causal reasoning to arrive at the pathogenesis or pathophysiology of a condition, based on the cause-and-effect relationships between clinical variables.[545, 551, 588] Causal reasoning involves an understanding of the anatomic, physiologic, and biochemical mechanisms that operate in the human body in health and disease.[545, 588] To establish causation, clinicians generate cognitive "causal models" for each patient, which are essentially a sequence of plausible cause-and-effect mechanisms.[545, 588, 589] Any reliable data or clinical variable that helps to explain a causal model is incorporated into it.[545, 590] Clinicians' causal models can be based on their pathophysiologic knowledge; statistical evidence of correlations between variables that form the basis for an epidemiologic assessment of causality; considerations of biological plausibility; case reports; temporal proximity between a suspected insult and an outcome; and idiosyncratic knowledge of the individual patient.[545, 590] Several causal models may be considered during the hypothesis refinement stage.[551] Such models are most helpful at the diagnostic verification stage when evaluating the coherency of a working diagnosis.[545] Definitive tests for causality are rare,[551] and causality in medicine can virtually never be proved,[545] thus, an opinion on cause is often arrived at through the process of ruling out or eliminating potential differential diagnoses.[551] Clinicians use both causal and probabilistic reasoning to establish causation.[551]

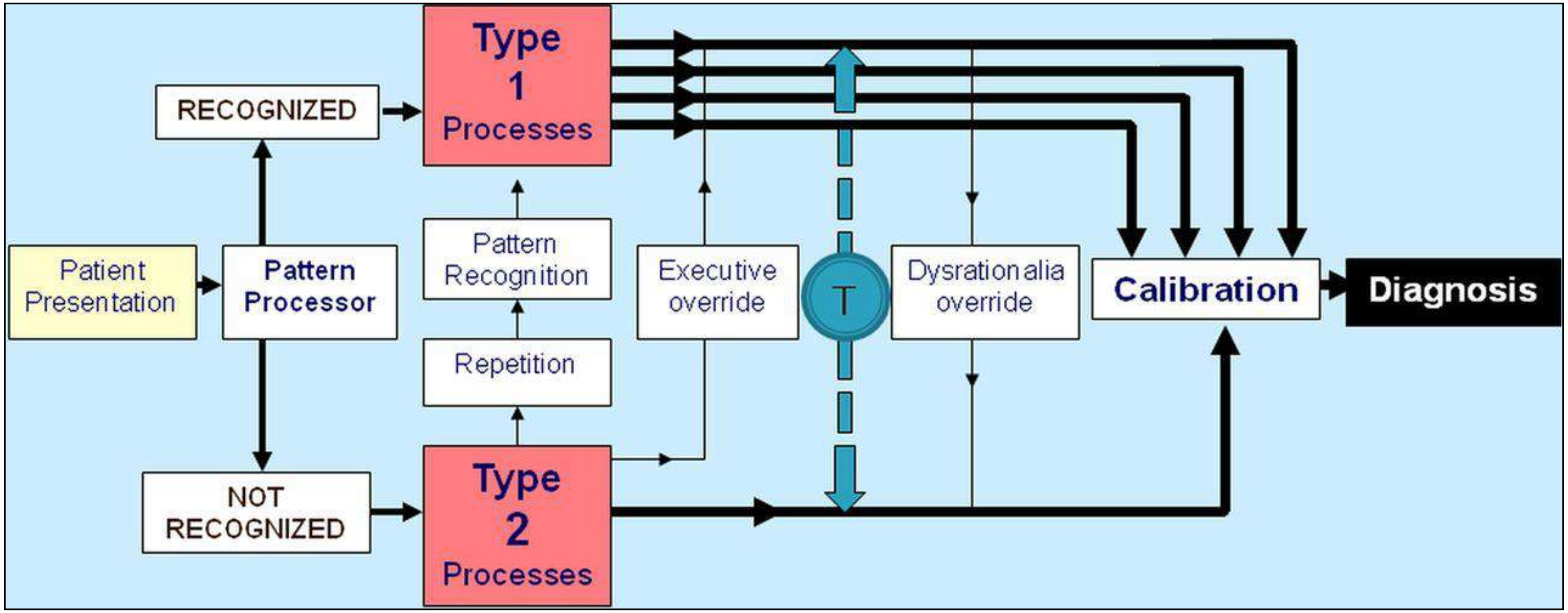
To determine whether a child's head injury was caused by accidental trauma, abusive trauma, or a medical condition, clinicians use their knowledge of the pathophysiological processes of disease and the human body to eliminate conditions from their list of differentials.[508] The diagnosis of AHT is also based in large part on determining whether a child's injuries are pathophysiologically inconsistent with the proffered history and mechanism of injury.[53] As described in Chapters 1 and 2, specialist paediatricians and pathologists have studied and tested numerous aspects of the diagnosis and identification of AHT, including the clinical presentation, historical, clinical and radiological features, injury mechanisms and pathophysiology, risk factors, and outcomes. Since AHT and nAHT can have similar clinical manifestations and there is no gold-standard diagnostic test for AHT, clinicians must rely on their expertise, their pathophysiologic and epidemiologic knowledge from the scientific literature, a forensic approach to injury assessment, and individual patient information in determining causation.[53, 551]

3.4.6 The dual process theory of clinical reasoning

The prevailing model of clinical reasoning is based upon the psychological dual-process theory of cognition, which distinguishes between two information processing systems: the intuitive, non-analytical System 1, and the rational, analytical System 2.[544, 591-593] System 1 processing is automatic; implicit; fast; unconscious; effortless; holistic; difficult to control; emotionally-charged; and context-bound. Conversely, System 2 processing is deliberate; explicit; slow; conscious; effortful; piecemeal; controlled; and potentially rule-governed. System 1 employs heuristics and other mental shortcuts, which represent an adaptive mechanism to facilitate our day-to-day decision-making,[594] as well as pattern recognition strategies described in section 3.4.3; System 1 therefore allows for rapid generation of diagnostic hypotheses. Emotional and affective responses may also be triggered by System 1. One such response is known as a "gut feeling", which has been defined as a specific type of intuition that something is wrong even in the absence of specific clinical indicators, or a sense of reassurance about a patient's condition and management despite the absence of a definitive diagnosis.[595] System 2 employs analytical strategies such as the hypothetico-deductive and probabilistic approaches outlined in sections 3.4.2 and 3.4.4 and is therefore useful for hypothesis generation when clinical patterns are not recognized, and for hypothesis refinement.[544, 592]

It is currently believed that the two systems do not operate discretely but are interactive and represented on a cognitive continuum.[596] In some circumstances, System 1 processes may be sufficient to generate and confirm a working diagnosis, and System 2 may not be activated. However, if the clinical presentation is not recognized or is ambiguous, or if there is considerable uncertainty involved, System 2 processes are required instead.[544] System 2 processes are preferred when the clinical stakes are high, when the situation is complex, and when clinicians have more time to consider a diagnosis.[592] It is likely that in most situations, optimal clinical reasoning requires a combination of both intuitive and analytical processes.[597-600] Indeed, Balla et al.[601] demonstrated that clinicians used their prior experience and clinical knowledge to generate initial hypotheses based on salient features of the clinical picture, and subsequently actively gathered and consciously weighted further information in order to test their hypotheses. System 2 is capable of monitoring and overriding System 1 by rejecting an incorrect initial hypothesis, however this ability is affected by variables such as clinician fatigue, distraction, or cognitive overload.[544, 594] System 1 can also override System 2, a phenomenon known as *dysrationalia*, akin to the state of acting against one's better judgment.[602] The dual process model of clinical reasoning is shown in Figure 3.5.

Figure 3.5 The dual process model of clinical reasoning (Croskerry, 2009)[544]



If a patient's clinical presentation is recognized, intuitive System 1 processes are engaged; if it is not, analytical System 2 processes are engaged instead. Repeated System 2 processing (i.e. clinical experience) leads to automatic System 1 processing. Both systems can override the other, and they may interact to generate the final working diagnosis. Figure reproduced from Croskerry et al., 2013[594]

3.4.7 Heuristics and biases

The heuristics and mental shortcuts which characterize System 1 processing are known to be vulnerable to a multitude of cognitive and affective biases.[594] Tversky and Kahneman[603] noted that "people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors" (p. 1124). The clinical literature generally cautions clinicians against intuitive reasoning to avoid errors resulting from cognitive biases.[560, 604, 605] However, cognitive biases have also been found to be associated with analytical reasoning.[598, 606, 607] Additionally, in some cases intuitive decision-making has been found to be superior to analytic deliberation,[608] perhaps because of the considerable cognitive resources required for analytical reasoning, such as the increased demands on working memory.[609] The majority of diagnostic errors are attributed to cognitive failings when processing and synthesising information, including at the hypothesis verification stage.[610] Diagnostic errors can involve inaccuracies in the assessment of the pre-test probability of disease, the strength of the evidence, or both, leading to inaccurate post-test probabilities of disease.[549]

Over 100 cognitive and around one dozen affective biases have been identified and described in the literature, arising from both System 1 and System 2 processing.[611, 612] Examples of biases associated with hypothesis generation, which is likely governed by intuitive responses, include availability bias and representative bias.[598] Availability bias refers to the tendency to judge things as being more likely if they easily come to mind, while representative bias refers to the tendency to judge a case based on its similarity to a diagnostic prototype.[549] Biases associated with hypothesis refinement include anchoring and adjustment bias, the tendency to lock onto salient features of the clinical presentation early on and failing to consider alternative diagnoses in light of new information, and premature closure, the tendency to stop considering other diagnoses or failing to adequately confirm a diagnosis once one is reached.[549, 610] A study of diagnostic error in medicine found the most common cognitive bias to be premature closure.[610]

Cognitive error may also be caused by biases that have become established through faulty or flawed probabilistic reasoning.[549, 613] A recent systematic review demonstrated that clinicians have difficulty understanding, interpreting and applying diagnostic information including commonly used measures of test accuracy.[581] Of 22 studies examining the ability of clinicians to estimate the post-test probability of disease by combining prevalence data with test accuracy data, only two studies found some evidence of success.[614, 615] Information with high diagnostic value is generally underestimated while information with low diagnostic value is generally overestimated.[616] As previously noted, diagnosis is inherently multivariable and diagnostic probabilities are estimated based upon combinations of multiple clinical features and test results, which to some extent may be correlated.[553-556] The diagnostic process involves updating existing information with new data and assessment of the relations among all of the clinical findings.[545] Unfortunately, applying Bayes' theorem when a number of correlated variables are involved in a diagnostic judgment is extremely complicated and impractical.[617] Accounting for possible correlation among variables and accurately estimating the relative contribution or independent diagnostic value of the multiple clinical features and test results for a diagnosis is a significant challenge for clinicians.[618, 619]

3.4.7.1 How susceptible is the diagnosis of abusive head trauma to heuristics and biases?

As the methodology employed in reaching a diagnosis of AHT is the same as that employed in reaching any other medical diagnosis, the diagnosis of AHT is not immune to heuristics and biases.[508] Many biases associated with the evaluation of AHT, such as racial and ethnic biases, are implicit, meaning that they operate outside of a clinician's conscious awareness and are therefore particularly difficult to overcome.[202, 620, 621] One recent study found that child abuse paediatricians without access to social intuition were around twice as likely to conduct a gold standard evaluation for neurotrauma compared with those paediatricians who had met the families.[622] Common cognitive errors that can occur during the forensic evaluation for child abuse, potentially leading to misdiagnosis, are presented in Table 3.4.[620, 621] Narang et al.[508] contend that one of the most important attributes that can alleviate the impact of biases and heuristics in the diagnosis of AHT is the utilization of EBM and statistical thinking. Similarly, other researchers advocate consistent application of evidence-based decision tools in order to best overcome the biases and practice disparities that have been demonstrated in the evaluation of AHT.[202, 622]

Table 3.4 Common cognitive errors that occur during the evaluation of suspected child abuse

Type of Cognitive Error	Definition	Example
Implicit stereotypes	Relying on generalizations to describe a family or caregiver	"Bad parent", "nice family"
Anchoring	Difficulty considering alternative diagnoses even if all the information does not fit as one would expect	New data is analysed in the context of whether it fits with the working diagnosis
Triage cueing	Sending a patient to a specialist based on a very specific symptom or finding	Referral to a haematologist to evaluate for a bleeding disorder based on bruising
Premature closure	Failing to appreciate that there is more to know before forming a view	Stopping a work-up for e.g. child abuse as one has already made up one's mind
Tunnel vision	Seeing an incident from a personal perspective or through a narrow lens	Failing to consider alternative explanations for a child's signs and symptoms
Contextual bias	The significance of the findings is dependent on the context it is thought to exist in	The investigation of a presenting symptom may vary depending on the specialty domain of the medical expert, e.g.[623]
Confirmation bias	The tendency to test a hypothesis by looking for instances that confirm it	Clinicians may be influenced by what is already considered by other professionals to be suspicious
Hindsight/outcome bias	Retrospective suspicion that was not present initially, in light of new findings	Has been found in radiology reporting
Groupthink	A strong compulsion within certain groups to reach unanimous decisions	Multidisciplinary team attempts to reach consensus at the expense of objectivity

Adapted from Laskey, 2014[620] with information from Skellern, 2015[621]

3.5 Clinical prediction rules as a strategy for improving clinical decision-making

Since most diagnostic errors are the result of flaws and biases in the way clinicians think,[610] efforts and strategies to minimise the cognitive shortcomings of clinicians have abounded in recent years.[560, 624-626] CPRs are one such strategy. CPRs are commonly used at the hypothesis refinement stage of diagnosis to quantify the probability of a target condition[569] (Figure 3.1). The use and application of CPRs is based on probabilistic or Bayesian diagnostic reasoning[627]: estimating a pre-test (prior) probability of the target condition, then applying a likelihood ratio derived from the presence or absence of the clinical features included in the CPR (similar to applying a diagnostic test result), which in turn enables the calculation of a post-test (posterior) probability[628] that is interpreted in relation to an individual clinician's test and treatment thresholds.[580] By explicitly combining and quantifying the contribution of aspects of the history, physical examination or diagnostic tests, CPRs can be readily applied to the diagnostic process.[292, 293]

3.5.1 What are clinical prediction rules?

As briefly outlined in Chapter 1, CPRs are algorithmic tools that estimate the probability of the presence of a clinical condition or the likelihood of an outcome by considering a small number of highly valid indicators.[629, 630] They include three or more predictors, from patients' clinical findings, history, or investigation results.[292] CPRs are most valuable when decision-making is challenging, when there is evidence that clinicians are failing to accurately diagnose a condition, and when there are serious consequences associated with an incorrect diagnosis,[293, 630] as is the case in suspected AHT. Their purpose is to assist clinicians in making decisions under conditions of uncertainty and enhance diagnostic, prognostic or therapeutic accuracy and decision-making, with the ultimate aim of improving the quality of patient care.[293, 629, 630] The predicted probabilities from a CPR allow clinicians to stratify patients into risk groups and help them to decide whether further assessment or treatment is necessary. Patients with a high probability of having a condition may be candidates for treatment or further investigations, while patients with a low probability of having a condition may not require further work-up.[631] Some CPRs can help to 'rule in' a condition by identifying patients who are very likely to have a condition and who thus require additional diagnostic testing or treatment, whilst others aim to 'rule out' a condition by identifying patients who are very unlikely to have a condition, thus reducing unnecessary testing without compromising patient care.[293, 630] CPRs that aim to predict the probability of a condition are termed *diagnostic* or *screening* rules.[630]

3.5.2 A note on terminology

Importantly, while the term "diagnostic prediction rule" implies that a definitive diagnosis can be made on the basis of the results, it is widely acknowledged that a CPR should never replace clinical judgment, but rather assist clinical reasoning by providing clinicians with more objective and evidence-based probability estimates to complement other relevant information.[296] Reilly and Evans[299] distinguish between *assistive* prediction rules that simply provide clinicians with predicted probabilities without recommending a specific clinical course of action, and *directive* decision rules that explicitly suggest additional diagnostic tests or treatment in line with the obtained score. Decision rules intend to directly influence clinician behaviour, while prediction rules intend to help clinicians predict risk without providing recommendations, with the assumption that accurate predictions will lead to better decisions. Some researchers also distinguish between prediction *models* that provide predicted probabilities along the continuum between certified impossibility (*Pi*=0) and absolute certainty (*Pi*=1),[632] and prediction *rules* that classify patients into risk groups by applying a clinically relevant cut-off that balances the likelihood of benefit with the likelihood of harm.[633, 634] Such cut-offs are known as decision thresholds; a threshold must be applied if a prediction model aims to influence decision-making.[633]

3.5.2.1 Terminology used in this thesis

Diagnostic/screening rules developed for the identification of AHT are the focus of this thesis. This includes those that estimate the probability of abuse (prediction models or rules), those that recommend a clinical course of action (decision rules), or those that do both. The rule developed and tested within this thesis can be described in its current form as a prediction rule rather than a decision rule, as it makes no recommendations to clinicians based on specific scores. It can also be defined as both a prediction model and a prediction rule, as it provides probability estimates on the continuum of risk, and its performance as a clinical rule is assessed using various thresholds. The term “clinical prediction tool” has been chosen to describe this rule, as it functions as a practical device akin to a clinical calculator. Specifically, the full name given to the tool is the “Predicting Abusive Head Trauma (PredAHT) tool”. For brevity this is simply referred to as “PredAHT”, throughout. In general the term “clinical prediction rule” will be used, however the term “model” may be used to describe the logistic regression coeffcients/formulae resulting from the statistical derivation of a rule. When describing or evaluating other specific rules, the terms that the authors of these rules have chosen to use themselves are adhered to.

3.5.3 Why clinical prediction rules may be more accurate than clinical judgment

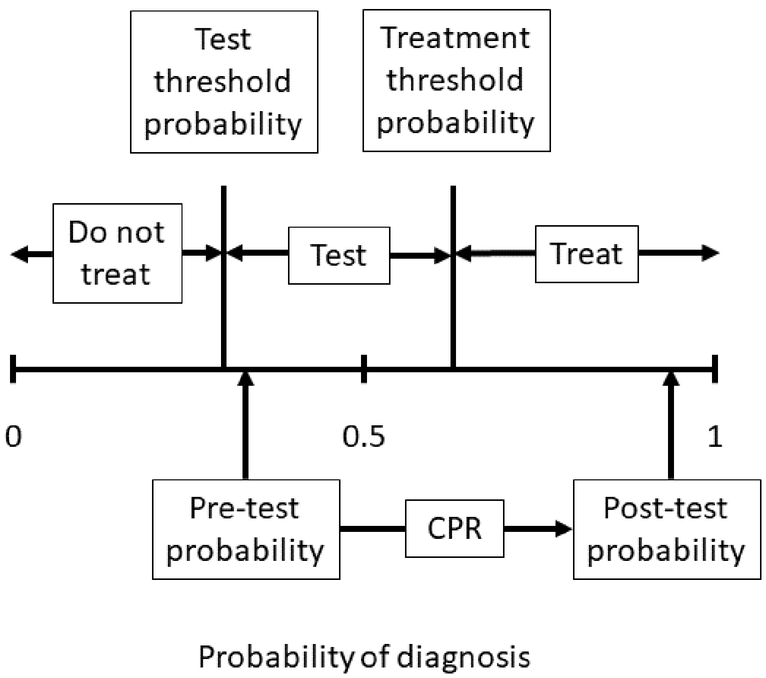
An underlying assumption is that the predictions of CPRs are objective, which helps to overcome the subjectivity inherent in unaided clinical judgment, and potential biases associated with intuitive decision-making.[293, 635, 636] Indeed, ‘statistical or actuarial’ methods of prediction, where clinical data are integrated based on “empirically established relations between data and the condition or event of interest”,[619(p.1668)] have been shown to be superior to clinical judgment in a number of comparative studies and reviews across a wide range of sectors including finance, education, psychology, medicine, and child welfare.[619, 637-641] However, an opposing body of work suggests that human judgment based on heuristics is on par with statistical models.[642] A recent systematic review of studies comparing CPRs with clinical judgment found that CPRs are rarely superior to clinical judgment, however the authors point out that to date there have been limited studies addressing this question and note the high or unclear risk of bias for the majority of the included studies in at least half of the risk of bias domains assessed.[643]

There are other reasons to believe that CPRs would be more accurate than clinical judgment. When CPRs are derived using multivariable regression modelling techniques, the joint effects of multiple variables from patients' clinical findings, history and investigation results are considered and variables with no independent predictive value are discarded.[619] The estimated regression coefficients of predictors included in the final model represent the independent contribution of each predictor to the outcome probability when mutually adjusted for one another.[294, 324, 619] Thus, unlike clinicians, CPRs can account for possible correlation between variables and estimate the independent predictive value of multiple pieces of clinical information when considered together. Furthermore, CPRs are more reliable than clinical judgment, because probabilities are calculated according to a fixed formula or algorithm, providing consistent results regardless of the expertise of the clinician. Conversely, since clinical judgment can be subjective there is likely to be variability between individual clinicians in predicting outcomes.[619]

3.5.4 Mechanisms by which clinical prediction rules may improve clinicians' decision-making

A CPR may conceivably improve clinical decision-making via a number of different mechanisms. A clinical decision rule that provides specific treatment, testing or management recommendations may change clinicians' decisions directly, by altering a decision they had already made, or by leading them to a different decision than they would have made without using the rule.[316, 644] Alternatively if the decision made by the clinician is the same as the one recommended by the rule, this may provide the clinician with additional confidence or reassurance in their decision-making.[627] A CPR that provides predicted probabilities of disease and/or stratifies patients into risk groups may alter clinicians' decisions via changes to the diagnostic accuracy of their clinical judgment (Figure 3.6). Within the context of the probabilistic and threshold approaches to clinical reasoning, the predicted probability provided by a CPR may 1) move a clinician's pre-test probability estimate above the treatment threshold probability, in which case a diagnosis can be made and treatment initiated, or in the context of AHT, a diagnosis of AHT suspected and a referral to social services made 2) move a clinician's pre-test probability estimate below the test threshold probability, in which case alternative diagnoses can be pursued, and AHT is effectively ruled out or 3) move a clinician's pre-test probability to within the intermediate range, in which case further diagnostic testing, such as ophthalmology, skeletal survey, or additional investigation, is required.[580, 585] Finally, the actual process of using the CPR may alter clinical decision-making. By collecting and inputting the data required to complete the CPR, a clinician may learn to focus on those clinical features that are truly predictive of the condition of interest and to discount other clinical variables that are less predictive.[618]

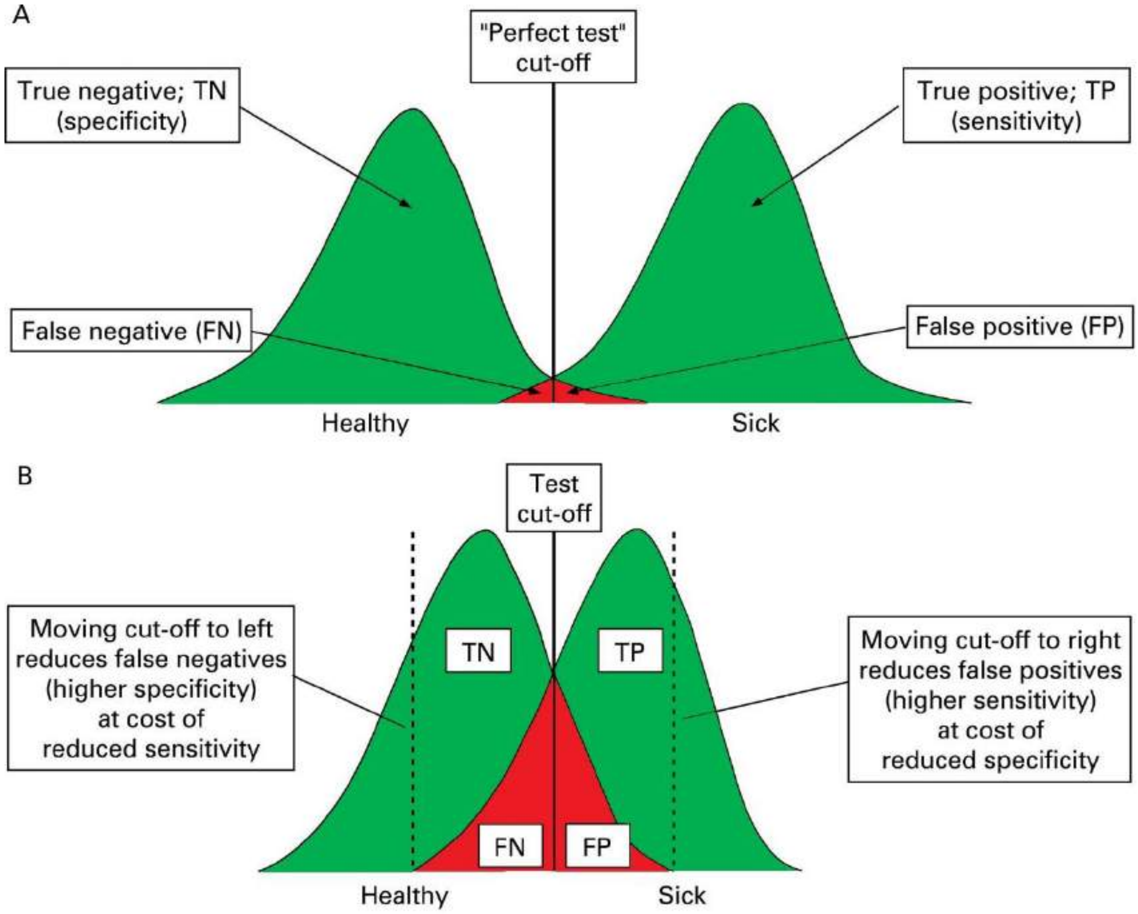
Figure 3.6 A clinical prediction rule raises the probability of a diagnosis past the treatment threshold probability



3.5.5 The ideal clinical prediction rule

The ideal CPR will demonstrate good concordance between the predicted probabilities calculated from the model and the observed patient outcomes, and will accurately discriminate between individuals with and without the condition of interest. As previously noted, if a CPR aims to influence decision-making, a cut-off threshold must be identified in order to classify patients as low or high risk; after this threshold is applied, sensitivity and specificity can be used to measure the performance of the CPR at different thresholds.[633, 634] Ideally a CPR will have both a high sensitivity and a high specificity, and therefore correctly identify the majority of patients who truly have the condition, as well as correctly exclude the majority of patients who truly do not have the condition (Figure 3.7). However, this scenario rarely occurs in clinical practice. More often than not the definition of a threshold is based on clinical considerations about the relative consequences of false positive and false negative classifications.[633, 634] Sensitivity and specificity are inversely proportional, so that as sensitivity increases, specificity decreases and vice versa.[645] Defining a high cut-off point will result in good specificity and few false positives but poor sensitivity and many false negatives. A test with a high specificity is useful for ruling in a disease if a person tests positive. This is because it rarely misdiagnoses those who do not have the condition of interest. Defining a low cut-off point will result in good sensitivity and few false negatives, but poor specificity and many false positives. A test with a high sensitivity is useful for ruling out disease if a person tests negative. This is because it rarely misdiagnoses those who have the condition of interest.[645] Receiver operating characteristic (ROC) curves display the sensitivity and specificity of a CPR across the full range of cut-off values and can be used to choose an optimal cut-off threshold.[646] Other approaches to determining clinical cut-offs have also been proposed.[647]

Figure 3.7 Distribution of test results in healthy and sick individuals and the trade-off between sensitivity and specificity

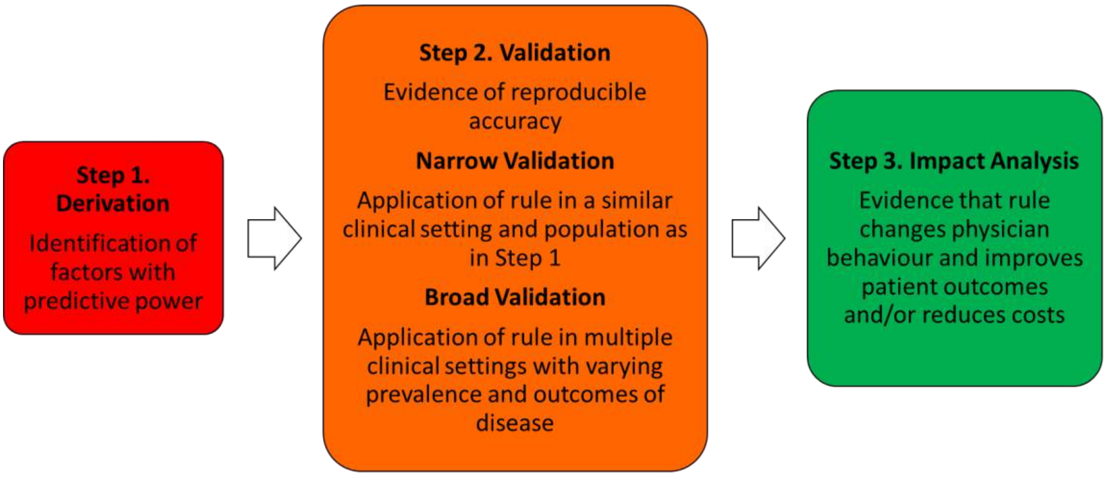


Panel A shows little overlap in the distribution of test results in healthy and sick individuals; this test discriminates almost perfectly between the two populations. Panel B shows the more likely situation where there is considerable overlap in the distribution of test results in the two groups; this test has reduced discriminatory ability. By altering the cut-off for a test, the sensitivity can be increased at the cost of decreased specificity and vice versa. Adapted with permission from BMJ Publishing Group Limited. [Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research, Søreide, 62, 2, 2009][646]

3.6 Stages in the development of clinical prediction rules

Methodological standards for the development of CPRs were originally outlined by Wasson and colleagues.[636] With the increase in popularity of CPRs inspired by the EBM movement, these standards have since been modified and updated by a number of authors over the years.[292, 293, 301, 633, 648] Guidelines for reporting clinical prediction research have also been developed.[326] As acknowledged in Chapter 1, there are three main stages in the development of CPRs (Figure 3.8); derivation; validation and refinement; and impact analysis to determine their impact on clinician behaviour and patient care.[293] Detailed methodological guidelines pertaining to each individual stage of development have been published, as each stage requires a different methodological approach.[294, 295, 300, 649] The three main stages of development correspond to increasing hierarchies of evidence (Table 3.5).[293, 299, 300] A CPR that has been derived but not externally validated corresponds to the lowest level of evidence and should not be used in clinical practice. If a CPR has been externally validated in a setting or population similar to the one from which it was derived, clinicians can use its predictions cautiously in similar future patients. If a CPR has been externally validated in multiple settings or populations, clinicians can use its predictions confidently in future patients. If a CPR has undergone an impact analysis and demonstrated improved patient care, it can be used as a decision rule for the management and treatment of patients. Ideally the impact of a CPR should also be tested in multiple settings; impact analysis studies correspond to the highest level of evidence.[299] Stiell and Wells[301] identified a further three important stages in the development of a CPR, namely identifying the need for a CPR, which was addressed in Chapters 1 and 2, determining the cost-effectiveness of a CPR, and long-term dissemination and implementation of a CPR. The following description is focused on methodological aspects related to the impact analysis of CPRs, as this is the stage that is relevant to the remainder of this thesis. The derivation and validation of PredAHT will be described and critically appraised in Chapter 4, alongside other CPRs for identifying AHT. The interested reader is referred to Appendix 3 for a comprehensive overview of the methodological stages involved in the development and evaluation of CPRs.

Figure 3.8 The three main stages in the development of clinical prediction rules



Reproduced from McGinn et al., 2000[650]

Table 3.5 Hierarchies of evidence in the development and evaluation of clinical prediction rules

Level of Evidence	Definitions and Standards of Evaluation	Implications for Clinicians
Level 1: Derivation of CPR	Identification of predictors using multivariable model; blinded assessment of outcomes.	Needs validation and further evaluation before it is used clinically in actual patient care.
Level 2: Narrow validation of CPR	Validation of CPR when tested prospectively in one setting; blinded assessment of outcomes.	Needs validation in varied settings; may use CPR cautiously in patients similar to derivation sample.
Level 3: Broad validation of CPR	Validation of CPR in varied settings with wide spectrum of patients and clinicians.	Needs impact analysis; may use CPR predictions with confidence in their accuracy.
Level 4: Narrow impact analysis of CPR used for decision-making	Prospective demonstration in one setting that use of CPR improves clinicians' decisions (quality or cost- effectiveness of patient care).	May use cautiously to inform decisions in settings similar to that studied.
Level 5: Broad impact analysis of CPR used for decision-making	Prospective demonstration in varied settings that use of CPR improves clinicians' decisions for wide spectrum of patients.	May use in varied settings with confidence that its use will benefit patient care quality or effectiveness.

CPR = clinical prediction rule. Adapted from Reilly and Evans, 2006[299]

3.6.1 Impact of a clinical prediction rule on clinical practice

Since the ultimate aim of a CPR is to improve the quality of patient care, the effect of a validated CPR on clinician behaviour and patient outcomes should be examined in what are known as impact analysis studies.[295, 296] Impact analysis studies clearly differ from validation studies as they must be comparative, typically requiring a control group of clinicians providing usual care.[295, 296, 299] It is possible to assess the impact of both assistive CPRs that simply provide predicted probabilities, and directive "decision" rules that suggest a specific course of action based on probability categories.[299] Assistive CPRs respect clinicians' individual judgment and leave room for intuition, whereas directive rules may be more likely to influence clinician behaviour.[299, 651, 652] However it is not guaranteed that clinicians will follow the logic of a CPR or the recommendations provided by directive rules.[299] Therefore an impact study must demonstrate that clinical behaviour can be altered and patient care improved by the CPR prior to widespread dissemination and implementation.[301] As described in Chapter 1, exploratory and preparatory work is essential before undertaking a formal experimental impact analysis study, to establish whether such a study is feasible, and to determine whether the CPR is acceptable to those professionals it is intended for.[300, 306]

3.6.1.1 Study design for an impact analysis

The optimal study design for an impact analysis is a cluster randomised trial with centres as clusters.[296] Randomising individual patients is not recommended as clinicians may learn the rule and apply it to patients randomised to the control group.[296] Randomising clinicians is preferable but requires more patients, and may lead to contamination of experience between clinicians in the same centre.[295, 653] An attractive variant of a cluster randomised trial is the stepped-wedge cluster randomised trial. In a stepped-wedge design, all centres apply care-as-usual, and then use the CPR at different, randomly allocated time periods.[654] This design allows for the comparison of outcomes both within and between hospitals, generates a wealth of data regarding potential barriers to implementation, and is particularly beneficial if the CPR turns out to have a promising effect.[655] When the outcome of interest in an impact study is clinician behaviour or decision-making, a cross-sectional randomised study without patient follow-up is sufficient, with randomisation at either the patient or clinician level. However, to determine the impact of a CPR on patient outcomes or cost-effectiveness, follow-up of patients is essential.[296] Given the significant practical, logistic and economic challenges associated with cluster randomised trials, non-randomised approaches are possible and are often used. A popular design is a before–after study, in which outcomes are assessed in a time period before a CPR is available and compared with outcomes measured in a time period after it is introduced; this design is susceptible to temporal confounding.[295] Finally, a relatively low-cost and simple design is a before–after study within the same clinicians. In this design, clinicians are asked to indicate their treatment or management decision or perceived risk of disease for the same patient both before and after receiving the CPR prediction.[295] This is the study design used in the feasibility study reported in Chapter 7.

3.6.1.2 Measures of impact of a clinical prediction rule

During an impact analysis study the sensitivity and specificity of the CPR should be recalculated to determine its accuracy in the new study population.[301] However, measures of CPR accuracy are not synonymous with measures of impact and only represent the *potential* impact of the CPR.[299] This is because clinicians are unlikely to follow the logic of the CPR or its recommendations in every case; they may not use the CPR at all, they may not use it correctly, they may deliberately disregard its predictions or suggestions, or they may be unable to use it for other reasons.[299] Measures of impact of a CPR include safety and efficiency. Safety is defined as the proportion of patients found to have the outcome of interest who received the appropriate intervention, while conversely efficiency is defined as the proportion of patients *not* found to have the outcome of interest who *did not* receive the intervention. The sensitivity and specificity of a CPR will only be the same as its safety and efficiency if clinicians follow the logic and recommendations of the CPR exactly.[299] Therefore, in an impact analysis study a CPR may demonstrate greater or less actual impact than its potential impact. The effect of clinicians' incorrect use of the CPR or their deviations from its logic or suggestions can provide important insights into its impact under specific circumstances and may reveal complex interactions between clinicians and the CPR.[299] For example, Reilly et al.[656] found that when clinicians did not consult a CPR for suspected acute cardiac ischemia at all or overruled its recommendations, their decisions were less efficient than if they had followed the CPR in every case. Although it was not possible to calculate the safety and efficiency of PredAHT following the feasibility study reported in Chapter 7, qualitative interview data enabled an in-depth exploration of the reasons why PredAHT did or did not influence clinicians' decision-making in specific cases.

3.6.1.3 Acceptability of the clinical prediction rule

If the use of a CPR is warranted but it is not used, the considerable time, money and effort that go into its development and evaluation is wasted. Assessing the acceptability of a CPR, its ease of use, and barriers to its uptake is therefore crucial for successful implementation.[296] Even valid and reliable CPRs may not be accepted or used by clinicians.[301] An in-depth qualitative study exploring the acceptability of PredAHT with a range of CP professionals is reported in Chapter 5.

3.6.1.3.1 Barriers and facilitators to the use of clinical prediction rules

Clearly, identifying the barriers and facilitators to the implementation of CPRs is crucial for the development of targeted implementation strategies that may encourage clinicians to use the CPR. The adoption of CPRs into clinical practice is influenced by a number of different factors including clinician characteristics, patient factors, features of the CPR itself and environmental factors.[299, 657-673] Table 3.6 provides an overview of the barriers to the adoption of CPRs identified in the literature, grouped according to their effect on clinician knowledge, attitudes or behaviours.[674] Barriers relating to knowledge include lack of awareness of the CPR or the burden of the clinical problem it applies to, unfamiliarity with the CPR, and a lack of understanding of the purpose of CPRs in general.[663, 671-673] Clinicians may also be unaware of a CPR due to the increasing volume of research studies about CPRs and the high number of CPRs that are sometimes developed for the same condition.[664, 675] Common barriers relating to clinician attitude include a conviction that clinical judgment is superior to the CPR and distrust of the accuracy of the CPR.[299, 657, 662, 663, 669, 671] Barriers relating to behaviour include organizational factors,[668] the complexity of the CPR and the time it takes to apply the CPR; survey studies suggest that clinicians much prefer a CPR that is simple to use and remember and saves time.[659, 660, 667] Other studies have demonstrated that clinicians will be unlikely to use a CPR if there are predictors missing which are deemed to be important, or if the predictor variables are not logically related to the outcome variable.[299, 673] Barriers and facilitators to the use of PredAHT are explored in the qualitative study reported in Chapter 5.

Table 3.6 Barriers to the use of clinical prediction rules in practice identified in the literature

Theme	Subtheme	Barrier
Knowledge	Awareness	Unaware: • That CPR exists • Of clinical problem or burden of clinical problem to which CPR applies Unable to choose from multiple CPRs
	Familiarity	Unfamiliar with CPR
	Understanding	Lack of knowledge and understanding of the purpose, development and application of CPRs in general
	Forgetting	Clinician forgets to use CPR despite best intentions
Attitudes	Negative beliefs about CPRs	Belief that: • CPRs threaten autonomy • CPRs are too 'cook-book', and oversimplify the clinical assessment process • Clinical judgment is superior to CPRs • Clinical judgment is not error prone • Use of CPRs causes intellectual laziness • The development of the CPR was biased • Patients will deem clinicians less capable if using a CPR • CPRs only apply to the less experienced • Probabilities are not helpful for decision-making Dislike of the term 'rule' Clinician had a false negative result when using a CPR in the past Existing CPRs are not ready for clinical application
	Outcome expectancy	Belief that: • CPRs will not lead to improved patient or process outcomes • The information provided by the CPR is not sufficient to alter clinical decisions Clinician: • Fears unintended consequences of use • Is uncertain about using the CPR in patients with an atypical presentation • Worries that improving efficiency threatens patient safety
	Self-efficacy	Belief that the CPR is too difficult to use Clinician uncertain how to interpret or use CPR output
	Motivation	Clinician lacks motivation to use the CPR
Behaviour	Patient factors	Patients expectations are not consistent with the CPR
	Features of the CPR	Clinician: • Finds CPR too complicated • Finds CPR 'too much trouble' to apply Perception that: • The CPR is not an efficient use of time

		• the CPR does not have face validity or that important predictors are missing • the CPR does not fit in with usual work flow or approach to decision-making • the CPR is not generalizable to the clinician's patient • the CPR is static and does not consider the dynamic nature of clinical practice • overruling the CPR is often justified Data required for the CPR is difficult to obtain
	Environmental factors	Lack of: • Time • Organisational support • Peer support for use Perceived increased risk of litigation Insufficient incentives or reimbursement for use of the CPR

Adapted from Haskins et al., 2014[669]

3.6.1.4 Comparison of a clinical prediction rule with unstructured clinical judgment

For a CPR to improve the diagnostic accuracy of clinicians, its performance in distinguishing between patients with and without the condition of interest should be superior to that of unstructured clinical judgment alone. Therefore a vital metric is the comparison of the accuracy of the CPR predicted probabilities of disease or recommended decisions with the accuracy of clinicians' own estimated disease probabilities or management decisions.[648] The sensitivity and specificity of clinicians' predictions or decisions are generally measured under usual practice and compared to the sensitivity and specificity of the CPR predictions or decisions when applied to the same patients.[676, 677] Some studies have used clinical vignettes[678] while others have used multivariable logistic models to assess the added value of a CPR over and above clinical judgment alone.[679] Understanding the performance of a CPR in contrast to clinician judgment may aid clinicians' acceptability of the CPR.[299] Although comparison of a CPR to clinician suspicion regularly takes place at the impact analysis stage, some researchers have recommended that this is carried out during the derivation or validation stages.[680] In addition, Finnerty et al.[681] recommend that comparison is undertaken in multiple settings, as the performance of a CPR may be superior to clinical judgment in certain settings but inferior or no different in other settings. A comparison of PredAHT with clinical judgment is reported in Chapter 7.

3.7 Tools for the identification of child physical abuse

Tools developed to assist in the identification of child physical abuse broadly fall into three categories: 1) generic screening tools based on checklists, flowcharts, educational interventions or multidisciplinary collaborative efforts, aimed at screening all children presenting with trauma or injury, usually to an ED 2) clinical prediction or decision rules, usually injury-specific and using statistical methods to predict the probability that abuse has occurred 3) actuarial risk assessment instruments for use in CP developed for predicting one or more types of maltreatment in the near future or the risk of recurrence of maltreatment. The discussion of actuarial risk assessment tools for the prediction of future or recurrent maltreatment is beyond the scope of this thesis; the interested reader is referred to a recent and comprehensive meta-analysis assessing the predictive validity of such instruments, which were found to outperform clinical instruments in predicting future maltreatment.[640]

Many generic screening tools for physical abuse or neglect have been developed over the years.[682-687] The majority are designed for use in the ED, however tools have also been developed for use in primary care.[688] Literature reviews and systematic reviews have demonstrated that such generic tools are ineffective at identifying maltreatment and that formal evaluation of their performance in clinical practice is lacking.[689-695] Some authors involved in the development of generic screening tools for the detection of physical abuse have concluded that injury-specific prediction rules are likely to be more promising in accurately identifying abuse.[696] Injury-specific CPRs have been developed for bruising,[697, 698] fractures,[387] and burns[699] as well as AHT.

3.7.1 Clinical prediction rules for identifying abusive head trauma

There are a number of points along the clinical pathway for infants and young children with head injury and suspected AHT where a CPR could be applied: to determine which children should undergo neuroimaging in the first instance,[700, 701] to assist in determining which children *do not* require further work-up and therefore to exclude abuse when the results of the CPR are negative,[351, 702] or to assist in determining the likelihood of abuse at different points in the clinical pathway based on specific clinical and/or radiological information.[59, 60, 467, 703] A CPR has also been used to predict outcomes in children hospitalized with confirmed AHT.[704] Four CPRs to assist in the identification of AHT have been derived and externally validated: the Pediatric Brain Injury Research Network (PediBIRN) four-variable CPR,[351, 702, 705] the Pittsburgh Infant Brain Injury Score (PIBIS),[701] the Biomarkers for Infant

Brain Injury Score (BIBIS),[700] and PredAHT.[59, 60] PredAHT is the main focus of this thesis. Therefore, its derivation, validation, and development into a computerised tool will be fully described in Chapter 4. A critical evaluation of the strengths and limitations of each of the four CPRs will also be presented in Chapter 4, in addition to a systematic review of CPRs for AHT conducted in collaboration with colleagues from Australia.[706]

3.8 Discussion

The review of clinical decision-making theories presented in the current chapter has demonstrated that clinicians use a range of strategies as part of the diagnostic reasoning process. The current prevailing model of clinical-decision making, dual process theory, suggests that pattern recognition, hypothetico-deductive, and probabilistic approaches are all used and that decision-making operates in two interactive information processing systems that are represented on a cognitive continuum.[544, 591-593, 596] When considering a diagnosis of AHT, clinicians must synthesize elements of their clinical experience and the scientific literature, and apply this knowledge to individual patients to establish injury plausibility.[53, 508, 621] Clinicians develop causal hypotheses for a child's injuries and utilize the differential diagnosis methodology to systematically eliminate and validate these hypotheses.[508] The inferential and deductive reasoning process is in some aspects Bayesian, and some aspects not; that is, when eliminating certain conditions from their list of differentials, clinicians do not, and indeed need not, always statistically quantify the singular or cumulative probabilities of various diagnostic features in the diagnostic process.[508] Narang et al.[508] emphasize that the use of some non-Bayesian approaches to the diagnosis of AHT does not imply unreliability, as such methods are rooted in an understanding of the pathophysiological processes of disease in children. However, they go on to state that "there are many aspects of the AHT/SBS differential diagnosis methodology that are conducive to Bayesian analysis" (p. 318). For example, clinicians can combine their knowledge of the discriminatory power of clinical and historical findings, diagnostic tests or CPRs from the published literature with the prevalence of AHT in their clinical setting, to estimate post-test probabilities of AHT.[508, 697, 707, 708] In practice however, clinicians are required to weight a complex composite of clinical variables which is hopelessly complicated in the absence of a CPR.[617, 619] In addition, even when a CPR is used, cognitive errors can occur when combining the CPR with a prior probability to produce a posterior probability.[577, 581, 616] Flaws in the application of both heuristics and probabilistic reasoning strongly raise the case for some type of decision support.[617] Indeed, experts in clinical decision-making have highlighted that "we need studies of conditional independence in clinical data, clinical prediction rules, and convenient methods (e.g. microcomputers or programmable calculators) to perform Bayesian calculations in clinical settings".[709(p.110)]

CPRs represent an ideal strategy for overcoming cognitive errors in clinical diagnostic reasoning when refining diagnostic hypotheses.[569] However, "ultimately the differential diagnosis methodology is a marriage of evidence-based literature and experience; a symbiosis of inferential and deductive reasoning; a synergy of linear and non-linear dynamic thought".[508(p.321)] It is important to bear in mind that the probabilistic approach is not the only approach to decision-making in suspected AHT, and therefore, it follows that some clinicians may not find a CPR useful in their decision-making. Research has highlighted a number of barriers to the use of CPRs in practice. In addition, due to the phenomenon of dysrationalia described in section 3.4.6, clinicians may ignore well-developed, validated CPRs even if they are known to outperform them.[602]

3.9 Conclusion and implications for this thesis

This chapter reviewed the literature on clinical decision-making theories relevant to the diagnosis of AHT, and the logic underlying the utility of CPRs for improving clinical decision-making and the identification of AHT. The findings from this review highlight that clinicians use a range of decision-making strategies when considering a diagnosis of AHT, including Bayesian reasoning, but nevertheless, diagnostic decision-making in AHT is susceptible to a multitude of cognitive and affective biases arising from both "intuitive" and "analytical" reasoning, emphasizing the need for decision support. In addition, aspects of the diagnostic process in suspected AHT cases are clearly conducive to Bayesian reasoning; taken together, this suggests that CPRs, which by nature are based on Bayesian reasoning, represent an ideal approach to aid clinicians in the identification of AHT. This scoping review was conducted in line with phase one of the MRC framework for the development and evaluation of complex interventions,[322, 323] as it has identified relevant theories to advance the theoretical understanding of the likely process of change that may be effected by introducing PredAHT into clinical practice. The clinical decision-making theories presented in the current chapter, and the information regarding the mechanisms by which CPRs may improve clinical diagnostic decision-making, will be used to guide the development of the computerised PredAHT reported in Chapter 4, and to help interpret the findings of qualitative interviews conducted with clinicians regarding the factors influencing their decision-making in suspected AHT cases, the acceptability of PredAHT, and the reasons why PredAHT did or did not influence their decision-making in specific cases, reported in Chapters 5 and 7.

4 Development of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool

Parts of this chapter have been presented in the following articles related to this thesis:

1. **Cowley LE**, Morris CB, Maguire SA, Farewell DM & Kemp AM (2015). Validation of a clinical prediction tool for abusive head trauma. *Pediatrics,* 136(2): 291-298.
2. Pfeiffer H, Crowe L, Kemp AM, **Cowley LE**, Smith AS, & Babl FE, on behalf of the Paediatric Research in Emergency Departments International Collaborative (PREDICT) (2018). Clinical prediction rules for abusive head trauma: A systematic review. *Archives of Disease in Childhood*, 103(8): 776-783.
3. Pfeiffer H, **Cowley LE**, Kemp AM, Dalziel SR, Smith A, Cheek JA, Borland ML, et al. (2019). Validation of the PredAHT prediction tool for abusive head trauma. Manuscript submitted to *Archives of Disease in Childhood*.
4. **Cowley LE**, Farewell DM, Maguire S, & Kemp AM (2019). Methodological standards for the development and evaluation of clinical prediction rules: A review of the literature. *BMC Diagnostic & Prognostic Research*, In Press.

4.1 Chapter overview

This chapter reports the systematic process used to create the computerised PredAHT clinical prediction tool, and a critical appraisal of PredAHT against other CPRs for the identification of AHT. The development of the computerised PredAHT adhered to the MRC framework for the development and evaluation of complex interventions, and methodological guidelines for the development of CPRs.[301, 322, 323] This chapter first describes the previous derivation of the PredAHT regression model following a systematic review of the clinical features associated with AHT and nAHT, and the subsequent external validation of PredAHT on novel data. Next, PredAHT is compared with three other CPRs developed to assist in the identification of AHT. A systematic review of validated CPRs for AHT is described, undertaken in collaboration with colleagues in Australia,[706] and the four CPRs for AHT are critically appraised. Then, the development of the computerised PredAHT is presented, informed in part by findings from Chapter 3. This was an iterative process, with the final version informed by the findings of the empirical studies reported in subsequent chapters. Finally, the computerised PredAHT is externally validated in an Australian/New Zealand population. The development of PredAHT is discussed in terms of methodological considerations, its validity and level of evidence, and its strengths and limitations.

4.2 Introduction

As discussed in Chapter 1, the MRC framework for the development and evaluation of complex interventions suggests that complex interventions should be developed and evaluated in an iterative manner, following identification of the evidence-base and relevant theories related to the intervention.[322, 323] Chapters 1 and 2 reviewed the literature on the identification of AHT, determining that identifying AHT is challenging for clinicians and that there is a need for a CPR in this field, to help prevent missed cases and help clinicians to overcome biases and practice disparities in their evaluation of children with ICI where AHT may be suspected. Chapter 3 reviewed clinical diagnostic decision-making theories, models and approaches relevant to the diagnosis of AHT, and reviewed the literature regarding the logic underpinning CPRs, in order to gain a theoretical understanding of clinical decision-making in suspected AHT cases and the mechanisms by which CPRs may improve clinical decision-making, and to identify potential features of a computerised tool that might best support clinicians in their decision-making. Another useful method of identifying the existing evidence-base for an intervention is to conduct a systematic review to identify similar interventions and what is already known about them.[322, 323] Therefore the current chapter describes a systematic review,[706] comparison, and critical appraisal of validated CPRs to assist in the identification of AHT.

In addition, as discussed in Chapter 3, guidelines for the development of CPRs recommend a systematic, multiphase process involving three main distinct stages: derivation, validation, and impact analysis.[293, 299, 301] PredAHT was derived following a systematic review to identify key candidate predictor variables,[24] and a pooled analysis of individual patient data,[59] and was externally validated on an independent dataset.[60] In order to facilitate the use of PredAHT in clinical practice at the bedside, and enable efficient evaluation of PredAHT in the empirical studies reported in subsequent chapters, a computerised version was created. This chapter aims to describe the previous derivation and validation of PredAHT, compare it with other CPRs for AHT, and describe its development into a computerised tool.

4.2.1 Aims of this chapter

1. To describe the previous derivation of the PredAHT regression model and its subsequent external validation
2. To describe a systematic review of validated CPRs for AHT conducted in collaboration with Australian colleagues, and present a critical appraisal of the four validated CPRs for AHT

3. To report the development of the computerised PredAHT, and its external validation in an Australian/New Zealand population

4.3 Derivation of the Predicting Abusive Head Trauma clinical prediction tool

The first stage in the development of a CPR is the derivation of the rule. This involves an examination of the ability of multiple potential variables from the clinical findings, history, or investigation results to predict the target outcome of interest.[292, 293] Predicted probabilities are derived from the statistical analysis of patients with known outcomes, and the outcome of interest serves as the reference standard by which the performance of the CPR is assessed. Often, relevant predictor variables are chosen based on a systematic review of the literature. Predictor variables for PredAHT were selected based on a systematic review of the clinical features associated with AHT and nAHT, described below.

4.3.1 Systematic review of the clinical features associated with abusive and non-abusive head trauma (Maguire et al., 2009)[24]

As detailed in Chapter 2, numerous studies have assessed the diagnostic value of individual features in distinguishing between children with AHT and nAHT. A systematic review of the scientific literature conducted by Maguire et al.[24] set out to identify the clinical features that are indicative of AHT and nAHT in children with ICI. Fourteen high quality comparative studies were included in the review. These studies included data on 1655 children, 779 of whom were victims of AHT. The items reported in each of the 14 primary studies allowed for an analysis of the following seven clinical features: apnoea, RH, rib fractures, seizures, long-bone fractures, skull fractures, and bruising to the head and/or neck. A multilevel logistic regression analysis derived ORs, their 97.5% confidence intervals and significance values for each feature in discriminating between AHT and nAHT, as well as PPVs and their 97.5% confidence intervals for each feature, that is, the predicted probability of AHT given that the feature is present in a child with ICI (Table 4.1). Apnoea and RH were statistically significantly associated with AHT and were the features most predictive of AHT. Rib fractures, seizures and long-bone fractures were more associated with AHT than nAHT, however these associations did not reach statistical significance. Skull fractures and head/neck bruising were more associated with nAHT than AHT although again not significantly so.

Table 4.1 Clinical features associated with abusive and non-abusive head trauma

Clinical Feature	PPV	(97.5% CI)	OR	(97.5% CI)	p value
Apnoea	93%	(73.3%–98.6%)	17.062	(5.018–58.011)	p<0.001
Retinal haemorrhage	71%	(48.3%–86.8%)	3.504	(1.088–11.280)	p=0.03
Rib fractures	73%	(5.00%–88.2%)	3.027	(0.716–12.799)	p=0.13
Seizures	66%	(45.4%–82.1%)	2.924	(0.731–11.694)	p=0.13
Long-bone fractures	59%	(48.0%–69.0%)	1.722	(0.824–3.601)	p=0.14
Skull fractures	44%	(22.3%–67.8%)	0.852	(0.316–2.301)	p>0.2
Head/neck bruising	37%	(3.50%–90.6%)	0.811	(0.070–9.410)	p>0.2

Data from Maguire et al. (2009).[24] Light grey features were significantly associated with AHT. Medium grey features were more associated with AHT than nAHT but not significantly so. Dark grey features were more associated with nAHT than AHT but not significantly so. PPV = positive predictive value, CI = confidence interval, OR = odds ratio

This systematic review showed that apnoea, RH and rib fractures have a high PPV and OR for AHT, and identified key features that should be recorded in suspected AHT cases. However, as described in Chapter 3, the diagnostic process in clinical practice is inherently multivariable, and the diagnostic work-up is inherently hierarchical, proceeding in a sequential fashion.[553-556] While knowledge of the individual clinical features associated with AHT and nAHT is useful, in reality children with head trauma may present with different combinations of multiple features which will become apparent to clinicians following appropriate clinical assessment and further investigation.[710] In a clinical scenario, the diagnosis of AHT is therefore based on the integration and interpretation of the combined features. Several single-centre studies have explored the association of combinations of features with AHT.[86, 711, 712] However, these studies analysed different features, and due to the rarity of AHT in any given population,[147] single-centre studies often do not have enough power for researchers or clinicians to make statistically conclusive or clinically meaningful inferences for multiple combinations of features. Therefore, Maguire et al.[59] derived PredAHT to assist in determining the probability of AHT in children less than three years of age with ICI, based on combinations of six clinical features.

4.3.2 Derivation study (Maguire et al., 2011)[59]

Maguire et al.[59] obtained anonymised individual patient data on 1053 children (348 of whom had AHT), from the authors of six comparative studies included in their earlier systematic review.[24] These data were used to conduct a pooled analysis and construct a multivariable logistic regression model which can be used to estimate the probability of AHT in children with ICI and different combinations of clinical features.[59] ICI was defined as "any combination of SDH, SAH, EDH, intraparenchymal injury, cerebral contusion, DAI, hypoxic ischemic injury and/or associated cerebral oedema" while children with isolated skull fracture but no ICI were excluded.[59(p.e551)] AHT was defined as ICI where abuse had been confirmed as the cause. Confirmed cases of AHT included only those ranked 1 or 2 for abuse according to Maguire and colleagues' "ranking of abuse" criteria (Table 4.2). This was to minimise circular reasoning, the risk that AHT was diagnosed based solely on the presence of the clinical features recorded.

Table 4.2 Quality standards for confirmation of abusive injury, only studies ranked 1 or 2 included

Ranking	Criteria used to define abuse
1	Abuse confirmed at case conference or civil, family, or criminal court proceedings, or admitted by perpetrator or independently witnessed
2	Abuse confirmed by stated criteria, including multidisciplinary assessment
3	Diagnosis of abuse defined by stated criteria
4	Abuse stated as occurring, but no supporting detail given as to how it was determined
5	Abuse stated simply as "suspected", no details on whether it was confirmed

4.3.2.1 Model development

4.3.2.1.1 Statistical analysis

Logistic regression models are one way to predict binary events such as the presence or absence of a condition or outcome, like AHT. Such models estimate regression coefficients (e.g. log odds ratios) of each predictor. Regression coefficients are mutually adjusted for the other predictors, and thus represent the contribution of each predictor to the probability of the outcome.[294] As PredAHT includes only binary categorical predictors, the probability of AHT is computed for a patient by multiplying the regression coefficients corresponding to the predictors by 0 or 1, depending on whether the predictors are absent (0) or present (1); the resulting values are then summed, along with the model intercept, to give a linear predictor value.[294, 636] Exponentiating this linear predictor value gives the odds, and so the probability (absolute risk) is calculated by use of the inverse logistic link function,[713] as defined by:

$$\exp(X\beta)\,/(1-\exp(X\beta)) \tag{1}$$

where $X\beta$ is the linear predictor value. As such, the probability of AHT can be estimated from any combination of the predictors.[326] In the PredAHT derivation study, multilevel logistic regression was used, to account for the different populations and prevalence's of AHT in each of the six primary studies. With six predictors included in the model, the number of possible combinations totals 64.

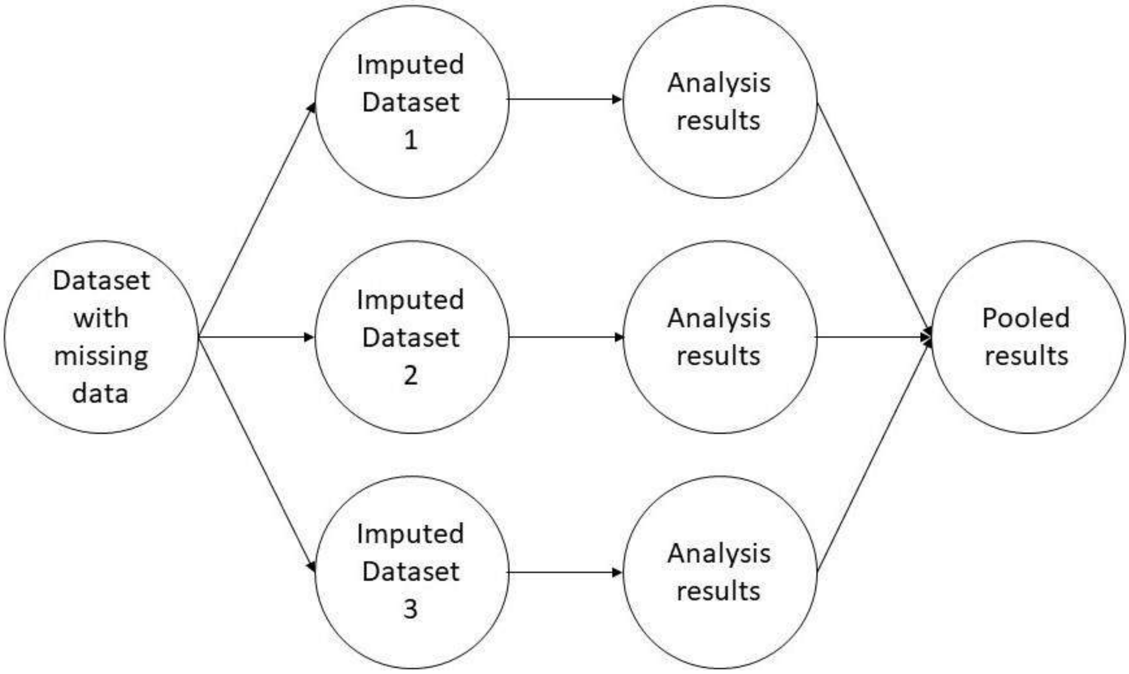
4.3.2.1.2 Missing data strategy

In clinical research, investigators almost always encounter missing observations, even in carefully designed studies and in spite of their best efforts to maximize data quality.[714] There are three types of missing-data mechanisms: 1) Missing completely at random (MCAR) 2) Missing at random (MAR), and 3) Missing not at random (MNAR).[715] When data are MCAR this means that there are no systematic differences between the missing and observed values; for example, laboratory tests may be missing because of a dropped test tube or broken equipment. When data are MAR this means that the probability of a missing value depends on the observed values of other variables (but not the unobserved values); for example, missing blood pressure measurements may be lower than observed measurements because younger people may be more likely to have missing measurements; in this case data can be said to be MAR given age.[716] When data are MNAR this means that the probability of a missing value depends on the unobserved values or other unobserved predictors; for example, people with high blood pressure may be more likely to miss a doctor's appointment due to headaches.[716] Missing values are rarely MCAR, that is, their "missingness" is usually directly or indirectly related to other subject or disease characteristics, including the outcome.[294, 297] Missing data is frequently addressed with case-wise deletion, which excludes all participants with missing values from the analysis.[716] However, when data are plausibly MAR, this reduces sample size and statistical power and biases the results,[716] leading to inaccurate estimates of predictor-outcome relationships and the predictive performance of the model, since the participants with complete data are not a random subsample of the original sample.[715, 717, 718]

Alternatively, missing values can be replaced with imputed values, based on the distribution of the observed data.[718] Multiple imputation is recommended over single imputation, as single imputation fails to account for uncertainty about the missing values and thus underestimates standard errors.[716, 718, 719] Multiple imputation quantifies the uncertainty in the imputed values by generating multiple different plausible imputed datasets, and pooling the results obtained from each of them.[716, 719] Multiple imputation involves three stages.[716, 719-722] First, multiple imputed datasets are created, based on the researcher's chosen imputation strategy. This is arguably the most challenging stage, as the researcher must model the relationship between the observed and missing components of the dataset.[723] Parametric models are often chosen to represent these relationships, such as a multivariate normal model, or an approach known as multiple imputation by chained equations (MICE), which formally models each missing variable through regressions on all other variables.[721, 722, 724] Alternatively, a non-parametric strategy can be used, such as hot-deck imputation,[715, 723, 725] which involves replacing missing values with observed values of similar participants.[726] This first stage accounts for uncertainty in estimating the missing values by adding variability into the values across the imputed datasets. In the second stage, standard statistical techniques are used to fit the models that are of interest in the substantive analysis to each of the imputed datasets. Estimated associations in each of the imputed datasets will be different, due to the variability introduced in stage 1. In the final stage, the multiple results are averaged together, and standard errors are calculated using Rubin's combination rules,[719] which account for both within- and between-imputation variability and the number of imputed datasets, and therefore the uncertainty of the imputed values (Figure 4.1).

In the derivation study, missing data were handled with a bespoke hot-deck multiple imputation strategy informed by inspection of the raw data, under the assumption that data were MAR.[59] Missing values were replaced with a random draw from children who were matched to the child with missing data with respect to their observed clinical features. Ten imputed datasets were generated for analysis. A multilevel logistic regression model was fitted to each of the imputed datasets, and the results were combined according to the established procedures described above. An attractive feature of hot-deck imputation is that since the imputed values are derived from observed responses in the dataset, unrealistic values cannot be imputed.[725, 726]

Figure 4.1 The three stages involved in multiple imputation, portraying three multiply imputed datasets



In stage 1, multiple imputed datasets are created by drawing plausible values from a distribution specifically modelled for each missing value. In stage 2 each imputed dataset is analysed using standard statistical techniques. In stage 3 the results are pooled, enabling calculation of a corrected standard error. Figure adapted from van Buuren & Groothuis-Oudshoorn (2011).[721]

4.3.2.1.3 Modelling strategy

Candidate predictors for inclusion in the multivariable model consisted of the seven clinical features identified in the systematic review,[24] listed in Table 4.1, as well as age and gender. During multivariable modelling, predictors were selected by one-step backward elimination; predictors that were nonsignificant at the 5% level, namely age, gender, and skull fractures, were dropped from the model. The regression coefficients of the final model and their associated 95% CIs are presented in Table 4.3. Figure 4.2 shows a graphical representation of the coefficients. Definitions of the six clinical features included in the final model are given in Table 4.4.

Table 4.3 Logistic regression coefficients of the final multivariable model

	Regression Coefficient	Lower 95% CI	Upper 95% CI	OR	Lower 95% CI	Upper 95% CI	p value
Intercept	-3.213	-4.358	-2.067	0.040	0.013	0.127	<0.001
Rib fractures	3.800	2.036	5.565	44.720	7.659	261.098	<0.001
Retinal haemorrhage	3.519	2.882	4.157	33.764	17.855	63.848	<0.001
Long bone fractures	2.621	1.235	4.006	13.747	3.440	54.937	<0.001
Apnoea	1.931	0.732	3.129	6.893	2.079	22.858	0.001
Seizures	1.624	0.695	2.553	5.072	2.003	12.843	0.001
Head/neck bruising	1.451	0.072	2.830	4.268	1.075	16.951	0.038

Data from Maguire et al. (2011).[59] *CI = confidence interval, OR = odds ratio*

Figure 4.2 Graphical representation of the logistic regression coefficients of the final multivariable model

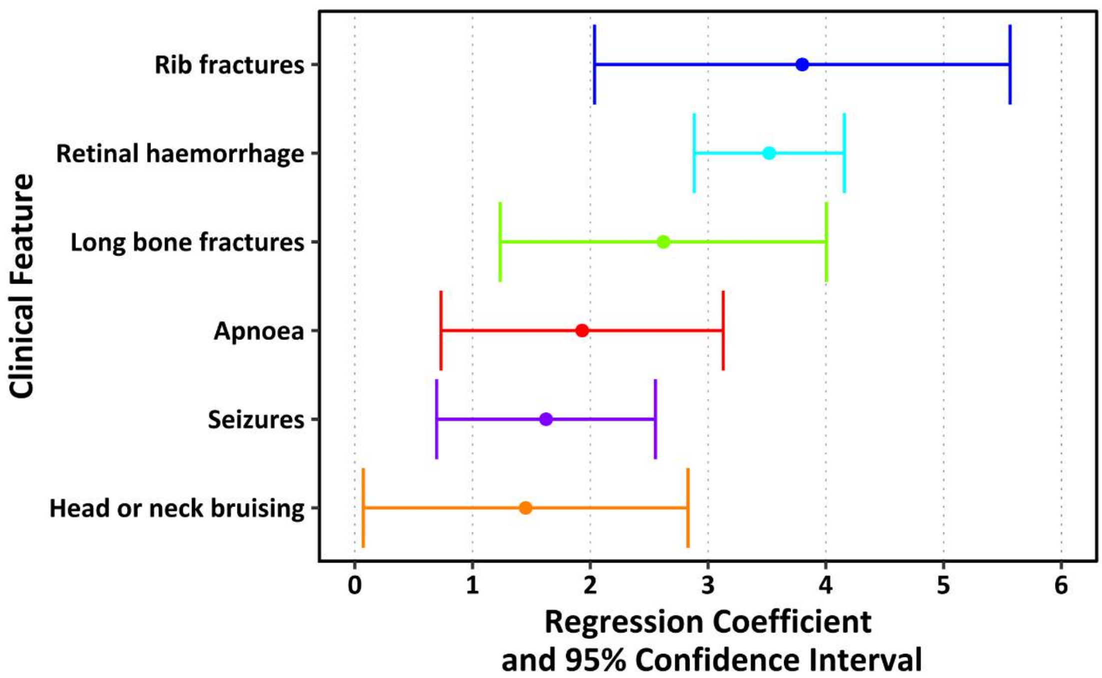


Table 4.4 The six clinical features included in the Predicting Abusive Head Trauma clinical prediction tool

Feature	Description
Head or neck bruising	Any documented bruising to head or neck
Seizures	Any documented seizures from a single seizure to status epilepticus
Apnoea	Any apnoea documented in the initial history or during inpatient stay
Rib fracture	Any rib fracture documented after appropriate radiologic imaging
Long-bone fracture	Any long-bone fracture documented after appropriate radiologic imaging
Retinal haemorrhage	Any retinal haemorrhage documented after indirect ophthalmologic examination by a paediatric ophthalmologist

4.3.2.2 Model predictions

In order to compute the probabilities for each of the 64 different combinations of features included in the model, firstly the linear predictor values (Xβ) were calculated, as described in section 4.3.2.1.1. The formula appropriate for the calculation of linear predictor values for this model is:

$$X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6,$$

where Xβ is the linear predictor value, β_0 is the intercept, X_i are the individual predictors and the β_i values are the model coefficients:

$$X\beta = \beta_0 + \text{Rib fractures}(1/0) + \text{Retinal haemorrhage}(1/0) + \text{Long bone fractures}(1/0) + \text{Apnoea}(1/0) + \text{Seizures}(1/0) + \text{Head/neck bruising}(1/0)$$

$$X\beta = -3.213 + 3.800(1/0) + 3.519(1/0) + 2.621(1/0) + 1.931(1/0) + 1.624(1/0) + 1.451(1/0)$$

Once the linear predictor values for each combination of features were computed, probabilities between 0 and 1 were calculated using the inverse logistic link function given in Equation (1), section 4.3.2.1.1. When there are no predictors present, the terms of the linear predictor consist solely of the intercept. The value for each of the predictors will be 0; when each of these is multiplied by its relevant coefficient the value of 0 is retained.[713] To illustrate, in a child with ICI but none of the other six features present, the calculation is:

$$-3.213 + 3.800 \times 0 + 3.519 \times 0 + 2.621 \times 0 + 1.931 \times 0 + 1.624 \times 0 + 1.451 \times 0 = -3.213$$

$$\exp(-3.213)/(1+\exp(-3.213)) = 0.04023572/1.040236 = 0.039$$

$$0.039 \times 100 = 3.9\%$$

Similarly, when there is only a single predictor present the terms of the linear predictor consist of the intercept plus one coefficient value, since the coefficient of the predictor that is present is multiplied by one. In a child with ICI plus rib fractures but with none of the other five features present, the calculation is:

$$-3.213 + 3.800 \times 1 + 3.519 \times 0 + 2.621 \times 0 + 1.931 \times 0 + 1.624 \times 0 + 1.451 \times 0 = 0.587$$

$$\exp(0.587)/(1+\exp(0.587)) = 1.798585/2.798585 = 0.642$$

$$0.642 \times 100 = 64.2\%$$

When more than one predictor is present the coefficients of the predictors that are present are simply added together with the intercept. In a child with ICI plus rib fractures and long-bone fractures but with none of the other four features present the calculation is:

$$-3.213 + 3.800 \times 1 + 3.519 \times 0 + 2.621 \times 1 + 1.931 \times 0 + 1.624 \times 0 + 1.451 \times 0 = 3.208$$

$$\exp(3.208)/(1+\exp(3.208)) = 24.72958/25.72958 = 0.961$$

$$0.961 \times 100 = 96.1\%$$

Exponentiating the regression coefficient for each individual predictor gives the odds ratio for the feature being present. Since all included predictors are main effects, the assumption is that the effects of all predictors are additive on the log odds scale, implying multiplicative effects on the original odds scale.[727] For example, the odds ratio for AHT in a child with head/neck bruising and apnoea is simply 4.268 x 6.893 = 29.4 (see Table 4.3 for odds ratios for individual features). The additivity assumption means that the effect of each predictor does not depend on the values of the other predictors.[326] In a child with ICI plus one or two of the six clinical features, the estimated probability of AHT varied depending on the possible combinations (Table 4.5) and (Table 4.6). In a child with ICI and any combination of three or more of the six clinical features, the estimated probability of AHT was always greater than 85%.[59] The estimated probabilities for the 64 different combinations of features are displayed in Figure 4.3. Odds ratios are displayed in Figure 4.4.

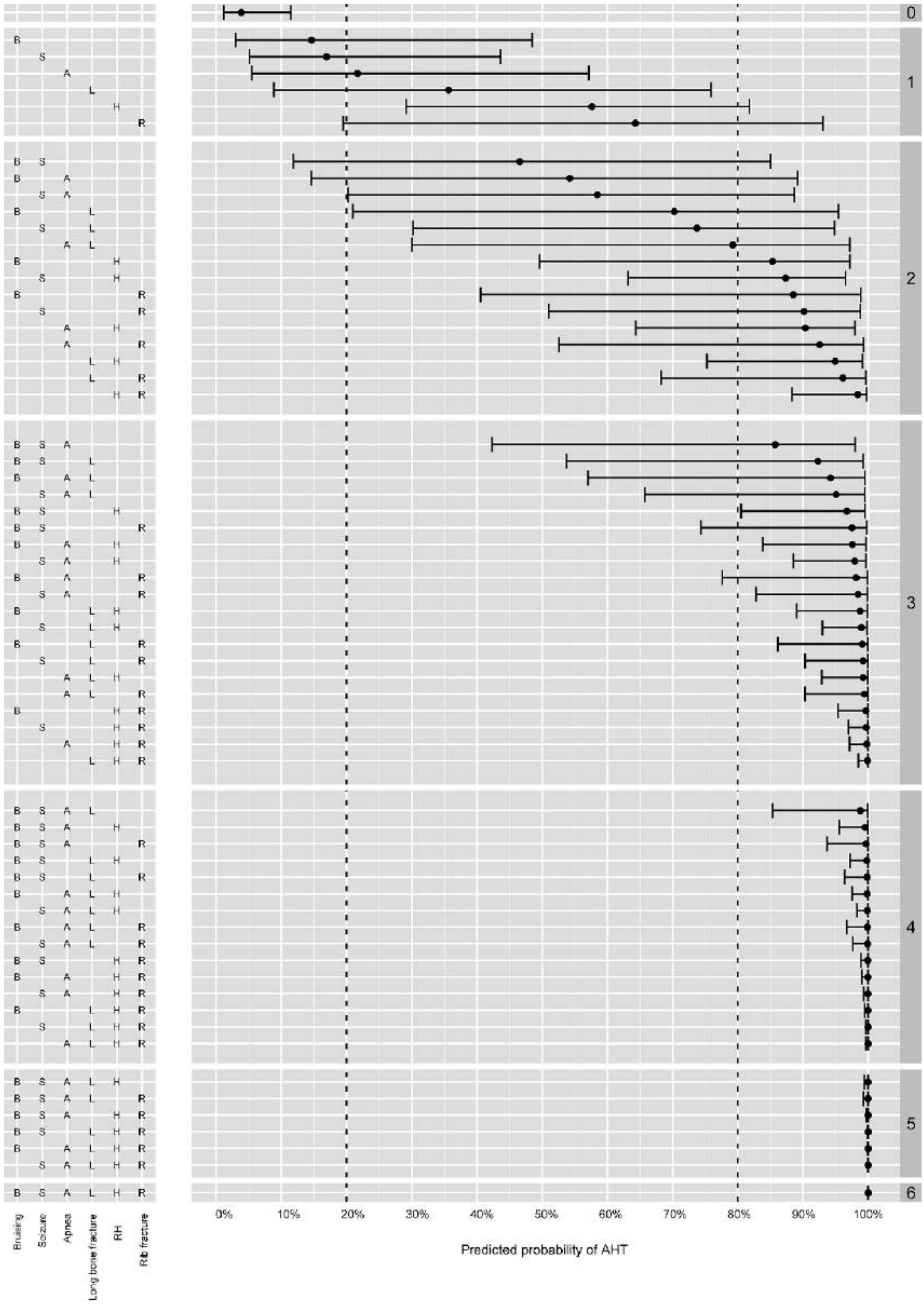
Table 4.5 Estimated probabilities of abusive head trauma in children less than three years of age with intracranial injury plus one clinical feature present

Feature	Probability of AHT
Head/neck bruising	14.7%
Seizure	17.0%
Apnoea	21.7%
Long bone fracture	35.6%
Retinal haemorrhage	57.6%
Rib fracture	64.3%

Table 4.6 Estimated probabilities of abusive head trauma in children less than three years of age with intracranial injury plus two clinical features present

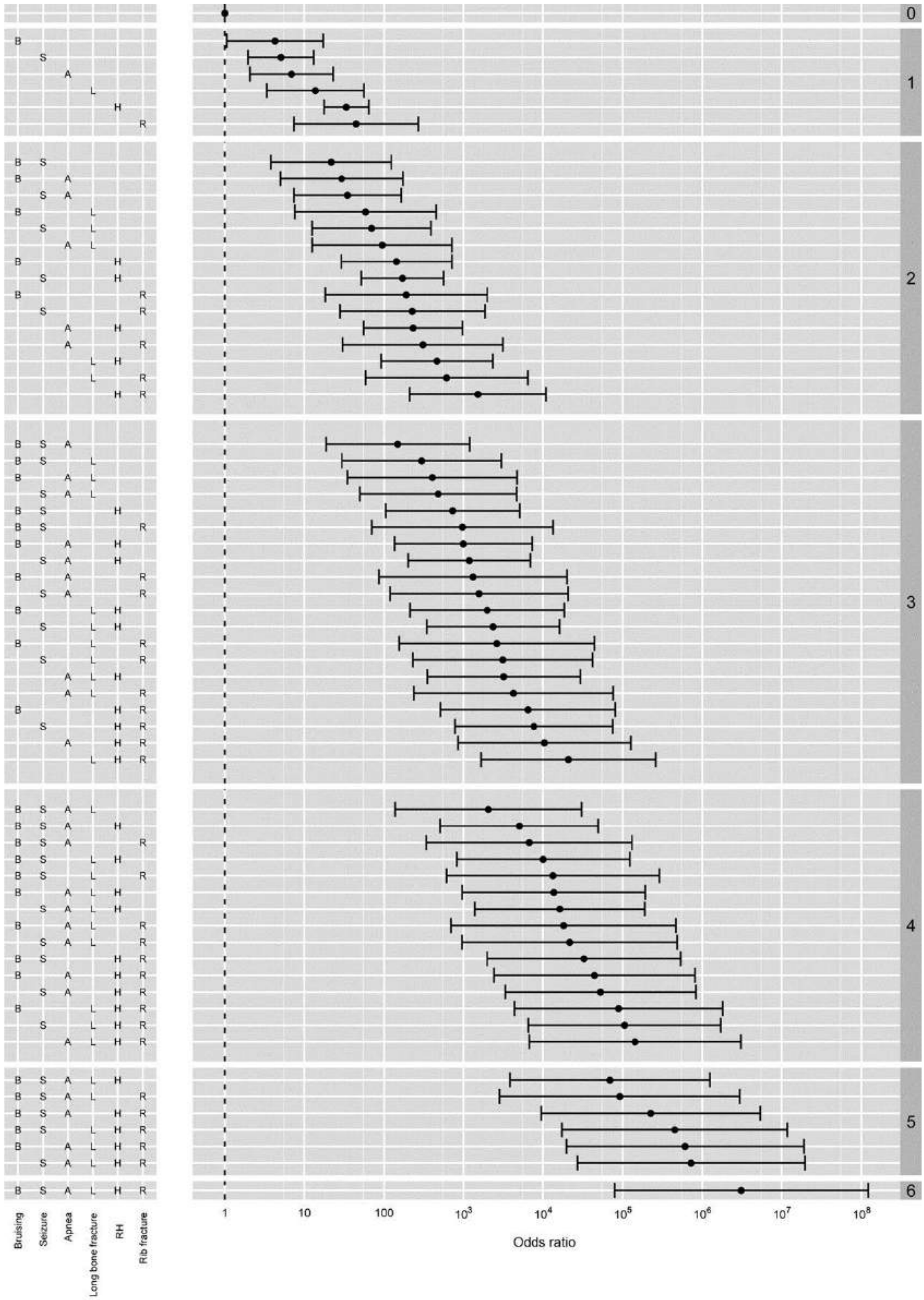
Features	Probability of AHT
Head/neck bruising & seizure	46.6%
Head/neck bruising & apnoea	54.2%
Seizure & apnoea	58.5%
Head/neck bruising & long bone fracture	70.2%
Seizure & long bone fracture	73.7%
Apnoea & long bone fracture	79.2%
Head/neck bruising & retinal haemorrhage	85.3%
Seizure & retinal haemorrhage	87.3%
Head/neck bruising & rib fracture	88.5%
Seizure & rib fracture	90.1%
Apnoea & retinal haemorrhage	90.4%
Apnoea & rib fracture	92.5%
Long bone fracture & retinal haemorrhage	94.9%
Long bone fracture & rib fracture	96.1%
Retinal haemorrhage & rib fracture	98.4%

Figure 4.3 Estimated probabilities of abusive head trauma in children less than three years of age with intracranial injury and combinations of six clinical features



Reproduced with permission from Pediatrics, Vol. 128, Page e556, Copyright © 2011 by the American Academy of Pediatrics

Figure 4.4 Odds ratios for abusive head trauma in children less than three years of age with intracranial injury and combinations of six clinical features

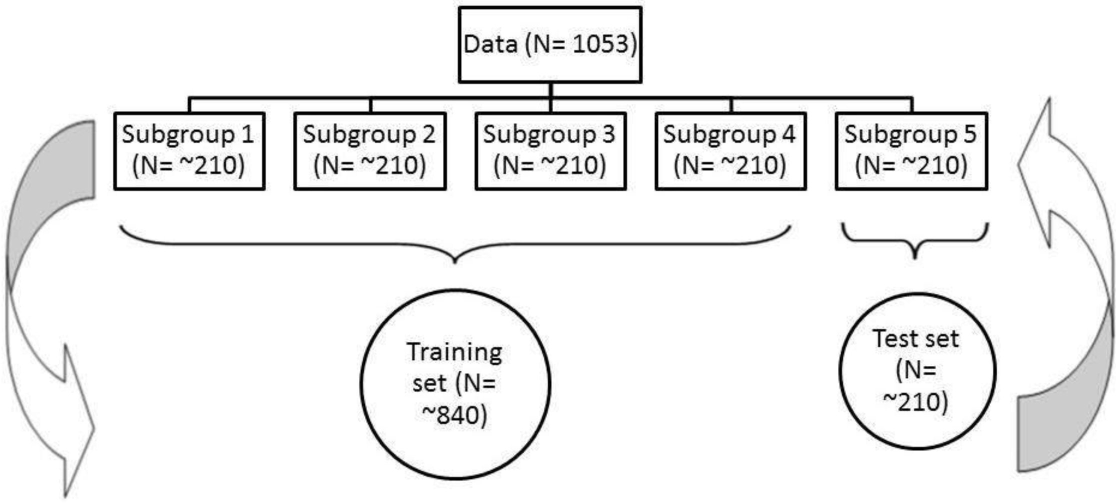


Reproduced with permission from Pediatrics, Vol. 128, Page e556, Copyright © 2011 by the American Academy of Pediatrics

4.3.2.3 Model accuracy

The accuracy of the model was assessed with five-fold cross-validation (Figure 4.5). The dataset was randomly divided into five equal subgroups. The model was then refitted using four of the five subgroups and its performance tested using the fifth; this process was repeated five times until each of the five subgroups was used as the test data, and an average of the estimated performance was taken. This meant that no child was used to both develop and test the model.[728] Predictions were considered to be correct if the probability of AHT or nAHT was greater than 80% and the predicted aetiology was true. Probabilities of less than 80% were considered as indeterminate. Given this, the model predictions were correct 80% of the time, indeterminate 15% of the time, and incorrect 5% of the time.[59]

Figure 4.5 Illustration of five-fold cross-validation



The dataset was randomly divided into five subgroups. Four of the subgroups were merged together to form a training dataset. The fifth subgroup was used as the testing dataset. This process was repeated five times.

4.4 External validation of the Predicting Abusive Head Trauma clinical prediction tool (Cowley et al., 2015)[60]

It is well known that CPRs perform better in the dataset from which they are derived compared to when they are applied to “plausibly related” individuals i.e. those who are suspected of having the same condition.[295, 324, 729] There is no guarantee that even well-developed CPRs will be generalisable to new individuals.[730] It is therefore essential to assess the performance of a CPR in individuals outside the derivation dataset; this process is known as external validation.[731] External validation involves taking the original fully specified model, with its predictors and regression coefficients as estimated from the derivation study; measuring and documenting the predictor and outcome variables in a new patient sample; applying the original model to these data to predict the outcome of interest; and quantifying the predictive performance of the model by comparing the predictions with the observed outcomes.[298]

4.4.1 Statistical analysis and missing data strategy

PredAHT was externally validated on a novel dataset consisting of cases ascertained from two centres that were included in the model derivation: Cardiff, Wales (UK) and Lille, France.[60] The sensitivity, specificity, PPV, and NPV of PredAHT were calculated. Discrimination was assessed with the area under the receiver operating characteristic curve (AUROC); discrimination reflects the ability of a CPR to discriminate between patients with, and without, the outcome of interest.[634, 732] The AUROC represents the chance that, given one patient with AHT and one without, the CPR will assign a higher predictive probability to the patient with AHT compared to the one without. An AUROC of 0.5 indicates predictions that are no better than random predictions, and a value of 1 represents perfect discrimination between patients with and without AHT.[733] As the presence or absence of the six clinical features was not always recorded, multiple imputation was required to estimate the probability that an unrecorded feature was present.[60] This was done using MICE,[721] otherwise known as fully conditional specification, or sequential regression multivariate imputation.[724] As described in section 4.3.2.1.2, this approach formally models each missing feature using regressions on all other variables,[721, 722, 724] and depends on the MAR assumption, i.e. that the probability of a missing feature depends on the observed values of the other features.

The MICE procedure is as follows[722]: given a dataset in which variables x_1, ….., x_k have missing values, all missing values are initially imputed by simple random sampling with replacement from the observed values. Then, the first variable to have one or more missing values (x_1) is regressed on all of the other variables (x_2, ….., x_k), limited to individuals with observed x_1. Missing values in x_1 are then imputed by simulated draws from the posterior predictive distribution of x_1. This process is repeated for all other variables with missing values x_2, …., x_k, which are regressed on all of the other variables, including the imputed values of the previous variables.[722] To stabilise the results, this cycle is repeated several times to produce one imputed dataset, and the whole process is repeated until the desired number of imputed datasets is created.[722] In the validation study,[60] missing features were imputed either as absent (0) or present (1), and ten imputed datasets were generated. The values from the ten imputed datasets were then averaged, to provide an estimated probability that an unrecorded feature was present. Finally, these estimates were combined with knowledge of observed features in the regression model to obtain a predicted probability of AHT for each child.[60] Specifically, when calculating the linear predictor value for each child's combination of features, when a feature was unrecorded, the regression coefficient corresponding to the unrecorded feature was multiplied by the estimated probability of the feature being present. Several alternative statistical imputation strategies were explored in the validation study, and MICE was found to be the best available approach.[60]

4.4.2 Model accuracy

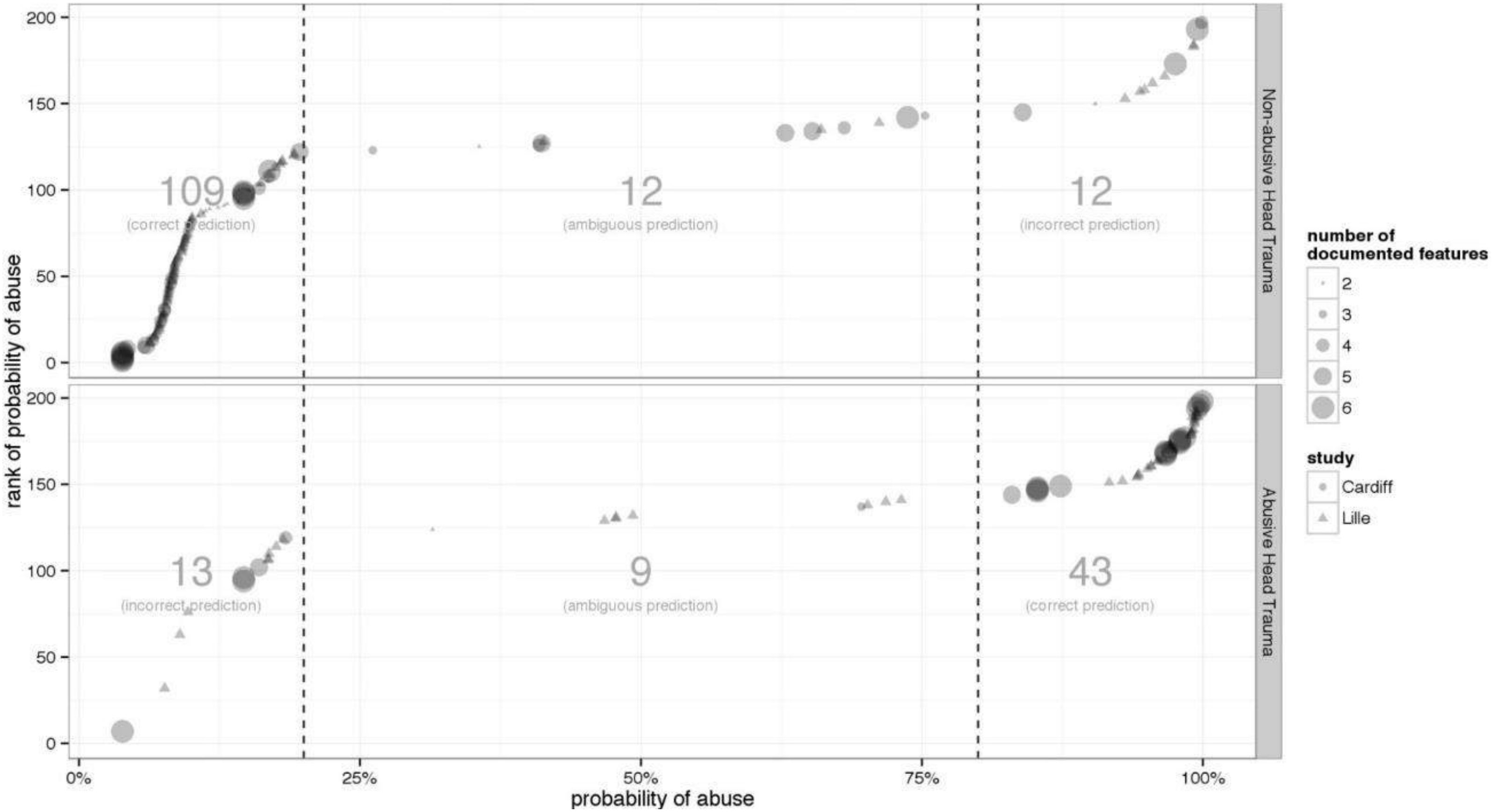
Figure 4.6 shows how PredAHT performed based on the predicted probabilities of AHT. Taking a probability cut-off of greater than 80% to define predicted AHT, and a probability cut-off of less than 20% to define predicted nAHT, PredAHT was correct for 82% (109/133) nAHT cases and 66% (43/65) AHT cases. Twelve accidental cases were predicted as AHT and 13 cases that were abused were predicted as nAHT. The size of the data points in Figure 4.6 reflect the number of documented features. Importantly, the cases with a predicted probability of between 20% and 80% had fewer recorded features, emphasizing the importance of recording the presence or absence of each of the key features included in PredAHT. Performance measures of PredAHT are reported in Table 4.7. Using a 50% probability cut-off, the sensitivity of PredAHT was 72.3% (95% CI 60.4%–81.7%) and the specificity was 85.7% (95% CI 78.8%–90.7%). The estimated probability of AHT when three or more of the six features were present was 81.5% (95% CI 63.3%–91.8%), consistent with the findings of the derivation study.[59] The AUROC was 0.88 (95% CI 0.823–0.926). The ROC curve (Figure 4.7) displays the sensitivity and specificity of PredAHT given different probability cut-off values. The ROC curves corresponding to the 95% confidence limits for the AUROC are also displayed.

Table 4.7 Performance of the Predicting Abusive Head Trauma clinical prediction tool in the validation dataset

Applying PredAHT	50% cut-off	
	Outcome	
	AHT	nAHT
Higher risk	47	19
Lower risk	18	114
	Value	**95% CI**
Sensitivity	72.3%	60.4%–81.7%
Specificity	85.7%	78.8%–90.7%
Positive predictive value	71.2%	59.4%–80.7%
Negative predictive value	86.3%	79.5%–91.2%

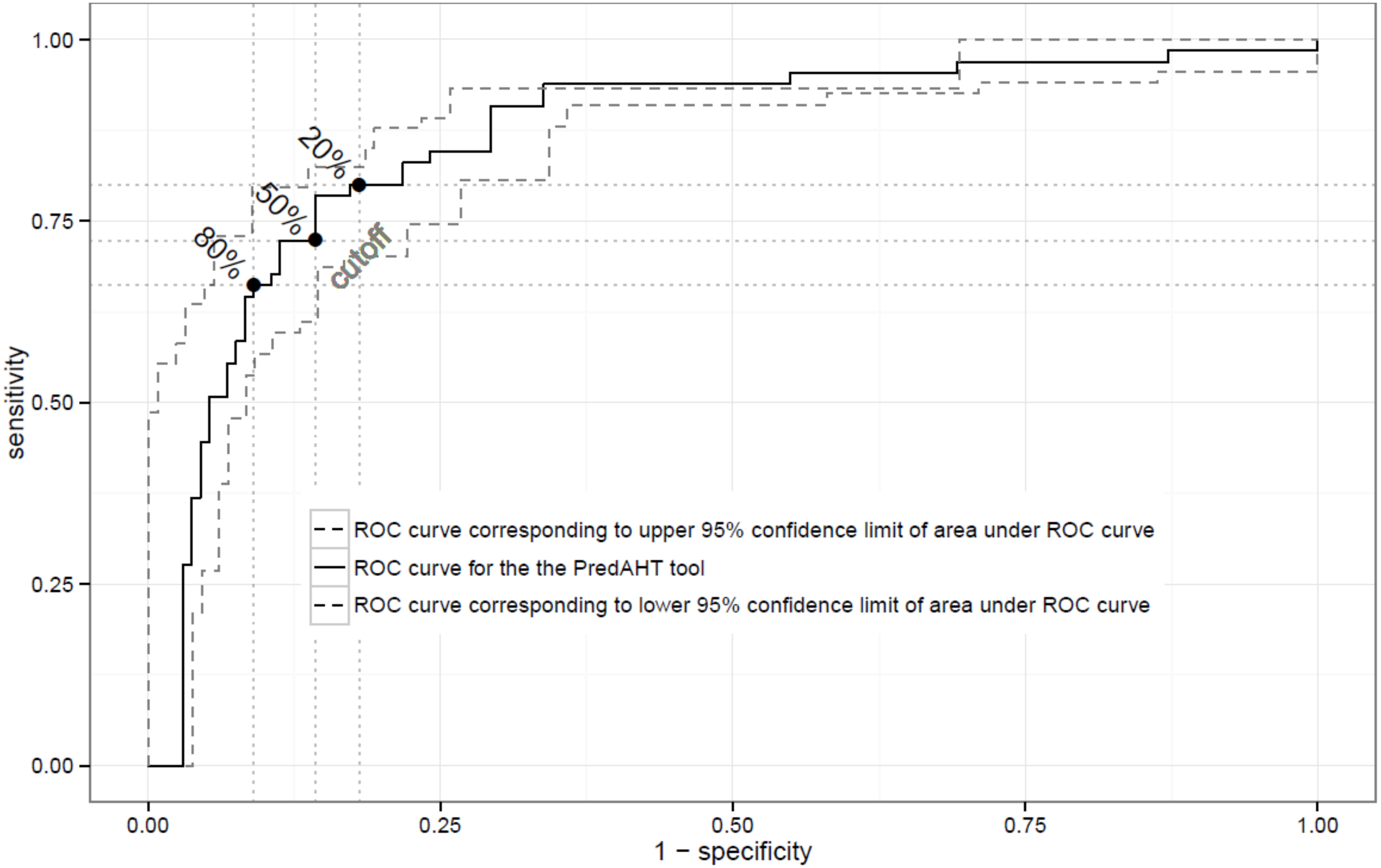
Data from Cowley et al. (2015)[60]

Figure 4.6 Predicted probability of AHT, by aetiology, study, and number of documented features



Circles represent data from the Cardiff site; triangles represent data from the Lille site. The size of the data points reflect the number of documented features. Predicted probabilities for the AHT cases are depicted in the lower half of the plot, predicted probabilities for the nAHT cases are depicted in the upper half of the plot. Reproduced with permission from Pediatrics, Vol. 136, Page 294, Copyright © 2015 by the American Academy of Pediatrics.

Figure 4.7 Receiver operating characteristic (ROC) curve for the Predicting Abusive Head Trauma clinical prediction tool



Reproduced with permission from Pediatrics, Vol. 136, Page 295, Copyright © 2015 by the American Academy of Pediatrics

4.5 Systematic review of validated clinical prediction rules for abusive head trauma (Pfeiffer et al., 2018)[706]

As discussed in Chapter 3, there are a number of points along the clinical pathway for infants and young children with head injury and suspected AHT where a CPR may be of use. Following the validation of PredAHT, the researcher and supervisory team were approached by colleagues in Australia and invited to collaborate on a formal systematic review of validated CPRs designed to assist in the identification of AHT. The search strategy, databases searched, methodological details and PRISMA flowchart are available in the published paper (Appendix 4).[706] Three validated CPRs for AHT were identified, critically appraised and compared in terms of their derivation, population, definition of AHT, variables used, external validation and performance: the PIBIS, the PediBIRN CPR, and PredAHT.[706] Following the publication of the systematic review, the BIBIS was derived and validated, and is thus included in the following discussion.

4.5.1 The Pittsburgh Infant Brain Injury Score

The PIBIS was derived in a retrospective study of 187 infants presenting to a tertiary care children's hospital with nonspecific symptoms, to determine which high-risk infants would benefit from neuroimaging for evaluation of brain injury.[701] Five predictor variables were identified in the derivation study, which was not published: age ≥3 months, head circumference percentile >90%, serum haemoglobin <11.2 g/dL, abnormality on neurologic or dermatologic examination, and a previous ED visit for a high-risk symptom.[701] The AUROC was 0.87 (95% CI 0.80–0.95).

A subsequent multicentre prospective study including 1040 infants was published and described as a validation and refinement of the PIBIS.[701] The analyses were conducted as part of a larger parent study deriving and validating a CPR using serum biomarkers to detect ICH,[700] described in the next section. Ascertainment criteria included well-appearing infants aged 30–364 days presenting to the ED with a temperature <38.3°C, no history of trauma and one or more symptoms associated with a high risk for AHT, including: ALTE/apnoea; vomiting without diarrhoea; seizures/seizure-like activity; soft tissue swelling of the scalp; bruising; or other non-specific neurologic symptoms such as lethargy, fussiness or poor feeding. The study outcomes were abnormal neuroimaging at enrolment or follow-up, and diagnosis of AHT as assessed by the hospital CP team, defined as a brain injury due to definite or probable abuse. Infants with no neuroimaging at enrolment or follow-up were classified as having no brain abnormalities. Missing data were handled with listwise deletion. The five original predictor variables were re-evaluated and the variables "abnormality on neurologic examination" and "previous ED visit for a high-risk symptom" were excluded from the CPR due to non-significance. The final four-variable CPR was generated using logistic regression and converted to a 5-point scoring system (Table 4.8). In infants with a score of 2 or more, the recommendation is that neuroimaging should be performed, while infants with a score of 0 or 1 can be safely discharged without neuroimaging. However the authors also present the sensitivity and specificity of the CPR at each score and suggest that consideration of the cut-off may depend on the imaging modality used. At a cut-off score of 2, the sensitivity of the CPR to detect abnormal neuroimaging was 93.0% (95% CI 89.0%–96.0%) and the specificity was 53.0% (95% CI 49.0%–57.0%). The positive predictive value (PPV) was 39.0% (95% CI 34.8%–43.6%) and the negative predictive value (NPV) was 96.0% (95% CI 93.6%–97.9%). The AUROC was 0.83 (95% CI 0.80–0.86). These measures were calculated for all brain abnormalities rather than for brain abnormalities due to AHT specifically.[701]

Table 4.8 The Pittsburgh Infant Brain Injury Score

Variable	Points
Abnormality on dermatologic examination	2
Age ≥ 3 months	1
Head circumference > 85^{th} percentile	1
Haemoglobin <11.2 g/dL	1

Reproduced with permission from Pediatrics, Vol. 138, Page 6, Copyright © 2016 by the American Academy of Pediatrics

4.5.2 The Biomarkers for Infant Brain Injury Score

Berger et al.[700] derived and validated the BIBIS to identify ICH in well-appearing infants at increased risk of AHT. Candidate predictors for inclusion in a multivariable logistic regression model were chosen prior to derivation based on the literature and the authors' experience, and included brain-specific and nonbrain-specific biomarkers as well as clinical variables. The BIBIS was derived in a retrospective study of sera samples from a serum databank, and comprised 99 well-appearing infants aged 30–364 days with a temperature <38.3°C, presenting to the Children's Hospital of Pittsburgh with nonspecific symptoms including vomiting or fussiness. Multiplex immunoassays for biomarker measurement were performed on the Ziplex System, which has potential as a point-of-care screening system. The technician measuring biomarkers was blinded to clinical data, while the CP team were blinded to biomarker data. Multiple models with different combinations of predictor variables were evaluated, and the model with the greatest AUROC was chosen. The median values of the patients in the derivation dataset with no ICH serve as the baseline for all other samples. The regression formula for the final model including three serum biomarkers and one clinical variable was presented (Box 4.1). BIBIS was internally validated using 20-fold cross-validation. At a cut-off of 0.182, sensitivity for predicting AHT was 95.8% (95% CI 94.4–97.0) and specificity was 54.9% (95% CI 50.9–58.9). The AUROC was 0.906 (95% CI 0.893–0.919).

BIBIS was then prospectively validated in the cohort of 1040 infants described in Berger et al.,[701] plus an additional 38 infants from the serum databank with rare intracranial abnormalities, who may present with similar symptoms to those patients with ICH owing to AHT. BIBIS was applied to only 561/1040 (54%) of the infants enrolled to the prospective study, mainly due to missing data for the predictor variables. At a cut-off of 0.182, the sensitivity of BIBIS for predicting AHT specifically was 86.4% (95% CI 84.1–88.7) and the specificity was 48.9% (95% CI 47.9–49.8). The sensitivity of BIBIS for predicting acute ICH of any aetiology was 89.34% (95% CI 87.7–90.4) and the specificity was 48.0% (95% CI 47.3–48.9). The PPV for any acute ICH was 21.3% and the NPV was 95.6%.[700]

Box 4.1 The Biomarkers for Infant Brain Injury Score

−2.442 + 0.000430 × [MMP-9 − Median MMP-9] +0.1058 × [Adjusted NSE − Median Adjusted NSE] −1.306 × [Haemoglobin − Median Haemoglobin] −0.004165 × [VCAM1 – Median VCAM1]

MMP-9 = matrix metallopeptidase-9, NSE = neuron-specific enolase, VCAM1 = vascular cellular adhesion molecule-1. MMP-9, NSE, and VCAM-1 are measured in nanograms per millilitre, and total serum haemoglobin is measured in grams per decilitre. Formula reproduced from Berger et al. (2017).[700]

4.5.3 The Pediatric Brain Injury Research Network clinical prediction rule

Hymel et al.[351] derived a CPR in a multicentre prospective study of 209 acutely head-injured children less than three years old admitted to ten PICUs, to “inform clinicians’ early decisions to launch (or forego) an evaluation for abuse” (p. 2). Candidate predictors for inclusion in the CPR were chosen following bivariate analysis of 45 historical, clinical and radiological variables, which identified 20 variables that both discriminated significantly between AHT and non-AHT and demonstrated high inter-rater reliability. The authors created a number of classification trees using binary recursive partitioning to derive four different CPRs with maximum sensitivity, to help clinicians exclude AHT if negative. AHT was defined based on a priori study criteria designed by the authors in an attempt to minimise circular reasoning. Classification of patients based on this definitional criteria was compared to the final diagnosis given by the treating clinicians at the point of hospital discharge. The CPR chosen for external validation had a sensitivity of 96%, a specificity of 36%, a PPV of 56%, an NPV of 91%, a positive LR of 1.50, and a negative LR of 0.12. Internal validation was not conducted. The CPR recommends that if one or more of four clinical or neuroradiological variables are present in a child without a history of a motor vehicle collision, the clinician should undertake a thorough abuse evaluation (Table 4.9).

PediBIRN was externally validated in a further prospective multicentre study of 291 children presenting to 14 PICUs, ten of which had participated in the derivation study.[702] The sensitivity of PediBIRN in the validation study was 96% (95% CI 90%–99%), specificity was 43% (95% CI, 35%–50%), PPV was 55% (95% CI 48%–62%), NPV was 93% (95% CI 85%–98%), the positive LR was 1.67 (95% CI 1.46–1.9) and the negative LR was 0.09 (95% CI 0.04–0.23). The AUROC was 0.78. The CPR demonstrated similar performance when defining AHT according to treating clinicians' final diagnoses instead of the a priori study definitional criteria. The PediBIRN investigators have also undertaken a retrospective, theoretical analysis of the potential impact of the CPR using the derivation and validation datasets, by comparing actual screening for AHT with screening guided by consistent and accurate application of the CPR.[705] Their analysis suggests that use of PediBIRN could theoretically increase AHT detection and the overall diagnostic yield of abuse evaluations while recommending marginally fewer children for abuse evaluation. In addition, a retrospective cost analysis using claims data and based on the theoretical analysis of the potential impact of the CPR suggests that applying PediBIRN accurately and consistently could lower the cost for each correctly identified child with AHT by 15.1% and reduce health systems costs arising from missed AHT by 72.4%.[734] An impact analysis of PediBIRN is currently being conducted in 8 PICU sites in the United States (http://www.pedibirn.com/).

Finally, PediBIRN has recently been externally validated in an Australian/New Zealand population in another collaborative study of which the researcher was a part of.[735] The study externally validated PediBIRN as designed (in the PICU population only), as well as using broader inclusion criteria (all children admitted to hospital with head injuries). The performance of the CPR in the broader sample comprising all head injured children admitted to hospital was similar to the performance of the CPR in the original validation study conducted by Hymel and colleagues.[702] This suggests that the CPR may have a role in screening for AHT in all children less than three years of age admitted to hospital with abnormal neuroimaging (excluding motor vehicle crashes or unintentional injuries and pre-existing abnormality on neuroimaging).[735] The sensitivity of PediBIRN in the PICU population was 100% (75%−100%) and the specificity was 11% (0%−48%). The sensitivity in the broader sample was 96% (82%−100%) and the specificity was 43% (32%−53%). This extends its level of evidence to a Level 3 rule with broad validation, whereby it is claimed by some that predictions can be used with confidence in their accuracy.[299]

Table 4.9. The Pediatric Brain Injury Research Network clinical prediction rule

To minimize missed cases, every acutely head-injured infant or young child <3 years of age hospitalized for intensive care (excluding motor vehicle collisions) who presents with ≥1 of these 4 predictor variables should be thoroughly evaluated for abuse:
• Any clinically significant respiratory compromise at the scene of injury, during transport, in the emergency department, or before admission • Any bruising involving the child's ears, neck, or torso • Any subdural haemorrhages or fluid collections that are bilateral or involve the interhemispheric space • Any skull fractures other than an isolated, unilateral, nondiastatic, linear, parietal skull fracture

4.5.4 Critical appraisal of validated clinical prediction rules for abusive head trauma

The four CPRs for AHT were critically appraised using a checklist of 17 items proposed by.[303] These items assess the methodological quality of the development and validation of CPRs for children. An overall score was calculated for each CPR according to the number of quality standards achieved (Table 4.10). The CPRs were also assessed using the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis checklist[326] (Table 4.11).

Table 4.10 Assessment of methodological quality of validated clinical prediction rules for abusive head trauma as proposed by Maguire et al. (2011)[303]

Quality Item	PediBIRN	PredAHT	PIBIS	BIBIS
Prospective Validation	Yes	Only DS 2	Yes	Yes
Study site well described	Yes	Yes	Yes	Yes
Population well described	Yes	Yes	Yes	Yes
Rule applied to all patients at risk	>90%	N.s.	No	No
Predictive variables				
Clear definition	Yes	Yes	Yes	Yes
Blind assessment	Yes	N.s.	N.s.	Yes
Reproducible	Yes	N.s.	N.s.	N.s
Outcome variable				
Definition	Yes	Yes	Yes	Yes
Blind assessment	Yes	N.s.	N.s.	Yes
Adequate follow-up	N.s.	N.s.	Yes	N.s
Sensibility				
Clinically sensible	Yes	Yes	Yes	Yes
Easy to use	Yes	Yes	Yes	No
Course of action	Yes	No	Yes	Yes
Statistical analysis				
Mathematical technique reported	Yes	Yes	Yes	Yes
Adequate calculated power reported	No	No	No	No
Adequate reporting of results	Yes	Yes	Yes	Yes
95% CIs reported on rule properties	Yes	Yes	Yes	Yes
Score	15	9	12	12

Present=score of 1; not specified/no=score of 0; DS = data set; NS = not specified; PediBIRN = Pediatric Brain Injury Research Network clinical prediction rule; PredAHT = Predicting Abusive Head Trauma clinical prediction tool PIBIS = Pittsburgh Infant Brain Injury Score; BIBIS = Biomarkers for Infant Brain Injury Score. Adapted from Pfeiffer et al., 2018.[706]

Table 4.11 Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis checklist: Prediction model development and validation

Section/Topic	Item		Checklist Item	PediBIRN	PredAHT	PIBIS	BIBIS
Title and abstract							
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	✓;✓however target population not stated in either	✗;✓however target population not stated	✓however target population not stated	✓however target population not stated
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results and conclusions.	✓other than sample size & predictors; ✓	✓;✓	✓other than objectives	✓
Introduction							
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	✓;✓but PredAHT derivation not discussed in either	✓;✓	✓but existing models not discussed	✓but existing models not discussed
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	✓;✓	✓;✓	✓	✓
Methods							
Source of data	4a	D;V	Describe the study design or source of data (e.g. randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	✓;✓	✓;✓	✓	✓
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	✓;✓	N/A; ✓	✓	✓

Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	✓;✓	✓;✓	✓	✓
	5b	D;V	Describe eligibility criteria for participants.	✓;✓	✓;✓	✓	✓
	5c	D;V	Give details of treatments received, if relevant.	N/A; N/A	N/A; N/A	N/A	N/A
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	✓;✓	✓;✓	✓	✓
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	✓;✓	✗;✗	✗	✓
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	✓;✓	✓;✓	✓	✓
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	✓;✓	✗;✗	✗	✓
Sample size	8	D;V	Explain how the study size was arrived at.	✗;✗	✗;✗	✗;✗	✗;✗
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	✓;✓	✓;✓	✓	✓
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	N/A	✓	✓only for refined rule	✓
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	✓however internal validation not conducted	✓	✓only for refined rule	✓
	10c	V	For validation, describe how the predictions were calculated.	✓	✓	✓	✓
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	✓;✓	✓;✓	✓	✓

	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N/A	N/A	Model updated *prior to* validation	N/A
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	✓;✓	✓;✓	✓	✓
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	✓	✓	✓	✓
Results							
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	✓;✓	✓;✓	✓	✓
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	✓;✓	✓;✓	✓	✓
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	✓	✗	✗	✗
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	✓	✓	✗ number of outcome events not specified for the 862 subjects with complete	✓

						data for refined rule	
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	✓	N/A	N/A	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	✓Classification tree presented	✓	✓However only refined model presented	✓
	15b	D	Explain how to the use the prediction model.	×	✓	✓	✓
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	✓;✓	✓;✓	✓	✓
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A	N/A	Model updated prior to validation	N/A
Discussion							
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	✓;✓	✓;✓	✓	✓
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	✓	✓	✓	✓
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	✓;✓Results from similar studies not discussed in either	✓;✓	✓Results from similar studies not discussed	✓
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	✓;✓	✓;✓	✓	✓

Other information							
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	N/A; ✓	N/A; ✓	N/A	N/A
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	✓;✓	✓;✓	✓	✓

Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. It is recommended that the TRIPOD Checklist be used in conjunction with the TRIPOD Explanation and Elaboration document (Moons et al., 2015).[326]

4.5.5 Comparing four validated clinical prediction rules for abusive head trauma

The four CPRs for AHT were compared in terms of their objectives, inclusion and exclusion criteria, definitions of AHT and nAHT, and performance, as summarised in Table 4.12. The CPRs are aimed at different populations and time points within the clinical assessment.[706] PIBIS and BIBIS are targeted at a population of well-appearing infants in the ED who might benefit from a head CT scan to identify intracranial abnormalities.[700, 701] PredAHT applies to children less than three years of age admitted to hospital with ICI, where children have been examined and may have had an ophthalmology exam and/or radiological investigations.[59, 60] PediBIRN as originally derived applies to a narrower population of children less than three years old admitted to the PICU with intracranial or cranial injury,[351, 702] although as previously noted, a recent external validation suggested that it may be applicable to a broader population of all children admitted to hospital with head injury.[735] Indeed, the high pre-test probability of AHT in the PICU setting raises the question whether all children in this high-risk group should be screened.[736] In addition, there is little overlap between the predictor variables included in each of the CPRs.[706] However, some of the features incorporated in the inclusion criteria for the BIBIS and PIBIS (apnoea, seizures and bruising) appear among the predictor variables of PediBIRN and PredAHT.[706] While the PIBIS, BIBIS and PediBIRN are directive decision rules that recommend a specific course of action (PIBIS and BIBIS – head CT scan; PediBIRN – thorough abuse evaluation), PredAHT is an assistive prediction rule that provides a probability of AHT, in order to “assist frontline professionals when deciding whether to refer a child for specialist clinical and multiagency investigation of possible AHT”,[59(p.e558)] “assist clinicians in their discussions with social welfare, law enforcement, or other professionals involved in the child protection process”,[60(p.291)] and “assist clinicians offering medical testimony in civil or criminal proceedings, in demonstrating why certain combinations of features are more or less predictive of an abusive etiology”.[59(p.e558)]

Due to the lack of gold standard diagnostic criteria for AHT, different approaches were taken to minimise circular reasoning, i.e. the possibility that AHT was diagnosed based on the presence of the predictor variables included in the CPRs.[706] Hymel et al.[351] devised a list of a priori study definitional criteria that excluded ICI and injury severity and attempted to exclude references to any of their PediBIRN predictor variables, although of note “bruising” appears in both the CPR and the definitional criteria.[737] The PIBIS, BIBIS and PredAHT used the multidisciplinary CP team assessment decision for confirmation of AHT.[59, 60, 700, 701]

In terms of the performance of the CPRs, PediBIRN, PIBIS and BIBIS all have very high sensitivities and comparatively low specificities, while PredAHT performs with a relatively high sensitivity of 72% and a higher specificity of 86%, applying a 50% probability cut-off.[60, 700-702, 735] However, PredAHT provides predicted probabilities of AHT along the full continuum of risk depending on the presence or absence of the six clinical features, and its sensitivity and specificity, as with the PIBIS and BIBIS, is contingent on the cut-off point applied.[59, 60] The PIBIS variables are available from physical examination, history and blood tests, while the BIBIS variables are available from blood tests but require analysis using the Ziplex system, which has potential as a point-of care test.[700, 701] Neuroimaging is required for PediBIRN, and further investigations such as ophthalmology exam and radiology investigations are desirable for PredAHT.[59, 60, 351, 702] This places PredAHT as a useful tool for assessing the significance of these investigations at a later stage in the diagnostic process.[706] However, the computerised version described later in this chapter can also account for missing information regarding the predictor variables, and could therefore contribute to decision-making at multiple stages in the assessment pathway e.g. by assisting clinicians in deciding whether to perform further investigations to obtain the missing information. The PediBIRN CPR received the highest score for methodological quality (Table 4.10) as it was the only CPR where the study investigators recorded that they assessed the predictor and outcome variables independently of one another and undertook an evaluation of the inter-rater reliability of the predictor variables.[706]

In summary, the four validated CPRs for AHT focus on different populations, and have different inclusion criteria. They include different predictor variables available at different stages in the clinical assessment pathway, and different outcome variables. None of the CPRs are diagnostic tools, rather, all aim to act as aids or prompts to clinicians to seek further clinical, social or forensic information within the context of a multidisciplinary CP team assessment.[706] The CPRs are discussed and appraised further in the discussion section of this chapter.

Table 4.12 Comparison of four validated clinical prediction rules for abusive head trauma

Name	**Pediatric Brain Injury Research Network (PediBIRN) clinical prediction rule**	**Predicting Abusive Head Trauma (PredAHT) clinical prediction tool**	**Pittsburgh Infant Brain Injury Score (PIBIS)**	**Biomarkers for Infant Brain Injury Score (BIBIS)**
Country	USA	UK	USA	USA
Derivation paper	Hymel et al (2013). Derivation of a clinical prediction rule for pediatric abusive head trauma. *Pediatric Critical Care Medicine*, 14(2): 210-220.	Maguire et al (2011). Estimating the probability of abusive head trauma: a pooled analysis. *Pediatrics*, 128(3):e550-e564.	Unpublished data	Berger et al (2017). Derivation and validation of a serum biomarker panel to identify infants with acute intracranial haemorrhage, *JAMA Pediatrics*, 171(6):e170429.
	Prospective: N = 209	Prospective: N = 133 Retrospective: N = 920	Retrospective: N = 187	Retrospective: N = 99
Validation paper	Hymel et al (2014). Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics*, 134(6): e1537-e1544.	Cowley et al (2015). Validation of a prediction tool for abusive head trauma. *Pediatrics*, 136(2):290-298.	Berger et al, 2016. Validation of the Pittsburgh infant brain injury score for abusive head trauma. *Pediatrics*, 138(1):pii:e20153756.	Berger et al (2017). Derivation and validation of a serum biomarker panel to identify infants with acute intracranial haemorrhage, *JAMA Pediatrics*, 171(6):e170429.
	Prospective: N = 291	Prospective: N = 138 Retrospective: N = 60	Prospective: N = 1040	Prospective: N = 476 for AHT Prospective: N = 511 for acute ICH of any aetiology
Clinical prediction rule	Every acutely head-injured infant/child meeting the inclusions criteria and presenting with **≥ 1 of these 4** predictor variables should be thoroughly evaluated for abuse	Estimated probability of AHT varies from 4% when none of the features are present to close to 100% when all six features are present and	Children with a score of **≥ 2** should undergo neuroimaging to check for abnormal findings	Children with a score of ≥ 0.182 on the mathematical model should undergo neuroimaging to look for

	• Any clinically significant **respiratory compromise** (infrequent/laboured respirations, apnoea or any need for intubation or assisted ventilation) at the scene of injury, during transport, in the ED or before admission • Any **bruising** involving the child's **ears, neck and torso** (including chest, abdomen, genitourinary region, back or buttocks) • Any **subdural haemorrhages** or fluid collections that are bilateral or involve the interhemispheric space • Any **skull fractures** other than an isolated, unilateral, nondiastatic, linear parietal skull fracture	>81.5% (63.3% - 91.8%) when **≥ 3 of these 6** features are present • **Head or neck bruising** • **Seizures** • **Apnoea** (documented in initial history or during inpatient stay) • **Rib fracture** (documented after appropriate radiologic imaging) • **Long-bone Fracture** (") • **Retinal Haemorrhage** (documented after indirect ophthalmologic examination by a paediatric ophthalmologist)	• **Abnormality on dermatologic examination** (2 points) • **Age ≥ 3.0 months** (1 point) • **Head circumference >85th percentile (**1 point) • **Haemoglobin <11.2 g/dL** (1 point)	acute intracranial haemorrhage −2.442 + 0.000430 × [MMP9 −Median MMP-9] + 0.1058 × [Adjusted NSE − Median Adjusted NSE] − 1.306 × [Hemoglobin − Median Hemoglobin] −0.004165 × [VCAM1 − MedianVCAM1]
Objective	Detection of AHT among acutely head-injured children admitted to PICU	Prediction of the likelihood of AHT in head-injured children	Detection of abnormal neuroimaging in well-appearing children with non-specific symptoms	Detection of acute intracranial haemorrhage in infants at increased risk of AHT
Inclusion	• Children < 3 years of age • Admission to PICU • Symptomatic, acute, closed, traumatic, cranial or intracranial injuries confirmed by CT or MRI	Dataset 1 (Cardiff, UK): • Children < 3 years of age • Hospital admission • ICI (combination of extraaxial haemorrhage, diffuse or focal parenchymal injury,	• 30 – 364 d of age • Well-appearing • Temperature <38,3°C • No history of trauma • Seeking medical evaluation for 1 of the following symptoms	• 30 – 364 d of age • Well-appearing • Temperature <38,3°C • No history of trauma • Seeking medical evaluation for 1 of the following symptoms

		cerebral oedema, cerebral contusion, hypoxic ischemic injury or diffuse axonal injury) confirmed on neuroimaging Dataset 2 (Lille, France): • Children < 2 years of age • Craniocerebral traumatic lesions diagnosed based on at least 1 CT [20] • Referred alive to the neurosurgical department, the PICU or the ED	○ ALTE/apnoea ○ Vomiting without diarrhoea ○ Seizures or seizure-like activity ○ Soft tissue swelling of the scalp ○ Bruising ○ Other non-specific neurologic symptom not described above, such as lethargy, fussiness or poor feeding	○ ALTE/apnoea ○ Vomiting without diarrhoea ○ Seizures or seizure-like activity ○ Soft tissue swelling of the scalp ○ Bruising ○ Other non-specific neurologic symptom not described above, such as lethargy, fussiness or poor feeding
Exclusion	• Children ≥ 3 years of age • Head injury resulting from a collision involving a motor vehicle • Initial neuroimaging revealed clear evidence of pre-existing brain malformation, disease, infection or hypoxia-ischemia	• Children ≥ 3 years of age (Dataset 2: ≥ 2 years of age) • Normal neuroimaging • Underlying structural abnormality or pre-existing disease (hydrocephalus, cystic lesion/tumour, metabolic cause,	• Previous abnormal CT scan of the head	• Samples with a hemocue of at least 0.5 g/dL • Patients without all 4 variables measured

		malformation, abnormal brain development) • Injuries caused by neglect • Birth injuries		
Definition of AHT	• The primary caregiver [25] admitted abusive acts • Abusive acts by the PC were witnessed by an unbiased, independent observer • The PC specifically denied that the preambulatory child in his/her care had experienced any head trauma • The PC provided an account of the child’s head injury event that was clearly historically inconsistent with repetition over time • The PC provided an account of the child’s head injury event that was clearly developmentally inconsistent with the child’s known (or expected) gross motor skills • Further workup confirmed the presence of two or more categories of extracranial injuries considered moderately or highly suspicious for abuse	Confirmed cases on AHT (ranked 1 or 2 for abuse) • Rank 1: > Abuse confirmed at case conference or civil, family, or criminal court proceedings > admitted by perpetrator > independently witnessed • Rank 2 > Abuse confirmed by stated criteria including multidisciplinary assessment	Brain injury due to definite/probable, but not possible, abuse as assessed by the hospital-based CP team at each enrolled site (Cases = abnormal neuroimaging)	Acute intracranial haemorrhage and the assessment of probable or definite abuse by each site’s hospital-based CP team

	> classic metaphyseal lesion fracture or epiphyseal separation > rib fracture, fracture of the scapula or sternum > fractures of the digits > vertebral body fractures > dislocation/fracture of spinous process > skin bruising/abrasion/ laceration in two or more distinct locations other than knee, shins or elbows > patterned bruising or dry contact burns > scalding burns with uniform depth, clear lines of demarcation, and paucity of splash marks > confirmed intra-abdominal injuries > retinoschisis confirmed by an ophthalmologist > retinal haemorrhages described by an ophthalmologist as dense, extensive, covering a large surface area and/or extending to the ora serrata			
Definition of nAHT	• All remaining patients	• Witnessed accidental mechanisms	• N.s.	• N.s.

		• Confirmed organic causes • Abuse excluded after CP investigations			
Validation study Sensitivity* Specificity Prevalence PPV NPV LR+ LR- Area under the curve	N = 291 **0.96** (0.90–0.99) 0.43 (0.35–0.50) **0.43** (0.37–0.49) 0.55 (0.48–0.62) 0.93 (0.85–0.98) 1.67 (1.46–1.91) 0.09 (0.04–0.23) 0.78 * Accuracy of detecting AHT cases among children with head injury admitted to PICU	N = 198 **0.72** (0.60–0.82) 0.86 (0.79–0.91) **0.33** (0.27–0.40) 0.71 (0.59–0.81) 0.86 (0.80–0.91) 5.06 (3.25–7.88) 0.32 (0.22–0.48) 0.88 (0.82–0.93) * Accuracy of detecting AHT cases among admitted children with head injury	N = 862 **0.93** (0.89–96.0) 0.53 (0.49–0.57) **0.26** 0.39 (0.35–0.44) 0.96 (0.94–0.98) 1.98 0.13 0.82 * Accuracy of detecting cases with abnormal neuroimaging in well-appearing children with at least 1 non-specific symptom, that is common in AHT	N = 476 **0.86** (0.84–0.89) 0.48 (0.48–0.50) * Accuracy of predicting AHT	N = 511 0.89 (0.88–0.90) 0.48 (0.47–0.49) 0.21 0.96 * Accuracy of predicting acute intracranial haemorrhage

AHT = abusive head trauma, ALTE = apparent life-threatening event, CP = child protection, ED = emergency department, ICI = intracranial injury, LR+ = positive likelihood ratio, LR− = negative likelihood ratio, nAHT = non-abusive head trauma, NPV = negative predictive value, NS = not specified, PC = primary caregiver, PICU = paediatric intensive care unit, PPV = positive predictive value. Adapted from Pfeiffer et al., 2018.[706]

4.6 Development of the computerised Predicting Abusive Head Trauma (PredAHT) tool

The PredAHT derivation study described in section 4.3.2 provided predicted probabilities (Figure 4.3) and odds ratios (Figure 4.4) of AHT for children less than three years old with ICI and each of 64 possible combinations of the presence or absence of six clinical features.[59] However, as described in Chapter 2, due to the variability exhibited by clinicians in the extent of their evaluation and investigation of AHT, in the clinical setting some investigations may not be undertaken to identify key clinical features e.g. X rays or ophthalmology for fractures or RH, respectively, and at different time points along the clinical assessment pathway, clinicians may not have access to or knowledge of all six clinical features included in PredAHT. In addition, at some points in time children with head injury may be too ill to undergo the necessary investigations. In the event of an unknown clinical feature, the model's probability of AHT for a particular patient cannot be estimated at the time of decision-making, unless the unknown feature is presumed to be absent, and thus a strategy is needed to help clinicians applying a CPR to deal with such missing values.[315, 738, 739] Therefore, the probability of AHT was estimated when one or more of the six clinical features was unknown (see section 4.6.2, below). This increases the number of possible combinations of clinical features to 729.

A key consideration when developing CPRs is the format in which to present the predictions.[294, 324] A CPR is often presented using a simplified scoring system to facilitate use, a paper-based nomogram or graphical decision tree,[294, 324, 733, 740] or a figure such as the one presented in Figure 4.3. Clearly, a CPR that provides predicted probabilities for 729 permutations of variables is too complex to be presented in paper-based format or as a simplified scoring system. Alternatively, a CPR can be developed into a simple web-based calculator or application, which is an ideal format for complex CPRs such as PredAHT.[324] Therefore, the decision was made to develop PredAHT into a simple web-based calculator, with the aim of facilitating its adoption into clinical practice.[741]

Chapter 3 reviewed the literature on clinical decision-making theories relevant to the diagnosis of AHT and the logic underpinning CPRs, which enabled the identification of potential features of a computerised tool that might best support clinicians in their decision-making. The literature review revealed that clinicians use probabilistic reasoning as one decision-making strategy when refining clinical hypotheses, whereby an initial opinion (prior probability) is updated with new information (clinical data) by applying a likelihood ratio, to produce a posterior probability of an outcome.[569, 580] In addition, the literature review identified that by explicitly quantifying the discriminatory power of combinations of clinical features, CPRs are conducive to probabilistic diagnosis and can be readily applied to the probabilistic reasoning process.[569, 627] Therefore, likelihood ratios of AHT were calculated for each of the 729 combinations of clinical features, to enable clinicians to incorporate their own prior probabilities of AHT, based on factors unaccounted for by PredAHT, e.g. purported history, clinical presentation or psychosocial features. In addition, given the finding from Chapter 3 that clinicians may have difficulty combining their prior probability with a CPR to produce an accurate posterior probability,[577, 581, 616] a sliding scale was incorporated into the computerised PredAHT, which allows clinicians to enter their own prior probability, and facilitates automatic calculation of the posterior probability.

4.6.1 Programming

The computerised tool was built using Shiny, a Web application framework for the R language and environment for statistical computing.[742, 743] R was used with the RStudio Integrated Development Environment.[744] The Shiny package for R allows researchers to interactively show the output for R programs to Web-browsers.[745] Contrary to other Web-page design methods, researchers only require previous experience with the R programming language in order to build a Shiny application. Due to the flexibility of the R language and the many extension packages available, researchers have complete control in coding all aspects of a model, the appearance of the application's user interface, and the output produced.[745] Output is reactively updated in response to changing input by using widgets. Shiny widgets are interactive elements such as sliders, radio buttons or drop-down boxes that enable users to enter different values or categories of parameters or variables. When input from a widget changes, the generated output is updated to reflect the change.[745, 746]

Shiny applications are built using two R scripts that communicate with one other: a user-interface script (ui.R), and a server script (server.R). The ui.R script controls the application's layout and appearance, and the server.R script controls the processing of user-input, creating the reactive output that is sent to the user interface for display.[745, 746] Shiny applications can be hosted on Web servers and accessed via the Internet without an R installation or files containing R code.

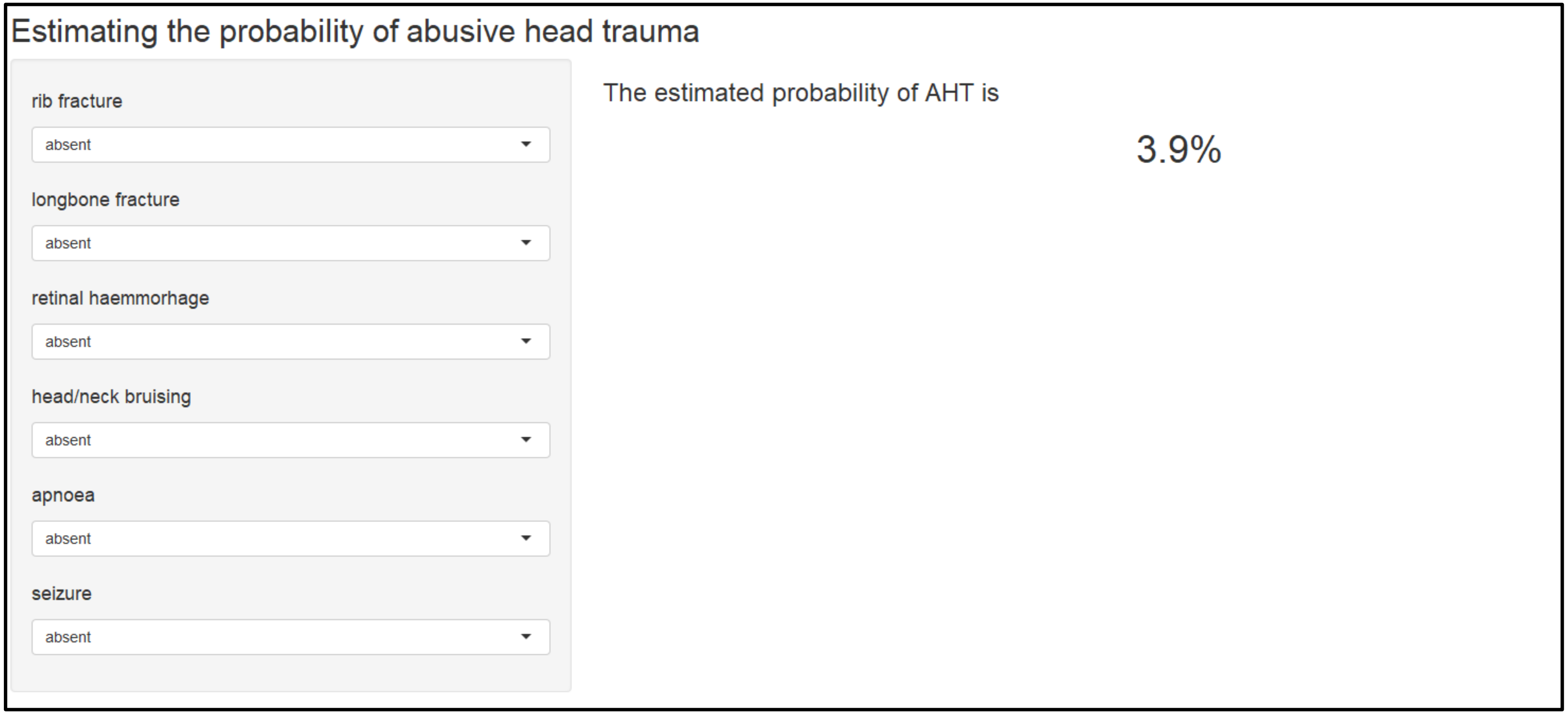
4.6.2 Estimating the probability of abusive head trauma when one or more of the six clinical features are unknown

Using data from the derivation study,[59] the probability of AHT was estimated when one or more of the six clinical features are unknown, using MICE.[721] The MICE procedure is fully described in section 4.4.1. To determine an estimated probability of AHT when one or more of the six clinical features are unknown, each possible combination with one or more unknown features was added to the derivation dataset and the unknown values were multiply imputed.[738] Unknown features were imputed either as absent (0) or present (1), and ten imputed datasets were generated. For each combination of features, the predicted probability of AHT was then estimated by calculating and exponentiating the linear predictor value, as previously detailed. The predicted probabilities of AHT from the ten imputed datasets were then averaged.[738] PredAHT thus estimates the probability of AHT for all 729 permutations of the six clinical features depending on whether each is present, absent or unknown, and can provide a probability calculation when certain investigations such as ophthalmology or skeletal survey have not yet been undertaken. PredAHT can therefore be applied at multiple points along the assessment pathway, according to the extent of information available about each of the six features.

4.6.3 Version 1 of the computerised Predicting Abusive Head Trauma clinical prediction tool

Figure 4.8 shows a screenshot of the user interface of Version 1 of the computerised tool. The six clinical features appear on the left-hand side. Each clinical feature is accompanied by a drop-down box, in which the user is required to indicate whether the feature is present, unknown or absent. Once this is completed, the estimated percentage probability of AHT is automatically generated on the right. This was the version that was shown to pathologists, CPSWs, police officers and legal practitioners in the qualitative study reported in Chapter 5.[747]

Figure 4.8 Screenshot of Version 1 of the Predicting Abusive Head Trauma clinical prediction tool



The screenshot depicts the scenario when all of the six clinical features are absent in a child less than three years old with intracranial injury.

4.6.4 *Calculation of likelihood ratios*

As described in Chapter 3, a likelihood ratio derived from the presence or absence of clinical features included in a CPR can be combined with a clinician's prior probability of a condition to produce a posterior probability.[627, 628] To understand why the likelihood ratio represents a convenient and powerful way to characterize clinical information, it is useful to understand the notation of conditional probability and the derivation of Bayes' theorem. The probability of AHT given that a particular combination of clinical features is present in a child is an example of conditional probability. Conditional probability is "the probability that an event is true given that another event is true (i.e. conditional on the second event being true)".[547(p.64)] The conditional probability of event A, given that event B is true, is written: $P[A|B]$ meaning the probability of event A conditional on event B, where the vertical line is read "conditional upon". The formal definition of conditional probability is

$$P[A|B] = \frac{P[A \text{ and } B]}{P[B]} \tag{2}$$

which may be translated as "the conditional probability that A is true given that B is true, is the ratio of the probability that both A and B are true divided by the probability that B is true".[547(p.64)] The posterior probability of AHT can be calculated using Bayes' theorem and the following quantities: 1) the prior probability of AHT, 2) the probability of a particular combination of clinical features conditional upon the patient having suffered AHT, 3) the probability of the particular combination of clinical features conditional upon the patient not having suffered AHT. As we are trying to calculate the probability of AHT (*AHT*) given that a child has a particular combination of clinical features (*F*), using the notation of conditional probability, we must calculate

$$P[AHT|F]$$

where *F* represents the combination of clinical features. This notation reads "probability of AHT conditional upon the combination of clinical features being present." We know from the definition of conditional probability (Equation 2) that

$$P[AHT|F] = \frac{P[AHT \text{ and } F]}{P[F]}$$

$P[F]$, the probability of the combination of clinical features, is simply the sum of the probability of the combination of clinical features in patients with AHT and its probability in patients without AHT:

$$P[F] = P\ [F \text{ and } AHT] + P[F \text{ and no } AHT]$$

Thus, we get

$$P[AHT|F] = \frac{P[AHT\ and\ F]}{P[F\ and\ AHT] + P[F\ and\ no\ AHT]} \quad (3)$$

By the definition of conditional probability (Equation 2)

$$P[F|AHT] = \frac{P[F\ and\ AHT]}{P[AHT]}$$

and

$$P[F|no\ AHT] = \frac{P[F\ and\ no\ AHT]}{P[no\ AHT]}$$

Rearranging these expressions, we get

$$P[F\ and\ AHT] = P[AHT] \times P[F|AHT] \quad (4)$$

$$P[F\ and\ no\ AHT] = P[no\ AHT] \times P[F|no\ AHT] \quad (5)$$

Substituting equations 4 and 5 into 3, we obtain Bayes' theorem:

$$P[AHT|F] = \frac{P[AHT] \times P[F|AHT]}{(P[AHT] \times P[F|AHT] + P[no\ AHT] \times P[F|no\ AHT])} \quad (6)$$

A major disadvantage of Bayes' theorem is that the majority of clinicians require a calculator; however, the solution is to use the odds ratio form of Bayes' theorem to calculate the post-test odds, which can be converted to the post-test probability.[547] The odds ratio form of Bayes' theorem is simple to remember, and involves multiplying only two numbers. Expressing the pre-test probability of a condition in terms of the pre-test odds achieves a compelling simplification of Bayes' theorem. Using the odds ratio form of Bayes' theorem, clinicians can update their prior probabilities upon learning of new information or data. In addition, expressing Bayes' theorem in its odds ratio format enables the effect of new diagnostic information to be conveyed in a simple and intuitive manner.[547]

To derive the odds ratio form of Bayes' theorem, we must start with Bayes' theorem in its familiar form (Equation 6). Next, convert *P*[*AHT*], a probability, to odds using the following relationship between probability and odds:

$$P[AHT] = \frac{odds[AHT]}{1 + odds[AHT]}$$

Instead of odds[*AHT*] we write *O*[*AHT*]. Substituting *O*[*AHT*]/(1 + *O*[*AHT*]) where we see *P*[*AHT*] in Bayes' theorem, we get the expression:

$$P[AHT|F] = \frac{\left(\frac{O[AHT] \times P[F|AHT]}{1 = O[AHT]}\right)}{\frac{O[AHT] \times P[F|AHT]}{1 = O[AHT]} + \left(1 - \frac{O[AHT]}{1 + O[AHT]} \times P[F|no\ AHT]\right)}$$

Remembering that

$$P[AHT|F] = \frac{O[AHT|F]}{1 + O[AHT|F]}$$

This long expression for $P[AHT|F]$ simplifies to:

$$\frac{O[AHT|F]}{1 + O[AHT|F]} = \frac{O[AHT] \times P[F|AHT]}{O[AHT] \times (P[F|AHT] + P[F|no\ AHT])}$$

Multiplying the numerator and denominator of the right side by $1/O[AHT] \times P[F|AHT]$),

$$\frac{O[AHT|F]}{1 + O[AHT|F]} = \frac{1}{1 + \left(\frac{P[F|no\ AHT]}{O[AHT] \times P[F|AHT]}\right)}$$

Cross multiplying,

$$O[AHT|F] + O[AHT|F] \times \frac{P[F|no\ AHT]}{O[AHT] \times P[F|AHT]} = 1 + O[AHT|F]$$

$$O[AHT|F] \times \frac{P[F|no\ AHT]}{O[AHT] \times P[F|AHT]} = 1$$

Rearranging terms, we obtain the odds ratio form of Bayes' theorem:

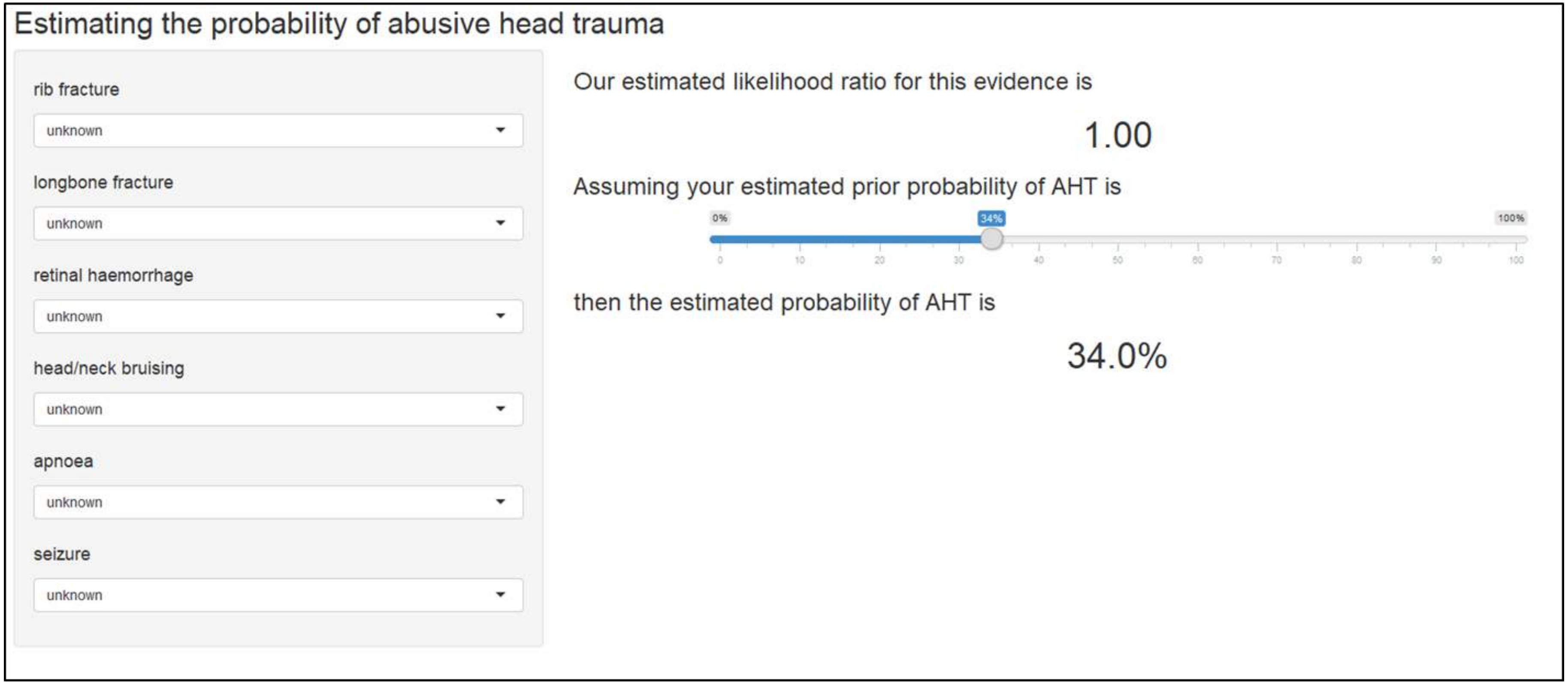
$$O[AHT|F] = O[AHT] \times \frac{P[F|AHT]}{P[F|no\ AHT]} \quad (7)$$

Equation 7 states that the post-test odds of AHT conditional upon the combination of clinical features F ($O[AHT|F]$) are equal to the pre-test odds ($O[AHT]$) multiplied by the likelihood ratio $P[F|AHT]/P[F|no\ AHT]$). The likelihood ratio is the amount that the odds change after the combination of clinical features is known. Another way to state the odds ratio expression of Bayes' theorem is: post-test odds = pre-test odds × likelihood ratio.[547] In order to calculate LRs of AHT for the 729 possible combinations of clinical features, firstly the predicted probabilities of AHT for each combination of features were converted to odds of AHT. The odds of AHT for each combination of features were then divided by the pre-test (prior) odds. The pre-test odds are simply the prevalence of AHT in the dataset; this is approximately 34%, therefore the pre-test odds are 0.34/0.66 = 0.52. The resulting LRs of AHT for each of the 729 possible combinations of features can be multiplied by the "new" pre-test odds (i.e. a clinicians' own prior probability of AHT) to obtain the post-test odds, which are finally converted into a post-test probability of AHT.

4.6.5 *Version 2 of the computerised Predicting Abusive Head Trauma clinical prediction tool*

Figure 4.9 shows a screenshot of the user interface of Version 2 of the computerised tool. Once again the six clinical features are presented on the left-hand side, and the user is required to use drop-down boxes to indicate whether each is present, unknown or absent. The likelihood ratio of AHT for the combination of features entered is then displayed on the right-hand side, at the top. In addition on the right-hand side, there is a sliding scale representing probabilities in percentage terms between 0% and 100%. This is where the user can choose to move the slider to enter their own estimated prior probability of AHT based on other factors pertinent to each case unaccounted for by PredAHT e.g. purported history, clinical presentation, or psychosocial features. The default/baseline prior is set at 34% which is the prevalence of AHT in the data used to derive the tool. When the slider is moved, behind the scenes, PredAHT converts the clinician's prior probability to odds, multiplies this with the calculated LR for the specific combination of features entered, and converts the resulting posterior odds to a posterior probability, which is automatically displayed as an estimated percentage probability of AHT below the slider. This version was shown to clinicians in the qualitative study reported in Chapter 5[747] the vignette study reported in Chapter 6,[748] and the feasibility study reported in Chapter 7.

Figure 4.9 Screenshot of Version 2 of the Predicting Abusive Head Trauma clinical prediction tool



The screenshot depicts the scenario where all six clinical features are unknown in a child less than three years old with intracranial injury.

4.6.6 Modifications

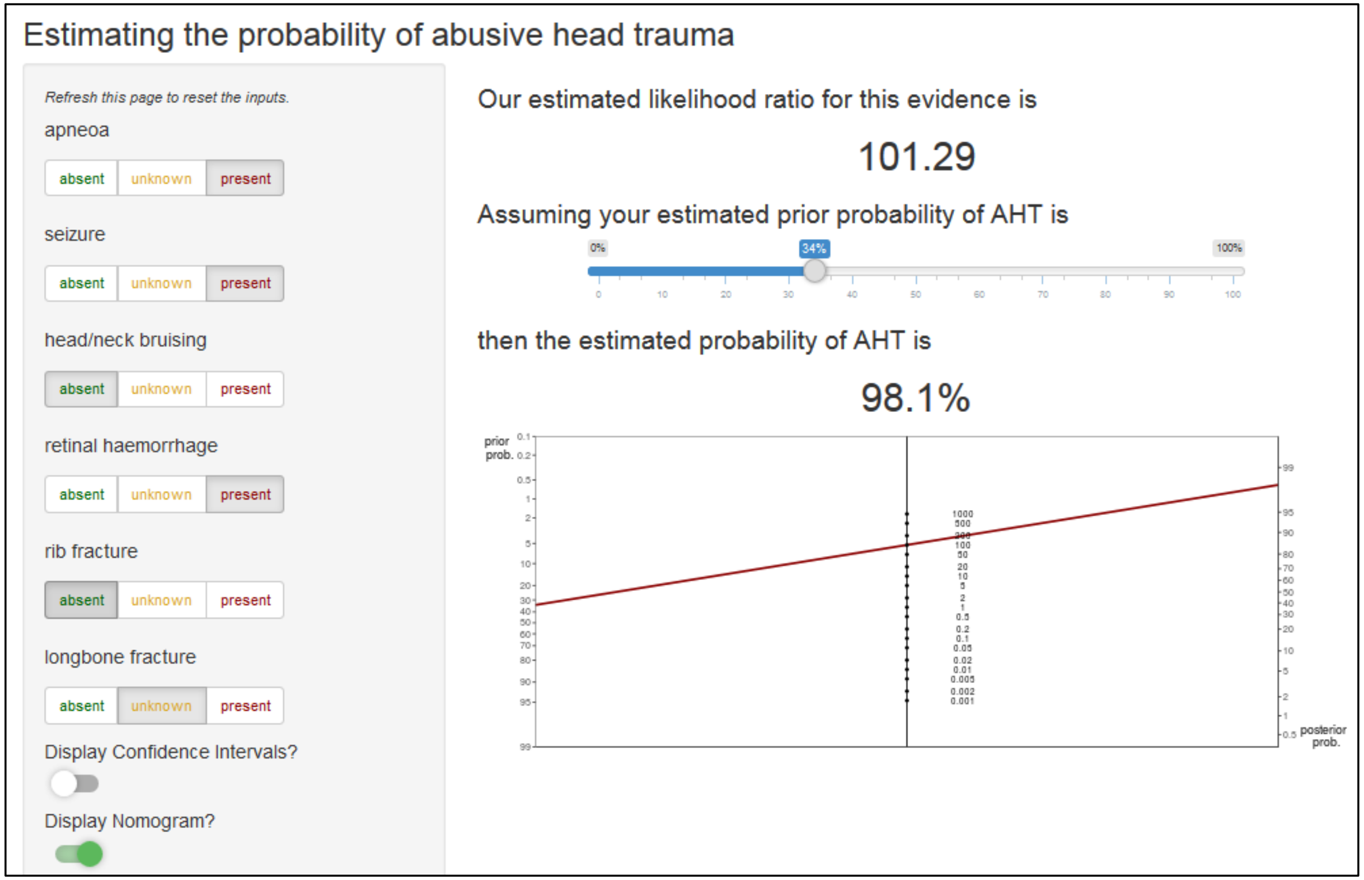
Based on the findings of a qualitative study reported in Chapter 5,[747] a vignette study reported in Chapter 6,[748] and a feasibility study reported in Chapter 7, a number of modifications were made to Version 2 of the computerised PredAHT to produce the final version. During qualitative interviews participants offered suggestions for improvements to the tool to enhance its usability and clinical utility.[747] Other modifications were made following the researcher's observations of participants' interacting with the tool in the three studies. The tool also underwent "in-house" testing by the researcher and supervisory team, who checked the output for accuracy, suggested improvements, and reported any practical issues or difficulties with its use. Modifications that were made are outlined in Table 4.13. A screenshot of the user interface of the final version of the computerised tool is shown in Figure 4.10.

Table 4.13 Modifications made to the computerised Predicting Abusive Head Trauma clinical prediction tool following empirical studies

Feature suggestion	Source	Modifications made
Confidence intervals	Qualitative study	Option to display confidence intervals if required
Graphical representation of the impact of the user's estimated prior probability on the PredAHT score	Qualitative study Vignette study Feasibility study	Option to display Fagan's nomogram to depict how the user's estimated prior probability and the likelihood ratio for the different combinations of features interact to produce the PredAHT score
Traffic light colour coding system	In-house testing	Absent is now colour coded as green, missing is colour coded as orange and present is colour coded as red
Radio buttons instead of drop-down boxes to avoid potentially selecting the wrong option	In-house testing Qualitative study Vignette study Feasibility study	Drop-down boxes changed to radio buttons
To ensure that none of the options (absent, unknown, present) are highlighted or selected on the default interface for each feature, to guarantee that the correct options are selected by the user for each feature	In-house testing Qualitative study Vignette study Feasibility study	When the PredAHT tool is loaded, none of the options (absent, unknown, present) are selected or highlighted for each feature, so that the user must actively select an option for each feature themselves

Ability to reset the options back to the default "no option selected"	In-house testing Qualitative study Vignette study Feasibility study	An instruction has been added to refresh the web page to reset the inputs
Probability and likelihood ratio only displayed once an option is selected for each feature (absent, unknown, or present), to avoid confusion and potential errors	In-house testing Qualitative study Vignette study Feasibility study	The probability and likelihood ratio is now only displayed once an option for all six features is selected (absent, unknown, or present)
Alter the order of the six clinical features in line with their order of availability in clinical practice	Qualitative study	The six features are now listed in the following order: Apnoea Seizures Head/neck bruising Retinal haemorrhage Rib fracture Long-bone fracture

Figure 4.10 Screenshot of the final version of the Predicting Abusive Head Trauma clinical prediction tool

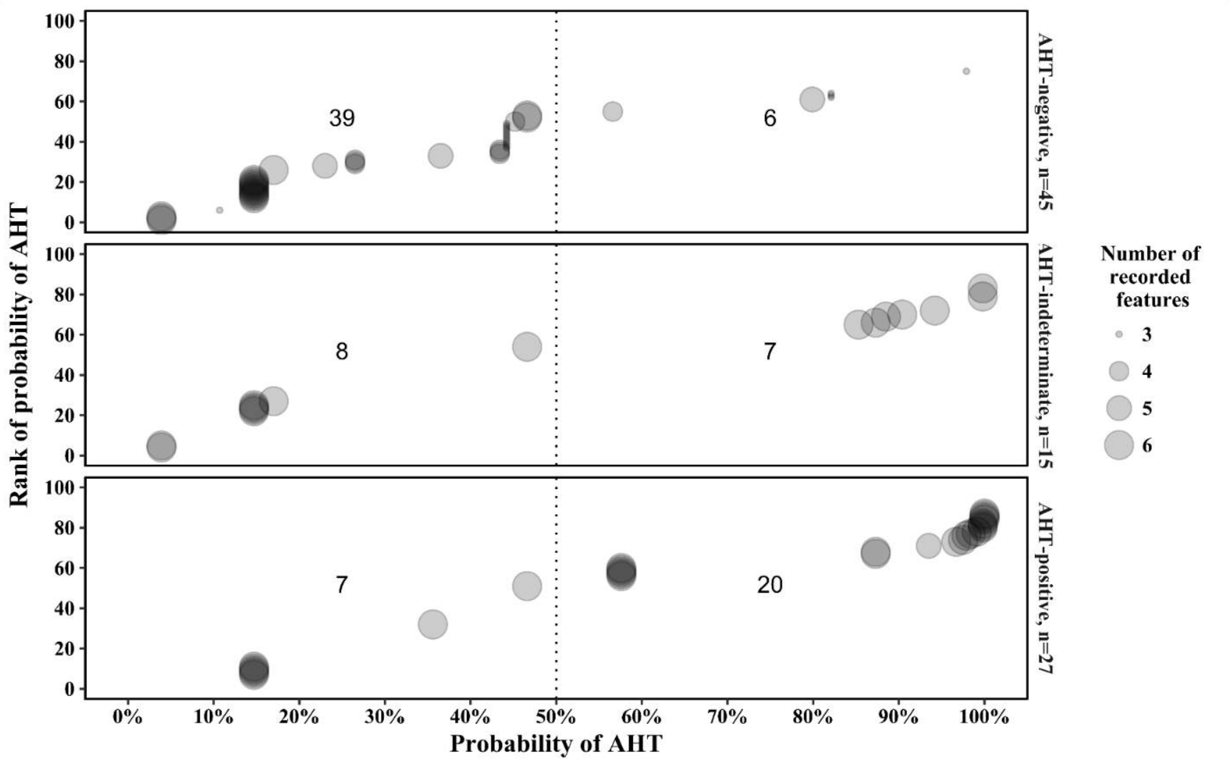


The screenshot depicts the scenario where apnoea, seizures and retinal haemorrhages are present, head/neck bruising and rib fractures are absent and long-bone fractures are unknown.

4.7 External validation of the Predicting Abusive Head Trauma clinical prediction tool in an Australian/New Zealand population

When applied to the validation dataset[60] in-house, using a 50% probability cut-off, the performance of the version of PredAHT that can account for unknown features was identical to the performance of the original PredAHT in this dataset; its sensitivity was 72.3% and its specificity was 85.7% (see Table 4.7, above). The latest version of PredAHT has also recently been externally validated in an Australian/New Zealand population. The researcher and supervisory team were approached by colleagues in Australia and invited to collaborate on a formal external validation of PredAHT using data collected in a prospective multi-centre observational study of children with head injuries in ten Australian and New Zealand paediatric EDs.[749, 750] In this secondary analysis, the medical records of children with possible AHT were reviewed at five of the participating sites; the strategy to identify possible AHT cases has been described in the external validation of the PediBIRN CPR by this group, discussed previously,[735] and in the manuscript submitted to *Archives of Disease in Childhood* (Appendix 5). Cases were categorized as AHT, indeterminate or nAHT based on the decision of the multidisciplinary CP team. PredAHT was applied to 87 children less than three years of age with ICI confirmed on neuroimaging (Figure 4.11). The probability of AHT was calculated for each child based upon whether the six clinical features were present, unknown or absent. As before, as a primary analysis, a 50% probability cut-off was used to categorize all patients with a probability of ≥50% as higher risk for AHT and those with a probability of <50% as lower risk for AHT. In addition, as individual clinicians' interpretation and application of probability thresholds for AHT differ,[490, 529, 530, 532, 533] as a secondary analysis, the implications of using different probability cut-offs to categorize cases as AHT were explored. This was done by using a 20% probability cut-off and an 80% probability cut-off (Table 4.14). The performance of PredAHT was very similar to that in the original validation study[60]; using a 50% probability cut-off, sensitivity was 74% and specificity was 87% (Table 4.14).

Figure 4.11 Predicted probability of abusive head trauma assigned by the Predicting Abusive Head Trauma clinical prediction tool for all 87 children with intracranial injury, by outcome and number of recorded features



The circles represent the calculated probability of AHT for each of the 87 children with intracranial injury included in the study. The numbers on the figure correspond to the number of children categorized as higher or lower risk for AHT based on a 50% probability cut-off. The size of the circles shows how *many of the six features are recorded as present or absent. The smaller the circle, the more information is unknown and the less likely that a skeletal survey and/or an ophthalmology examination was undertaken.*

Table 4.14 Performance of the Predicting Abusive Head Trauma clinical prediction tool in an Australian/New Zealand population using three different probability cut-offs

Applying PredAHT (indeterminate excluded)	**20% cut-off**		**50% cut-off**		**80% cut-off**	
	Outcome		Outcome		Outcome	
	AHT	nAHT	AHT	nAHT	AHT	nAHT
Higher risk of AHT	22	30	20	6	15	4
Lower risk of AHT	5	15	7	39	12	41
	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	81%	62%–94%	74%	54%–89%	56%	35%–75%
Specificity	33%	20%–49%	87%	73%–95%	91%	79%–98%
Positive predictive value	42%	29%–57%	77%	56%–91%	79%	54%–94%
Negative predictive value	75%	51%–91%	85%	71%–94%	77%	64%–88%
LR +	1.22	0.93–1.61	5.56	2.55–12.1	6.25	2.31–16.9
LR −	0.56	0.23–1.36	0.30	0.16–0.57	0.49	0.32–0.75

Confidence intervals for the sensitivity, specificity, positive predictive value and negative predictive value are exact binomial confidence intervals.

Confidence intervals for the positive likelihood ratio and the negative likelihood ratio are based on the risk ratio.

4.8 Discussion

The aims of this chapter were threefold. The first aim was to describe the previous derivation and validation of PredAHT.[24, 59, 60] The second aim was to describe a systematic review, critical appraisal and comparison of validated CPRs for AHT conducted in collaboration with Australian colleagues.[706] The third aim was to report the systematic process used to create the computerised PredAHT for use in clinical practice, and present an external validation of this in an Australian/New Zealand population. The development of the computerised PredAHT followed an iterative process and consisted of several stages, including selection and development of key features, programming, design of the user interface, in-house user testing, and modifications based on suggestions and observations from the subsequent empirical studies presented in this thesis. Key features were selected based on findings from the scoping review of the literature reported in Chapter 3, and based on the appreciation that if clinicians were to apply PredAHT in practice, they may face the problem of an unknown clinical feature.[315, 738, 739] Key features of the computerised PredAHT included an "unknown" option for each of the six clinical features, LRs of AHT for each of the possible combinations of clinical features, and a sliding scale to enable clinicians to incorporate their own prior probability of AHT and facilitate automatic calculation of the posterior probability of AHT. Additional key features incorporated following feedback and observations from the empirical studies included an option to display confidence intervals around the posterior probabilities, and the addition of Fagan's nomogram to illustrate how the user's own prior probability of AHT and the LRs of AHT for the different combinations of features interact to produce the posterior probability of AHT. The development of PredAHT will be discussed below in terms of methodological considerations, its validity and level of evidence, and its strengths and limitations.

4.8.1 Methodological considerations in the development of the Predicting Abusive Head Trauma clinical prediction tool

An important consideration in the development of CPRs is the approach to handling missing data.[714] As described in section 4.3.2.1.2, multiple imputation is a widely accepted method for dealing with missing data and was the approach taken in the derivation study,[59] the validation study[60] and when estimating the probability of AHT when one or more of the six clinical features are unknown. However, such approaches are statistically valid provided the data are MAR.[715] Importantly, the MAR assumption is just that; an assumption, rather than a property of the data.[716] The MCAR assumption can be tested, but it is not possible to differentiate between MAR and MNAR from the observed data.[324, 716] Most missing data are expected to be at least partly MNAR.[716, 751, 752] For example, it may be reasonable to assume that in children aged greater than two years for whom the variable "long-bone fracture" is unknown, there may have been no clinical indication to perform a skeletal survey,[512] and therefore they may be less likely to have a long-bone fracture. Nevertheless, the MAR assumption is reasonable in the current context, as clinical decisions about whether to perform a skeletal survey or ophthalmology exam are usually determined by the "measured" presence or absence of other features, that is, the observed data.[59] In addition, if the MAR assumption is violated, this violation also affects older procedures such as listwise deletion; in short, techniques based on multiple imputation are always at least as good as statistically unprincipled procedures, and are often far better.[752] One myth that abounds in the literature is that imputation is akin to "making up the data", however the aim of the process is not to obtain the individual values themselves, but to preserve important characteristics of the dataset.[752] By using all of the information available in a dataset, analyses with multiply imputed data benefit from an increased sample size and are likely to be less biased than estimates resulting from statistically unprincipled methods used to address missing data.[716] Crucially, when exploring alternative imputation methods in the validation study,[60] when each unknown clinical feature was deemed to be absent, PredAHT was more likely to miss AHT cases than when MICE was used.

One assumption of the derived PredAHT regression model is the additivity assumption i.e. the assumption that the effect of each predictor does not depend on the values of the other predictors.[326] This assumption can be formally tested by examining statistical interactions between some or all of the predictors.[733] For the purpose of simplicity and because of the sample size of the derivation data, only the parameters for the main effects were estimated, and no interaction terms were included in the model.[59] Thus, it is impossible to infer whether the effects of each of the clinical features on the probability of AHT are dependent on the presence of any of the other clinical features. However, only a minority of CPRs contain interactions, and it appears that few investigators assess the interactions between predictors during the modelling process; it has been suggested that this approach is usually reasonable as interaction terms seldom add to the predictive ability of a CPR.[326] Model overfitting may occur if many interactions are tested and only the strongest retained in the model, leading to overly optimistic performance measures.[324]

An assumption underlying the use of Bayes' theorem and the LR in clinical diagnosis, is the stability of the LR across contexts.[753] While predictive values directly depend on disease prevalence and thus cannot be directly translated from one setting to another, sensitivity and specificity are theoretically mathematically unaffected by prevalence,[754] and it is therefore commonly believed that they do not vary with disease prevalence.[755-757] As LRs are a function of sensitivity and specificity, it is assumed that they also remain stable when prevalence varies.[753] However, a number of studies have demonstrated that this may not always be the case in clinical practice and that sensitivity, specificity and LRs may not be as stable as once thought.[753, 758-764] While it is not always feasible to identify the mechanisms responsible for this phenomenon, it is thought that the "spectrum effect" plays a significant role.[627, 753, 754, 758, 765] The spectrum effect describes the variation in the performance of a predictive test between different settings and population subgroups.[765] Thus, bias may occur when applying a CPR to a population of patients with a different clinical spectrum of disease i.e. more or less severe or advanced disease, than the population of patients from which the CPR was derived.[627] Bossuyt[766] emphasizes that potential variability of sensitivities and specificities between subpopulations does not justify a "burial of Bayes' rule", as these measures can be considered as average likelihoods of a positive and negative test result in a given (sub)population. While the variability in likelihoods may complicate the use of Bayes' rule, it does not invalidate it, as "Bayes' rule only puts consistency in quantitative statements of uncertainty" (p. 5). In addition, absolute certainty on the presence of a condition and absolute precision in probability statements are not usually necessary for clinical practice, where action is required only when the strength of suspicion exceeds a clinicians' probability threshold[766]. Nevertheless, the existence of the spectrum effect emphasizes the importance of conducting external validation studies of CPRs in patient populations with different clinical spectrums of disease, to enhance the generalizability of the CPR.[767] In addition, clinicians and investigators should discuss possible differences between the settings of the previous derivation and validation studies and the setting of a potential impact study prior to conducting an impact analysis.[315] If possible, external validation should first be carried out in the impact setting and if necessary the CPR should be updated in order to tailor it to the new setting.[315] However, this requires individual patient data regarding the predictors and the outcome from the potential impact setting to be available.[315]

4.8.2 *Validity, performance and level of evidence of clinical prediction rules for abusive head trauma*

When applied in-house to the validation dataset,[60] the performance of the computerised PredAHT, that provides predicted probabilities for 729 possible combinations that arise when one or more clinical features may be unknown, was identical to the performance of the original PredAHT. In addition, the tool demonstrated slightly better performance in an external validation study in an Australian/New Zealand population. This strengthens the evidence for the utility of PredAHT, and raises its level of evidence from a Level 2 CPR with narrow validation in similar patients as in the derivation study,[60] to a Level 3 CPR with broad validation in multiple settings by independent investigators.[299] Some researchers propose that clinicians may use the predictions of a Level 3 CPR with confidence in their accuracy.[299] While PediBIRN is also a Level 3 CPR, the BIBIS is a Level 2 CPR, and the PIBIS remains a Level 1 CPR. This is because, although the published study is described as a validation and refinement of PIBIS,[701] in reality it is a re-derivation of the CPR,[768] as two previous variables were excluded from the model.[701] Therefore PIBIS requires validation before it can be applied to future patients.[326, 706, 768] None of the CPRs have yet undergone an impact analysis in clinical practice, although as previously noted, a multicentre trial of PediBIRN is currently being conducted in the States. According to Wallace et al.,[300] a CPR that has been derived and broadly validated using pre-defined methodological criteria can be considered ready for impact analysis.

The PediBIRN CPR is the only other CPR for AHT aside from PredAHT that is intended for use in an inpatient setting. Taking an arbitrary probability cut-off of 50%, PredAHT has a lower sensitivity than PediBIRN but a much higher specificity, and will categorize fewer nAHT cases as higher risk for AHT than PediBIRN. PredAHT is designed to assign a specific probability of AHT to each individual case, and three of the features (RH, long-bone fracture, rib fracture) will only be available at the point in the clinical pathway when ophthalmology and skeletal radiology investigations have been completed. As PredAHT can be used at a later point in the diagnostic process, this explains and allows for differences in sensitivity and specificity.[706] At the outset, it is paramount to ensure that cases undergo sufficient investigation, to minimise missed AHT, therefore high sensitivity is important. When more investigations have been undertaken, a higher specificity is desirable to ensure that an incorrect diagnosis of AHT is avoided. As each of the CPRs for AHT apply to different points along the clinical pathway, this suggests that they may be used complementarily in clinical practice.[706] Of note, while

PediBIRN provided probability estimates for each of 16 possible combinations of the four features included in their CPR,[705] these were calculated by implicitly fitting a saturated model, and could have alternatively been estimated by fitting a simple logistic regression model, allowing for information borrowing across categories and therefore improved estimation of the variance.[737] In addition the predicted probabilities arising from the PediBIRN CPR have not been validated in their own right.

4.8.3 Strengths and limitations of the Predicting Abusive Head Trauma clinical prediction tool

Strengths of the computerised PredAHT tool include the highly systematic and rigorous approach to its development, its ability to provide a real-time probability estimate of AHT when one or more features may be unknown, and the inclusion of LRs and a "sliding scale" feature, that allows users to incorporate their own prior probability of AHT and facilitates accurate calculation of posterior probabilities. PredAHT was derived and validated based on methodological guidelines for the development of CPRs.[293, 301] The computerised version was developed in line with the MRC framework for the development and evaluation of complex interventions,[322, 323] following the identification of theories relevant to decision-making in AHT and literature regarding the logic of CPRs in Chapter 3, and identification of the evidence-base for the use of CPRs for assisting in the identification of AHT.[706] The computerised version was then externally validated in-house and using an Australian/New Zealand dataset. Roll-out of the computerised PredAHT would enable simple application at the bedside, as new information is collected.

With its added capacity to provide a probability of AHT for an individual case where one or more of the six features are unknown, PredAHT is currently the only CPR that has the potential to contribute to decision-making at multiple points along the assessment and referral pathway. It is becoming increasingly apparent that a real-time strategy to impute missing values is desirable when applying a CPR in clinical practice.[315, 738, 739] Such a task is not typically straightforward, as it requires access to the derivation dataset via, for example, a website.[738] By estimating *in advance* the probability of AHT when one or more features are missing, the PredAHT tool circumvents this issue. The inclusion of probability estimates in incompletely investigated cases offers an opportunity for clinicians to consider the probability of AHT at different stages of the clinical assessment and whether or not further investigations are indicated within the context and circumstances that surround the case. To illustrate, in the external validation study described in section 4.7, six nAHT cases were assigned a probability of

>50% (Figure 4.11). Five of these cases did not have an ophthalmology examination or skeletal survey. Completing the investigation would identify whether RH, rib or long-bone fractures were present, and refine the probability estimate. For example, in children with ICI and head/neck bruising but no information about RH or fractures, the calculated probability of AHT is 44.2%. If skeletal survey and ophthalmology were normal this would decrease to 14.7%. Conversely, if either long-bone fracture, RH or rib fracture were identified, the probability would increase to 70.2%, 85.3%, and 88.5%, respectively. This highlights the importance of considering an ophthalmology examination and skeletal survey for those children presenting with ICI in the absence of an independently witnessed accident, and suggests that PredAHT may encourage the standardisation of these investigations in suspected AHT cases.

A unique feature of the computerised PredAHT is the "sliding scale" that allows clinicians to input their own prior probability of AHT, which is multiplied by the LRs calculated for each combination of clinical features to automatically produce a posterior probability of AHT. This feature was included following a review of clinical decision-making theory, which demonstrated that the revision of probability is the central clinical strategy underlying diagnostic reasoning, and highlighted the concordance between this strategy and the mathematical properties of LRs.[769] This is the only computerised CPR that the researcher is aware of that includes such a feature, and it is currently unclear how the incorporation of prior probabilities may impact on clinicians decision-making and the use of the CPR. This is explored in subsequent chapters. Of note, one of the variables included in the Wells CPR for predicting pulmonary embolism is the clinicians' judgment of whether an alternative diagnosis is more likely than a pulmonary embolism diagnosis.[770] This CPR has been criticised by some due to the subjectivity and non-standardisation of this "clinical judgment" variable,[771-773] and because this variable is influenced by other factors in the CPR,[774] thereby potentially reducing the CPR's diagnostic value and reproducibility.[775] However, others have found that this component of the CPR has a high predictive value in comparison to the other variables, and that the CPR therefore remains valuable for the management of patents with suspected pulmonary embolism.[776] It is important that a clinician's prior probability is not based on variables that are already included in PredAHT so that the same variables are not used twice in decision-making.[577]

One possible limitation is that, unlike the PediBIRN, PIBIS, and BIBIS, PredAHT does not recommend a direct course of action based on the predicted probabilities. Current literature suggests that CPRs that provide direct recommendations to clinicians have a greater impact on decision-making and health outcomes than CPRs that simply provide assistive probabilities.[295, 299, 652, 777] However, there is limited empirical evidence for this suggestion as there are few studies that directly compare a directive and assistive format of the same CPR.[315] Kappen et al.[315] recently undertook a direct comparison of a directive and an assistive format of the same prediction model within a single setting, and demonstrated that the assistive CPR altered clinician behaviour but did not improve subsequent patient outcomes, while the directive CPR improved both decision-making and patient outcomes. One of the objectives of the feasibility study reported in Chapter 7 is to explore how different probability predictions relate to clinicians' CP actions, to determine whether predictions can be translated into decisions.

Another limitation is that the final version of the computerised PredAHT has not been formally user-tested. Guidelines suggest that technological interventions should be tested with a sample of users to elicit their perceptions and reactions to the content and design of the intervention.[778, 779] User-testing provides an opportunity to optimise the usability of a CPR for future users prior to implementation,[779] which may encourage its adoption in clinical practice.[780] The final version of the CPR could be formally tested using "think-aloud" techniques, where participants are required to use the CPR in front of the researcher while discussing out loud their thoughts and opinions.[781] This would provide an opportunity to understand how clinicians actually interact with and use the CPR in practice. The "think-aloud" technique was employed in the vignette study presented in Chapter 6, however this was used to capture participants reasoning behind their decision-making in a series of clinical vignettes describing suspected AHT, rather than to test the usability of the PredAHT tool *per se.*[748] Similarly, participants in the empirical studies reported in subsequent chapters suggested that if PredAHT were to be made available online or as a mobile application, additional content would be required, to explain how the tool was developed and describe its current level of validation, with links to relevant publications and literature.[747] Others suggested the addition of disclaimers, and graphics to display the data on which PredAHT is based.[747] The PredAHT Shiny app could be embedded into a website containing additional information and resources. Due to limited time and resources, it was not possible to develop a website during the time-scale of this PhD, however if PredAHT were to be implemented in clinical practice, a website containing more comprehensive information regarding the tool should be considered for development.

One concern in all studies investigating the diagnosis of AHT is the issue of circularity.[782] The confirmation of AHT in the systematic review,[24] derivation[59] and validation studies[60] on which PredAHT is based, was established using quality standards designed by the research team to minimise circularity, whereby AHT was confirmed only by witnessed AHT, perpetrator admission, court proceedings or a thorough assessment by a multidisciplinary team of CP professionals (see Table 4.2). Arguably the decision of the multidisciplinary team will include a consideration of clinical features, however, this is the case in any clinical diagnosis. There are many diseases and diagnoses that are based upon a collection of symptoms, signs and clinical history where a gold-standard "diagnostic test" does not exist, e.g. Kawasaki syndrome, asthma or indeed the majority of mental health conditions.[782] The process of identifying such features and formulating a probability of an illness or disease, to then seek further information from additional investigations etc. is fundamental to the diagnostic process where diagnostic decisions must be made based on clinical criteria and the exclusion of differential conditions for the benefit of the patient. One simply cannot make any diagnosis without including an assessment of the physical findings.[782] In addition, as described in Chapter 2, within the CP process there are multiple additional forensic, social and historical factors that are included in decisions about the balance of probability of AHT and the risk of future harm to the child. Hymel et al.[351, 702] chose to design a priori definitional criteria to define AHT and minimise circularity; when challenged on the issue that bruising was included both in their definitional criteria and their CPR, potentially introducing incorporation bias,[737] the authors responded that of 73 patients with bruising, 61 met other definitional criteria and 12 remaining patients were subsequently diagnosed with definite/probable AHT by the treating clinicians.[783]

4.9 Conclusions and implications for this thesis

This chapter described the previous derivation and validation of PredAHT, critically appraised PredAHT alongside other validated CPRs for the identification of AHT, and reported the systematic and iterative process used to create the computerised PredAHT tool. PredAHT is currently the only CPR for AHT designed for use in an in-patient hospital setting that could be used at multiple points in the assessment pathway, and provides predicted probabilities and LRs for 729 possible permutations of six clinical features depending on whether each is present, unknown, or absent. PredAHT has been validated in multiple settings and is thus ready for impact analysis.[300] The computerised tool will be used in the following three empirical studies investigating the acceptability and potential impact of PredAHT and the feasibility of evaluating its impact in clinical practice.

5 Acceptability of the Predicting Abusive Head Trauma clinical prediction tool: A qualitative study with child protection professionals

The results from this chapter were published in two articles (see Appendix 6):

1. **Cowley LE**, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO & Kemp AM (2018). Factors influencing child protection professionals' decision-making and multidisciplinary collaboration in suspected abusive head trauma cases: A qualitative study. *Child Abuse & Neglect,* 82: 178-191.
2. **Cowley LE**, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO & Kemp AM (2018). Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals. *Child Abuse & Neglect,* 81: 192-205.

5.1 Chapter overview

This chapter presents a novel qualitative interview study with 56 CP professionals involved in suspected AHT cases, including clinicians, CPSWs, police officers, pathologists and lawyers.[187, 747] Factors influencing decision-making and multidisciplinary collaboration in suspected AHT cases, and attitudes towards PredAHT were explored. Data were analysed using thematic analysis facilitated by the Framework Method. The findings are placed within the context of the current literature, and implications for the further development, evaluation and implementation of PredAHT are discussed.

5.2 Introduction

As described in Chapter 3, a crucial aspect of CPR development is determining its acceptability to the population for whom it is intended.[299, 301, 784] Phase two of the iterative four-phased framework for the impact analysis of CPRs,[300] involves assessing the acceptability of a CPR and identifying barriers to its uptake and implementation, while the "feasibility and piloting" phase of the MRC framework for the development and evaluation of complex interventions includes initial testing of the acceptability of an intervention prior to full-scale evaluation.[322, 323] If a CPR proves to be acceptable in addition to demonstrating a positive impact on clinician behaviour and patient care, its long-term and widespread dissemination and implementation would be justified; if not, the CPR could undergo modification and further evaluation.[784]

The evaluation and investigation of suspected AHT is unique in that it requires multidisciplinary collaboration, and consultations must be informative to outside agencies as well as clinical colleagues.[785, 786] Clinicians, police officers and CPSWs must work together to jointly determine the likelihood of AHT. All have different responsibilities towards the child yet need an understanding of the clinical findings and their relevance. Recent guidelines for the evaluation of suspected physical abuse recommend that medical records include a clear opinion about the likelihood of physical abuse and should elucidate specific levels of concern to aid police and CPSWs' investigations,[333] so that they can gauge the degree of certainty of AHT in each case.[334] It is proposed that PredAHT may assist clinicians in their discussions with social services, law enforcement and other professionals involved in the CP process, and support clinicians offering medical testimony in civil or criminal proceedings.[59, 60] Therefore, it is important to know if PredAHT is acceptable to the range of professionals involved in the CP process, including legal practitioners. Assessment of the acceptability of CPRs to those who may use them has mostly been limited to training sessions and engagement with local stakeholders,[306] survey methodology,[663, 787] and studies using Likert scales,[788, 789] or simulation exercises,[667, 790] however qualitative methods have the potential to yield more detailed information about the acceptability of a CPR.[673]

In addition, in order to determine whether PredAHT is likely to be acceptable and useful in clinical practice, it is important to understand professionals' decision-making processes in suspected AHT cases. The clinical decision-making theories reviewed in Chapter 3 contributed to a deeper understanding of clinical diagnostic decision-making in suspected AHT cases, however to date no study has sought to directly explore how CP professionals make decisions in these cases using qualitative methods. Much of the evidence regarding the barriers or facilitators to multidisciplinary working or the perceptions of professionals working in multidisciplinary teams in suspected abuse cases has been anecdotal, or has relied on case studies or surveys.[786, 791] Furthermore, while survey studies have assessed the factors affecting clinicians' decisions to *report* suspected abuse,[490, 531, 792] these were all conducted in North America, and do not address decision-making processes in suspected AHT specifically. There is a need to better understand decision-making in suspected AHT cases, the working relationships between the different professional groups, whether PredAHT is acceptable to CP professionals and whether or not PredAHT it is likely to be used in clinical practice.

5.2.1 Aims of the qualitative study

This study aimed to explore the factors influencing decision-making and multidisciplinary collaboration with CP professionals involved in suspected AHT cases, and to determine the acceptability of PredAHT amongst these professionals.

5.3 Methods

This was a qualitative, face-to-face, in-depth, semi-structured interview study with clinicians, CPSWs, legal practitioners, police officers, and pathologists involved in suspected AHT cases.

5.3.1 Rationale

The purpose of qualitative methods is to develop an understanding of participants' experiences, attitudes, meanings, behaviours and views of social phenomena in natural rather than experimental contexts.[793, 794] In contrast with quantitative methods, qualitative methods allow for rich, in-depth description of the phenomenon under study.[795] Qualitative research is used to gather and analyse data that cannot be epitomised by numbers; it involves the collection, management and interpretation of textual information gained from discourse or observation.[796, 797] It is increasingly being recognised as an essential component to medical research and has been described as compatible with the philosophy of evidence-based medicine.[794, 798] There are many approaches to the collection and analysis of qualitative data that are based on various epistemological perspectives and methodologies.[799] The main data collection methods that are used are interviews and focus groups.

Qualitative face-to-face semi-structured interviews were deemed the most appropriate data collection method to explore the opinions and views of the participants included in this study. It was thought that focus groups were not a practical method of data collection for this work due to the busy schedules of the participants. In addition, interviews allow for more detailed responses than focus groups, as there is more time for each participant to speak, and no interference from other group members.[800] Interviews are the preferred method when the objective of the research is to explore individual decision processes and opinions, when the research is of a sensitive nature, and when probing techniques are included as part of the data collection process,[800] as is the case in the current study. Semi-structured interviews enable the interviewer to build rapport with the participants,[801] and lead to richer data than might be gained from a more structured approach, a survey or a questionnaire. In addition participants can discuss topics pertinent to them, that the researcher may not have anticipated.[802] In the context of this research, a semi-structured approach enabled exploration of the factors influencing professionals' decision-making and multidisciplinary working during cases of suspected AHT, and their views and opinions about PredAHT, allowing issues that were important to the professionals to arise.

5.3.2 Participant recruitment

In qualitative research the selection of participants differs from quantitative sampling procedures because the aim is to explore the full range of opinions about a phenomenon rather than to count opinions or people.[803] An adequate sample size is not measured by frequencies but by the richness of the data acquired; the purpose is to gather sufficient depth of information in order to fully describe the phenomenon under study.[804, 805] Therefore, purposive and snowball sampling were used to recruit participants for this study. Purposive sampling is a non-probabilistic sampling technique, the objective of which is to produce a sample that can be "logically assumed to be representative of the population".[806] Participants are selected based on their knowledge and experience with the phenomenon of interest.[807] Snowball sampling is another non-probabilistic sampling method whereby participants recommend colleagues who are eligible to take part in the study.[808] Clinicians, pathologists, legal practitioners, police officers and CPSWs likely to be involved in suspected AHT cases across south west United Kingdom were targeted, as these are professionals who may potentially benefit from the use of PredAHT. The sampling frame was generated by compiling a list of 175 potential participants, who were identified through personal contacts of the research team and organisational websites. Contacts were sent an information sheet to explain the study (Appendix 7), and were asked to suggest suitable participants for interview. A modified version of the information sheet was sent to clinicians who were also recruited for the vignette study reported in Chapter 6 (Appendix 8). To generate a sample to approach for participation, the sampling frame was stratified by profession and seniority. Clinicians were also stratified by hospital type (teaching versus district) and specialty (community paediatrics, general paediatrics, radiology, emergency medicine, neurosurgery). Within these strata, a random sample of individuals was drawn, roughly in proportion to the size of the stratum. For example, a greater number of clinicians were approached as there were a greater number of clinicians in the sampling frame compared to individuals from the other targeted professions. Individuals were excluded if they were unlikely to ever be involved in a suspected AHT case, as PredAHT would be of no relevance to them. Individuals were contacted via email, with the exception of judges who were sent formal letters of invitation.

Although purposive sampling is the most frequently used form of non-probabilistic sampling, guidelines for determining sample sizes using this method are scarce.[809] The concept of "thematic saturation" or "data saturation" is often referred to in the qualitative literature, whereby interviews cease at the point when no new insights are obtained or there are no new patterns identified in the data, however it is often poorly operationalised.[810, 811] Kuzel[812] recommended six to eight interviews for each homogeneous sample but provided no evidence to support this recommendation. In an experiment with data saturation, Guest et al.[809] found that 94% of their high frequency codes were identified within six interviews. Based on this, in the present study, the aim was to recruit a minimum of six professionals in each group; however, recruitment and interviews continued until no novel insights from each group were generated from the data. In the interests of transparency, the procedures and criteria used to determine and achieve thematic saturation are fully described and explained in section 5.3.6.3.

5.3.3 Interview schedule

The semi-structured interview schedule was developed by the researcher and a visiting student, discussed within the supervisory team and revised accordingly (Appendix 9). Questions were derived from the scientific literature on the identification of AHT and the acceptability of CPRs, discussed in Chapters 2 and 3, respectively. The schedule was piloted with one police officer and one clinician, regarding the appropriateness of the length and content, and amended accordingly. The schedule was comprised of core open-ended questions, prompts and probing or clarifying questions.[813] Open-ended questions allow respondents to explain their own views and experiences as fully as possible, prompt questions ensure that key issues are addressed should the participant fail to provide initially detailed responses, and probing questions are used to elicit further insight.[813] Interviews explored four main themes 1) factors influencing participants' decision-making and multidisciplinary collaboration in suspected AHT cases[187] 2) participants' evaluations of PredAHT 3) participants' opinions about how to present the calculated probabilities, and 4) participants' interpretation of probabilities in the context of suspected AHT.[747]

Although a specific set of topics was explored during the interviews and the questions were planned well, in semi-structured interviewing the schedule need not be strictly followed.[814] The researcher is able to deviate from the interview schedule to investigate other areas that may be raised by participants and to explore the concepts that arise as each interview evolves.[815] Responses from earlier interviews influenced the questions of later interviews, and the interview schedule was continually revised and updated as data collection and analysis progressed.

5.3.4 Data collection procedure

All interviews were conducted by the researcher, who has received extensive training in qualitative research methods and qualitative interview techniques. No relationship was established between the interviewer and participants prior to the study. All participants were given the participant information sheet to read and keep (Appendix 7 and 8). Informed consent was obtained, including permission for audio recording (Appendix 10). Participant demographic data were collected in order to describe the characteristics of the sample, as well as information regarding participants' CP experience and training (Appendix 11). Prior to commencing the interview, the interviewer explained the purpose of the interview and digital recorder; clarified the topic under discussion and the length and structure of the interview; and gave assurance of confidentiality. The interviewer also gave a short demonstration of PredAHT to each participant on a laptop computer, briefly described how it was developed and explained its intended purpose. All participants were told that the tool is for clinicians to complete, and that the results are not intended to be used in isolation from other information gathered as part of an investigation into suspected AHT.

Interviews were conducted in the participants' office or a meeting room at their workplace. A visiting student was present to record relevant field notes such as participant non-verbal behaviour, response to the interview, and critical reflections about the interview. Two of the interviews were small group interviews, one including five lawyers (two barristers and three solicitors) and one including three circuit judges, as these participants were available for interview at the same date and time. During group interviews, personal interaction between the participants was minimised, to maximise individual contributions from each participant. When two participants declined to be audio recorded the interviewer made detailed notes of their responses and sent these to the participants to check that they were a fair reflection of their views. Both participants responded by annotating and editing the notes to provide clarification. All participants were given a debriefing form explaining the purpose of the study at the end of the interview (Appendix 12). Interviews lasted approximately 45 minutes. Interviews with CPSWs, legal practitioners, police officers and pathologists took place between June and September 2015; these professionals were shown Version 1 of the computerised PredAHT, as described in Chapter 4. Interviews with clinicians took place between May and September 2016; all clinicians were shown Version 2 of the computerised PredAHT, as described in Chapter 4. No repeat interviews were conducted.

5.3.5 Ethical and governance issues

5.3.5.1 Ethical approval

This study received ethical approval from the Cardiff University School of Medicine Research Ethics Committee (Ref: 15/35) on Wednesday 27th May 2015 (Appendix 13). An amendment to the study protocol was submitted on 21st August 2015 to request recruitment of clinicians in addition to the other professional groups already included. The amendment was approved on Friday 9th October 2015 (Appendix 14).

5.3.5.2 Data management

Data were stored and managed in accordance with the Data Protection Act 1998. Audio recordings and transcripts were securely stored on a password-protected, confidential Cardiff University server. All audio recordings, transcripts, consent forms and demographic data will be held securely for 15 years, in line with Cardiff University research data policies. After this, all data will be destroyed.

5.3.5.3 Ethical considerations

Child abuse and particularly AHT is an emotive and sensitive subject and there were some ethical issues around discussing suspected AHT cases with participants. Participants were informed that the recorder could be stopped at any time and were reminded of their right to withdraw from the study at any point during the interview or beyond. Building rapport and gaining the trust of participants was imperative due to the highly emotive and often personal topics raised. For example, some participants reflected on past cases of AHT that they or their colleagues had misdiagnosed, and the consequences of this. It was important that the participants did not feel that they were being tested or judged on their knowledge or decision-making practices. Participants were assured that they did not have to answer any questions that made them feel uncomfortable.

Prior to commencing the interview, participants were informed that should they wish to discuss a particular case, they were not required to impart any identifiable details of the persons involved, and were given assurance of strict confidentiality. The ethical principles of Cardiff University were upheld at all times and the study was carried out according to the principles of the Declaration of Helsinki.[816]

5.3.5.4 Reflexivity

In qualitative research it is imperative to adopt a reflexive approach to the collection and analysis of data.[797, 817, 818] Reflexivity is the process of critical self-evaluation of the researcher's background and positionality, including characteristics such as age, gender, race, personal experiences, values, knowledge, beliefs, and biases, and explicit acknowledgement that this position may impact the research process and findings.[819-823] By accounting for the researcher's assumptions and preconceptions, reflexivity is used as a means to enhance the rigour, quality, accuracy and ethics of a research study.[819, 821, 824] Reflexivity can help to ensure the research process is ethical by minimising potential power imbalances between the researcher and the participants.[821] Spencer et al.[818] identified three components of reflexivity: being aware of how biases may arise, attempting to minimise the impact of the researcher on the data collected, and attempting to address bias through systematic and comprehensive analysis and reflection on the research methods used.

Throughout the study, the researcher considered how her own characteristics, assumptions and preconceptions as a PhD student researching decision-making in AHT and developing PredAHT might influence her interview responses or the interpretation of the findings. The researcher was familiar with the literature on decision-making in AHT at the outset of the study, and therefore had opinions regarding how decisions should be made in suspected AHT cases. Where the researcher's knowledge of the subject area exceeded that of the participants' knowledge, this had the potential to create a power imbalance between the researcher and participant. Going into interviews with clinicians, the researcher was acutely aware of her status as a non-medical student with no medical training but nevertheless conducting PhD research in a medical topic, and how this may affect the power relationship between the researcher and the participant. Similarly, going into interviews with so-called 'elite' participants such as judges and legal practitioners, the researcher was very conscious that these professionals are often accustomed to exerting control and influence over others and considered the impact this status imbalance may have on the interview schedule and procedure.[825] Finally, going into all interviews, the researcher had a vested interest in PredAHT, believing and hoping that it would be useful for the participants when investigating suspected AHT cases. It was therefore difficult to receive criticism or negative opinions about PredAHT. The researcher kept a reflective journal in an attempt to minimise any potential bias, to develop her personal interview skills, and to capture the key points and essence of each interview immediately following completion.[813, 826] An excerpt from the journal is included in

Appendix 15. To break down power imbalances, every effort was made to build a rapport with the participants, and ensure that the interview was guided by them, but stayed on-topic. The researcher made sure not to ask any leading questions or impose their own views on the participants. During 'elite' interviews it was important that the researcher was flexible with regard to time and interview structure, and good etiquette was maintained to ensure high professional standards.[827] During data analysis, reflexivity was essential to avoid "unconscious editing" based on the researchers position.[828] The researcher inspected the data and the interpretations developed for possible competing conclusions and involved other qualitative researchers in the analysis to reduce subjectivity.

5.3.6 Data analysis

Data were analysed using thematic analysis[799] facilitated by the Framework Method[829] and the constant comparative method.[830]

5.3.6.1 Epistemological position

It is imperative within qualitative research to acknowledge the epistemological position underpinning the analysis of the data.[799] Crotty[831] defined epistemology as the theory of knowledge embedded in a methodology and a methodology as a strategy or process underpinning the choice of methods. Epistemological positions include realism and constructionism, while ethnography, discourse analysis and grounded theory are methodologies, although grounded theory may also be classified as a method.[799, 831] Data analysis methods include thematic analysis, content analysis, interpretive phenomenological analysis, conversational analysis and narrative analysis.

In this study, the most appropriate analysis was considered to be a thematic analysis facilitated by the Framework Method, conducted within a critical realist paradigm and based on certain principles of grounded theory, namely the constant comparative method. Thematic analysis is an analytic method that aspires to identify, interpret and describe patterns and themes across data.[799] Similar to grounded theory, thematic analysis facilitates the generation of codes and themes that are grounded in the data.[799] However, grounded theory seeks to develop an overarching theory concerning the social processes involved in the phenomenon under study,[832] while thematic analysis does not. Braun and Clarke[799] proposed that researchers conducting a thematic analysis need not adhere to the full theoretical commitments of grounded theory if they do not wish to guide their analysis towards theory development. The current study was not concerned with theory generation but with understanding factors influencing professionals' decision-making and multidisciplinary working in suspected AHT cases, and their views and opinions about PredAHT. A thematic analysis of participants' narratives was thought to lend itself well to developing an understanding of these issues. A general inductive approach was adopted, enabling the analysis to be guided by both the aims and objectives of the research, and the raw data.[833, 834]

Neither thematic analysis nor the Framework Method are bound to an epistemological position.[799, 829] The epistemological position for this study was critical realism.[835] This position sits between the two poles of realism and constructivism, theorizing that while language constructs our social realities, these constructions are constrained and influenced by extra-discursive elements in the material world.[835] Thus, critical realism acknowledges "the ways individuals make meaning of their experience, and, in turn the ways that the broader social context impinges on those meanings, while retaining focus on the material and other limits of 'reality'".[799(p.9)] From a critical realist perspective, the concept of child abuse is a social construction defined within a historical, cultural, social and economic context.[32, 836] Denial in child abuse is a global problem, operating at the individual, official and cultural level,[837] and, as described in Chapters 2 and 3, individuals exhibit bias and variability in investigating and identifying abuse. So-called extra-discursive elements that influence the social construction of abuse include physical elements such as bodily abnormalities, e.g. fractures and bruises; material resources and technologies such as microscopic cameras and cranial scans; and institutions and social infrastructures such as governments and laws.[838, 839] Medical professionals, CPSWs, police officers and legal practitioners each have their own knowledge and professional perspectives regarding the meanings ascribed to the material and physical aspects of abuse, which are not neutral but embedded in power relations.[839] Thus, critical realism was chosen as an epistemological position to acknowledge that "the reality of child abuse is made both in the physical happening of it and also in how we perceive and respond to it".[838]

5.3.6.2 The Framework Method

Data analysis began shortly after the first interview was conducted, and followed a highly systematic and rigorous procedure. The Framework Method was used to manage, summarise, display, and synthesise the data and to facilitate analysis.[829] The Framework Method is ideally suited for supporting thematic analysis as it provides a structured model for managing interview data, and it is particularly useful to support analysis of large datasets.

Quotes from individual participants are organised by themes and categories, generating a matrix style output.[840] This enables in-depth analysis across the whole dataset while also allowing for comparisons within and between individual participants and groups.

Analysis proceeded under seven phases: transcription, familiarisation, coding, developing an analytic framework, applying the analytic framework, charting data into framework matrices, and interpretation.[829] Audio-recordings of interviews were transcribed verbatim; the researcher checked each transcript for accuracy and ensured familiarisation with the data by listening to the audio-recordings. Initial codes were generated independently by the author of this thesis and two other researchers trained in qualitative methods; a code is a word or short phrase that captures the essence of a portion of qualitative data.[841] Data were first coded manually using hard copies of the transcripts. Manual coding is recommended alongside the use of assistive software in order to permit researchers to explore the data in different ways.[841]

The three researchers compared codes during joint analysis sessions, and grouped initial codes into clearly defined categories using axial coding, the process of relating codes to each other.[841] Categories were then further arranged under the four main overarching themes explored in the interviews (described in section 5.3.3). Discrepancies between coders were resolved by discussion and consensus. This process was undertaken in an attempt to minimise individual biases in the results; 38% of the transcripts were independently double-coded. The joint analysis also enabled the development of a preliminary analytic framework. Categories and themes from the preliminary analytic framework were then applied to subsequent interviews by indexing each transcript. Transcripts were imported into the data analysis software package NVivo.[842] NVivo was used to organise and manage the data and assist with data analysis. Quotes pertaining to each main theme were retrieved and summarised or 'charted' into thematic framework matrices. Each theme had a separate matrix, with categories as column headings and participants as rows (e.g. Appendix 16).

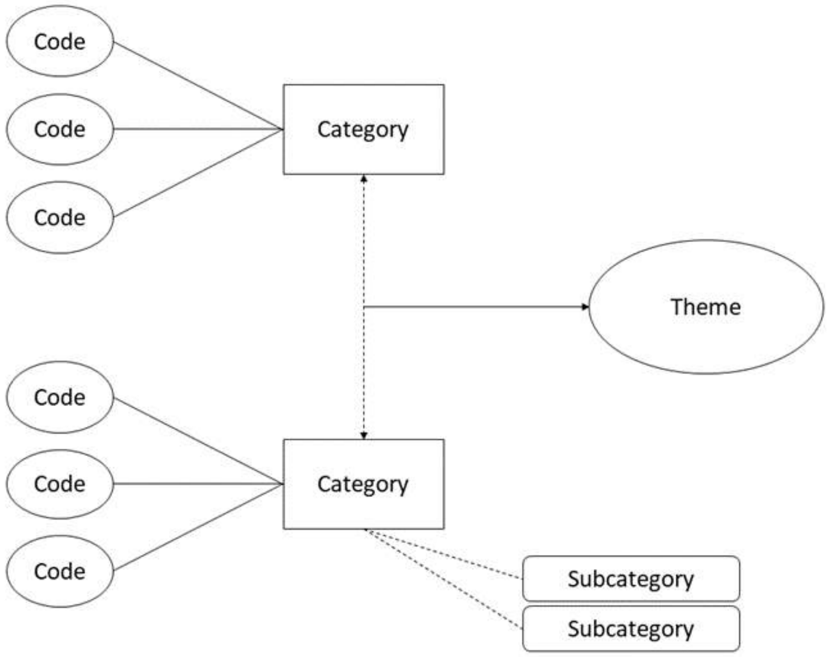
5.3.6.3 Constant comparison

As more interviews were conducted, new codes and categories were generated from the data. In order to verify thematic saturation, the constant comparative method was employed, a method originally stemming from grounded theory.[843] Constant comparison involves systematically comparing cross-sections of your data in order to hone your categories and themes, and is greatly facilitated by the Framework Method through the review of data across thematic matrices.[829] A purposeful approach to constant comparison was undertaken.[830] Previously coded transcripts were continually revisited and compared with new transcripts to check whether any of the newly generated codes applied to the data and whether existing categories could be developed or refined. Data within and between the professional groups were compared to search for patterns and commonalities or differences. Categories were compared with each other to explore potential relationships between them and assess whether they could be collapsed or integrated into other categories, and quotes under each category were compared and reviewed for consistency and coherence. Throughout the analysis process, analytic memos were written to capture any thoughts, ideas or reflections about the developing categories and their relationships. When additional interviews failed to generate novel categories and data were easily assigned to existing categories, thematic saturation was reached and recruitment ceased. As new categories were developed, the analytic framework was amended accordingly (Appendix 17).

5.3.6.4 Themes, categories and subcategories

The final phase of the analysis involved abstraction and interpretation of the data. Factors that were perceived to influence decision-making in cases of suspected AHT (Theme 1) were categorised as either ‘professional’, ‘medical’, ‘circumstantial’, ‘family’, ‘psychological’, or ‘legal’ factors. These factors were reviewed to identify barriers and facilitators to decision-making. Participants’ evaluations of PredAHT (Theme 2) were categorised as ‘potential benefits of PredAHT’, ‘potential risks of PredAHT’, ‘provisos for the use of PredAHT’, ‘use of PredAHT in court’ and ‘clinicians’ views about the practical use of PredAHT’. Participants’ opinions about how to present the calculated probabilities (Theme 3) were categorised as ‘percentage probabilities versus broad risk categories’, ‘confidence intervals’ and ‘additional suggestions.’ Finally, participants interpretation of probabilities in the context of suspected AHT (Theme 4) were categorised under ‘threshold criteria’ and ‘comments about PredAHT scores’. The systematic synthesis of the data excerpts into thematic matrices enabled a final inspection of the categories across cases, to identify subcategories, i.e. the range of different elements being described under each category. All subcategories and their definitions are detailed in the analytic framework (Appendix 17). The relationship between codes, categories, themes and subcategories is depicted in Figure 5.1. Samples of analysed transcripts from each of the professional groups are included in Appendix 18. Participants were not asked to provide feedback on the study findings.

Figure 5.1 Relationship between a code, category, theme and subcategory



Adapted from Saldana [841].

5.3.7 Reporting

This study is reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines[844]; a checklist is included in Appendix 19.

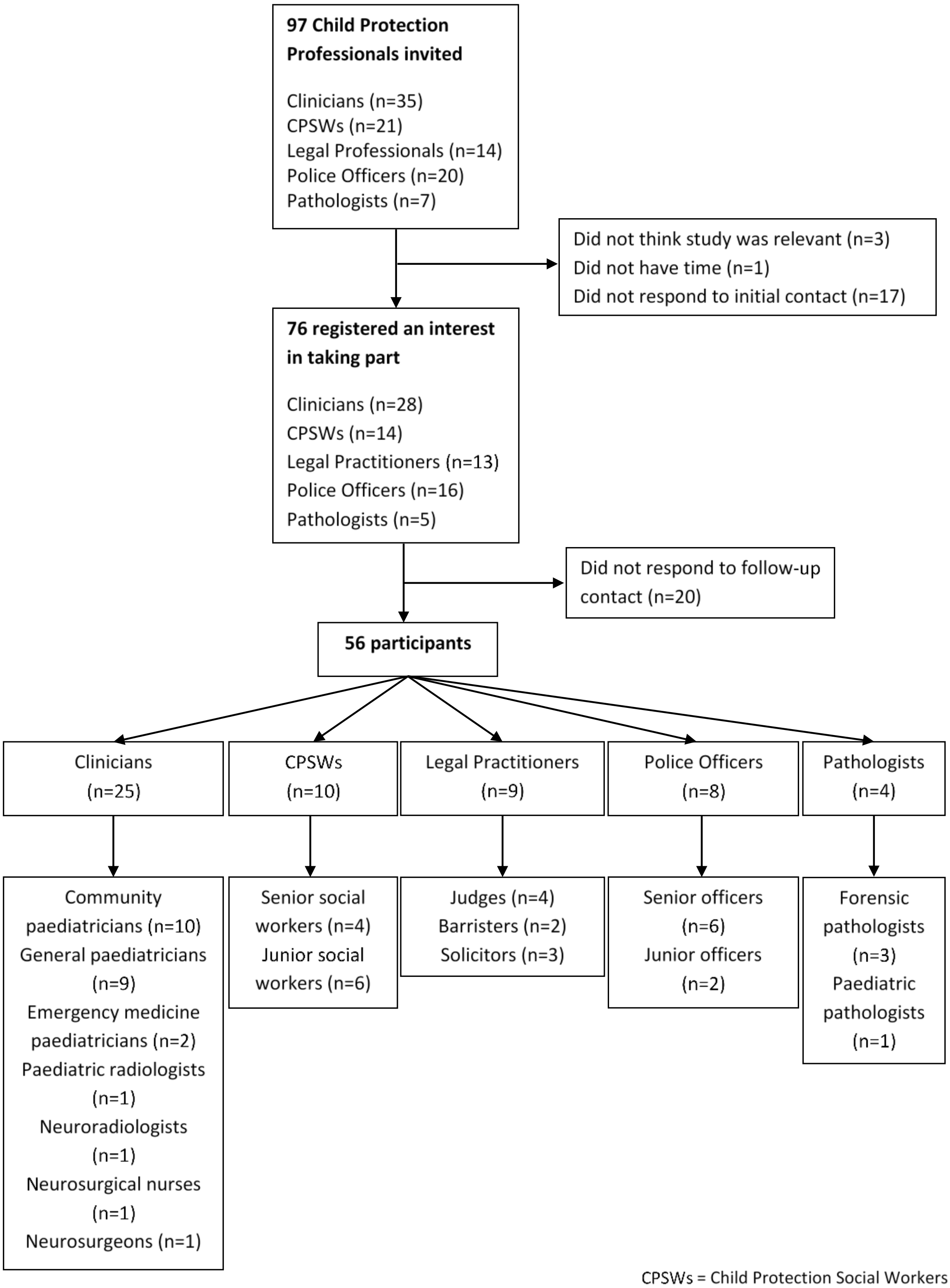
5.4 Results

Data are presented using quotations, selected as examples of the categories that were generated from the data. Within the quotations, square brackets represent text inserted by the researcher for clarification. Word repetitions and irrelevant sections were removed and denoted by '…'. Subcategories are highlighted in bold text. Additional quotations are presented in Appendix 20.

5.4.1 Response rates

Response rates to interview invitations for the five professional groups are shown in Figure 5.2. Initial response rate across the five groups was 78% (76 of a total of 97 professionals approached). Of those who initially responded, 74% (56 of 76) participated, including 25 clinicians, ten CPSWs, nine legal practitioners (including four judges), eight police officers and four pathologists. Most participating clinicians were experienced consultants or associate specialists, two were trainee doctors and one was a nurse. Senior CPSWs were team managers, who had a greater number of years of CP experience than their junior counterparts. Senior police officers were inspectors or chief inspectors, and junior police officers were constables or sergeants. Judges had the greatest legal seniority with the responsibility of delivering the judgment on the evidence submitted by barristers or solicitors, and forensic pathologists had more experience of CP investigations than the paediatric pathologist. Nine participants were recruited from south west England, while 47 were recruited from south west Wales. Data regarding the demographics and characteristics of non-responders and non-attendees were not collected.

Figure 5.2 Flowchart of child protection professionals participating in an evaluation of the acceptability of the Predicting Abusive Head Trauma clinical prediction tool



5.4.2 Participant demographics and characteristics

Participant demographics, and their CP experience and training are reported in Table 5.1.

5.4.3 Participant familiarity with clinical prediction rules

Altogether 15/25 (60%) clinicians, 1/4 (25%) Judges and 2/4 (50%) pathologists were familiar with CPRs and overall had a positive opinion of them. None of the CPSWs, barristers, solicitors, or police officers were aware of them. Six of 25 clinicians had previously used a CPR.

Table 5.1 Demographics of child protection professionals participating in an evaluation of the acceptability of the Predicting Abusive Head Trauma clinical prediction tool

	Clinicians (N=25)		CPSWs (N=10)		Legal Practitioners (N=9)		Police Officers (N=8)		Pathologists (N=4)	
	n	%	n	%	n	%	n	%	n	%
Gender										
Female	16	64	7	70	7	78	3	37.5	0	0
Male	9	36	3	30	2	22	5	62.5	4	100
Age group										
25–34	2	8	2	20	2	22	0	0	1	25
35–44	11	44	5	50	1	11	5	62.5	1	25
45–54	8	32	1	10	4	45	3	37.5	1	25
55–64	4	16	2	20	2	22	0	0	1	25
Ethnicity										
White British	19	76	10	100	8	89	8	100	4	100
White Other	4	16	0	0	1	11	0	0	0	0
Indian	2	8	0	0	0	0	0	0	0	0
Years in CP										
<5	0	0	2	20	1	11	3	37.5	0	0
5–9	6	24	3	30	1	11	2	25	2	50
10–20	7	28	4	40	4	45	3	37.5	0	0
>20	12	48	1	10	3	33	0	0	2	50
CP training										
Yes	25	100	10	100	3	33	7	87.5	4	100
No	0	0	0	0	6	66	1	12.5	0	0
Pediatric HI training										
Yes	18	72	1	10	3	33	4	50	3	75
No	7	28	9	90	6	66	4	50	1	25

CPSWs = child protection social workers, CP = child protection, HI = head injuries.

5.4.4 Theme 1. Factors influencing decision-making and multidisciplinary collaboration in suspected AHT cases

Participants discussed a wide range of factors perceived to influence their decision-making and multidisciplinary collaboration in suspected AHT cases. 'Barriers' to decision-making were defined as factors that were thought to complicate the decision and contribute to remaining undecided about whether AHT had occurred. 'Facilitators' on the other hand were defined as factors that made the decision easier and could lead to a diagnosis being made.

5.4.4.1 Professional factors

Participants' **perceived role in the decision-making process** differed by professional group. All community and general paediatricians agreed that it is within their remit to come to a decision as to whether a child has suffered AHT, as part of a multidisciplinary team. Emergency medicine specialists, radiologists, the neurosurgeon and the paediatric pathologist would raise concerns with other colleagues, but not make the final diagnosis. Two forensic pathologists would provide a steer to other professionals, while one stated that it was not their job to make decisions about abuse. Barristers and solicitors, the neurosurgeon and one forensic pathologist believed that it is the role of the court and ultimately the judge to decide whether a child has suffered AHT.

> "I suppose in every case you wonder whether that's happened, but it's not for us to ultimately make that decision, we just have to present the evidence and it's for the judge to make the decision at the end of the day."
> ***Legal Practitioner 2***

Judges and CPSWs reported that their role in suspected AHT cases is to protect the child from future harm, rather than to determine whether AHT has occurred *per se*.

> "What the [family] court has to decide is…has this child suffered significant harm? Or, does the evidence disclose, based on facts that you can find that there is a real possibility of significant harm in the future…so in terms of us deciding was this a non-accidental injury or an accidental injury, in some cases it won't make any difference to a decision that we have to make because you can have a very serious accident that will occur as the result of an inappropriate carer, or an unsafe carer, or a lack of supervision." ***Judge 3***

Many participants reported that they are **reliant on other professionals** throughout the decision-making process. CPSWs and police officers stated that they are heavily reliant on medical professionals to come to a decision as to whether AHT has occurred, and to guide their decision-making. This is due to their own lack of medical training and knowledge. The more experience these participants had investigating suspected AHT, the more knowledge they appeared to have.

> "To support my decision-making I would rely heavily on what consultants are telling us, what the experts are telling us about those injuries and what the likely cause is, and what's acute, what's not, what's explained, what could potentially cause this. So yeah, major, major reliance on that clinical information, I can't emphasize that enough…I would rely quite heavily on that expert view, and the views of those medical professionals with child protection experience." ***Police Officer 7***

> "We're not trained medically to know whether something's accidental or non-accidental. We can have an opinion on it, but it won't be based on research and training." ***CPSW 9***

> "I'm not a doctor, I'm not a medical expert, I'd want some clear guidance from the doctors about what they think, but having some experience now, no training, but some experience of dealing with these cases, I'm able to ask some relevant questions of the doctors." ***Police Officer 6***

Judges explained that they rely on medical professionals to conduct a timely and high quality clinical investigation in suspected AHT cases, to facilitate the decision-making process in a court environment. In addition, legal practitioners and CPSWs said that they expected medical professionals to be able to categorically determine the cause of the child's injuries by the clinical features alone.

> "We have experience of saying to the medics to pin down to an absolute, this is the way it happened...and they will always say...I can't tell you that."
> ***Legal Practitioner 1***

However, clinicians and pathologists highlighted that other professional groups shouldn't be relying solely on them to come to a decision about suspected AHT.

> “What I do with the police with these cases is actually say to them look, you shouldn’t be relying on me. A case depends on lots of different bits of evidence.” ***Pathologist 3***

Many CPSWs, police officers and legal practitioners, including judges, said that decision-making is more difficult when medical professionals are unable to provide them with a clear answer as to whether AHT has occurred or not, or when they will not commit to a view either way.

> “Often if there are clear injuries and the medics are actually saying that it is non-accidental then there is a clear process for us to follow. That makes it easier, it makes it a lot harder when health professionals are sitting on a fence.” ***CPSW 3***

However, CPSWs and police officers explained that the majority of the time medical professionals will express their suspicions to other agencies, which facilitates the investigative process, and if clinicians remain unsure, CPSWs and police officers said that they would continue their investigations regardless.

> “I mean, 9 times out of 10 it’s fairly self-evident. I was able to have generally an open and honest discussion with the paediatrician, that paediatrician would say, ‘In my opinion, this is what you’ve got. Either it’s non-accidental or I’m concerned its non-accidental’. In which case they’re both dealt with in pretty much the same way and investigated appropriately. It’s fairly straight forward…But if they say, ‘I’m unsure’. Then we still run with it anyway.” ***Police Officer 3***

> “We usually get perhaps an initial medical report to say it’s felt that these injuries are non-accidental…so initially you do get a concern that it is non-accidental.” ***CPSW 6***

Although clinicians stated that they rely on other agencies to assist them in making decisions in suspected AHT cases, they reported that they seek support and advice from clinical colleagues to a greater extent.

> “We can always speak to colleagues and we’re never in it by ourselves…we’re always in discussion with colleagues. I have never been in

the situation where I've said, 'Right, I'm the only one making that call.' You're always discussing it with other people and so that certainly lessens the burden on you when you have to make those decisions. Even though you may be the person who is called to go to a strategy meeting or the case conference or actually to court you will have had those discussions with colleagues as well…you become confident because we reassure each other that that's the diagnosis." ***Clinician 5***

CPSWs and police officers said that overall, they have had positive experiences with **multidisciplinary collaboration**.

"We had, and we still do have, very good working relationships with health professionals, with paediatricians in particular." ***Police Officer 3***

"Generally I find it's quite positive working with other agencies around safeguarding children." ***CPSW 2***

"I think we've generally got a great relationship with the forensic pathology team." ***Police Officer 7***

However, a handful of participants felt that multidisciplinary working is a barrier to decision-making due to competing interests and disagreements between professionals both within and across agencies.

"Working with other agencies [is difficult] really, sometimes coming from competing backgrounds and also from here even you know, decision-making by managers is not always the same it can be varied." ***CPSW 3***

CPSWs and police officers also reported that delays can occur while the other agencies are carrying out their own assessments, which impacts on the overall investigation.

"We're guided a lot by medical staff; waiting for their statements to come through…It can take a long time. It can take months sometimes, you get an initial report but it's very much really not until towards the end where you really know what you've got." ***Police Officer 5***

Police officers reported that they are often influenced by the amount of **resources** they are able to put into an investigation. They reported that more severe cases are better resourced.

> "Do you say, 'This is definitely non-accidental?' In which case you're going to put a lot more resources in it. Or is it one of those really difficult ones to gauge and you know you're not going to have the budget to do absolutely everything." ***Police Officer 2***

Some participants felt that coming to a decision about whether AHT has occurred is extremely **difficult**, while others did not.

> "I think it is probably the most stressful and difficult set of decisions that one has to make in medicine. Partly due to the difficulty of coming to conclusions." ***Pathologist 3***

> "It's not that it's difficult, I think it just needs a lot of consideration and a lot of thought and weighing of the evidence rather than the actual decision being difficult. If the evidence is strong enough, I think the decision to be made is not that difficult." ***Clinician 6***

Participants reported that their **confidence** when investigating suspected AHT cases is strongly related to the amount of experience they have.

> "I'm confident in dealing with the family, knowing my role, knowing the role of other professionals, but that might just be because I've been doing this for such a long time." ***Police Officer 5***

> "I'd say I was not very confident working on these cases, without a doubt, because I haven't worked on many physical abuse cases." ***CPSW 9***

In terms of professional decision-making strategies, participants' discussed the importance of '**seeing the bigger picture'** in suspected AHT cases, and piecing together evidence from various different sources.

> "It's a bit like a jigsaw puzzle to put together a number of different pieces of evidence to see if you can get any closer to the truth." ***Judge 3***

5.4.4.2 Medical factors

Participants mentioned a range of **clinical features** that they understood to be concerning for abuse, including intracranial injuries, encephalopathy, retinal haemorrhages, fractures, apnoea, seizures, spinal injury, and evidence of external injury such as bruising, bites or burns. They also recognized that specific patterns of injury are suspicious for abuse, including posterior rib fractures, metaphyseal fractures, patterned bruising and certain distributions of intracranial and retinal haemorrhages. Some also mentioned that fractures or intracranial bleeds of different ages are indicative of previous abuse or multiple incidents of abuse.

> "If we are beginning to be concerned about abusive head trauma we would get an eye examination, so the presence of any retinal haemorrhages would be corroborative evidence, but particularly multi-layer widespread dot, blot and flare haemorrhages, other evidence of intraocular bleeding…I would be expecting or might see multiple focal thin layer subdural haemorrhages in different brain compartments." ***Clinician 10***

Clinicians and pathologists explained that they refer to the literature and evidence-base on the different types of injury seen in abusive and non-abusive trauma when investigating suspected AHT cases, which gives them more confidence in their decisions.

> "I've been through a lot of the literature about it…so that has helped me in feeling maybe more confident about these cases." ***Clinician 7***

> "First of all I see whether there is any injury and decide what sort of injury it is, whether it's a blunt force injury or sharp force injury etcetera and then the distribution of the injuries on the body, and then relate the distribution that I find with what I know about the literature on different patterns of injury for assault or accident, falls." ***Pathologist 1***

Some participants, particularly police officers, said that they have a high suspicion of AHT when the "triad" of subdural haemorrhages, encephalopathy and retinal haemorrhages is present. However, importantly, these were not the only features that these participants said they considered when coming to a decision about AHT.

> "Once you get the triad of injuries and everything else, if you've got some attending injuries that appear to be evidence of abuse, in my view that would fuel the theory that there has been abusive head trauma. So for example if I had a child present in hospital with a head injury and they had bruising elsewhere on their body, that would make me more concerned about the head injury because of the evidence of abuse elsewhere." ***Police Officer 3***

One CPSW alluded to the "triad" of injuries, demonstrating a lack of training on the clinical indicators of AHT, an out-dated view of the features of the "triad" as diagnostic for AHT, and a lack of knowledge of the potential differential diagnoses of retinal haemorrhages.

> "We haven't had proper training on this...I had training when I was studying my degree, but I had it drummed into me that if there's subdural haemorrhaging, retinal haemorrhage, it's abuse. Am I right in thinking that there can be no other organic cause for retinal haemorrhage?" ***CPSW 1***

Participants said that it is important to rule out organic medical conditions in children with suspected AHT, listing a variety of **differential diagnoses** they would consider, including blood clotting disorders, birth trauma, and glutaric aciduria among others.

> "We would need the bloods, we would want to be screening for a significant coagulation disorder you know these kids often have a coagulation disorder after the event, so you need to confirm whether the coagulation disorder returns to normal after the child has been resuscitated. You would probably want to go back and re-examine the child looking for evidence of connective tissue disorders, you would want to review the family history, is there anybody in the family with a coagulation or connective tissue disorder." ***Clinician 10***

One pathologist pointed out that abuse can still occur even when a child has an organic condition.

> "I think that sometimes it is forgotten that even with natural pathology, it doesn't preclude there being something deliberate to go with it." ***Pathologist 2***

Similarly, participants reported that they often **link the injuries to the mechanism** purported by the carers, and deliberate over the plausibility of this.

> "If I saw a head injury where there was not obviously impact, I would be looking for corroboration of the application of force somewhere, because as soon as you find a bruise or a fracture, or a graze or a split at whatever level in the body skin, soft tissues, the skeleton, you have what is undeniably the application of force and that helps you…the pattern of rib fractures indicates that there has been compression of the chest…the head injury may indicate impact or deceleration." ***Pathologist 3***

> "They could be toddling couldn't they, if they fell down the stairs from the top floor to the bottom they could have a brain injury couldn't they depending on the floor downstairs, they might have got a stone floor…but you wouldn't expect to have the other stuff there." ***CPSW 8***

Participants reported that a clear factor influencing their decision-making is the **severity of the injuries** sustained. They believed that an intracranial bleed or rib fracture in a young child are serious injuries, and the more clinical features a child has, or the more impaired they are, the more likely participants said they are to suspect AHT.

> "My decision really would be based on the fact that I think you're talking about trauma here, a brain injury, if that's where we're looking at it, it's a really serious condition isn't it? So you've got to do everything you can to make sure that that child is safe, and there's no risk that this could happen again." ***CPSW 9***

Participants discussed **dealing with uncertainty** in medicine, and in the CP arena in particular, stating that so-called "grey" cases, where there is considerable uncertainty surrounding a diagnosis of AHT, are the most difficult.

> "Medicine is rarely black and white, there are shades of grey in the middle of it, and often these are quite dark grey. You're pretty certain, there was a while where I felt slightly less sure, but most of them I felt reasonably confident, given if there's that constellation of injury, in the absence of an adequate explanation." ***Clinician 17***

> “Regularly in child protection we find ourselves in a situation where it’s not 100% clear the person’s guilty or the offence has happened, neither is it 100% clear that there is an innocent explanation, and left this grey area in between. Well there’s still a risk, something possibly, or probably happened, but we can’t prove it, can’t rule it out, so where do we go with that…you know the grey area ones.” ***Police Officer 6***

Clinicians reported that it is hard for them to convey to other agencies that a case may not be clear-cut.

> “We are often trying to explain things to people who don't necessarily understand the uncertainty in medicine like police and social workers.”
> ***Clinician 11***

5.4.4.3 Circumstantial factors

Participants discussed the specific **circumstances surrounding the incident** in suspected AHT cases, including the explanation given for the child’s injuries and details of their presentation to the hospital.

> “What I’d be looking to do is looking at the accounts that have been given to the attending officers, the accounts given to the paramedics, what’s been said on the 999 call [emergency number], what’s been said when they first attend, usually they speak to the Accident & Emergency paediatrician, what they then say to the community paediatrician and my officer when they get there.” ***Police Officer 1***

Some participants considered a delay in presentation to hospital to be an important risk factor for AHT.

> “The other thing we always worry about is a delay. So we have had the odd few children that have presented a few days later because of a significant swelling and while that is possible that would raise a flag in your head.”
> ***Clinician 16***

Participants discussed the behaviour and appropriateness of the parents and the interactions between the parent and the child throughout the investigation.

> "We look at exploring the family dynamics, the response of the parents during the immediate child protection enquiry, the interview." ***CPSW 1***

Participants reported that the single most important factor that influences their decision-making when AHT is suspected is the **history**, in particular whether the mechanism of injury is consistent with the type and severity of the injuries or the developmental stage of the child.

> "Probably the single most important thing, is the detailed history. And the features of that, the description of what's happened, how possible and plausible that is, is it compatible with the injuries, is the child capable of what's described in terms of their developmental stage? So I think the detail of the history is really, really important." ***Clinician 12***

However, participants stated that they find decision-making difficult when the history is consistent with the injuries, but the mechanism could nevertheless be either abusive or accidental.

> "The difficult ones are where they come in and say 'I've fallen down the stairs with my baby' because you think if somebody has just lost it with a baby and smacked them against a wall, and is switched on and intelligent and actually quite manipulative, they probably would come up with a story of 'I've just dropped my baby'. So those ones are always a bit more difficult because you think it might be true, on the other hand it might not be." ***Clinician 13***

When there is no history of a traumatic event whatsoever, participants said that they have a very high index of suspicion for AHT.

> "The lack of disclosure is a biggie, you know the child who presents with collapse and then you subsequently find that they have subdual haemorrhages or a fracture or broken ribs, that makes you very concerned that it's not the whole story being told to you." ***Clinician 24***

Participants reported that another influential factor is whether the history is consistent over time and/or between caregivers.

> "I might be wanting to speak to a nurse, so 'You spoke to the parents when they first arrived, now you and the paediatrician have spoken to the parents here', is there any discrepancy between the two stories, or are they consistent, or is mum saying one thing and dad saying another thing?"
> ***Police Officer 6***

5.4.4.4 Family factors

Participants discussed the importance of the families' **social history** when investigating suspected AHT cases. They talked about a wide range of issues including parental drug and alcohol use; parental mental health; domestic violence; previous involvement with social services; parent-child interactions; level of supervision of the child; neglect; socioeconomic status; and parental criminal history.

> "I would be concerned if there was also then a family history of family violence, if I was getting background social history that there was known abuse in the past, or I guess if this baby had been more vulnerable for whatever reason, was maybe a pre-term or indeed if this was a mother who's quite young, not supported, new partner, and partner's not the biological father of this baby. They are things that I would actually…they'd be helping with the diagnosis. It wouldn't necessarily tip it but they would obviously add to my concern that my feeling is this is likely to be the case."
> ***Clinician 2***

> "Obviously if there's domestic violence, substance abuse, a history of neglect, that's obviously going to shoot up in terms of our assessment."
> ***CPSW 1***

CPSWs and police officers said that they place more emphasis on factors within the family than clinical factors.

> "I would probably have a better understanding of the context in terms of the family scenario, in terms of levels of supervision and what it's actually like within the household." ***CPSW 5***

However one police officer mentioned that they would give less weight to the social history of the family during the investigation.

"The social background is less important, because if we're investigating if there's been abuse or not, it's determined principally by the injuries, by an explanation, by the evidence, not by whether the parents are employed, whether the parents are smokers, or if the dad is an alcoholic, whether there has been domestic violence in the house, those are interesting background features, perhaps more likely to be prevalent in some cases than others, but it's not going to tell me abuse has happened or hasn't happened." ***Police Officer 6***

Similarly, participants believed that a lack of history with social services or a lack of a criminal record does not rule out AHT.

"It's not always families that come revolving door, we have families as well that have not been known to us for years, or never been known and then they've harmed a child." ***CPSW 4***

Participants stated that a major factor influencing their decision-making in suspected AHT cases is the **impact on the family**. They discussed the impact of removing a child from the family home, and how intervening in a child's home life could be damaging for the child and family, particularly where a head injury is found to be non-abusive. Participants explained that the decision to remove a child from their parents is not taken lightly, as it may not be the best thing for the child.

"I know accidents happen with babies and children with the best will in the world and what you don't want to do is if a family is already traumatised by something that the child has experienced and they're doing the best for them, to add in the trauma of querying the abuse factor could just tear the family apart." ***CPSW 9***

Clinicians, CPSWs and police officers reported that they find it difficult **working with the family** and having to treat parents as potential suspects or perpetrators when they are grieving or coping with a seriously unwell child. Participants talked about the need for sensitivity and the potential repercussions of falsely accusing a family of abuse.

> "It's not so much the clinical diagnosis it's managing it and being the one who talks to the parents and is having to deal with their anxieties, their uncertainties and all their anger." ***Clinician 10***

5.4.4.5 Psychological factors

Participants stated that their decision-making in suspected AHT cases is influenced by their own **personal biases**, such as a disbelief that parents or carers from 'nice, middle-class families' are capable of inflicting injuries on their children.

> "Well they shouldn't but if it looks like a really nice family that you couldn't imagine doing anything like that and that shouldn't influence you but it makes you think. People say 'Oh I've seen a case like this before' or they say 'No, no the family is too nice'. And other people will be saying 'But don't be fooled by it', all this goes on, I hear it all the time." ***Clinician 9***

However, most participants reported that they acknowledge these biases and attempt to remain objective in their assessments.

> "We always keep an open mind, we always continue to gather information and if there is new information, it will change our decisions." ***Clinician 10***

CPSWs and judges explained that they find cases difficult when they only have medical evidence to rely on, and there are no other risk factors that they are able to identify within the family.

> "There have been cases where we've removed children begrudgingly because of medical evidence and genuinely from the way the parents are with the child, their backgrounds you just don't think they did it. So that's very difficult ethically having to remove a child on the basis of a medical decision where there's nothing else to substantiate that." ***CPSW 1***

However this CPSW also stressed that even in the absence of other risk factors, she would remain suspicious and continue with her investigations.

> "I would be led very much by medical evidence and even if there was no other risk factors identified for that child, I would not be willing to take any risk on a case like that." ***CPSW 1***

Clinicians, CPSWs, and particularly police officers reported that they are influenced by their "**gut instincts**" when conducting their investigations and assessments.

> "I'd probably rely on my professional suspicion…my gut feeling…If I had an inkling something was not right then we would be doing more." ***Police Officer 3***

Some participants reflected that investigating suspected AHT is **emotionally demanding** and can be a barrier to remaining objective in these cases.

> "There is emotion attached to them, so seeing children who are injured whether it is accidentally or deliberately, there's an emotional component to that. I find it difficult because I am intrinsically a relatively trusting and non-suspicious person and I've had to train myself to just take the emotion out of it, and deal with whatever facts are available." ***Clinician 12***

5.4.4.6 Legal factors

Police officers and legal practitioners disclosed that **identifying the perpetrator** in suspected AHT cases is particularly difficult.

> "The difficulty in my experience isn't identifying it, it's in establishing who's done it." ***Police Officer 3***

Legal practitioners and especially judges, reported that they rely on **expert witnesses** to provide an interpretation of the clinical features, but noted that there are often disagreements and conflicting opinions between expert witnesses coming from different disciplines.

> "There will sometimes be subtleties, particularly in the expert evidence that we get and you will have two extremely eminent experts sometimes from different disciplines, sometimes the neurosurgeon has a different view from the radiologist. I can remember doing a case in which they'd looked at the same scan and said I don't think we can agree what's there…So those are the difficulties that you have to encounter when you get a range of opinion on the interpretation of the medical evidence." ***Judge 3***

Judges and police officers referred to the various theories that are put forward by the parents or the defence in an attempt to disprove cases of suspected AHT.

> "Something needs to be looked at because this hasn't happened because the child has got gastro-oesophageal reflux, which was one theory which used to be propagated at one stage because if a child had gastro-oesophageal reflux it might stop breathing and that would lead to a rise in intracranial pressure which would then give rise to the bleed and we had that theory at one stage, not from the medics but that was one that was often propagated." ***Judge 3***

5.4.5 Theme 2. Participants' evaluations of PredAHT

Participants discussed how PredAHT might have a positive or negative impact on the investigation of suspected AHT, the decision-making process, the child or the family, whether it would be useful for them or not and why, their willingness to use it, potential problems that may arise from its use, and how it might enable them to overcome specific barriers encountered when working on suspected AHT cases. Many participants would only use the tool with a proviso. Participants talked about the implications of using PredAHT in court, and clinicians discussed the practical implications of its use.

5.4.5.1 Potential benefits of PredAHT

Participants believed that one potential benefit of PredAHT is its **objectivity**. They felt that PredAHT would be useful to support decision-making as it is not influenced by personal feelings or opinions and could help reduce subjectivity in the assessment of risk.

> "I think they're a good idea because they can be completely evidence based, so it takes all your feelings out of it because it's…child protection, there's lots of emotions." ***Clinician 9***

> "It would be helpful if a medical professional would have some confidence in saying it's an 85% chance because we would all understand what the chances were, because sometimes at strategy discussions you might get a paediatrician who will give an opinion, but as we all know, we take in messages in a different way. I might go back and record it in some way, the social worker might go back and record it in a slightly different way." ***Police Officer 3***

Participants felt that PredAHT would be helpful for heightening **awareness** of AHT, and reinforcing or increasing concerns about possible AHT.

> "Where the number sits would help us to articulate that suspicion and perhaps work as a bit of a check. Probably in pushing it up, and highlighting to some people actually, you should be more suspicious because this is really unusual to get this combination." ***Police Officer 1***

Participants thought that PredAHT could work both ways, encouraging participants to consider the possibility of non-abusive injury if PredAHT gave a low probability score for AHT.

> "It would also be helpful for us not to panic too much in the sense of there is the idea of accidental head injury as well. So it's also helpful for us to take a step backwards and not think right it's abusive trauma." ***CPSW 1***

Clinicians, CPSWs and police officers said that PredAHT would provide them with **reassurance** or confidence that their concerns, suspicions or investigations were justified and that it would be useful to support their professional opinions.

> "I think as you used it more it would give you more confidence that actually, yes this is confirming that my level of suspicion is appropriate for the case…It would give you more confidence in making those decisions clinically." ***Clinician 15***

The majority of clinicians said that they would use PredAHT to back up their clinical opinion rather than to direct their decision-making. However, some said that they would find the score helpful if it did not agree with them.

> "I wouldn't use it just to sway my opinion, but if I had an opinion of whether it's abusive or not, and then, using this validated tool, with the injuries found and the presence of head injury, it is likely, so that helps back up your opinion and hopefully then would add more weight to what you're saying." ***Clinician 21***

> "If there was a mismatch between my clinical opinion and the risk assessment tool that would cause me to stop and think and seriously consider whether I have gone down a bit of a blind alley with this and

whether I need to stop and think again…It would be helpful just to reassure us that we are doing the right thing or maybe to cause us to stop and think actually perhaps we shouldn't walk away from this one." ***Clinician 10***

Participants felt that PredAHT would be useful for explaining, justifying or **rationalizing decision-making** in suspected AHT cases.

"Family courts, criminal courts might want to know how have you arrived at this decision and if I was asked, well, these are the facts...and I've documented clearly why I've made a decision about something. Any tool I think that helps give some…statistical interpretation for police, for social services, for the medical professionals, it's robust and trustworthy I suppose then, I only see that as a good thing because we're all accountable for the decisions we make." ***Police Officer 7***

Clinicians and pathologists suggested that PredAHT may help to **standardise** or modify the clinical assessment of suspected AHT cases by prompting clinicians to perform investigations such as a skeletal survey or ophthalmoscopy in line with international standards, and to review the results of investigations already undertaken.

"Is this patient a patient that may have been abused and if so [the tool] triggers safeguarding procedures for siblings and it flags up this is a child who is going to need an ophthalmologist to look in her eyes, a skeletal survey, and a child protection paediatrician on call. If that triggers all of those things to take place that would be great." ***Pathologist 1***

"We would just do all those investigations on anyone under 1, but it's in that 1-3 [age group], where you're just that bit more unsure, whereas should we be doing these things and it might actually guide us." ***Clinician 1***

Pathologists and legal practitioners, including judges, said that they could appreciate the value of PredAHT for advocating further investigations, even if they would not find it useful themselves.

"I think if it can be used to ensure that front line clinicians are actually encouraged to undertake exploration of what they've got at a better level then we'd be saying yippee absolutely that's the best that you can do for us…If you can get clinicians to actually do what they should be doing when they should be doing it and triggering the protocols that need...because we see quite a lot of missed opportunities with the initial investigation and you can't go back and do it again." ***Judge 2***

Similarly, police officers and CPSWs said that PredAHT may help to **justify further action** within their respective agencies.

"If we've got a figure that says actually there's an 80% chance that there's abusive trauma, then that child isn't safe at all…we need to be taking pro-active action and that would I think be supportive." ***CPSW 3***

Participants felt that PredAHT could contribute to **"the bigger picture"**, as part of a wider information gathering process. Many described PredAHT as a useful addition to the "toolbox", or "a piece of the jigsaw puzzle".

"It would form one part of your prosecution case wouldn't it? It wouldn't be an enormous part but it could form a part of the evidence you'd built generally…I can't see for a minute that it wouldn't be useful." ***Police Officer 4***

Participants suggested that PredAHT could be used at multi-agency meetings or as part of information sharing to facilitate **communication** about the likelihood of AHT.

"It would be valuable for talking to the police, social workers...just to say, 'Listen we've got this…' Because they will always say to you, 'Is there anything else it could be? Are we getting this wrong? Are we missing something medical?' I think when you're able to say with a degree of certainty, 'No, this is what it is because this is a validated tool. With this combination of injuries this is how confident we can be' I think it is going to be valuable for them as well...and you share it in the strategy meeting that would be very useful for me." ***Clinician 5***

Participants thought that PredAHT would be useful for peer review or **training**. CPSWs in particular thought it was helpful to know that the six clinical features included in PredAHT are potential indicators of AHT.

> "I think it will be very good for all child protection social workers dealing with these to know about these six things." ***CPSW 1***

Participants believed that PredAHT may be most beneficial for so-called **'grey' cases**, where there is considerable uncertainty surrounding the diagnosis, and most beneficial for those with the **least experience** in CP.

> "You get those ones where you think 'this is really not likely to be this' but you've got to go through the steps, and the ones where there clearly is likely to be a problem, so it's those grey ones in the middle where this might come in more useful than the clear cut ones." ***Clinician 1***

> "It might be helpful for someone who's never done paediatrics before and doesn't have the experience and the benefit of having done child protection work before and knowing these things…I think it is helpful for a very specific group of people." ***Clinician 20***

5.4.5.2 Potential risks of PredAHT

Participants were concerned that professionals may be **over-reliant** on PredAHT when making decisions in suspected AHT cases.

> "If there was too much reliance placed on it at the beginning of an investigation and someone with little knowledge simply populated those fields present, absent, features, and came out with a low probability you know 14% or whatever and decided not to investigate, regardless of the presence of other factors not in your fields, then that would be foolish and dangerous, so the tool itself is not a disadvantage, it's how its uses could be." ***Police Officer 6***

> "I think there's a potential for people to make it the single most significant part of the decision-making process so we'd end up sat in meetings and people would ignore most of what I said and say 'What does the tool say? Oh 67%, right that's the decision made'. That would be my worry that people would over-use it or overstate its importance." ***Clinician 12***

Some participants thought that a low score could instil **false reassurance**, and that appropriate investigations might not be carried out in the face of a low probability score.

> “If somebody is uncritically using this tool and they have got a child with an intracranial head injury, and head and neck bruising they are not worried then because the score says that it’s less than 15% then that would be an incorrect use of this tool.” ***Pathologist 1***

However, reassuringly, all participants said that they would still carry out appropriate investigations if they received a low score from PredAHT, as there may be other features of the case that warrant further enquiry.

> “If it was low like that, but there were previous allegations of abuse, dad had a violent background, that sort of stuff. That would sway me towards being quite concerned about this. So, I guess it’s about the attendant circumstances around it. So if it was low, it would help me, but I would still look at the bubble around it and what’s going on.” ***Police Officer 3***

Participants expressed concern that PredAHT wouldn’t be **used as intended**, alongside other known information about each case, and they agreed that it should never be used in isolation from other factors. Several clinicians, pathologists and CPSWs said that PredAHT was too reductionist and crude, comparing it to a box-ticking exercise.

> “My main concern is people not taking into account the history or the other facts because they’ve got a big number on this score.” ***Pathologist 2***

Participants questioned the **accuracy of PredAHT**. The sensitivity and specificity that participants would be willing to accept was discussed, together with the implications of false positive and false negative predictions.

> “The key thing is in how many cases is this wrong? And if it’s wrong in any, then you’ve just got to ask yourself is this safe?” ***Pathologist 4***

> “Hopefully this will help us find all cases of abusive head trauma but there is a chance that we might label a non-abusive one as an abusive one as well, but I would probably weigh the benefit more than the risks, because if this is helping me to identify the really vulnerable children, I would rather use it…as long as I am protecting the vulnerable children, I would find it useful.” ***Clinician 22***

A few pathologists and clinicians, including a neurosurgeon, stated that PredAHT is **irrelevant** for them as they do not need a tool to make decisions about suspected AHT, and would therefore not use PredAHT.

> "That's what we do in our brains, we put all the information together and spit out the probability based on our experience." ***Clinician 19***

> "It's something that is an irrelevance to me in that, one might take the view that this is an attempt to make my task less onerous by placing in my mind the conclusions of others, or their interpretation of the evidence, when it is my role to look at that evidence and the literature myself. This I fear might be regarded as a substitute for individual thought." ***Pathologist 4***

Two clinicians were unsure about how much PredAHT would add to the investigative process, and could not say whether they would use it in practice or not.

> "I have to say my initial thought looking at it is I'm not sure how much more it would add if you've done all the investigations already." ***Clinician 13***

Some participants thought that there are important clinical and historical **features not included in PredAHT**, e.g. skull fractures, bruising patterns, spinal injury, or a history of trauma, and questioned why they were not included.

> "You don't have a history of extraordinary trauma as an option. So the other thing is a non-declaration of the history that would be massive wouldn't it. Or no history of any injury…just woke up and the baby was like that. That would be a massive predictor I would imagine." ***Clinician 24***

Participants pointed out that PredAHT cannot take into account specific details of the clinical features, including the **age, number, location, pattern and severity of injuries**, some of which they believed may be highly specific for AHT.

> "What about a healing fracture as opposed to a recent fracture and position of the fracture, and particularly rib fractures are they at the front of the chest of a child who has had resuscitation or are they posterior ribs, and it's this granularity that we are grilled on and we have to take into account when we are giving our overall opinion but for a quick and dirty or

‘Should I be contacting child protection services, should I be contacting our paediatrician on call for child protection?’ This sort of thing is great.” ***Pathologist 1***

“Not all RHs are the same, you can have one in one eye and five in the other but if they’re not in the layers that you’d expect them to be…You’d have to make some allowance for, not only the categories but subcategories of that...It needs to be more refined.” ***Judge 2***

Some pathologists and judges said that PredAHT may condition their decision-making or inadvertently **introduce bias** into the decision-making process.

“You would almost make it more difficult for the judge because the judge would then have to disentangle the expert opinion from either an apparent bias or an unconscious bias that might be established by the fact that the expert had looked at the clinical tool.” ***Judge 3***

5.4.5.3 Provisos for the use of PredAHT

Many participants stated that they would only use PredAHT **with a proviso**; e.g. alongside their professional judgment, with more information about the definition of the six features, if it was kept up to date, with knowledge of the quality of the data on which PredAHT is based, with an understanding of how it works and how it is to be used, after agreeing acceptable risk thresholds with multi-agency colleagues, and if it was accepted by their colleagues (Table 5.2)

Table 5.2 Child protection professionals' provisos for the use of the Predicting Abusive Head Trauma clinical prediction tool

Alongside professional judgment	"I guess it's probably a combination of that along with a bit of professional judgement tied in…If you look at it in combination with other professional opinion, what else you know, what information you found out, then it could inform part of that pool of information." ***Police Officer 3*** "Yeah I think it's useful, it should not stop you from thinking I think you should still think outside the box and not 100% rely on it but I think as an additional tool to your clinical decision-making, I think it is supportive and helpful." ***Clinician 7*** "This in addition to the rest of our assessment is really, really helpful." ***CPSW 1***
Definition of features	"I would need a bit more information about some of the…you know I don't really know what retinal haemorrhages are." ***CPSW 6*** "Apnoea, presumably you'd have a definition of how long that's for and stuff like that?" ***Pathologist 2***
If kept up to date	"There's issues of keeping it up to date, you can't just do it once and then not revisit it, can you?" ***Clinician 3*** "To keep its credibility it would have to evolve with current thinking, so it's a continual process isn't it?" ***Police Officer 2***
Quality of the data	"I'd want to look at the original research and how the original cohort of patients were diagnosed with abusive head trauma and what's the robustness of that diagnosis in the first place, that the tool is then based on." ***Clinician 12*** "I would never use something like this without reviewing the publication and looking at the statistics and checking out that I was personally happy with the statistical analysis, because otherwise I'm just putting stuff into boxes." ***Pathologist 2***
Understanding how it works	"I would have to understand it and be able to explain it in court, so I'd need to come and have a little training session." ***Clinician 3*** "You would need to explain the unknown parts of it…as well you could do a small tutorial based on four or five cases if people want to get experience on how to use it." ***Clinician 14*** "It's important to understand what informs the figure, because otherwise it becomes a checklist…I think people need to have an understanding of what the tool is and how it is to be used." ***CPSW 5***
Agreeing accepted risk thresholds	"We can all have that figure and we can all explore then what that figure means to each independent agency, and what it means for that child and actually what safeguards need to be in place because of it." ***CPSW 3*** "I think within a team, there needs to be consistency as to what it's meaning at that point in that time…if we were using unknown, then it's giving a steer towards it…but what I would be concerned about is interpretation of what this means." ***Clinician 2*** "You want that consistency and agreement as to what the results could mean." ***Police Officer 1***
If accepted by colleagues	"It would be only useful for us if it's accepted by the medical profession as being acceptable for them." ***Legal Practitioner 2*** "Yeah, I think you'd want, whether it be the safeguarding board or the child death overview panel, you'd want something where the social workers and the paediatricians and the police all come together and agree that this is useful." ***Police Officer 1*** "Yeah I think that would give me the most confidence really if the medical professionals were on board with it." ***CPSW 6***

5.4.5.4 Clinicians' views about the practical use of PredAHT

All clinicians thought that PredAHT is **simple to use**, and not too time-consuming to complete.

> "Time is always a disadvantage in getting people to fill these out sometimes, but this is relatively simple and straightforward so I don't imagine it being a huge issue." ***Clinician 14***

Participants discussed the **hospital settings** in which PredAHT would be useful. The majority thought that PredAHT would be most useful for inpatients admitted to a ward or Paediatric Intensive Care Unit (PICU), and less useful in the Emergency Department (ED), where information about fractures and RHs is unlikely to be available. However, one emergency medicine paediatrician thought that PredAHT may have a role in the ED to prompt an initial referral to the safeguarding team.

> "If they were unwell enough to go to PICU you may use that tool much less in the ED. If they were somebody that was going to a ward then you would probably use it more. I think it would depend on the patient and how sick they were." ***Clinician 16***

There were different views regarding the **stages of the assessment process** that PredAHT would be most useful. Some clinicians said that they would only use it once all relevant investigations were completed, to assist with report writing or reaching their final conclusions.

> "In my opinion, there's not much point in using it if you have too many unknowns there...I personally would certainly like it for when I have to write my report." ***Clinician 7***

However, other clinicians said that they could see the value of PredAHT at **multiple stages** of the assessment process, and that they would use it more than once during a case to support their decision-making.

> "I would probably use it as soon as I knew about the case, just to give me some idea, and then as more data is collected you could add it in and see how it changes your figure, and then you've got your last kind of figure

then is what is going to be the most important one right at the end." ***Clinician 1***

Participants discussed **who should complete the tool**. They thought that PredAHT could be completed by general and community paediatricians and intensivists, but most agreed that it should be completed by a consultant. One clinician thought that it should be a team exercise.

"Whether it would be something that would be used by the lead consultants in PICU or a general paediatrics consultant where they are thinking do we need to get the safeguarding team involved or not and then potentially I suppose used by a safeguarding consultant when it came to writing up." ***Clinician 8***

Clinicians reflected on how PredAHT could be **integrated into the clinical workflow** and implemented in clinical practice. Although they acknowledged that each hospital has its own way of working, most thought that it would not be too difficult to incorporate the computerised version of PredAHT into existing hospital intranet systems. Some suggested including reminders or specific references to PredAHT on existing departmental or safeguarding paperwork.

"It could well go on to the intranet as an app…and maybe a reference to it as a little reminder on our safeguarding proforma." ***Clinician 4***

Some clinicians could see the value of including their own **prior probability** score in the calculation, but the majority felt that this would introduce too much subjectivity into PredAHT and that they would need guidance on how to use this element of the tool.

"I think that's an important element to bring in because actually a lot of our decisions are often based around the history and does the history fit, is it consistent, stuff about whether they're presenting late, stuff about what the family background and social history is." ***Clinician 13***

"I'd feel happier with the six features on its own, because I know that's very evidence based, isn't it, so that's fine. I do have an issue with the prior probability without some objectivity around it, because it's easy to think oh well, a child on the Child Protection Register and that ups my concerns and whereas a middle class family doesn't, and I think it is something you need to be really objective about." ***Clinician 3***

5.4.5.5 Use of PredAHT in court

Most clinicians, CPSWs and police officers thought that PredAHT would be useful in court, because it is **evidence-based** and validated. CPSWs in particular felt that it would be useful in the family courts for future safeguarding of children, where the **standard of proof** is based upon the balance of probabilities.

> "In the court arena I think it's going to be really very useful because it's not our hunch against the next doctor's hunch, you know? And I think people's general opinion that babies aren't injured by their care-givers and their parents...people don't want to hear that and they certainly don't want to believe it and acknowledge that this is happening, but if you've got a validated tool saying, 'Actually this is what has happened to this baby because of the other injuries that we've seen' then I think it's going to be very valuable indeed." ***Clinician 5***

> "It helps when going to court with the balance of probabilities if you can prove over 51%, that's the number I have in my head…that's what we've got to convince evidence of a judge of." ***CPSW 2***

Pathologists and legal practitioners, including judges, expressed caution regarding its use in court, particularly in the criminal courts.

> "What the criminal standard which is beyond reasonable doubt would make of that, because the decision in these circumstances would be that of a jury, again huge caution in thinking this is effectively steering a jury into saying it's 85%, it's beyond reasonable doubt therefore we've got no choice but to convict." ***Legal Practitioner 1***

However, one judge thought that PredAHT would be useful, in both the family and criminal courts.

> "It will help to remind the courts and the experts that a certain combination of features does make abusive head trauma a more likely explanation...I think it would have the same role in the criminal courts.

Even though the standard of proof is different, it would still be useful at the fact finding stage.” ***Judge 4***

Some felt that PredAHT would be **irrelevant** because it cannot account for every detail of every case, and each case must be considered based on the entirety of the evidence.

“As lawyers we would probably want to treat it with extreme caution. Simply because this tool cannot cover every factor in every situation that we have to deal with.” ***Legal Practitioner 1***

Others remarked that PredAHT may not stand up under **cross-examination**, or that the defence will claim that their case falls into the reverse probability of non-AHT given.

“You would have to prove the tool in every case. You’ll be cross-examined about how it’s been put together, how you’ve weighted the factors. There’s always something that somebody can find if you’re really trying to pull something apart. Then it goes out the window really evidentially.” ***Judge 2***

“We would be arguing well why isn’t this one in the 15% of cases that suggests that it isn’t non-accidental?” ***Legal Practitioner 1***

In addition, some participants believed that PredAHT will not help to **identify the perpetrator** in suspected AHT cases.

“That doesn’t help us with who caused it, it just says ‘what’s the probability of it being an abusive trauma’ so there is that other element we have to consider.” ***Legal Practitioner 1***

Some clinicians and judges discussed **historical CP court cases involving statistical evidence**, and the impact and implications of such cases on the subsequent acceptance of statistics in the courtroom.

“A slight worry any paediatrician will have, a study putting statistics up like that, is the way that [Roy] Meadow [UK paediatrician] was chopped down with statistics.” ***Clinician 17***

Despite their reservations, the majority of the legal practitioners interviewed, including judges, said that they would probably take the PredAHT score into account if it was included as part of a **medical court report**.

> "The way the courts see these matters from a child protection point of view is an analytical approach where you need the best evidence possible. If this is something that feeds into a medical report, by an expert who understands it, then I'm delighted to have that. Hopefully it can help to make the right decision for the family, because it is life changing." ***Judge 4***

5.4.6 Theme 3. Participants' opinions about how to present the calculated probabilities

Participants said that they preferred precise percentage probabilities of abuse, rather than broad risk categories such as low, medium or high likelihood of abuse. While some suggested presenting both, they were unsure as to what percentages would equate to low, medium and high. Some participants felt that confidence intervals would be unnecessary, however others felt that they would be an important addition. Additional suggestions included background information/data about PredAHT, disclaimers, and visual aids (Table 5.3).

Table 5.3 Child protection professionals' preferences for the presentation of the calculated probabilities from the Predicting Abusive Head Trauma clinical prediction tool

Percentage probabilities versus broad risk categories	"For a decision-making tool a percentage is spot on from my point of view, because I think everyone will understand it." ***Police Officer 2*** "A percentage just makes it a little bit more tangible, doesn't it, it's very real, because I find it difficult, I always have done, to quantify risk, so if it can be done for me, then yeah…" ***Clinician 6*** "I am happy with the percentage, I am confident with that because that is where we are at with safeguarding children, we are talking about the balance of probabilities and so we are talking anything over 50 per cent probability should lead to further action and further evaluation where you may still get information that shifts it the other way." ***Clinician 10*** "I'd personally prefer it if captured with likeliness and you might have very suspicious, strongly indicative or not likely." ***Clinician 12*** "I prefer it like that because low, medium, high can mean anything." ***CPSW 4*** "A percentage wouldn't be helpful in court, because it would actually lead to more uncertainty. There would always be an argument to have which would detract away from the purpose." ***Judge 4*** "I appreciate what it is, there could be a range but what would low, medium, high then be? I suppose that's the problem." ***Police Officer 7***
Confidence intervals	"I would like to know the variance, that's what we would like to know, that's what we would need to know…" ***CPSW 3*** "I do like the number as a percentage but I do like to know the confidence interval as well…" ***Clinician 7*** "I think it might be a bit too much information." ***CPSW 6***
Additional suggestions	"There should be a little caveat statement there saying that it can go both ways, the higher it is the more likely it is to be, but a low one doesn't exclude it." ***Clinician 23*** "I think you need some sort of disclaimer on it about this needs to be used as part of a full assessment." ***Clinician 8*** "You want to know where it's come from, what's the research basis behind it, what's the evidence behind it, how much can you trust it. Now that could be a short blurb and then links to the publications, the literature that supports this." ***Clinician 2*** "I would want to know the data behind it because it is obviously chunking and splitting the data in different ways, maybe if you have got all the individual data, you could list everyone who fell outside the non-accidental injury bracket?" ***Clinician 24*** "If you can say here's a big block of how many of these kids were deliberately injured compared to a little smidge of kids, you could almost support it perhaps with a quick graphic to go big block is battered kids, small block is unfortunate accident." ***Police Officer 4***

5.4.7 *Theme 4. Participants' interpretation of probabilities in the context of suspected AHT*

Participants had very different interpretations of probabilities and risk categories in the context of suspected AHT.

5.4.7.1 Threshold criteria

Although all participants maintained that they have a very low threshold for suspicion of AHT in young children with intracranial injuries, their **probability thresholds for suspecting abuse varied widely**. Some participants said that they would only feel confident to completely rule out AHT if the percentage probability was less than 1%, while others said that they had higher thresholds.

> "If it's something like in the thirties, gosh that's a really hard thing to factor in isn't it then, in terms of decision-making, it still sounds quite high to me. You almost want it to be a 0.1% chance of it being an abusive head trauma to feel confident in your decision, because even at 30%, that's like one out of three families, that was abusive isn't it?" ***CPSW 9***

> "If that said to me there's a 1% chance then there's still a 1% chance. It's helping me it's not telling me there's no chance is it?" ***Police Officer 8***

> "I want to say sometimes I'm not happy about not taking any further action at times, but I'd have to be I don't know maybe 20%?" ***CPSW 2***

Many participants said that they simply **could not put a figure on their threshold** for abuse, stating that if there was any chance at all that it could be, then they would investigate further, and commenting that each case is dependent on the attendant circumstances around it. One CPSW indicated that often her risk judgements are very different to her colleagues', highlighting inconsistencies within the assessment process.

> "I don't think I can put a figure on my threshold because it depends sometimes I look at something and think 'Why are we going out on this?' And then something else, 'We should've looked at that, why did we have all of this and we haven't done anything with it?'" ***CPSW 4***

Participants also had **different perceptions** of what the expressions 'low', 'medium' or 'high' likelihood of abuse might mean in percentage terms.

> "I say less than 50 and it would be low. Maybe 50–70 medium and then over 70 high." ***CPSW 6***

> "I suppose low, medium and high can mean anything can't it, I'm guessing its maybe up to 30%, 60%, 90%." ***CPSW 8***

5.4.7.2 Comments on PredAHT scores

After receiving a demonstration of PredAHT, participants offered their **opinions on the probability scores** that it gives for different combinations of features. They were told that intracranial injury with head/neck bruising alone gives a score of 14.7%. Some participants thought this score was low. Others interpreted the score to be an unacceptably high level of risk, while others still thought this figure could be interpreted in both ways.

> "I think that's really low 14.7%." ***CPSW 4***

> "From my point of view 14.7%, what do I think of that? I still think it's bloody high." ***Police Officer 1***

> "It might make people think, but then that means there's an 85% chance that it was accidental, doesn't it?" ***CPSW 9***

Participants stated that they were **uncomfortable** that PredAHT can give scores at or very close to **100%** for certain combinations of features.

> "99.6% and I've ticked rib fracture present, head and neck bruising present, apnoea present, seizure present. That I find hard you're saying essentially that's definitely abusive head trauma…I agree, I'd be very worried if I had that combination of features but I wouldn't say it's nearly a hundred." ***Pathologist 2***

Others stated that PredAHT scores **too low** for certain combinations of features.

> "So if you're fitting and have subdural haemorrhages but you don't have retinal haemorrhages or any other markers, it's saying it's not non-accidental injury. OK. I would be a lot more hawkish than that." ***Clinician 24***

> "I think that my findings were stronger than that score…the RH was very strong, in all layers…I might've hesitated if I'd seen 57%." ***Judge 2***

5.5 Discussion

The current chapter investigated factors influencing decision-making and multidisciplinary collaboration in suspected AHT cases and the acceptability of PredAHT amongst a range of CP professionals.

5.5.1 Overview of the main findings

The findings from this study suggest that CP professionals diagnose AHT based on knowledge of a wide range of clinical features described in the literature, features in the history, and risk factors within the family, after exclusion of potential differential diagnoses and discussion with colleagues from other specialities and disciplines. Participants viewed the diagnosis of AHT as a "jigsaw puzzle" that could only be solved with multiple different pieces of evidence. Barriers to identifying AHT included lack of experience, uncertainty, emotional factors, personal biases, the impact on the family and the fear of making an incorrect diagnosis, disagreements between professionals including expert witnesses, and alternative theories of causation proposed in court. Participants' experiences with multidisciplinary collaboration were reported as generally positive, however CPSWs and police officers reported being heavily reliant on clinicians to guide their decision-making, due to their own lack of medical training and knowledge. Facilitators to identifying AHT included support from colleagues, multidisciplinary working, knowledge of the literature and evidence-base, and "gut instinct".

The study findings also suggest that PredAHT would support the decision-making of clinicians, CPSWs, and police officers investigating suspected AHT, and provide them with greater confidence in expressing their opinion in the CP and court setting. PredAHT was viewed as a piece of the "jigsaw puzzle" of evidence, to be used in combination with professional judgment, subject to training and other provisos. Benefits were perceived by junior and senior practitioners with different levels of CP experience, and across all specialities with the exception of a neurosurgeon, although it was acknowledged that PredAHT may be most useful for those with the least CP experience. Potential risks included over-reliance and false reassurance from a low score. Pathologists and legal practitioners, including judges, thought PredAHT to be useful as a screening tool for ruling in further clinical or multidisciplinary investigations, however with the exception of one judge, they expressed caution regarding its use in court. In practical terms, clinicians found PredAHT to be simple to complete and thought it would be straightforward to implement into existing hospital systems. Participants interpretations of probabilities and risk categories in the context of suspected AHT varied greatly, but most would prefer PredAHT to give a precise percentage probability of AHT over terms such as ‘low’, ‘medium’ and ‘high’ as it helps them to quantify risk.

5.5.2 Comparison with the existing literature

The findings are consistent with barriers and facilitators influencing detection of physical abuse generally, and clinicians’ decisions to report suspected abuse identified in previous studies. Flaherty et al.[490] found that the decision to report suspected abuse was primarily influenced by the child’s clinical and social history and physical examination findings, particularly if their injuries were inconsistent with the history or their developmental stage. Barriers to detecting[845] and reporting[792] abuse described previously include personal biases, the fear of being wrong and the subsequent impact on the family, uncertainty about the level of suspicion and the difficulty of establishing a diagnosis, while facilitators include support from colleagues and other agencies.

The findings clearly refute the claims of some recent literature that AHT is diagnosed based on the “triad” alone[251, 252] and echo the categorical statements made by experienced clinicians who do not diagnose AHT solely on the presence of the “triad”.[56, 262, 265, 846, 847] The misconception was the subject of a meeting convened by the Royal College of Pathologists in 2009 to consider the issues appertaining to the “triad” and the “unified hypothesis” in non-accidental head injury cases, following which legal guidance was issued from the UK Crown Prosecution Service[57] on the prosecution approach to non-accidental head injury. This states that “the expert evidence finding of typical triad pathological features *might not be considered as diagnostic in itself* but simply as *strong evidence* that the injuries were non-accidental” (emphasis added). This view was reflected by clinicians and police officers in the current study. However, one senior CPSW described being taught at undergraduate level that the features of the “triad” are diagnostic for AHT. Although this may have been some time ago, this highlights how misconceptions become established, the differences between agencies and training gaps for social worker education in the clinical indicators and differential diagnoses of AHT, and suggests that their training should be regularly updated in line with the evolving evidence-base.

While many studies have evaluated the relationship between law enforcement and child protective services in suspected child abuse cases,[526, 791, 848] comparatively few studies have assessed health professionals’ perceptions of multidisciplinary working. Previous studies have described a hostile relationship between police officers and CPSWs, due to conflicting priorities and agendas, assumptions regarding the other’s role, and time delays.[526] Clinicians have also criticized social workers, describing them as unresponsive or inconsistent.[845] In contrast, the majority of the participants in the current study described positive relationships and experiences with other agencies. Only a very small number of participants felt that multidisciplinary working is difficult due to competing interests and disagreements, while a handful of participants brought up time delays as significant barriers to the investigation. Overall, participants' views of multidisciplinary working indicated that police officers and CPSWs consider AHT to be a medical diagnosis, and are heavily reliant on clinicians decisions; many believe that medics can determine the cause of injures by clinical features alone, and it is difficult for other agencies when clinicians "sit on the fence". Conversely, clinicians find it difficult to convey medical uncertainty to other agencies. This finding echoes the results of a recent study exploring collaboration between paediatricians and CPSWs, which demonstrated that CPSWs rely on paediatricians' opinions regarding accidental and abusive bruising, but that paediatricians felt CPSWs harboured unrealistic expectations about the diagnostic value of a CP medical examination to identify abusive bruising.[527] Clearly, joint training that provides knowledge about the individual roles and limitations of each agency would be valuable. Meanwhile, improved communication within and between agencies is critical for identifying patterns and preventing further injury[849] and has been recommended in the UK Safeguarding Children Research Initiative report.[850] The current study confirmed that PredAHT would be useful in strategy meetings to facilitate interagency communication about the likelihood of AHT.

Participants are more likely to suspect AHT and put greater resources into a case when the child's injuries are severe. However, it is well known that children can suffer repeated and escalated instances of abuse that eventually result in severe injury, and can sustain comparatively minor "sentinel" injuries such as isolated bruising or intra-oral injuries prior to a catastrophic injury.[180, 277, 278, 281] Sheets and colleagues[277] found that 30% of children diagnosed with AHT had previous sentinel injuries; where clinicians were aware of these injuries, either abuse was not suspected or was suspected but unsubstantiated, and their significance also went unrecognized by clinicians during the subsequent abuse evaluation.

If a comprehensive evaluation reveals no other medical explanation for a child's injuries, clinicians must decide whether the injuries are accidental or abusive.[53, 289] In determining this, participants reported that one of the most important factors influencing their decision-making is whether the history of the mechanism of injury is consistent with the type and severity of the injuries seen, or the developmental stage of the child. This approach has been deemed both medically and legally valid,[53] having been first described in a landmark article on the diagnosis of "battered child syndrome".[39] Participants were also influenced by a changing history over time or between caregivers, or no history of trauma. Studies have demonstrated that a history of no trauma is highly predictive of AHT, as is a change in history.[9] However, in a recent simulation study of accidental and abusive femoral fractures, eight participants gave a "changing history" as justification for their diagnosis, when the history did not change, while others incorrectly believed that the injury was inconsistent with the child's development.[535] While the majority of the participants in the study (39/43) considered whether the injury matched the mechanism, 15 nevertheless arrived at the incorrect diagnosis. Anderst and colleagues[535] suggest that a changing history provided by the caregivers should be distinguished from multiple histories taken by different individuals who may phrase their questions differently, and that clinicians should receive specific training in assessing injury plausibility.

Participants discussed a range of social risk factors within the families, usually regarded as facilitators to reaching a decision about AHT. However, some felt that these factors impeded their decision-making, since a family without risk factors could be abusive, while a family with multiple risk factors may never harm their child. Previous research identified the presence of risk factors as a complicating factor in detecting child abuse for some clinicians[851] although a recent study found that children referred for abuse evaluations without certain risk factors were just as likely to be diagnosed with AHT as those with risk factors.[62] PredAHT allows clinicians to factor in other features that are not included in the predictors, such as social risk factors and aspects of the history, by incorporating their own prior probability of AHT into the calculation. Clinicians felt this element of PredAHT was subjective, and were unsure whether they would be comfortable estimating a prior probability of AHT in light of potential racial and socioeconomic bias.[201] This suggests a lack of knowledge amongst clinicians of the evidence base regarding psychosocial risk factors for AHT. A recent survey found that less than half of health care professionals are adequately trained or prepared to identify risk factors associated with maltreatment.[852]

Some participants stated that they are sometimes influenced by their "gut feeling" when investigating suspected AHT cases. As described in Chapter 3, "gut feelings" are an affective response triggered by intuitive System 1 thinking processes, which are susceptible to a multitude of cognitive biases.[560] The evidence-based medicine literature therefore generally advises doctors against intuitive reasoning, and instead promotes the use of analytical models, clinical guidelines and decision tools.[563] A recent study demonstrated that child abuse paediatricians who had met the family and therefore had access to social intuition or "gut feelings" associated with a face-to-face encounter, were significantly less likely to perform adequate abuse evaluations for neuro-trauma and long-bone fracture compared to those who had not met the family.[622] However, studies have shown that "gut feelings" may trigger the process of diagnostic reasoning, prompting clinicians to perform further investigations.[853] One qualitative study exploring the identification and management of child abuse found that Dutch healthcare professionals' intuitive "gut feelings" often formed the basis of a more objective investigation and triggered a systematic process of evidence gathering.[851] Studies investigating the diagnostic value of features indicative of serious infection in children have found that clinicians "gut feeling" has added value over and above an overall holistic assessment of the child's presenting symptoms,[854] and that in a setting with a low prevalence of serious infection, it is an important diagnostic "red flag" in itself.[855] Dhaliwal recommends that clinicians adhere to the principles of evidence-based medicine while also understanding when it is appropriate to "go with their gut".[856] Of note, UK police guidelines for the investigation of child deaths contain evidence-based research which aims to "advance what is sometimes referred to as the 'detective's gut instinct' based on their assimilation of psychological factors and anecdotal information".[857]

The participants in this study did report that their decision-making in suspected AHT cases is influenced by their personal biases and emotions, such as a disbelief that parents from 'nice, middle class families' could harm their children. However, that they are aware of these biases and their potential pitfalls is encouraging, as it provides opportunities for monitoring, reflection and deliberative efforts to minimise their negative effects.[620] Participants described the application of strategies recommended in the literature to avoid errors resulting from bias, including attempting to remain objective, consciously considering differential diagnoses, and collaborating with multidisciplinary colleagues.[620] Many participants felt that PredAHT would help them to remain objective when assessing the likelihood of AHT.

An important issue influencing CP professionals' decision-making in suspected AHT cases is the proposal of scientifically unsupported alternative theories of causation for AHT in court. Judges and police officers alluded both to genuine diagnoses that lack scientific evidence to explain the injuries associated with AHT (e.g. Vitamin D deficiency) and unproven speculative hypotheses with no scientific evidence-base (e.g. that gastro-oesophageal reflux causes intracranial pressure leading to intracranial haemorrhage). The use of these flawed theories has created controversy in the courtroom and the media regarding the diagnosis of AHT, and has serious consequences for the upholding of justice and the protection of children.[269]

Clinicians stated that PredAHT would give them more confidence in expressing their opinions about the likelihood of AHT in their court reports and in court settings. While pathologists and legal practitioners, including judges, appreciated the value of PredAHT for encouraging standardisation of clinical investigations, and further clinical or social assessment, overall they expressed caution regarding its potential use in court. However, the majority would incorporate the probability score with all of the other evidence if it was provided as part of a medical report and PredAHT was accepted by the medical community. Previously, high profile cases involving the misinterpretation of statistics and probabilistic evidence in the courtroom have caused controversy[858, 859] and led to the development of a working group within the Royal Statistical Society in the UK, to improve the use of statistics in the administration of justice. They recommended a broad programme of education for judges, lawyers, and expert witnesses in probability theory and statistics.[860] One judge and one pathologist were concerned that PredAHT would introduce cognitive bias into their decision-making, suggesting that they may not trust PredAHT to be a valid piece of evidence for their decision-making.

The scientific literature confirms that there are specific patterns of intracranial injury, haemorrhagic retinopathy,[82] and spinal injuries associated with AHT.[26] Various psychosocial variables may also be influential.[198] Some participants wanted these additional features incorporated into PredAHT. However, two very large scale multi-centre prospective studies would be needed in order to add further variables and externally validate the updated tool.[295] Meanwhile, the six clinical features in PredAHT are easily identifiable in the early phase of clinical assessment and recent studies have further confirmed their association with AHT.[226, 861]

Previous research has identified barriers to the use of CPRs, some of which were identified in the current study, such as scepticism of “cook-book” medicine, belief that clinical judgment is superior to the tool, distrust of the accuracy of the predictors and concern that the CPR does not address all relevant factors.[299] Reilly and Evans[299] offer a number of strategies to overcome these barriers, including comparing clinical judgment with the CPR, and checking whether any excluded factors affect the CPRs predictions. Skull fracture was analysed within the original derivation study and did not discriminate between AHT and non-AHT.[59] An analysis comparing PredAHT with clinical judgment has been conducted and is reported in Chapter 7.

Participants had varying opinions about what percentage probabilities equate to the terms 'low', 'medium' or 'high' likelihood of abuse, and diverse probability thresholds for suspecting abuse, although participants stated that PredAHT helped them to quantify risk. This is consistent with previous studies that demonstrated that pediatricians struggle to define "reasonable suspicion" or "reasonable medical certainty" of abuse.[529, 530] Thresholds in CP social work have been the subject of much debate in recent years and are affected by a wide range of organisational factors, relationships with other professionals, and individual biases, heuristics and value systems.[862] Similarly, participants postulated that PredAHT may be most useful for "grey" cases, where there is significant uncertainty surrounding a diagnosis of AHT. In reality, the interpretation of the PredAHT score will depend upon individual perception of risk. PredAHT is designed to be an assistive tool rather than a decision rule, which typically recommends a direct course of action based on the results[299]; PredAHT provides no recommendations for professionals on what to do based on specific scores. Despite this, the majority of participants thought PredAHT would be useful for supporting their opinions and decision-making.

5.5.3 Strengths and limitations

This study used in-depth qualitative interview methods, which allowed for rich insight into the factors influencing CP professionals' decision-making and multidisciplinary collaboration in suspected AHT cases, as well as their detailed evaluations of PredAHT. The strengths of this study lie in the wide range of professionals interviewed, the detail, richness, and depth of the data, and the robustness of the data analysis. Survey-based methods do not allow for such a detailed exploration of participants views.[801] While other CPRs have been developed for the identification of AHT, as described and evaluated in Chapters 3 and 4, respectively, whether clinicians or other practitioners would be prepared to use these CPRs in practice is unknown. Investigators very rarely determine the acceptability of CPRs prior to their use. To the best of the researcher's knowledge, the current study is one of only two studies[747, 790] exploring the acceptability of a CPR developed for use in CP, and the first study to have done so with a range of professionals.

There are limitations associated with participant recruitment and the potential generalisability of the findings. Despite the inclusion of a range of professional groups and clinical subspecialties, additional groups could have made valuable contributions, for example neurologists, intensivists, or ophthalmologists, and their views and experiences may have differed compared to the groups included. Most clinical participants were consultants based in teaching hospitals with considerable CP experience, and less experienced clinicians were under-represented. While the aim was to recruit a minimum of six participants in each professional group, only four pathologists were recruited. The fourth pathologist interview revealed new insights that the other pathologists had not brought up, suggesting that data saturation was not reached with this professional group. This does not necessarily invalidate the findings for this group but rather means that further exploration of the topic may be warranted with these professionals.[811] In addition only a small number of specialists in radiology and neurosurgery participated. Since participation was voluntary, the participants may have had a particular interest in the identification and investigation of AHT compared to other professionals who did not take part, and they may not be representative of the general population of professionals involved in suspected AHT cases. However, it should be noted that probabilistic representativeness is not a goal of qualitative research.[863, 864] While non-probabilistic sampling methods are often criticised due to the risk of selection bias and therefore non-representative samples, qualitative research does not aim to make probabilistic generalisations to a population, but to arrive at logical, contextualized generalisations regarding the phenomenon under study.[863, 864] In addition, participants were randomly approached and selected from a larger list of potential participants, which increases the credibility of the findings.[808] Finally, non-probabilistic sampling methods such as snowball sampling have been demonstrated to be successful, powerful and commonly used methods for recruiting busy clinicians[865] and 'elite' individuals such as members of the judiciary,[866] who may otherwise be difficult to reach.

Two of the interviews were group interviews conducted with five legal practitioners and three judges, respectively. Interviewing multiple people at the same time has limitations, as some participants may have withheld information, particularly when discussing a sensitive subject such as AHT. In addition, while every effort was made to maximise the individual contribution of each participant, by using prompts and probing questions directed to each individual in the group, it was clear that two participants dominated the interview with the legal practitioners, while in the other interview one judge was more responsive than the other two. However, group interviews were only conducted when necessary due to the availability and schedules of the professionals, and without the flexibility of allowing this, these individuals could not have participated in the study.

The data represent the views, experiences and attitudes of professionals as recounted to the interviewer rather than observations of their practice, and given the sensitive nature of the research, some respondents may have shown a social desirability bias, the propensity to respond in a manner likely to be viewed as favourable by the researcher. This may have been further confounded by the fact that some of the participants were already known to the wider research team, which could have biased the sample. However, despite this, a number of participants, including some of those known to the team, exhibited an unfavourable view of PredAHT and were open about their opinions and intentions not to use it. Finally, qualitative research inevitably relies on the researcher's interpretations, however, the researcher remained reflexive throughout the study and acknowledged and considered how her positionality may have affected the research process and outcome. In addition, subjective bias was minimised by using three trained qualitative researchers to double-code the data and resolve disagreements through discussion and consensus.

5.5.4 Implications for research and practice

All professionals who come into contact with children and families have a duty to safeguard children and young people and should receive regular training to ensure that they are competent in their respective roles.[524] Two UK government reports on social work interventions required in the CP arena stated the need for the development of an evidence-based approach and learning culture, to inform good practice; one explicitly recommended the use of standardised tools to support decision-making and analysis of information about whether a child is suffering, or likely to suffer significant harm.[867] The other highlights the importance of improving the skills and knowledge of CPSWs.[868] This study highlighted gaps in the training and knowledge of professionals working in CP; many were unaware that some of the clinical features included in PredAHT were indicators of AHT and one CPSW expressed an antiquated view that the features of the "triad" are diagnostic for AHT due to out-dated training. This means that CPSWs and police officers rely on medical professionals to come to a decision as to whether AHT has occurred, while medical professionals believe the expectations placed upon themselves are naïve and unrealistic. Taken together, this suggests the need for training on the medical aspects of AHT for police officers and CPSWs; the need for multidisciplinary training that provides knowledge about the roles of each agency; and regular training that is continually updated in line with the evidence. Interestingly, many participants felt that the use of PredAHT itself would be helpful for training and peer review purposes.

The findings have implications for the further development and implementation of PredAHT. Given that some participants wanted confidence intervals while others did not, it may be sensible to include an option to display these. Participants' identified a number of conditions under which they would use PredAHT, including if it was used alongside their professional judgment, with knowledge of the data on which it is based and the precise definitions of the six features, if they understood how it works, and if they received multi-agency training to agree accepted risk thresholds with colleagues. These are consistent with a study exploring the acceptability of a tool to identify abusive or neglectful burns.[790] Clinical participants also expressed a desire for training on how to incorporate their prior probability of AHT. Any training on PredAHT would need to encompass all of these elements. In addition, discussions about over-reliance on PredAHT, concern that it may be used improperly or failure to investigate appropriately if a low score is given, and the potential ramifications of "false positives" or "false negatives" emphasise the importance of providing clear guidance to professionals about how PredAHT is intended to be used, namely as an assistive CPR, and *not* a diagnostic tool.

5.6 Conclusions

The findings contradict recent literature claiming that AHT is diagnosed based on the "triad" alone.[251, 252] Rather, decision-making in AHT cases is complex and nuanced, and a diagnosis is arrived at only when all potential variables are carefully explored and considered, including clinical, historical, forensic and social features and potential differential diagnoses. The findings suggest that CPSWs and police officers may benefit from additional, regular training in the medical aspects of physical abuse, and that joint training might provide a better understanding of the roles, expectations and limitations of each agency, thereby facilitating more effective collaboration.

This evaluation has demonstrated that PredAHT is acceptable to CP professionals across a range of disciplines assessing suspected AHT cases, and that they would be willing to use it as an adjunct to their decision-making. Although it may be most useful for those with the least CP experience or knowledge, it is applicable to all professionals working in this area to help reduce missed cases of AHT. These results confirm that the addition of a precise and objective evidence-based probability score that calculates the risk of AHT for CP professionals is acceptable and potentially useful. This tool, when used in conjunction with a full clinical and social history, has the potential to standardise clinical assessment, and minimise subjectivity when weighing up the clinical features in cases of possible AHT. Chapter 6 therefore describes a vignette study carried out to determine the potential impact of PredAHT on clinicians' probability estimates of AHT and hypothetical CP actions,[748] and Chapter 7 describes a multisite feasibility study, conducted to determine whether it is possible to evaluate the impact of PredAHT when it is applied in clinical practice. These studies will inform the planning and design of a formal impact analysis study and a long term implementation and dissemination plan to maximise uptake.[301]

6 Potential impact of the validated Predicting Abusive Head Trauma clinical prediction tool: A clinical vignette study

The results from this chapter were published in the following article (see Appendix 21):

Cowley LE, Farewell DM, & Kemp AM (2018). Potential impact of the validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A clinical vignette study. *Child Abuse & Neglect,* 86: 184-196.

6.1 Chapter overview

This chapter presents a novel clinical vignette study conducted with clinicians involved in suspected AHT cases.[748] Participants included twenty-nine clinicians from different specialties, at teaching and community hospitals. Clinicians estimated the probability of AHT and indicated their CP actions in six clinical vignettes. One vignette described a child with AHT, another described a child with non-AHT, and four represented "grey" cases, where the diagnosis was uncertain. Clinicians calculated the PredAHT score, and reported whether this altered their probability estimate or CP actions. The "think-aloud" method was used to capture the reasoning behind their responses. Analysis included linear modelling, linear mixed-effects modelling, chi-square tests, Fisher's exact tests, intraclass correlation, Gwet's AC1 coefficient and thematic analysis. The findings are placed within the context of the current literature, and the implications for research and practice and the strengths and limitations of the approaches taken are discussed.

6.2 Introduction

As detailed in Chapter 1, the iterative four-phased framework for the impact analysis of CPRs recommends that preparatory work is undertaken to assess the acceptability of a CPR and the feasibility of evaluating its impact in clinical practice, and to set the groundwork for and inform the study design of a formal experimental impact study.[300] The qualitative interview study presented in Chapter 5 demonstrated that PredAHT is acceptable to a range of CP professionals, including clinicians, and that they would be willing to use it in practice.[747] However, before attempting to evaluate the potential impact of PredAHT in clinical practice, it is important to understand the likelihood that it would influence clinicians' judgments and decision-making, and the reasons why clinicians may or may not follow the logic of the CPR when applying it to specific cases. Considering the potential impact of a CPR is a key step in translating CPRs into clinical practice.[299] A useful strategy to engage clinicians in the preparatory phase of CPR evaluation is the use of a simulation exercise.[300] Such studies may provide preliminary evidence regarding the effect of CPRs on clinicians' decision-making processes.[667, 678, 790, 869] For example, a clinical vignette study exploring whether clinicians would follow the recommendations of a CPR to identify paediatric burns due to child maltreatment demonstrated that while clinicians were willing to use the CPR, there was substantial variation among them regarding the extent to which they would follow its recommended actions.[790] Therefore, prior to conducting a study in clinical practice, a preliminary exploratory study was conducted using six carefully designed clinical vignettes, in order to assess the impact of PredAHT on clinicians' probability estimates of AHT and their hypothetical CP actions.

6.2.1 Aims of the vignette study

Using six clinical vignettes, this study aimed to explore the impact of PredAHT on clinicians' probability estimates of AHT, and their proposed CP actions, assessing the degree of agreement between clinicians' opinions both before, and after, using PredAHT.

6.3 Methods

This was a vignette-based cross-sectional survey study with clinicians involved in suspected AHT cases. The concurrent "think-aloud" method was used to capture participants reasoning behind their responses.[870] The study therefore adopted a mixed methods approach, using qualitative methods to gain a comprehensive understanding of the quantitative results by incorporating the perspectives of the participants.[871] The convergent mixed methods design was used, whereby the quantitative and qualitative data were collected simultaneously, analysed separately, and integrated to determine whether the findings confirmed or disconfirmed each other.[871]

6.3.1.1 Philosophical underpinning for mixed methods research

Pragmatism provides a philosophical basis for conducting mixed methods research.[871] Pragmatism is a research paradigm that "sidesteps the contentious issues of truth and reality."[872(p.8)] Within a pragmatic paradigm, the epistemological stance is that any way of thinking or doing that leads to pragmatic solutions, is useful. Therefore, pragmatism "focuses instead on 'what works' as the truth regarding the research questions under investigation",[873(p.173)] and advocates the use of multiple approaches to deriving knowledge about a research question.[871, 872, 874] Thus, in this study, both quantitative and qualitative approaches were used to provide a broad and in-depth understanding of the potential impact of the PredAHT tool on clinicians' judgments and decision-making.[871]

6.3.2 Rationale

Vignettes are short, carefully created and realistic descriptions of a person or scenario, representing systematic combinations of characteristics, and designed to simulate key features of real-world situations.[875, 876] In the experimental vignette methodology, vignettes are presented to participants in order to elicit their attitudes, intentions, judgments and decision-making processes regarding the scenarios.[875-878] There are a number of advantages of vignette studies. They are ideal for analysing clinical judgments and decision-making under uncertainty, as they allow researchers to systematically manipulate certain variables while controlling for others, and therefore to assess how clinicians judgments are affected by a variety of experimental factors.[875-880] Put another way, the systematic manipulation and control of variables allows researchers to measure multiple predictors of clinician behaviour, and to determine the relative importance of different variables on clinician behaviour and thus how multiple pieces of information are combined during decision-making.[596, 877] Vignettes can be constructed in a way that overcomes the correlation between variables as they occur in clinical practice, while still remaining realistic.[881, 882] Furthermore, the effects of participant demographics and characteristics can be examined in the analysis,[875, 877, 881, 882] in addition to variability in clinicians' judgments and decision-making.[883] Vignettes are inexpensive and provide standardized data that can be consistently and meaningfully interpreted.[877, 880, 883]

Vignettes are an ideal medium to explore judgments and decision-making around sensitive topics, as they are non-personal, and they are a useful alternative to traditional experiments when there are ethical issues associated with the experimental manipulation of variables in practice.[878, 884] Finally, different forms of response can be used in combination with vignette-based surveys.[885] The concurrent "think-aloud" method is particularly useful to examine professionals' thought processes and rationale as they read through the vignettes and answer the survey questions.[870, 886-888] The "think-aloud" method explicitly instructs participants to articulate their thoughts and feelings as they perform a task.[870, 880, 889-891] It is a process tracing method that examines the cognitive processes involved in decision-making, including how judgments change over time as new information is acquired, and the aspects of information that are focussed on during decision-making.[870] The "think-aloud" technique is traditionally used in cognitive interviewing to validate survey or questionnaire items,[892-894] however it is increasingly being used to study clinicians' diagnostic reasoning and clinical decision-making, often alongside vignettes.[895-898]

6.3.3 Vignette design

Six clinical vignettes were designed according to methodological recommendations and best practices described in the literature and reported in Appendix 22.[877, 878] All vignettes described children <3 years old with ICI evident on neuroimaging, as this is the population the PredAHT tool is intended for. They were derived from a combination of the scientific literature, the findings from the qualitative study reported in Chapter 5,[187, 747] the clinical experience of the supervisory team, and actual cases of suspected AHT encountered during previous research.[60] The vignettes consisted of experimental features that were systematically manipulated to explore their effect on clinicians' judgments and decision-making, controlled features that were kept consistent to minimize extraneous variance, and contextual features to enhance the verisimilitude and realistic quality of the scenarios.[877] This ensured that the vignettes were structurally similar, and equivalent in all major respects apart from the intentional manipulation of experimental variables.[877]

Table 6.1 lists the key features of each vignette including the corresponding calculated PredAHT percentage probability and likelihood ratio; Table 6.2 includes the full vignettes as presented to clinicians. Each vignette comprised two sections. Section one included the child's age, gender, any history of trauma or social history, and the characteristics of the ICI. Section two included the clinical information required to complete PredAHT, namely; whether the six clinical features (head/neck bruising, apnoea, seizures, rib/long-bone fractures, RH) were present, absent or unknown. Two vignettes were based on real cases encountered in the original PredAHT validation study,[60] where the outcome was confirmed after thorough clinical evaluation and by multidisciplinary CP assessment. One described a child with confirmed AHT ("V1:AHT"), the other a child with confirmed nAHT ("V2:nAHT"). Demographic details, that would be unlikely to influence the diagnosis, were altered to protect the identity of the children. It was hypothesized that PredAHT would have the greatest impact on clinicians judgments and decision-making in "grey" cases, where there is uncertainty surrounding the diagnosis.[899] The remaining four vignettes were designed to represent such cases. Based on the findings of Chaiyachati et al.,[899] who found that there was no one component of the injury, incident or history associated with uncertainty around the perceived likelihood of physical abuse, two grey cases were created ("V3:AHT*" and "V4:nAHT*") by altering elements of the history and social history from "V1:AHT" and "V2:nAHT" but keeping the clinical features the same. Changes are indicated in italics (Table 6.2). Similar approaches have been taken in previous vignette studies evaluating clinicians' perceived likelihood of physical abuse.[533, 535] In addition, the scientific literature and the findings from the qualitative study reported in Chapter 5 suggest that the history and social history are key factors influencing clinicians' decision-making in suspected AHT cases.[9, 187, 198] These factors may influence how PredAHT impacts upon clinicians' AHT probability estimates and proposed CP actions.[747]

In "V3:AHT*" the child is older than in "V1:AHT", and it is developmentally plausible that a short fall occurred. There was no delay in presentation and the explanation was consistent; delay in seeking medical care and inconsistencies within the history are both suggestive of child abuse and it is important to rule these out during history-taking.[333] However, the incident was unwitnessed, and the clinical features and severity of the injuries appear discordant with the mechanism of injury; the evidence base in support of the theory that short falls can cause the signs and symptoms of AHT is weak.[268] It is widely accepted in the literature that severe ICI or RH very rarely result from minor head trauma.[900] In addition, in a rigorous systematic review of the differential diagnoses of RH in children with features concerning for abuse, no RH were found in children with prolonged apnoea associated with an apparent life threatening event, and only 3% of children with seizures had associated RH.[409] In "V4:nAHT*", there are inconsistencies within the history, a delay in presentation, plus social concerns within the family that may increase suspicion of AHT in comparison to "V2:nAHT".[198, 333] In addition, linear parietal skull fractures are neither indicative of AHT or nAHT, since a skull fracture resulting from impact is the commonest cause of accidental head trauma but can also occur in AHT.[374, 901]

Two further grey cases ("V5:ICI-only" and "V6:missing") were developed around one of the most challenging clinical scenarios whereby a baby has ICI with no additional clinical features suggestive of abuse, making confirmation or exclusion of abuse more difficult. In the absence of confirmed accidental trauma or a medical aetiology, multiple subdural haemorrhages are suspicious for child maltreatment[586] and three months of age is thought to be on the developmental cusp of when an infant starts to roll over.[902] "V6:missing" is almost identical to "V5:ICI-only", but neither skeletal radiology or ophthalmology examination were undertaken. Changes are indicated in italics (Table 6.2). These vignettes were created to explore the effects of missing investigation results and the imputation feature of PredAHT on clinicians' judgments and decision-making.

The vignettes were reviewed by the supervisory team, two of whom are paediatricians with substantial clinical and academic expertise in the field of child abuse. The vignettes were revised accordingly to improve their clarity and construct validity i.e. the degree to which they approximate a real-life clinical case.[876, 877] Finally, they were pilot tested with one clinician, who felt they were clear, and reflected cases that may plausibly be encountered in clinical practice.

Table 6.1 Key features of each of the six clinical vignettes

	Information given in Section 1		Information given in Section 2						PredAHT Results	
Vignette	Presentation, History and Social History	CT Scan Results	B	A	S	RF	LBF	RH	Percentage Probability[a]	Likelihood ratio
1:AHT	3 months old Lethargy, vomiting No history of trauma	HII affecting both cerebral hemispheres, brainstem and thalami; hyperdense SDH at the vertex	No	Yes	Yes	?	?	Yes	98.4%	118.79
2:nAHT	23 months old No delay in presentation Fall from chair at a height of 1.5 metres onto tiled floor Consistent history between parents & over time	Frontal lobe hyperdense SDH Linear, undisplaced skull fracture of left frontal parietal bone	Yes	No	No	No	No	No	14.2%	0.32
3:AHT*	14 months old Lethargy, vomiting No delay in presentation Unwitnessed short fall onto wooden floor Consistent history over time	HII affecting both cerebral hemispheres, brainstem and thalami; hyperdense SDH at the vertex	No	Yes	Yes	?	?	Yes	98.4%	118.79
4:nAHT*	23 months old Six hour delay in presentation Initially no history of trauma Possible fall from chair at a height of 1.5 metres onto tiled floor Domestic violence concerns Previous children's services involvement	Frontal lobe hyperdense SDH Linear, undisplaced skull fracture of left frontal parietal bone	Yes	No	No	No	No	No	14.2%	0.32

5:ICI-only	3 months old Lethargy, vomiting Rolled off sofa onto the floor	Multiple small bilateral SDHs	No	No	No	No	No	No	3.7%	0.08
6:missing	3 months old Lethargy, vomiting Rolled off the sofa onto the floor	Multiple small bilateral SDHs	No	No	No	?	?	?	10.4%	0.22

B, head/neck bruising; A, apnoea; S, seizures; RF, rib fractures; LBF, long-bone fractures; RH, retinal haemorrhage's, PredAHT, Predicting Abusive Head Trauma tool; HII, hypoxic ischemic injury; SDH, subdural haemorrhage

[a] This was calculated using the "baseline" prior probability of 34%, the prevalence of abusive head trauma in the data

Table 6.2 The six clinical vignettes as presented to clinicians

V1:AHT	**Section 1:** A 3-month-old female infant presents to the hospital with lethargy and vomiting and no history of trauma. A CT scan of head reveals hypoxic ischaemic injury affecting both cerebral hemispheres, the brainstem and the thalami, and an acute (i.e. hyperdense) subdural haemorrhage at the vertex. **Section 2:** Apnoea and seizures are noted to be present, but there is no evidence of head or neck bruising. The ophthalmology exam reveals bilateral superficial and deep multi-layered retinal haemorrhages. Location: Zone 1 and outside Zone 1, posterior pole and periphery. Number: confluent. Additional features: macular detachment and retinal folds adjacent to the macular. A skeletal survey is not performed and so it is unknown whether the child has any rib or long-bone fractures.
V2:nAHT	**Section 1:** A 23-month-old female infant presents to the hospital immediately following a head trauma. Both parents, when interviewed separately, said that the child had fallen off a chair at a height of approximately 1.5 metres onto a tiled floor. Both parents' accounts remain consistent over time. Inflicted trauma is vehemently denied. A CT scan of head reveals a frontal lobe hyperdense subdural haemorrhage and a linear, undisplaced skull fracture of the left frontal parietal bone. **Section 2:** Apnoea and seizures are noted to be absent. Bruising to the scalp and cheeks is noted. The ophthalmoscopy exam and skeletal survey are both negative.
V3:AHT*	**Section 1:** *A 14-month-old male infant* presents to the hospital with lethargy and vomiting. *His father states that he left the room momentarily and found him on the wooden floor after falling indoors. He brought him to the emergency department immediately and his story remains consistent over time. He denies inflicted trauma and states that the child has recently began to walk independently*. A CT scan of head reveals hypoxic ischaemic injury affecting both cerebral hemispheres, the brainstem and the thalami, and an acute (i.e. hyperdense) subdural haemorrhage at the vertex.

	Section 2: Apnoea and seizures are noted to be present, but there is no evidence of head or neck bruising. The ophthalmology exam reveals bilateral superficial and deep multi-layered retinal haemorrhages. Location: Zone 1 and outside Zone 1, posterior pole and periphery. Number: confluent. Additional features: macular detachment and retinal folds adjacent to the macular. A skeletal survey is not performed and so it is unknown whether the child has any rib or long-bone fractures.
V4:nAHT*	**Section 1:** A 23-month-old female infant *presents to the hospital with her mother. Initially no history of trauma is provided but following questioning the mother states that the child may have fallen* off a chair at a height of approximately 1.5 metres onto a tiled floor. *There are concerns about domestic violence within the family and there has been previous involvement with social services. It emerges that the incident occurred approximately six hours prior to presentation to the hospital.* A CT scan of head reveals a frontal lobe hyperdense subdural haemorrhage and a linear, undisplaced skull fracture of the left frontal parietal bone. **Section 2:** Apnoea and seizures are noted to be absent. Bruising to the scalp and cheeks is noted. The ophthalmoscopy exam and skeletal survey are both negative.
V5:ICI-only	**Section 1:** A 3-month-old female infant presents to the hospital with lethargy and vomiting. The parents state that the baby rolled off the sofa onto the floor. A CT scan of head reveals multiple small bilateral subdural haemorrhages. **Section 2:** Apnoea and seizures are noted to be absent, and there is no evidence of head or neck bruising. The ophthalmology exam and skeletal survey are both negative.
V6:missing	**Section 1:** A 3-month-old female infant presents to the hospital with lethargy and vomiting. The parents state that the baby rolled off the sofa onto the floor. A CT scan of head reveals multiple small bilateral subdural haemorrhages.

	Section 2: Apnoea and seizures are noted to be absent, and there is no evidence of head or neck bruising. *An ophthalmology exam and a skeletal survey have not yet been performed, and so it is unknown whether the child has any rib or long-bone fractures, or retinal haemorrhages.*

6.3.4 Participant recruitment

Participants were recruited at the same time and using the same methods as described in the qualitative study reported in Chapter 5,[187, 747] namely purposive and snowball sampling. Clinicians were recruited from three teaching hospitals and two district general hospitals in south west UK. The researcher targeted clinicians from different specialties who are involved in the evaluation and investigation of suspected AHT and who therefore contribute to the final clinical diagnosis of AHT. Briefly, a list of potential participants was identified through personal contacts of the research team. A random selection of 40 clinicians from this list, across a range of specialities and with different levels of CP experience and seniority, were then invited to take part.

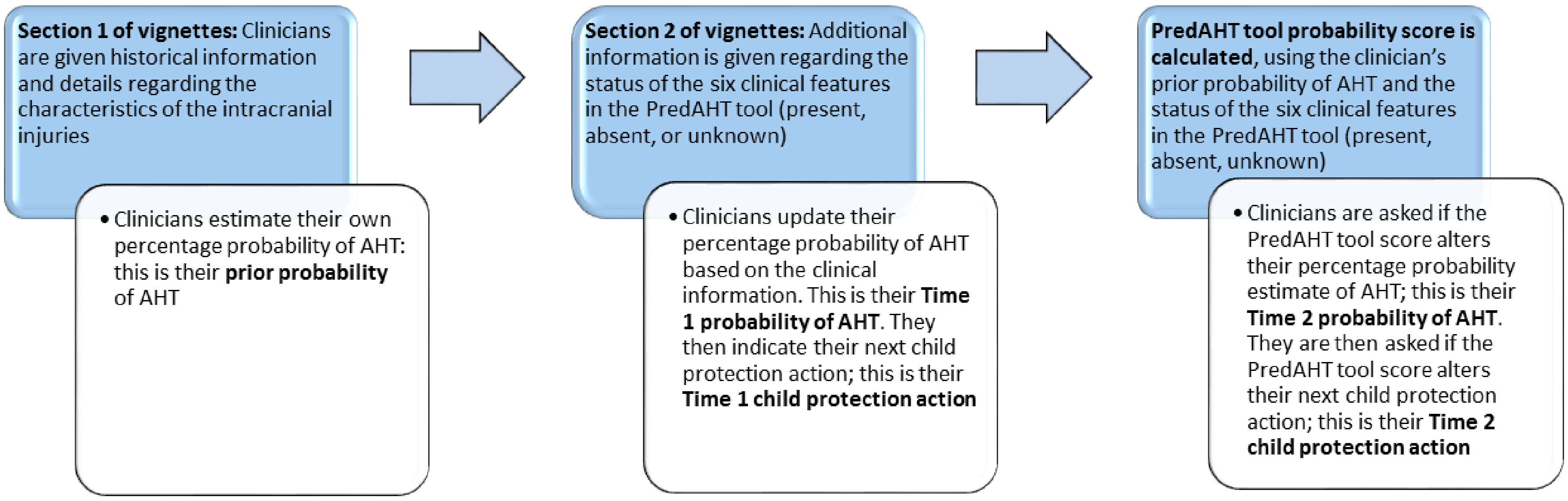
6.3.5 Data collection procedure

All participants were given the participant information sheet to read and keep (Appendix 8) and informed consent was obtained (Appendix 23). Demographic data and information regarding participants' CP experience and training were collected in order to describe the characteristics of the sample and for analysis purposes (Appendix 11). The researcher explained how PredAHT was developed and validated, and described its various features and intended purpose to each participant. The six vignettes were presented to each participant in a random sequence, to account for possible order effects. The data collection procedure is outlined in Figure 6.1 and took approximately 45 minutes. Participants first estimated their own "prior" probability of AHT for each vignette based on the information given in section 1 (see Table 6.2). They then estimated their Time 1 probability of AHT and indicated their Time 1 proposed CP action for each vignette, based on further information given in section 2. The PredAHT probability score was then calculated for each vignette using the clinicians' prior probability, and the clinical details in section 2. Finally, participants estimated their Time 2 probability of AHT and indicated their Time 2 proposed CP action for each vignette, after seeing the PredAHT probability score. CP actions were aligned with three categories of concern (Table 6.3), as per National Institute for Health and Care Excellence child maltreatment guidelines[586]; clinicians could choose multiple options. Participants were asked to respond as they would if they were to see the patient in clinical practice, and to mitigate observer effects they were assured that they were not being tested on their judgments or decision-making processes[883]. Free text boxes were included for comments at the end of each vignette. Data were collected between May and September 2016; all clinicians were shown Version 2 of the computerised PredAHT, as described in Chapter 4.

6.3.5.1 Think-aloud method

Participants were asked to verbalise their thought processes while completing the survey, using the concurrent "think-aloud" method.[870] This method is based on the assumption that an individual's cognitive processes are directly accessible as verbal data.[870] Participants were instructed to report their thoughts out loud when reading through the vignettes and deciding on their estimated probabilities of AHT and their proposed CP actions. If participants paused for longer than a few seconds, the researcher reminded them to keep thinking aloud. Otherwise, all interaction between the participant and researcher was minimised so as not to interrupt the participants' flow of thoughts. This enabled participants' verbalisations to be transcribed by the researcher in real time.

Figure 6.1 Flowchart of data collection procedure



AHT = abusive head trauma, PredAHT = Predicting Abusive Head Trauma tool.

Table 6.3 Possible child protection actions and associated categories of concern in line with National Institute for Health & Care Excellence (NICE) child maltreatment guidelines

Indicated child protection action		Category
No further child protection action		No concern (abuse excluded)
Investigate further:	Discuss with line manager	Concern (abuse considered)
	Discuss with child protection colleague	
	Gain collateral information from other agencies and health disciplines (e.g. health visitor)	
	Order further investigations (please specify)	
Refer to children's services		Suspicion (abuse suspected)

6.3.6 Ethical issues

6.3.6.1 Ethical approval

The researcher contacted the Research and Development department of Cardiff and Vale University Health Board who confirmed that ethical approval from a National Health Service Research Ethics Committee was not required. An amendment to the qualitative study reported in Chapter 5 was submitted to the Cardiff University School of Medicine Research Ethics Committee on August 21st 2015 to request the use of vignettes with clinicians; this was approved on Friday 9th October 2015 (Appendix 14). The Research and Development departments of all health boards of the participating hospitals were contacted prior to study commencement to obtain approval to carry out the study at each site.

6.3.6.2 Data management

Data were stored and managed in accordance with the Data Protection Act 1998. Transcripts and completed surveys were securely stored on a password-protected, confidential Cardiff University server. All transcripts, completed surveys, consent forms and demographic data will be held securely for 15 years, in line with Cardiff University research data policies. After this, all data will be destroyed.

6.3.7 Quantitative analysis

Analysis focused on determining the impact of PredAHT on clinicians' probability estimates of AHT and their proposed CP actions, and assessing the degree of agreement between their probability estimates and proposed CP actions both before, and after, seeing the PredAHT score. Statistical analyses were performed using R software version 3.2.3[742]; $p<0.05$ was considered statistically significant. Exploratory data analysis was conducted through graphical displays, to determine plausible models for the data and examine relationships between variables. All figures were produced using bespoke R scripts written by the researcher.

6.3.7.1 Impact of PredAHT on clinicians probability estimates of AHT

Firstly, the impact of PredAHT on clinicians' probability estimates of AHT was assessed for each vignette. Six linear models were fitted, using the formula:

$$y_2 = y_1 + \beta\ (t - y_1) + \varepsilon$$

where y_2 is the Time 2 probability estimate, y_1 is the Time 1 probability estimate, t is the PredAHT score with clinicians' priors, β is the slope, and ε is the error term. The slope represents the average proportion of the distance between y_1 and t that clinicians move. For example, if y_1 = 50%, t = 70% and y_2 = 60%, then $\beta = 0.5$. A slope of 0 indicates no difference between clinicians' Time 1 and Time 2 probability estimates on average (if $\beta = 0$, $y_2 = y_1 + \varepsilon$), while a slope of 1 means that clinicians' Time 2 probability estimates are the same as the PredAHT score on average (if $\beta = 1$, $y_2 = t + \varepsilon$). The intercept was not included, as the expected value of the independent variable given that the dependent variable is 0, is 0.[903] In other words, if $y_1 - t = 0$, then $y_2 - y_1$ will also equal 0.

Next, the overall impact of PredAHT across vignettes was assessed. Due to the multilevel nature of the data (vignette level and clinician level), analyses that focus on both levels simultaneously must be used.[875, 878] Several linear mixed models were fitted, with random effects at the vignette and clinician levels. To examine the influence of clinician demographics (hospital type, clinician specialty, clinician age, years of CP experience, paediatric head injury training, clinician seniority), a reduced model with the fixed effect described in the formula above and the random effects, was compared with six models allowing the average proportion moved to vary across the categorical demographic variables. The R package "lme4" was used for mixed model fitting.[904] Models were fitted using the maximum likelihood criterion, to enable comparison using likelihood ratio tests. Profile likelihood confidence intervals were computed for model parameters. The R package "multcomp" was used to obtain p-values for fixed effects coefficients,[905] and the contrast function within the "lsmeans" package was used for pairwise comparisons between factor levels.[906]

6.3.7.2 Impact of PredAHT on clinicians proposed child protection actions

To analyse whether certain clinicians were more likely to change their proposed CP action after seeing the PredAHT score, or whether clinicians were more likely to change their proposed CP action only in certain vignettes after seeing the PredAHT score, the chi-square test of independence and Fisher's exact test were used to examine associations between categorical variables (change in CP action vs. hospital type, clinician specialty, clinician age, years of CP experience, paediatric head injury training, clinician seniority, and vignette). In contingency tables with more than one degree of freedom, Fisher's exact test is recommended over the chi-square test if more than 20% of the cells have expected values of less than five or if any cell has an expected value of less than one.[907] A change in CP action was specified as one associated with an increase or decrease in the level of concern from Time 1 to Time 2 (see Table 6.3).

6.3.7.3 Inter-rater reliability of clinicians' probability estimates of AHT

Inter-rater reliability statistics were estimated to assess the degree of agreement between clinicians in their three probability estimates across the six vignettes. There are many inter-rater reliability indices available for different study designs; some, such as Cohen's kappa are only applicable for categorical data[908]; others, such as the Bland-Altman limits of agreement, are only applicable when there are two raters.[909] For this analysis, the intra-class correlation coefficient (ICC) was deemed the most suitable, since the data were continuous and multiple raters (29 clinicians) participated.[910] The ICC can be interpreted as "the proportion of observed variance in ratings that is due to systematic between-target differences compared to the total variance in ratings".[911] The ICC is a number ranging from 0 to 1 where values closer to 1 represent greater agreement.[912] McGraw and Wong[913] described ten different forms of ICCs, which are applicable for different scenarios and dependent upon different assumptions about the data. The choice of ICC used is determined by the study design and is guided by four principal considerations.[912, 914, 915] The first is whether all vignettes were rated by the same clinicians. The second is whether or not the clinicians were theoretically considered to be drawn from a larger population of clinicians with similar characteristics. If different clinicians rated each vignette, a one-way random effects model should be used. If the same clinicians rated each vignette and the clinicians were assumed to be a random sample of a larger population of clinicians, a two-way random effects model is appropriate. A two-way mixed effects model should be used if the focus is only on the agreement between the specific clinicians in the study sample. The third consideration is whether the purpose of the ICC is to quantify reliability based on single raters or the average of all raters. The fourth is whether good reliability should be characterised by consistency or absolute agreement.[912, 914, 915]

As the study design is fully crossed, i.e. with all clinicians each rating all six vignettes, and the intention was to generalise the results to a larger population of clinicians, a two-way random effects model was applicable. In addition, single measures ICCs were reported since the ratings were not averaged prior to the analysis. Finally, as the focus was on agreement between clinicians' absolute probability estimates rather than similarities in rank ordering, absolute agreement was chosen over consistency.[912, 914, 915] The R package "psych" was used to compute the ICC coefficients, with 95% confidence intervals, for clinicians' prior, Time 1 and Time 2 probability estimates.[916]

A number of biostatisticians have devised guidelines for the interpretation of agreement coefficients.[912, 917-921] For ICCs, the guidelines developed by Cicchetti and Sparrow[917] and Portney and Watkins[920] are commonly cited in the literature. Cicchetti and Sparrow[917] suggested that values less than 0.40 indicate poor reliability, values between 0.40 and 0.59 indicate moderate reliability, values between 0.60 and 0.74 indicate good reliability, and values between 0.75 and 1.00 indicate excellent reliability. The guidelines proposed by Portney & Watkins[920] are more conservative, with values less than 0.50 indicating poor reliability, values between 0.50 and 0.75 indicating moderate reliability, and values greater than 0.75 indicating good reliability. Koo and Li[912] support the use of Portney & Watkins[920] guidelines, with the slight modification that values between 0.75 and 0.90 indicate good reliability and values greater than 0.90 indicate excellent reliability. The 95% confidence interval of the ICC estimate was used as the basis to evaluate reliability, in line with published recommendations.[912]

6.3.7.4 Inter-rater reliability of clinicians' proposed child protection actions

For categorical data with multiple raters, Fleiss' kappa coefficient,[922] a chance-corrected agreement coefficient and an extension of Scott's pi statistic,[923] is a very popular index for estimating inter-rater reliability and is widely used. However, this agreement measure has been heavily criticized as it often yields coefficients that are paradoxically low in comparison to the observed percent agreement.[924, 925] This so called "kappa paradox" is widely acknowledged and described in the literature.[924-930] In addition, Fleiss's variance estimator assumes no agreement between raters beyond chance and so cannot be used for confidence interval construction to measure the precision of the coefficient, and it does not take into account variability due to the sampling of raters, meaning that the results cannot be generalised to a larger population of raters.[931] Finally, it has been suggested that Fleiss' kappa is inappropriate for fully crossed designs as it is assumed that a new sample of raters is selected to rate each subject.[914]

An alternative and more stable multiple-rater agreement coefficient, the AC_1, was consequently proposed by Gwet.[924] Gwet's AC_1 is suitable for fully crossed designs.[931] Gwet's estimator can be used for constructing confidence intervals, and allows for generalisation to a larger population of both subjects (vignettes) and raters (clinicians).[931, 932] Gwet's AC_1 has also been proven to be robust to the "kappa paradox" and to demonstrate plausible values in line with observed percent agreement values.[927, 930, 933-935] Its use is increasingly recommended[927, 930, 935, 936] but it is yet to appear widely in the medical literature. Inter-rater agreement between clinicians' proposed CP actions at Time 1 and Time 2 was therefore estimated using Gwet's AC_1 coefficient; the R code used is available from http://www.agreestat.com/r_functions.html. Jackknifing was used to estimate the variance due to the sampling of clinicians, as described in Gwet.[932] The jackknife is a resampling method particularly useful for variance estimation. In the simplest case, used here, jackknife resampling is accomplished by sequentially deleting single cases from the original sample.[937]

Regarding the interpretation of the AC_1 statistic, Gwet recommends an approach that takes into consideration the precision with which the coefficient was estimated.[938] This involves subtracting a "critical value" from the obtained coefficient before it can be compared to published guidelines. This critical value "represents the 95th percentile of the coefficient when the rating of subjects is performed randomly".[938(p.130)] The guidelines proposed by Landis and Koch[921] are the most commonly used to aid interpretation of this type of agreement coefficient.[938] These suggest that values less than 0.0 are poor, values between 0.0 and 0.20 are slight, values between 0.21 and 0.40 are fair, values between 0.41 and 0.60 are moderate, values between 0.61 and 0.80 are substantial, and values between 0.81 and 1.00 are almost perfect.[921]

6.3.7.5 *Impact of clinicians' priors on PredAHT score*

The PredAHT score was compared with and without clinicians' prior probabilities, to assess the impact of clinicians' priors for each vignette.

6.3.7.6 *Comparison of clinicians' likelihood ratios with the PredAHT likelihood ratios*

Clinicians' likelihood ratios for the clinical evidence were calculated by converting their prior and Time 1 probabilities of AHT back to odds and dividing their Time 1 odds by their prior odds. These were compared with the likelihood ratios calculated by PredAHT.

6.3.8 Qualitative analysis

Charters[891] argues that "think-aloud" data can be viewed through a qualitative lens. Participants' verbal data and free text comments were classified into themes using thematic analysis.[799] Thematic analysis has been used to analyse "think-aloud" data in a number of studies.[898, 939] Analysis entailed grouping codes into categories, and further arranging categories under overarching themes. This involved six phases including 1) familiarisation with the data 2) generating initial codes 3) searching for themes 4) reviewing themes 5) defining and naming themes and 6) writing up the results.[799] To enhance the trustworthiness and rigor of the thematic analysis, a purposeful approach was adopted.[940] The researcher developed an analytic framework that was amended as new data were collected; all categories and their definitions are detailed in the framework (Appendix 24). Codes, categories and themes generated from the data were discussed at research team meetings; disagreements regarding data interpretation were resolved by consensus. In the interests of reflexivity, the researcher considered how her own values and assumptions as a student involved in developing PredAHT might influence the interpretation of the findings.

6.3.9 Hypotheses

It was hypothesized that PredAHT would have a greater impact on clinicians judgments and decision-making in the four "grey" cases ("V3:AHT*", "V4nAHT*", "V5:ICI-only" and "V6:missing") compared to the two confirmed cases of AHT and nAHT ("V1:AHT" and "V2:nAHT"). Furthermore, PredAHT was expected to have a greater impact in "V3:AHT*" compared to "V1:AHT", where the clinical features are the same but the history is less concerning, and in "V4:nAHT*" compared to "V2:nAHT", where the clinical features are the same but the history and social history are more concerning. Finally, it was hypothesized that PredAHT would have a greater impact on younger, junior clinicians, those working in district general hospitals, those with no specific training in paediatric head injuries, and clinicians with the least experience of CP work.

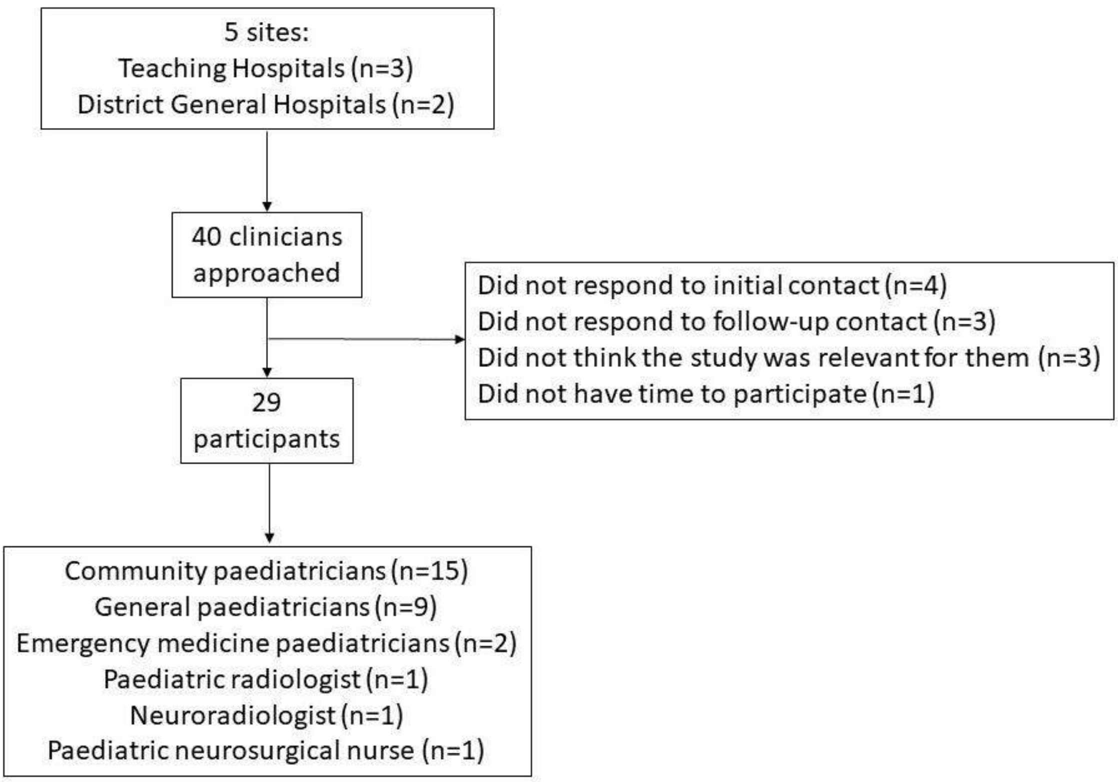
6.4 Results

6.4.1 Quantitative results

6.4.1.1 Response rates

Response rates are shown in Figure 6.2. Altogether 35 clinicians were invited to participate in both the qualitative study reported in Chapter 5[187, 747] and the current vignette study (see Figure 5.2, Chapter 5); an additional five clinicians were invited to take part in the vignette study only. All vignettes were completed by 29 clinicians in a fully-crossed design between April–September 2016. Twenty-four of the clinicians also took part in the qualitative study. Data regarding the demographics and characteristics of non-responders and non-attendees were not collected.

Figure 6.2 Flowchart of clinicians participating in the vignette study



6.4.1.2 Participant demographics and characteristics

Participant demographics, and their CP experience and training are reported in Table 6.4.

Table 6.4 Demographics and characteristics of clinicians participating in the vignette study

Demographics / Characteristics	Community Paediatricians (N = 15)		General Paediatricians (N = 9)		Other Specialty (N = 5)	
	n	%	n	%	n	%
Gender						
Female	15	100	2	22.2	4	80
Male	0	0	7	77.8	1	20
Age group						
25–34	0	0	1	11.1	1	20
35–44	5	33.3	4	44.4	3	60
45–54	6	40	3	33.3	0	0
55–64	4	26.7	1	11.1	1	20
Ethnicity						
White British	12	80	6	66.7	4	80
White Other	2	13.3	1	11.1	1	20
Indian	1	6.7	2	22.2	0	0
Years in CP						
5–9	3	20	2	22.2	2	40
10–20	4	26.7	3	33.3	1	20
>20	8	53.3	4	44.4	2	40
CP training						
Yes	15	100	9	100	5	100
No	0	0	0	0	0	0
Paediatric HI training						
Yes	11	73.3	4	44.4	5	100
No	4	26.7	5	55.6	0	0
Hospital Type						
Teaching	11	73.3	5	55.6	5	100
District general	4	26.7	4	44.4	0	0
Seniority						
Consultant	8	53.3	9	100	3	60
Associate specialist	5	33.3	0	0	0	0
Trainee doctor	2	13.3	0	0	1	20
Senior staff nurse	0	0	0	0	1	20

CP = child protection, HI = head injuries.

6.4.1.3 Exploratory data analysis

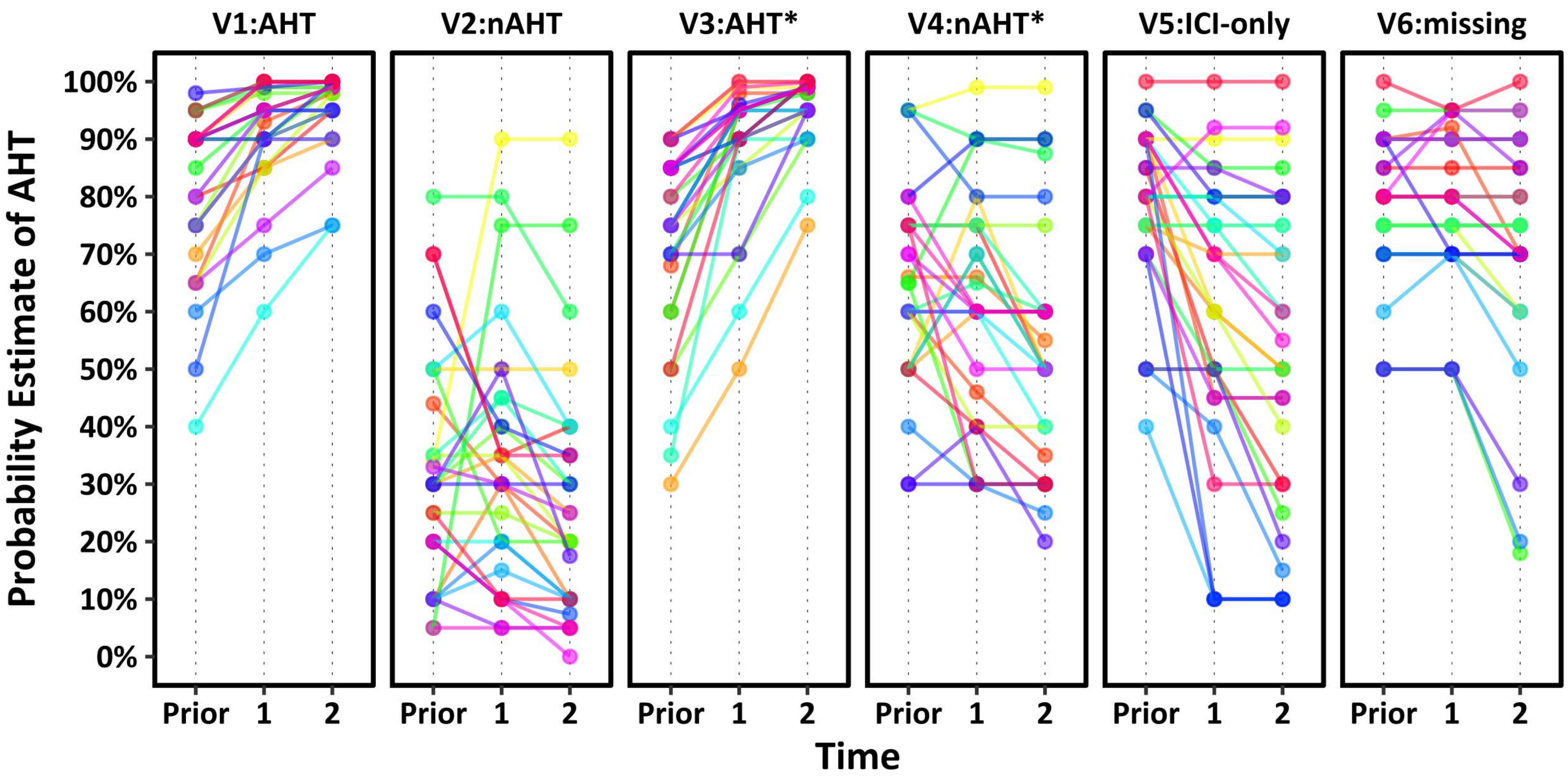
There were no missing data. A visual inspection of the raw data found no obvious order effects (Appendix 25). A scatterplot matrix (Appendix 26) showed a positive linear relationship between clinicians' prior, Time 1 and Time 2 probabilities of AHT, therefore linear models were deemed appropriate for the data. The correlation between clinicians' Time 1 and Time 2 probabilities was greater than the correlation between their prior and Time 1 probabilities, as measured by the Pearson correlation coefficient, suggesting greater differences between clinicians' prior and Time 1 probabilities than between their Time 1 and Time 2 probabilities. Clinicians' probability estimates of AHT were visualized with a parallel coordinates plot (Appendix 27). Parallel coordinates plots are useful for visualizing multivariate data and examining relationships among multiple variables. Each clinician is represented by a different coloured connected line that joins their probability estimates for each vignette, resulting in a multivariate "signature" for each clinician and therefore enabling comparison of AHT probability estimates both within and between clinicians. A boxplot comparing clinicians' Time 1 probabilities of AHT with the PredAHT score (incorporating clinicians' prior probabilities) showed that clinicians' Time 1 probabilities were lower on average than the PredAHT score in "V1:AHT" and "V3:AHT*", but higher on average than the PredAHT score in "V2:nAHT", "V4nAHT*", "V5:ICI-only" and "V6:missing" (Appendix 28). These differences were more pronounced when comparing clinicians' Time 1 probabilities of AHT with the PredAHT score using the baseline prior probability, in "V4:nAHT", "V5:ICI-only" and "V6:missing" (Appendix 29).

Figure 6.3 shows clinicians' prior, Time 1 and Time 2 probability estimates for each of the six vignettes while Table 6.5 reports the means, standard deviations, and minimum and maximum values. Higher estimates equate to a greater perceived likelihood of AHT. The highest mean prior probability estimates were given for "V1:AHT" and the lowest for "V2:nAHT". Clinicians' prior probabilities were higher in "V1:AHT" compared to "V3:AHT*" and in "V4:nAHT*" compared to "V2:nAHT", i.e. where the neuroradiological features are the same but there is a more concerning history and/or social history. At Time 1 and Time 2, the highest mean estimates were given for "V1:AHT" and "V3:AHT*". These were also the vignettes with the greatest agreement between clinicians. The lowest mean estimates were given for "V2:nAHT"; estimates for "V4:nAHT*" were higher than "V2:nAHT", with greater agreement between clinicians. Estimates for "V6:missing" were higher than for "V5:ICI-only", where the history and presentation are the same but fractures and RH are unknown about rather than absent. There was low agreement in both "V5:ICI-only" and "V6:missing", particularly at Time 2. Figure 6.4 shows each clinician's Time 1 and Time 2 probability estimates of AHT, with connected lines between the two showing the magnitude of change.

Table 6.5 Means, standard deviations, and minimum and maximum values of clinicians' probability estimates of AHT for each of the six vignettes

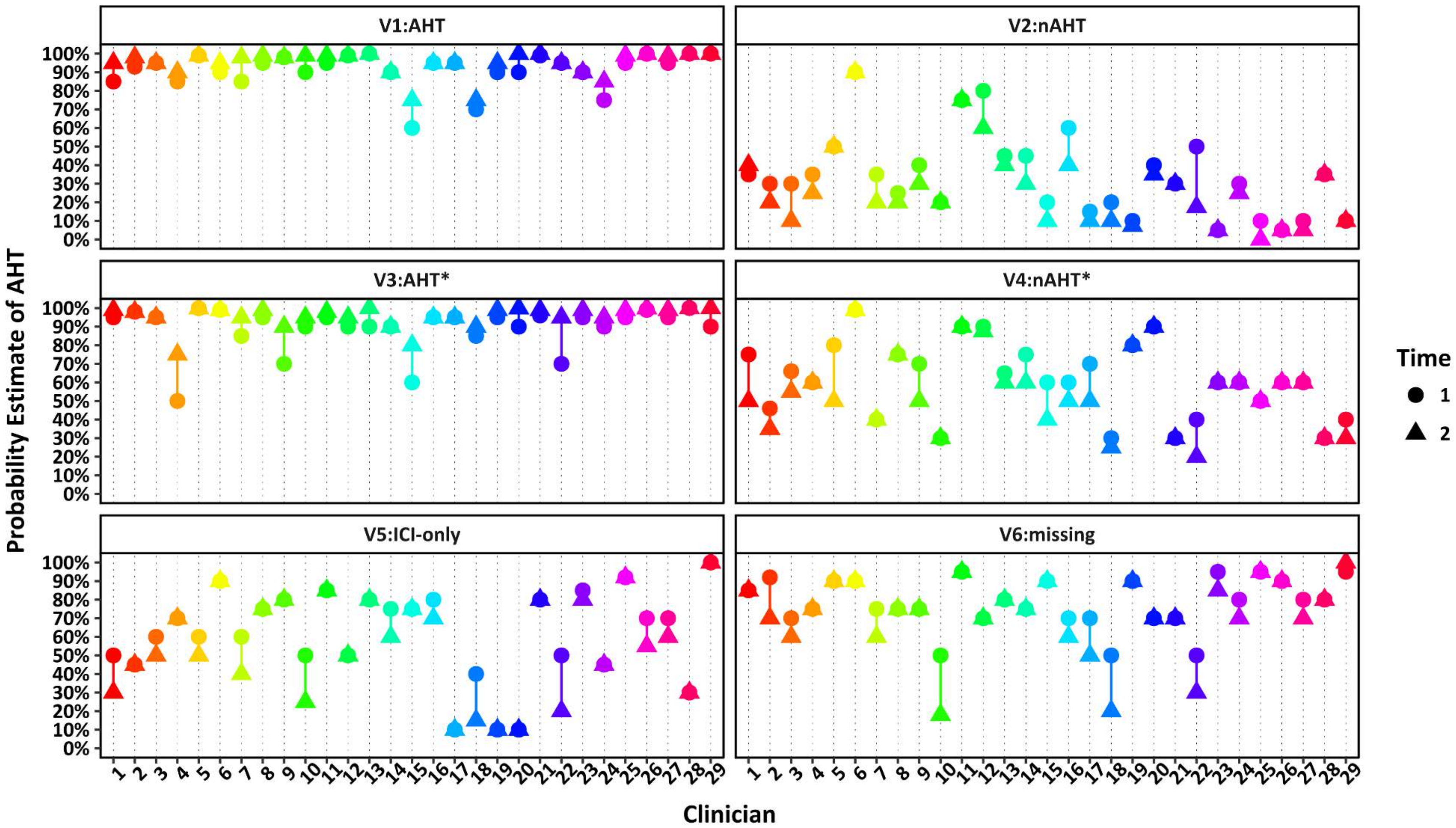
	Summary statistic	V1: AHT	V2: nAHT	V3: AHT*	V4: nAHT*	V5: ICI-only	V6: missing
Prior probability	Mean	80.28	32.45	72.34	64.34	78.28	77.93
	SD	(14.54)	(20.00)	(17.16)	(16.94)	(14.90)	(13.20)
	Min–Max	40–98	5–80	30–90	30–95	40–100	50–100
Time 1 probability	Mean	91.31	33.97	89.38	61.41	61.28	78.34
	SD	(9.38)	(21.97)	(12.02)	(19.82)	(24.61)	(13.25)
	Min–Max	60–100	5–90	50–100	30–99	10–100	50–95
Time 2 probability	Mean	95.06	26.72	95.61	54.36	54.55	72.00
	SD	(6.71)	(21.43)	(5.92)	(20.92)	(27.30)	(20.95)
	Min–Max	75–100	0–90	75–100	20–99	10–100	18–100

Figure 6.3 Clinicians' prior, Time 1 and Time 2 probability estimates of AHT for each of the six vignettes



Colours represent the 29 different clinicians.

Figure 6.4 Clinicians' Time 1 and Time 2 probability estimates of AHT for each of the six vignettes



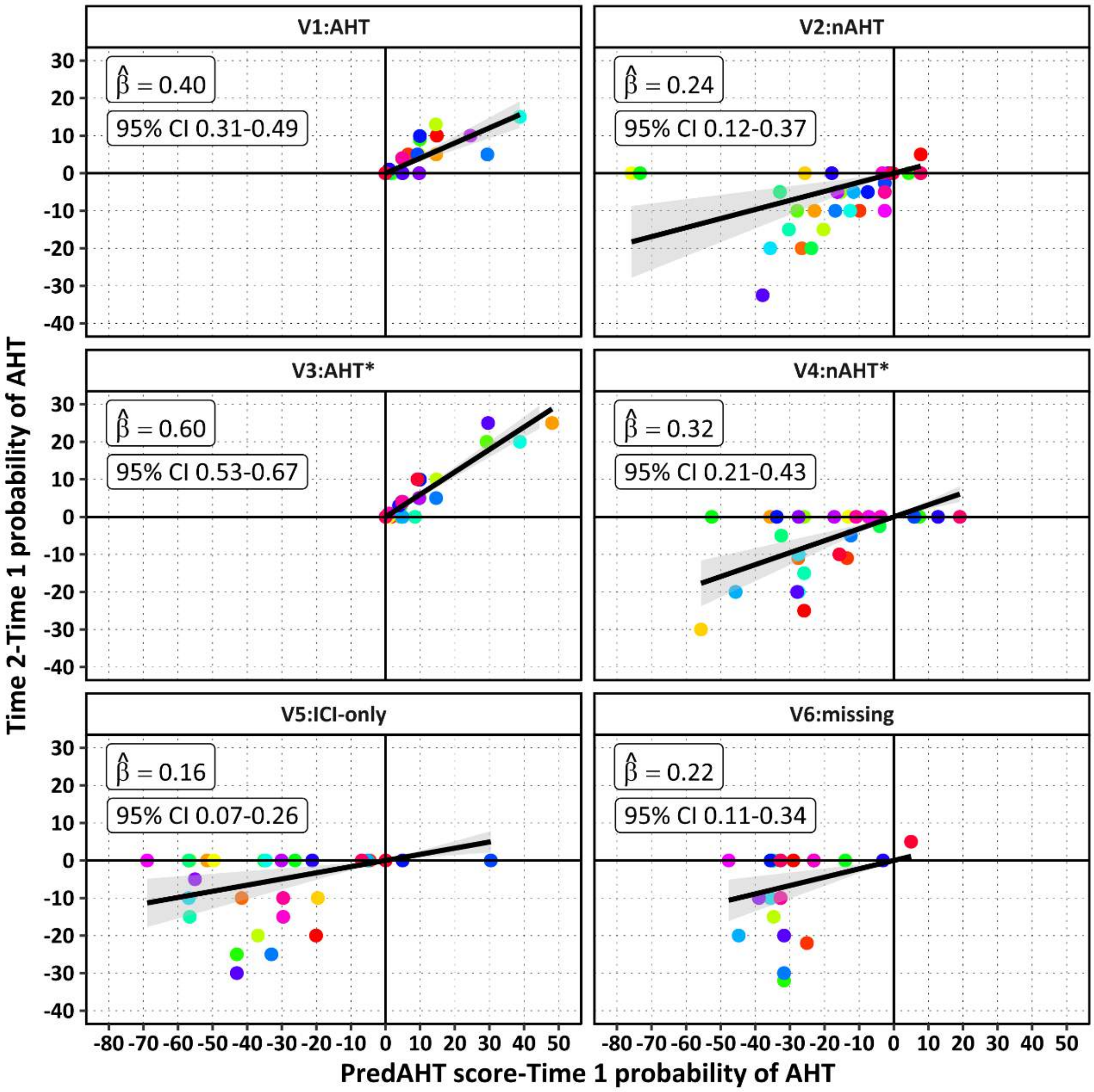
The lines connect each clinician's Time 1 and Time 2 probability of AHT. Longer lines represent a greater change in probability.

6.4.1.4 Impact of PredAHT on clinicians probability estimates of AHT

The PredAHT score significantly influenced clinicians' AHT probability estimates in all vignettes ($p<.001$). Figure 6.5 shows the estimated linear model slope coefficients $\hat{\beta}$ and 95% confidence intervals for each vignette. Higher slope coefficients $\hat{\beta}$ indicate a greater impact of PredAHT on clinicians' AHT probability estimates. Overall PredAHT had the greatest impact on clinicians' probability estimates of AHT in "V3:AHT*" and the least impact in "V5:ICI-only". Compared to the confirmed cases of AHT and nAHT ("V1:AHT" and "V2:nAHT"), PredAHT had a greater impact in the "grey" case "V3:AHT*", but a lesser impact in the "grey" cases "V5:ICI-only" and "V6:missing". As hypothesized, PredAHT had a greater impact in "V3:AHT*" compared to "V1:AHT" and a greater impact in "V4:nAHT*" compared to "V2:nAHT". Visual inspection of diagnostic plots did not reveal any major deviations from linearity, normality or homoscedasticity (Appendix 30). There were some influential outliers, however following inspection of the raw data these were not removed as they did not appear to have arisen from incorrectly entered or measured data. In addition, outliers are likely to be accounted for within the final mixed model described below.

Mixed modelling revealed a significant impact of PredAHT on clinicians' probability estimates of AHT overall across vignettes. Based on likelihood ratio tests, the six models that allowed the average proportion moved to vary across the categorical demographic variables were not significantly different from the reduced model at the level of $p=.05$, therefore the mixed model that best fit the data was the reduced model ($\hat{\beta}$ = 0.35, SE = 0.07, t = 5.05, $p<.001$, 95% confidence interval 0.21–0.50). PredAHT appeared most influential for those based at teaching hospitals, for those other than general or community paediatricians, for younger clinicians, for clinicians with the least CP experience and no formal training in paediatric head injuries, and for trainee doctors (Table 6.6), however none of these differences were statistically significant (Table 6.7). Variation in the slope coefficients $\hat{\beta}$ was greater between clinicians than between vignettes (Table 6.8 and Figure 6.6). This means that the impact of PredAHT was reasonably consistent across vignettes, but varied between individual clinicians. Comparison of the mean residuals with a simple linear model with no random effects revealed that the mixed model addresses clustering and reduces unexplained (residual) variance, as evidenced by the reduced variation in clinician/vignette-specific means and smaller standard error bars (Figures 6.7A and 6.7B). Visual inspection of diagnostic plots did not reveal any major deviations from linearity, normality or homoscedasticity, or any influential outliers (Appendix 31).

Figure 6.5 The impact of the Predicting Abusive Head Trauma tool on clinicians' probability estimates of AHT for each of the six vignettes



Coloured dots represent the 29 different clinicians. Higher coefficients $\hat{\beta}$ indicate a greater impact of PredAHT on clinicians' probability estimates of AHT. Points at 0 on the x-axis indicate no difference between the clinicians' Time 1 probability estimate of AHT and the PredAHT score. Points at 0 on the y-axis indicate no change in clinicians' probability estimates of AHT from Time 1 to Time 2. Points greater than 0 on the y-axis indicate an increase in clinicians' probability estimates of AHT from Time 1 to Time 2. Points less than 0 on the y-axis indicate a decrease in clinicians' probability estimates of AHT from Time 1 to Time 2.

Table 6.6 Impact of PredAHT on clinicians probability estimates of AHT, stratified by levels of clinician demographic variables

Demographics	Fixed Effects					
	$\widehat{\beta}$	Standard Error	t value	p value	95% Confidence Interval	
					Lower Bound	Upper Bound
District hospital	0.34	0.11	3.19	<.01	0.13	0.55
Teaching hospital	0.36	0.07	4.80	<.001	0.20	0.51
Community paediatrician	0.37	0.08	4.69	<.001	0.21	0.54
General paediatrician	0.25	0.09	2.68	<.05	0.06	0.44
Other specialty	0.46	0.12	3.95	<.001	0.22	0.69
Age 25–34	0.48	0.17	2.77	<.05	0.13	0.84
Age 35–44	0.31	0.09	3.62	<.01	0.14	0.49
Age 45–54	0.33	0.10	3.32	<.01	0.13	0.53
Age 55–64	0.41	0.11	3.83	<.001	0.20	0.63
Years in CP 5–9	0.42	0.10	4.09	<.001	0.21	0.62
Years in CP 10–20	0.27	0.10	2.74	<.05	0.07	0.47
Years in CP >20	0.36	0.08	4.35	<.001	0.19	0.54
Paediatric HI training	0.31	0.08	4.18	<.001	0.16	0.48
No paediatric HI training	0.43	0.09	4.53	<.001	0.24	0.63
Consultant	0.35	0.07	4.73	<.001	0.20	0.51
Associate/nurse specialist	0.26	0.11	2.42	<.05	0.04	0.48
Trainee	0.52	0.14	3.63	<.001	0.23	0.81

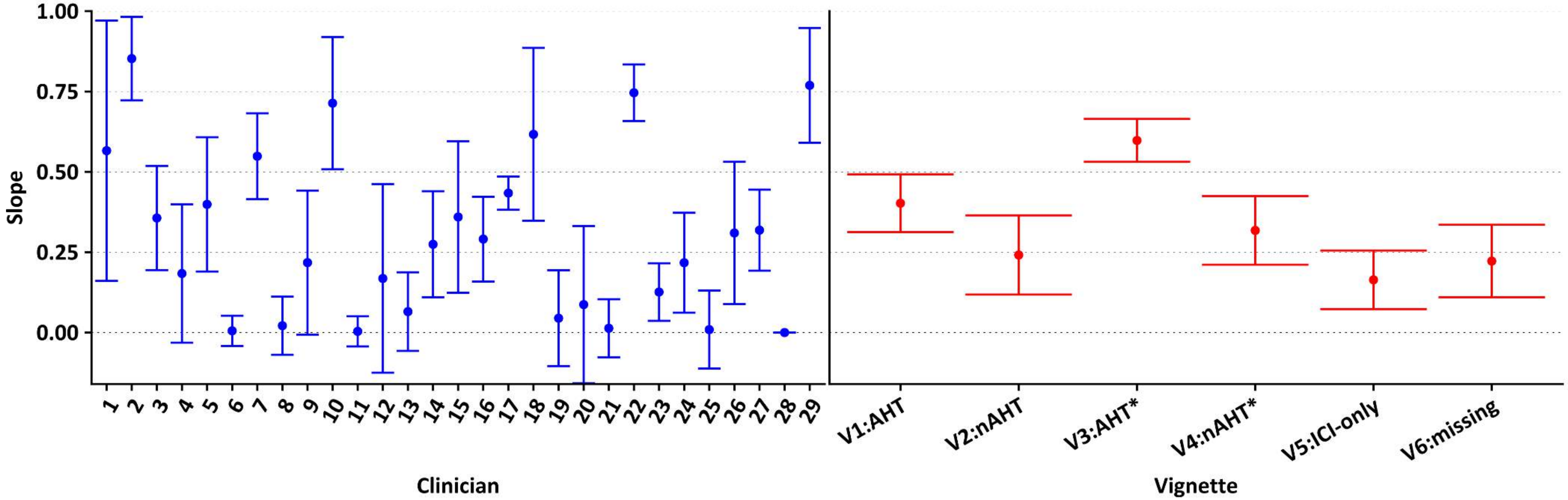
Table 6.7 Pairwise comparisons of the levels of clinician demographic variables

Contrast	Estimate	Standard Error	df	t ratio	p value
District - Teaching	0.28	1.42	29.30	0.19	.85
Community - General	-1.69	1.33	25.96	-1.27	.42
Community - Other	1.12	1.63	26.19	0.69	.77
General - Other	2.81	1.75	26.17	1.60	.26
Age 25–34 - 35–44	-2.34	2.44	26.06	-0.96	.77
Age 25–34 - 45–54	-2.08	2.54	27.29	-0.82	.84
Age 25–34 - 55–64	-0.94	2.60	25.69	-0.36	.98
Age 35–44 - 45–54	0.26	1.46	28.86	0.18	.99
Age 35–44 - 55–64	1.40	1.57	24.28	0.89	.81
Age 45–54 - 55–64	1.14	1.71	26.72	0.67	.91
Years in CP 5–9 - 10–20	-2.00	1.64	25.29	-1.22	.45
Years in CP 5–9 - >20	-0.72	1.49	26.44	-0.48	.88
Years in CP 10–20 - >20	1.28	1.43	26.50	0.90	.65
No HI training - HI training	-1.56	1.29	26.56	-1.20	.24
Consultant - Associate	-1.22	1.47	25.84	-0.83	.69
Consultant - Trainee	2.30	1.95	26.33	1.18	.48
Trainee - Associate	-3.51	2.22	26.00	-1.58	.27

Table 6.8 Random effects estimates from a linear mixed model of the impact of PredAHT on clinicians AHT probability estimates

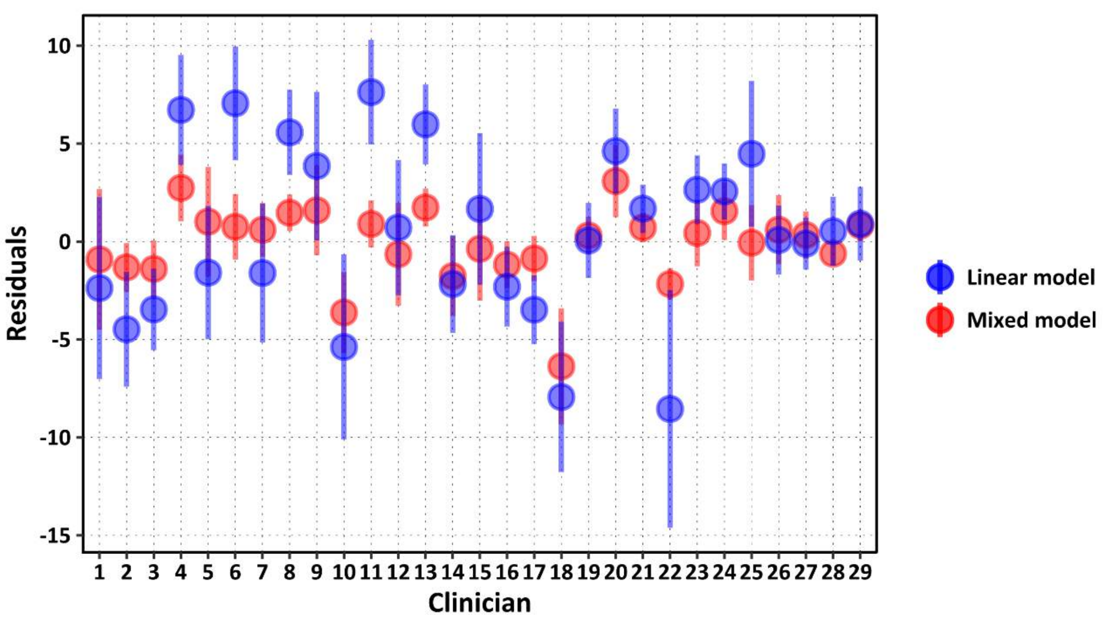
Random Effects				
Groups	Variance	Standard Deviation	95% Confidence Interval	
			Lower Bound	Upper Bound
Clinician	0.05	0.23	0.17	0.31
Vignette	0.02	0.12	0.07	0.23
Residual	23.72	4.87	4.35	5.50

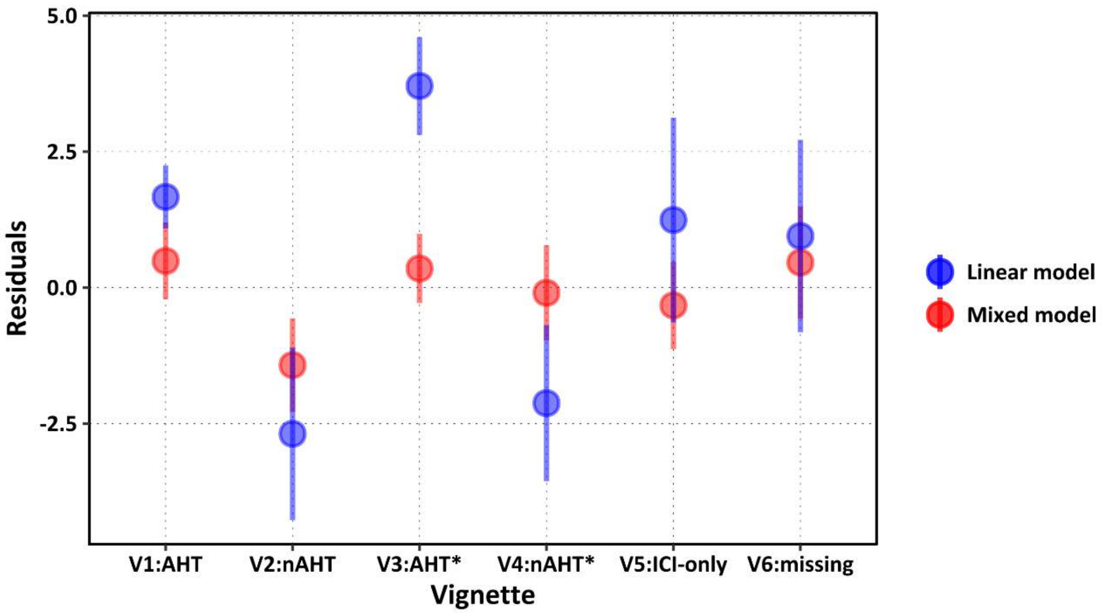
Figure 6.6 Slope coefficients $\widehat{\beta}$ for each of the 29 clinicians relative to the slope coefficients $\widehat{\beta}$ for each of the six vignettes



Slope coefficients $\hat{\beta}$ for the clinicians were derived by fitting 29 separate models, and slope coefficients $\hat{\beta}$ for the vignettes were derived by fitting six separate models. Error bars represent 95% confidence intervals. Higher coefficients $\hat{\beta}$ indicate a greater impact of the Predicting Abusive Head Trauma tool on clinicians' probability estimates of abusive head trauma.

Figures 6.7A and 6.7B Clinician-specific and vignette-specific mean residuals of the reduced mixed model versus a simple linear model with no random effects





Error bars represent standard errors.

6.4.1.5 Impact of PredAHT on clinicians proposed child protection actions

The majority of clinicians would have referred to social services at both Time 1 and Time 2 in all cases except "V2:nAHT", where most clinicians elected to investigate further (Figure 6.8). However, 9/29 (31%) clinicians changed their proposed CP action in 11/174 (6%) instances after seeing the PredAHT score (Figure 6.9). In four instances, clinicians escalated their proposed CP actions after seeing the PredAHT score, and in seven instances, they downgraded their proposed actions. Three clinicians changed their CP actions despite not altering their own probability estimate of AHT after seeing the PredAHT score. Figure 6.10 shows clinicians' proposed CP actions against their estimated probabilities of AHT. Their probability thresholds on which to act varied considerably; one clinician would have taken no further CP action in a case where they deemed the probability of AHT to be 45%, two clinicians would have referred to children's services at a probability threshold of 10%, and a number of clinicians would not have referred to children's services despite giving estimated probabilities of AHT of 85% or greater. There was no statistically significant association between a change in proposed CP action and hospital type, χ^2 (1, N = 174) = 1.64, p=.20, clinician specialty (p=.16, two-sided Fisher's exact test), clinician age (p=.14, two-sided Fisher's exact test), years of CP experience (p=.33, two-sided Fisher's exact test), paediatric head injury training, χ^2 (1, N = 174) = 0.60, p=.44, clinician seniority (p=.37, two-sided Fisher's exact test), or vignette (p=.94, two-sided Fisher's exact test). Clinicians' proposed CP actions at Time 1 and Time 2 are shown in Appendix 32, stratified by vignette.

Figure 6.8 Clinicians' Proposed Time 1 and Time 2 Child Protection Actions for each of the six vignettes

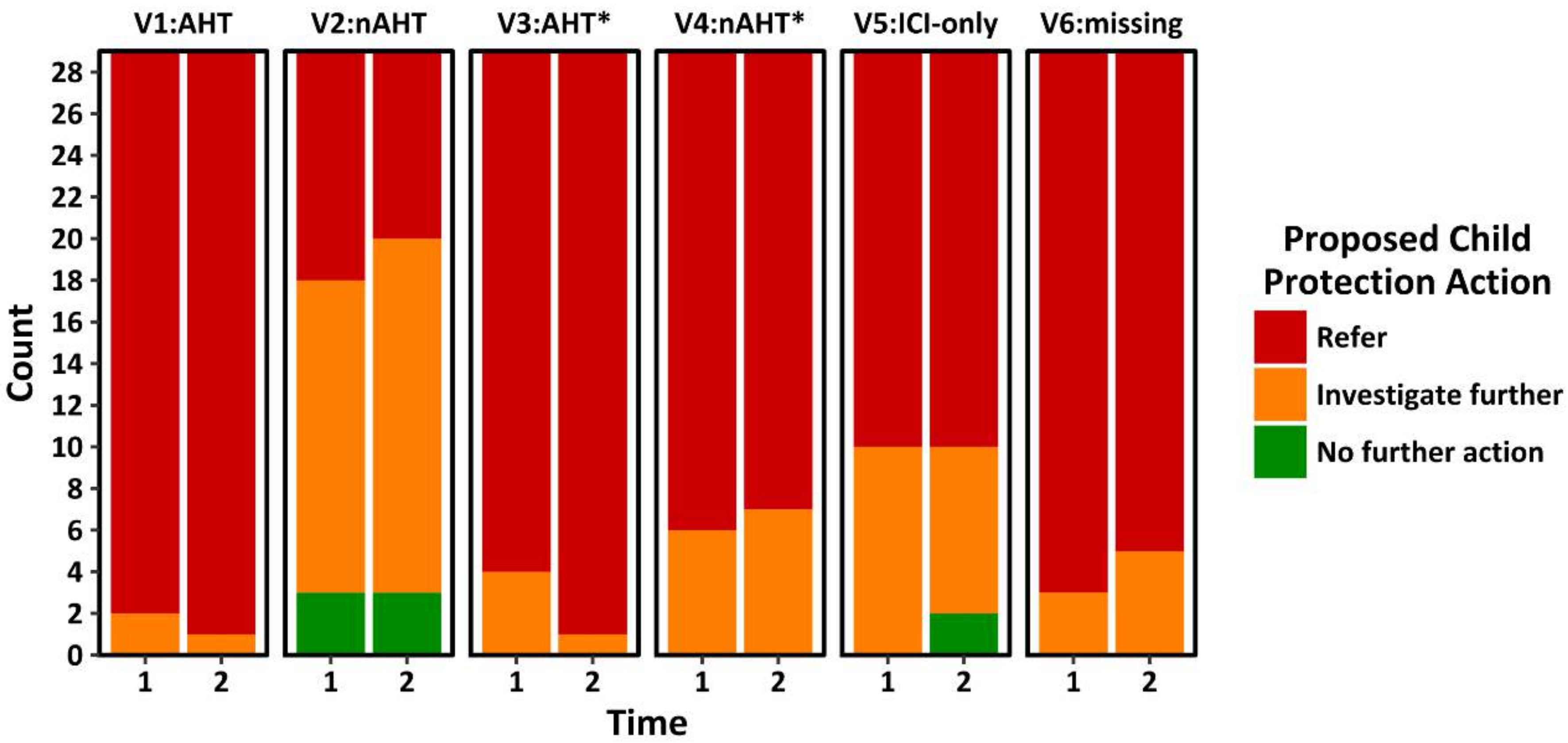
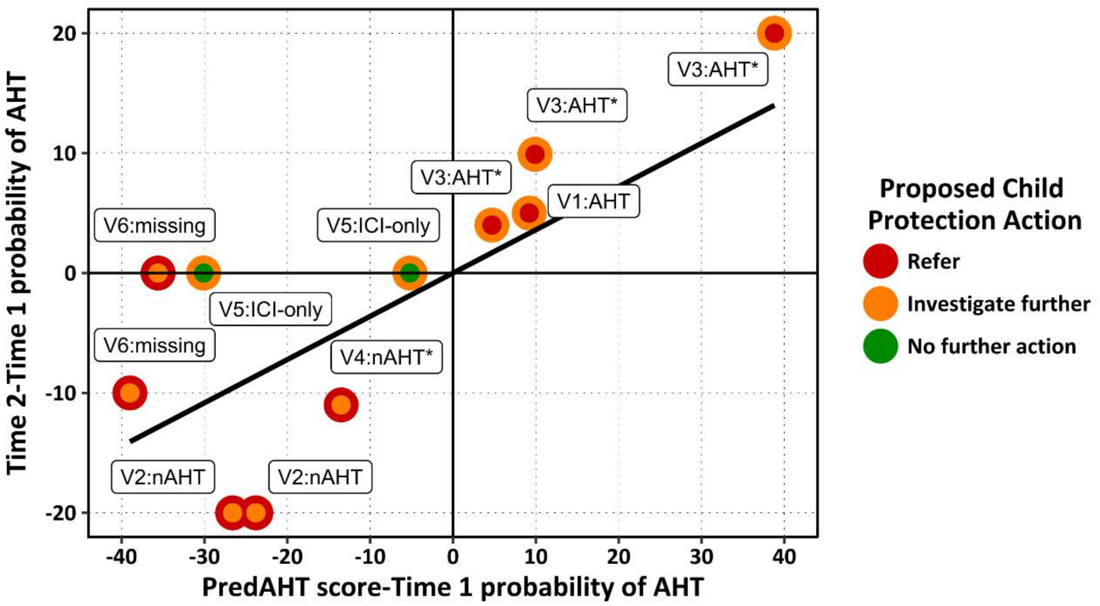
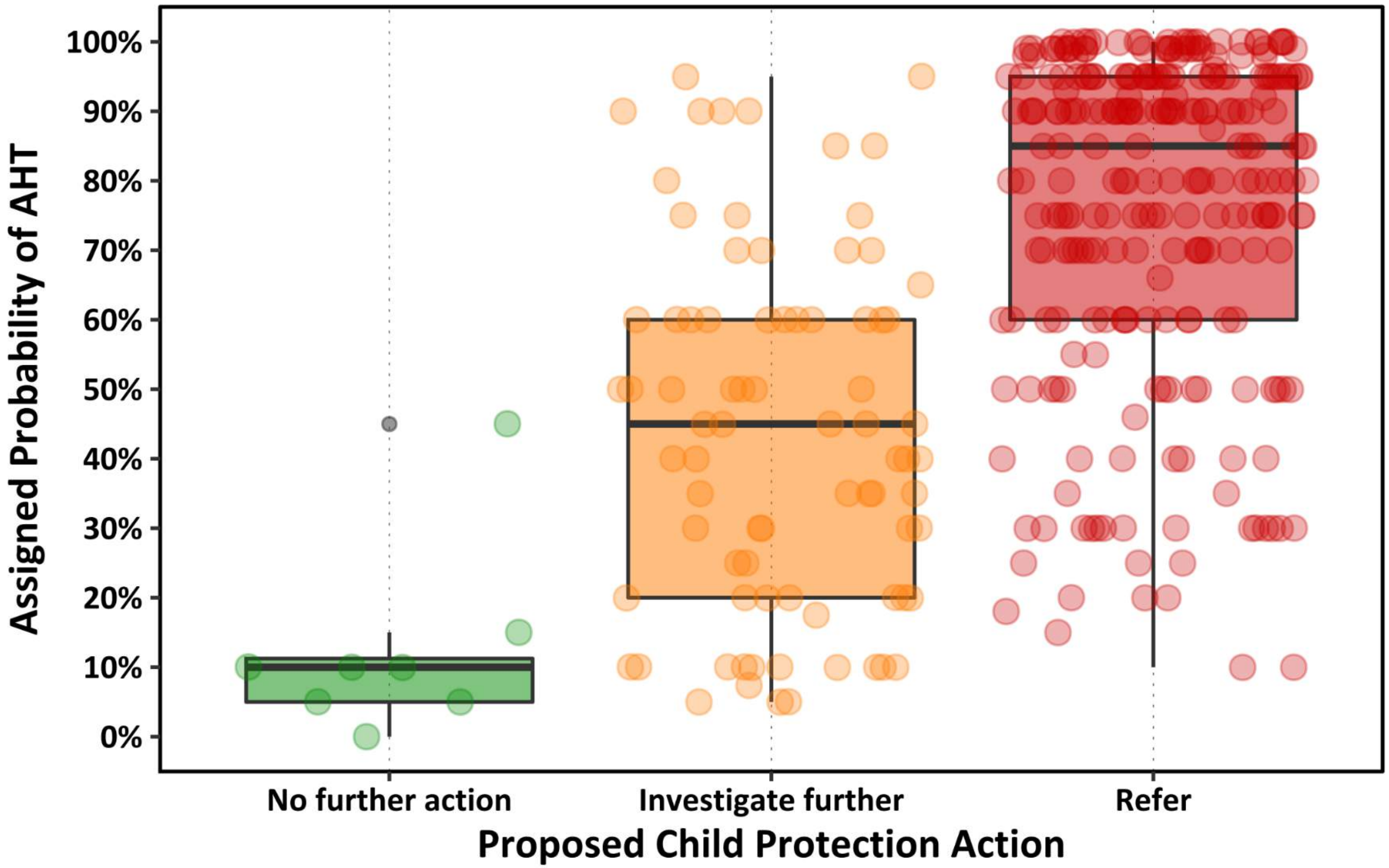


Figure 6.9 Impact of PredAHT on Clinicians' Proposed Child Protection Actions



Seven clinicians changed their proposed CP action for one of the six vignettes. Two clinicians changed their proposed CP action for two of the six vignettes. Clinicians' Time 1 actions (before PredAHT) are indicated by the larger circle, and their Time 2 actions (after PredAHT) are indicated by the smaller circle. Points at 0 on the y-axis indicate that clinicians did not change their probability estimate of abusive head trauma (AHT) from Time 1 to Time 2, despite changing their CP action. Points greater than 0 on the y-axis indicate an increase in clinicians' probability estimates of AHT from Time 1 to Time 2. Points less than 0 on the y-axis indicate a decrease in clinicians' probability estimates of AHT from Time 1 to Time 2.

Figure 6.10 Clinicians' proposed child protection actions against their estimated probabilities of AHT



Black dots represent outliers. Clinicians' probability thresholds on which to act varied considerably.

6.4.1.6 *Inter-rater reliability of clinicians' probability estimates of abusive head trauma*

The ICC coefficients are reported in Table 6.9. Inter-rater reliability of clinicians' prior and Time 1 probability estimates of AHT for the vignettes ranged from "poor" to "good", while their Time 2 probability estimates of AHT ranged from "poor" to "excellent".[912] The difference in agreement from Time 1 to Time 2 did not reach statistical significance, as indicated by the overlapping confidence intervals.

Table 6.9 Inter-rater reliability of clinicians' prior, Time 1 and Time 2 probability estimates of AHT for the six vignettes

	Intraclass Correlation Type	Intraclass Correlation	95% Confidence Interval		F Test With True Value 0			
			Lower Bound	Upper Bound	Value	*df*1	*df*2	Sig
Prior probability	Single measures	.55	.31	.88	41	5	140	.000
Time 1 probability	Single measures	.59	.35	.90	50	5	140	.000
Time 2 probability	Single measures	.66	.42	.92	75	5	140	.000

6.4.1.7 *Inter-rater reliability of clinicians' proposed child protection actions*

Gwet's AC_1 statistics and associated precision measures are reported in Table 6.10. Inter-rater agreement of clinicians' Time 1 and Time 2 CP actions was "fair".[921, 938] Therefore, PredAHT did not improve inter-rater agreement of clinicians CP actions.

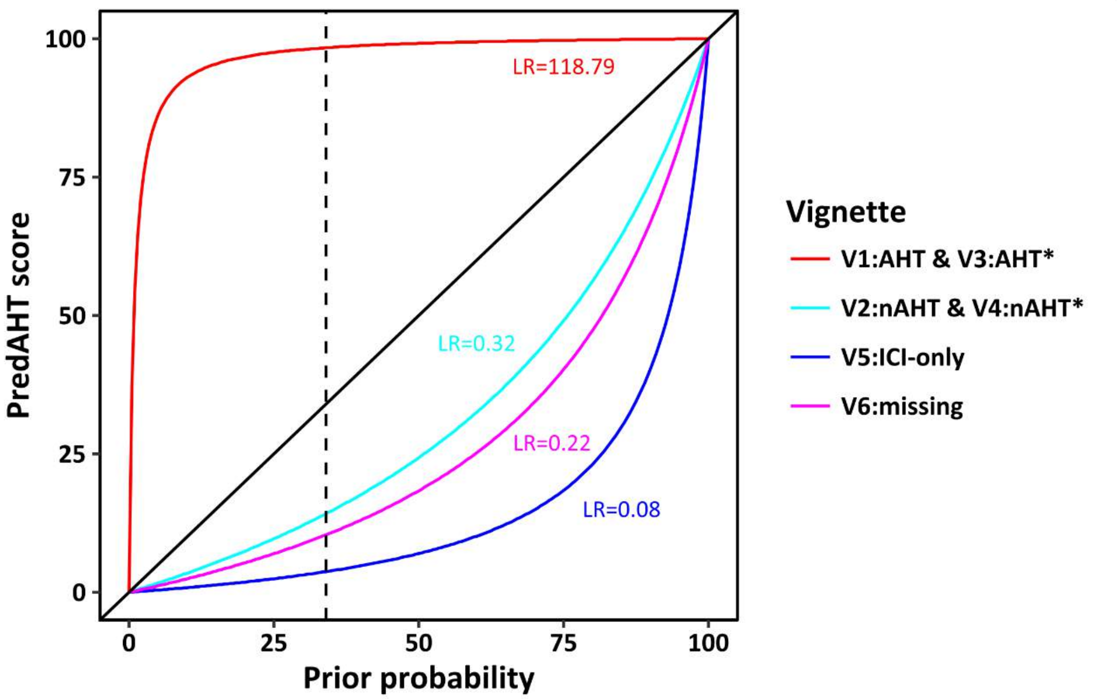
Table 6.10 Inter-rater reliability of clinicians' Time 1 and Time 2 proposed child protection actions for the six vignettes, measured by Gwet's AC_1 coefficient

Statistics	Time 1	Time 2
Percent agreement	67.0%	68.5%
Percent chance agreement	19.0%	19.5%
Gwet's AC_1 agreement coefficient	59.3%	60.9%
Standard error due to the sampling of vignettes	11.2%	13.3%
Standard error due to the sampling of clinicians	7.1%	6.6%
Unconditional standard error	13.3%	14.9%
95% CI conditionally upon the sample of clinicians	30.5%–88.1%	26.6%–95.1%
Unconditional 95% CI	33.3%–85.3%	31.7%–90.0%
Coefficient critical value	21.8%	24.6%
Coefficient recommended for interpretation	37.4%	36.3%

6.4.1.8 Impact of clinicians' prior probabilities on the PredAHT score

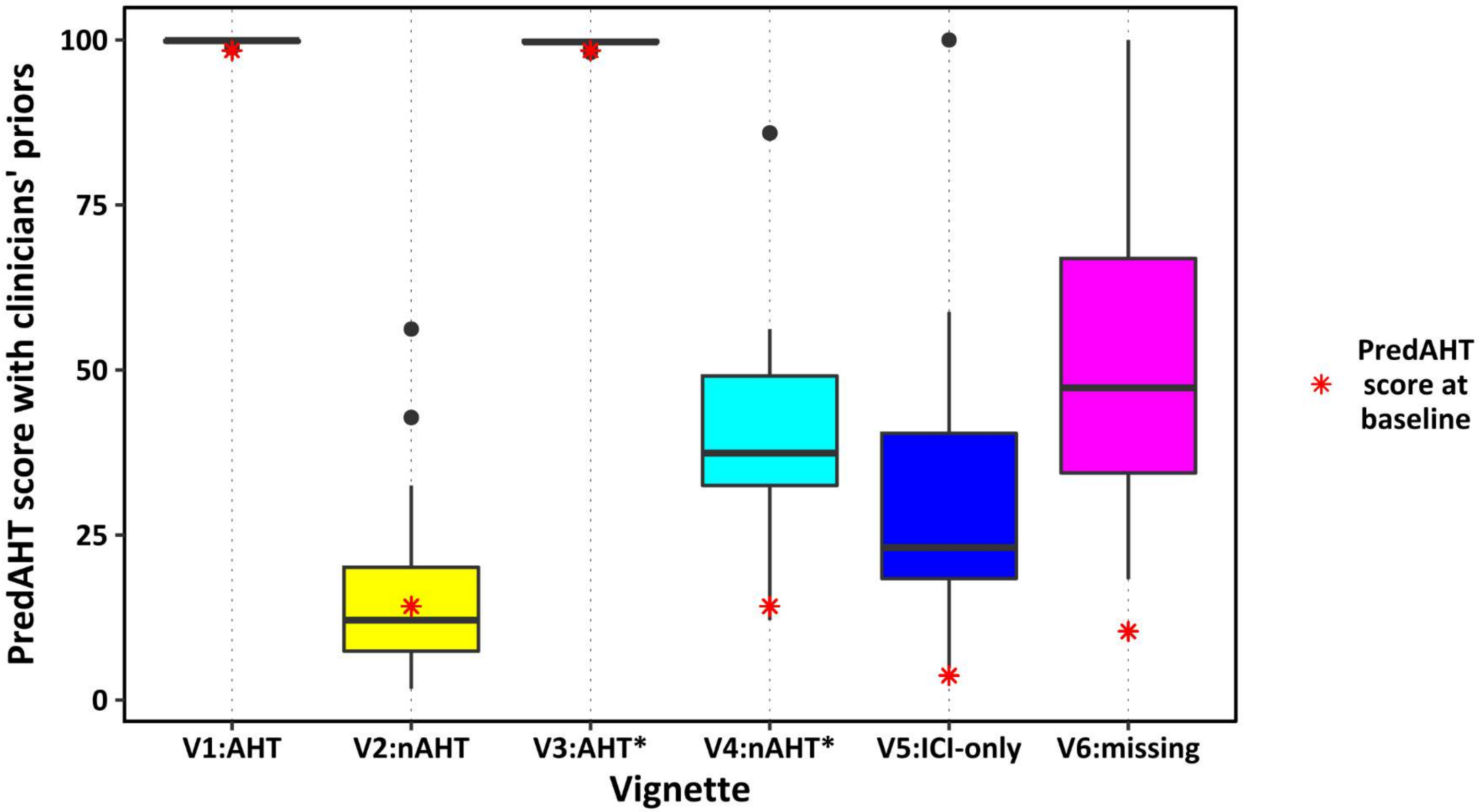
The relationship between prior probabilities and the PredAHT score (i.e. post-test probabilities) for each of the six vignettes is shown in Figure 6.11. For "V1:AHT" and "V3:AHT*", a high PredAHT score is obtained even if a very low prior probability is chosen. For all other vignettes, the PredAHT score increases by varying degrees as the prior probability increases. Figure 6.12 compares the PredAHT score, given a baseline prior probability of 0.34, with the scores obtained when clinicians' prior probabilities were incorporated. Scores incorporating clinicians' prior probabilities were similar to what would be obtained given the baseline prior for "V1:AHT", "V2:nAHT" and "V3:AHT*". However, PredAHT scores with clinicians' priors were higher than PredAHT scores using the baseline prior for "V4:nAHT*", "V5:ICI-only" and "V6:missing".

Figure 6.11 Relationship between prior probabilities and the PredAHT score for each of the six vignettes



LR = Likelihood ratio. The dashed line represents the baseline prior probability of 34%. Note that the combination of clinical features and therefore the likelihood ratio is the same in "V1:AHT" and "V3:AHT", and in "V2:nAHT and "V4:nAHT*", and thus the relationship between the prior probability and the PredAHT score is the same in these vignettes.*

Figure 6.12 Comparison of the PredAHT scores incorporating clinicians' prior's with the PredAHT scores using the baseline prior, for each of the six vignettes

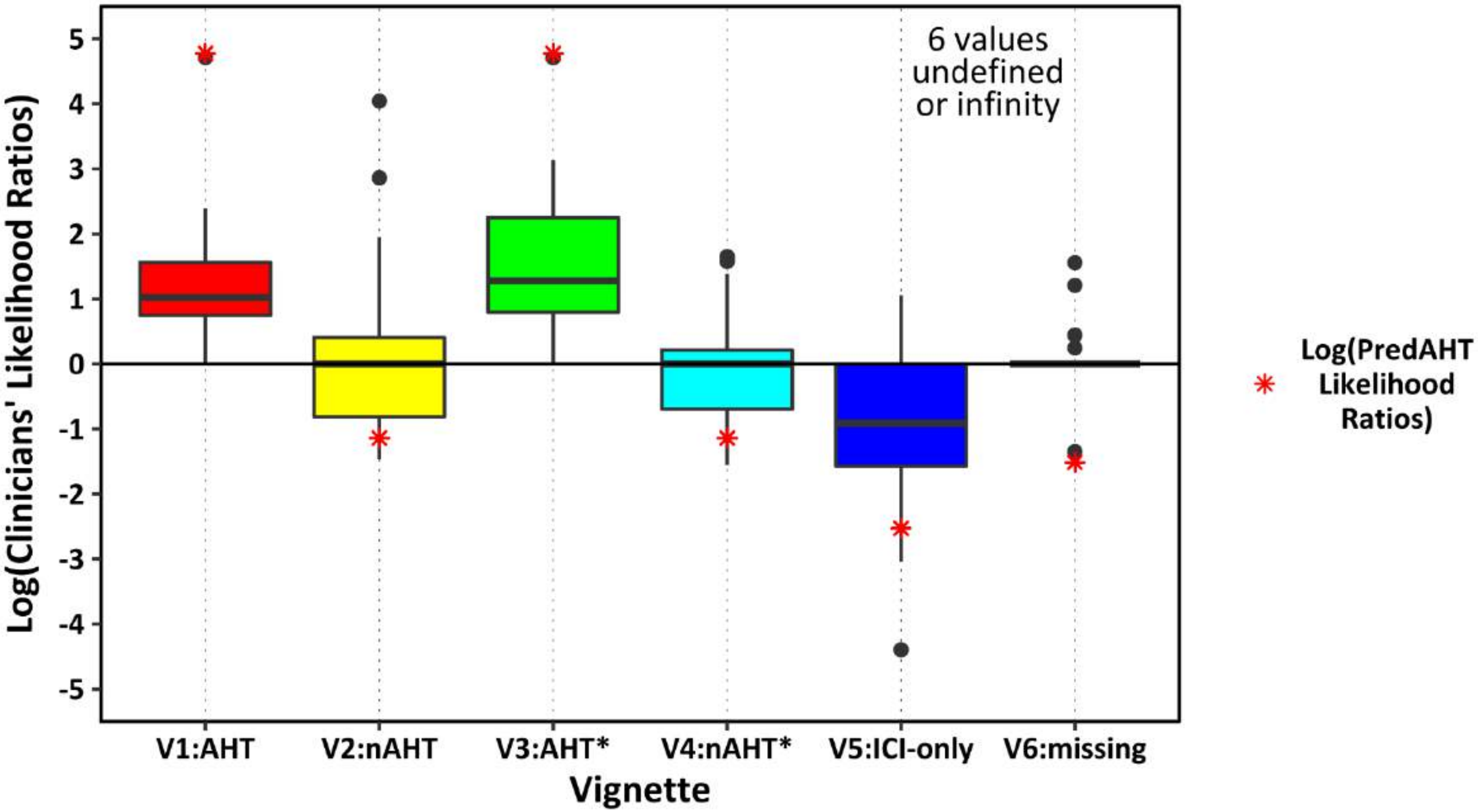


PredAHT scores incorporating clinicians' priors were higher than the PredAHT scores using the baseline prior in three of the six vignettes. Black dots represent outliers.

6.4.1.9 Comparison of clinicians' likelihood ratios with the PredAHT likelihood ratios

Figure 6.13 compares clinicians' likelihood ratios for the clinical evidence compared with the likelihood ratios calculated by PredAHT. Values were log-transformed using natural logarithms before being plotted, to aid interpretability. After seeing the clinical evidence presented in section two of the vignettes, on average clinicians' probability estimates of AHT increased in "V1:AHT" and "V3:AHT*", decreased in "V5:ICI-only" and stayed the same in "V2:nAHT", "V4:nAHT*" and "V6:missing". Clinicians' likelihood ratios were much smaller than those calculated by PredAHT in "V1:AHT" and "V3:AHT*", and larger than those calculated by PredAHT in "V2:nAHT", "V4:nAHT*", "V5:ICI-only" and "V6:missing".

Figure 6.13 Comparison of clinicians' (log) likelihood ratios with the PredAHT (log) likelihood ratios for each of the six vignettes



Values greater than 0 represent an increase in the probability of AHT after seeing the clinical evidence, i.e. from the prior probability to Time 1. Values less than 0 represent a decrease in the probability of AHT from the prior probability to Time 1. Values of 0 represent no change in the probability of AHT from the prior probability to Time 1. Black dots represent outliers.

6.4.2 Qualitative results

Four overarching themes were identified: clinicians' rationale for their responses, evaluations of PredAHT, interpretations of probabilities, and comments on the vignettes. Data are presented using quotations, selected as examples of the themes that emerged from the data. Within the quotations, square brackets represent text inserted by the researcher for clarification. Categories are highlighted in bold text.

6.4.2.1 Rationale for responses

The "think-aloud" method allowed clinicians to provide a rationale for their estimated probabilities of AHT and proposed CP actions for each vignette. Clinicians' comments confirmed that they thought the **"grey" cases were difficult** to classify.

> "I find these really difficult, the 3-month-old rolling off the sofa." ***Clinician 13, "V5:ICI-only" & "V6:missing"***

> "These are the difficult ones in practice." ***Clinician 16, "V5:ICI-only" & "V6:missing"***

> "I'm less confident with estimating the probability with this one." ***Clinician 10, "V5:ICI-only"***

Most clinicians reported that the presence of a **concerning social history increased their suspicion** of AHT in "V4:nAHT*" as compared to "V2:nAHT". This explains why participants' estimated probabilities of AHT were higher on average for "V4:nAHT*" than for "V2:nAHT" even though the clinical features were the same. However, some clinicians said that they placed more weight on the lack of additional clinical features concerning for AHT.

> "It shows what informs you in these cases, because for me the social services involvement and domestic violence are important." ***Clinician 20, "V4:nAHT*"***

> "All this [the history] is far more important, you know, it plays a big part in my assessment of these children." ***Clinician 4, "V4:nAHT*"***

> "The lack of clinical features is more important to me than the history here." ***Clinician 25, "V4:nAHT*"***

Some clinicians remarked that although the clinical features in these two cases are not particularly specific for abuse, the concerning history in "V4:nAHT*" indicates that the child may be a child in need and thus require further investigation and protection nevertheless.

> "Again I've got this cheek bruising that I don't like. I'm thinking this one might go down the child protection register route rather than the court process and removal and things, because even though the injury is more likely accidental, there are other concerns: bruising, domestic violence, delayed presentation – this is probably a child at risk." ***Clinician 16, "V4:nAHT*"***

Almost all of the clinicians reported that they were highly suspicious of AHT in "V3:AHT*" due to the concerning clinical features, although the history was potentially less concerning than in "V1:AHT" (no history of trauma). Clinicians stated that the **history did not match the severity of the injuries** sustained.

> "The history doesn't fit the level of trauma." ***Clinician 9, "V3:AHT*"***

> "A 14-month old is likely to be mobile, but they won't have a brain injury from falling, even on a floor like that." ***Clinician 16, "V3:AHT*"***

> "I am not happy with the history as 14-month old children fall a lot and don't get subdural haemorrhages." ***Clinician 8, "V3:AHT*"***

Importantly, one clinician stated that although the history does not affect whether or not she carries out the appropriate investigations for suspected AHT, it can unduly influence the final multidisciplinary diagnosis.

> "The history doesn't make you go down one path or another – it doesn't influence what you do and the decisions that you make but it does influence the outcome of the strategy meeting and your gut feelings." ***Clinician 5, "V3:AHT*"***

Clinicians gave reasons as to why they **disagreed with the tool** and confirmed why PredAHT had the lowest impact on their probability estimates for "V5:ICI-only" and "V6:missing".

> “This is where the tool takes away some of the subtlety in the history, this is where I would say I don’t care what it says.” ***Clinician 12, “V5:ICI-only”***

> “The tool is not useful for these two cases as the presence of a subdural alone in a baby that potentially cannot roll is very worrying. This is where clinicians will doubt the tool.” ***Clinician 14, “V5:ICI-only” & “V6:missing”***

> “The history is very suspicious which is why I am not changing my probability.” ***Clinician 8, “V5:-ICI-only”***

Although clinicians were informed about the imputation strategy built into PredAHT to account for missing investigations, they were reluctant to change their probability estimates from Time 1 to Time 2 in “V6:missing” because they didn’t have the full clinical picture, and stated that PredAHT might act as a prompt for ordering further investigations.

> “I would always do the skeletal survey so I can’t say if my probability would come down as I would need that information to make a decision.” ***Clinician 22, “V6:missing”***

> “That just goes to show how much you need the other information doesn’t it?” ***Clinician 19, “V6:missing”***

> “That’s the reason for doing the whole package isn’t it because if these things are absent it brings you right down again.” ***Clinician 10, “V6:missing”***

Similarly, clinicians explained why they *were* willing to change their probability estimates of AHT after seeing the PredAHT score.

> “Bruising to the cheeks made me worried but the tool would then reassure me to pull it back down.” ***Clinician 17, “V2:nAHT”***

> “Would the 24% make me bring my probability down? Yeah it would actually, I would give them the benefit of the doubt.” ***Clinician 17, “V4:nAHT*”***

Clinicians’ reasons for their estimated probabilities of AHT and proposed CP actions included **knowledge of the clinical features** indicative of AHT and non-AHT. Clinicians stated that their clinical knowledge sometimes increased their suspicions of AHT, sometimes it decreased their suspicions, and sometimes it contributed to their uncertainty about a case.

> "Retinal haemorrhages would increase my suspicion." ***Clinician 16, "V1:AHT"***

> "Bruises to the cheeks raises my suspicion." ***Clinician 18, "V2:nAHT"***

> "I would not be too concerned as the chair is very high, it is a linear undisplaced skull fracture and that type of floor is quite a hard floor." ***Clinician 16, "V2:nAHT"***

> "Left parietal skull fracture, the most common skull fracture in both abused and non-abused children." ***Clinician 21, "V2:nAHT"***

Clinicians reported that a lack of clinical knowledge also contributed to uncertainty in estimating the probability of AHT.

> "See this is going into detail about the eye findings some of which I don't know the significance of." ***Clinician 15, "V1:AHT"***

Clinicians considered the **age and developmental stage** of the child when estimating the probability of AHT.

> "I am less confident [that this is abuse] because they are older and could have fallen off a chair." ***Clinician 1, "V2:nAHT"***

> "A three-month old can't roll so the history is immediately suspicious." ***Clinician 9, "V5:ICI-only" & "V6:missing"***

> "I would think 3-month old rolling, really? And off a sofa is probably quite a significant height for a baby but I'd still be concerned that they might not be able to roll." ***Clinician 20, "V5-ICI-only" & "V6:missing"***

Some clinicians said that an important factor that influences their probability estimates of AHT is a **consistent history**. However, clinicians explained this is less important when there were other concerning features present.

"I would put much lower for this one [compared to "V4:nAHT*"] as it's an immediate presentation and a consistent history." ***Clinician 17, "V2:nAHT"***

"Unless there was some inconsistency in the history I would be erring on accidental." ***Clinician 17, "V6:missing"***

"Even though the story is consistent the history is still dodgy and the neuroimaging features are suspicious." ***Clinician 25, "V3:AHT*"***

When completing the vignettes, clinicians deliberated over the purported **mechanism of injury** and whether this was consistent with the features observed.

"I would be worried that there's no bruising because that means there's no impact." ***Clinician 21, "V3:AHT*"***

"They have no bruising but they have rolled onto the floor?" ***Clinician 23, "V6:missing"***

"I wouldn't expect multiple bilateral injuries from just one fall." ***Clinician 16, "V5:ICI-only"***

6.4.2.2 Evaluations of PredAHT

Participants talked about the **potential benefits** of PredAHT while completing the vignettes. Overall, 27/29 clinicians said that they would find PredAHT useful in their practice; two said that they were unsure whether PredAHT would be useful for them.

"This would undoubtedly be extremely useful." ***Clinician 6***

"I think this would be more useful for older children but I'm not sure it actually adds much." ***Clinician 15***

Clinicians reported that they would find PredAHT useful as they do not usually think in terms of probability when assessing risk.

"I never give percentages, even in court I would say that we don't talk in those terms, and that's why I think the tool is going to be helpful." ***Clinician 5***

Many clinicians stated that they felt reassured by PredAHT, and reported that it would give them more confidence in their opinions, even if they would not change their CP actions based on the score.

> "I would still need more information about the cheek bruising but the low score (24%) would reassure me." ***Clinician 16, "V2:nAHT"***

> "The history just doesn't fit with the level of trauma…the score helps to remind you that you are right to be concerned and helps you not to be too sensitive about the family." ***Clinician 9, "V3:AHT*"***

> "The 7% would make me much more confident that this is an accident." ***Clinician 10, "V5:ICI-only"***

> "My estimate is very close to PredAHT, so I wouldn't change my actions but my agreement with PredAHT would give me more confidence in expressing my opinion to multiagency colleagues." ***Clinician 16, "V3:AHT*"***

> "I think mostly where it helps is reassuring you." ***Clinician 27, "V1:AHT"***

> "It would be helpful at the end to validate my opinion that probably it is abuse." ***Clinician 25, "V3:AHT*"***

Participants also discussed the **potential risks** of PredAHT. Some thought that variables relating to the history should be included in the tool.

> "There's no factor for the lack of history is there which is key isn't it?" ***Clinician 3***

> "It must be used with caution. I think it would be useful in some cases, but for me the history plays a big, big part in the investigation of these cases." ***Clinician 4***

Others felt that the tool cannot account for the subtleties that are often present in individual cases, or that since it cannot account for all possible indicators for abuse, a low score may provide false reassurance.

> "The cheek bruising is really worrying, it shows that PredAHT can't take into account nuances with just yes and no answers." ***Clinician 9, "V2:nAHT"***

> "The downside would be when you get given a low score when actually there are other indicators of abuse." ***Clinician 5***

Some participants also discussed at length the need to understand how PredAHT works, and the importance of critically appraising the quality of the data that it is based on.

> "I think we should be thinking that way [in probabilities] as long as we can understand and explain it." ***Clinician 3***

> "We would need to know where the figures in the tool came from, and to make sure they are correct." ***Clinician 22***

6.4.2.3 Interpretation of probabilities

Participants talked about their **probability thresholds** for investigation and referral as justification for their proposed CP actions. One clinician said that she would refer all cases she considered to have a 50% risk or greater of AHT to social services, but said that she would investigate cases she thought carried a lower probability of AHT.

> "All that matters [for referral] is whether it's over 50% or not." ***Clinician 3***

Many clinicians were interested in exploring the estimated post-test probabilities that PredAHT provided based on different prior probabilities. Some reported that they were shocked by the **impact the prior probability had on the PredAHT score,** however other clinicians justified their high estimated prior probabilities due to the neuroimaging features in the vignettes.

> "I'm shocked by how much my prior probabilities have affected the scores. This makes me think I might be too hawkish about abuse." ***Clinician 26***

> "I can only say a 90% prior probability for all of these [vignettes] because if there is a subdural haemorrhage, to me that's a really high probability."
> ***Clinician 5***

Some clinicians questioned how they might estimate their prior probability in practice and mentioned that in reality some of the clinical features included in PredAHT may be incorporated in their prior probability estimates.

> "That's interesting then to see how my gut feeling is coming in. Really I'm estimating the [prior] probability without knowing all the information. What are we taking into account with our prior probability [in practice] and what is our evidence for that?" ***Clinician 10***

6.4.2.4 Comments on vignettes

Comments on the details of the vignettes themselves revealed important information about clinicians' behaviour when assessing suspected AHT. Some clinicians questioned **why certain investigations were or were not performed** e.g. why a skeletal survey was not performed in "V1:AHT" and "V3:AHT*, and why a skeletal survey and ophthalmology exam were ordered in "V2:nAHT".

> "You would still need to do a skeletal survey even if the probability is already high." ***Clinician 14, "V1:AHT"***

> "I'm not sure I would have done any of these tests in this case!" ***Clinician 15, "V2:nAHT"***

In addition one clinician reported that they would only order certain investigations based on the results of previous ones.

> "I would only do a skeletal survey if the ophthalmology exam was positive." ***Clinician 28, "V6:missing"***

Some clinicians talked about **additional investigations** they would perform.

> "I don't know why you keep missing the bloods out!" ***Clinician 16***

Many participants reported needing more **detail on the history** in order to make more informed probability estimates or CP decisions.

> "It's tricky as I would want to know more history." ***Clinician 15, "V3:AHT*"***

"In real life you would be asking the next question and the next question wouldn't you?" ***Clinician 11, "V2:nAHT"***

"The problem is you would want so much more information. I would assess if they could roll in the department." ***Clinician 19, "V6:missing"***

In addition some participants said that they wanted more **detail on the clinical features,** including the age of the injuries, in order to assess whether the mechanism was possible.

"What side is the cheek bruising and is the bruising to the scalp the same side as the head injury?" ***Clinician 19, "V2:nAHT"***

"Do we know if the subdural haemorrhages are the same ages?" ***Clinician 16, "V5:ICI-only"***

The majority of clinicians stated that they were concerned about the cheek bruising in "V2:nAHT" and "V4:nAHT*", and said that they wanted more information about the pattern and mechanism of the bruising. This explains why some participants' estimated probabilities of AHT were high for "V2:nAHT", despite the fact that this vignette represented a confirmed case of non-AHT.

"I am concerned about the cheek bruising, it depends on where it was." ***Clinician 13, "V2:nAHT"***

"It would depend on the pattern of bruising to the cheeks." ***Clinician 4, "V2:nAHT"***

While considering the probability of AHT, clinicians discussed a variety of possible **differential diagnoses**, not detailed in the vignettes, that they would rule out in practice.

"There's not enough to rule out a medical condition e.g. sepsis." ***Clinician 24, "V1:AHT"***

"Could this be birth related?" ***Clinician 18, "V1:AHT"***

"He wouldn't have hit his head that badly just falling on a floor, unless he has got some bleeding disorder or something." ***Clinician 10, "V3:AHT*"***

6.5 Discussion

In this study, statistical modelling demonstrated that PredAHT significantly influenced clinicians' probability estimates of AHT in all vignettes. Interestingly however, clinicians' proposed CP actions were only influenced by PredAHT in a minority of instances, and PredAHT did not significantly improve the overall agreement between clinicians' AHT probability estimates or their proposed CP actions. Despite this, the "think-aloud" data showed that 27/29 clinicians would find PredAHT useful in their practice, and that it provided them with greater confidence in their opinions in the vignette cases, even if they would not alter their CP actions after seeing the score. However, it was evident that clinicians were influenced by a variety of social, historical and clinical factors in each case, emphasizing the need to consider the PredAHT probabilities in the context of these associated factors.

It was hypothesized that PredAHT would have a greater impact in the four "grey" cases compared to both the confirmed AHT and nAHT cases. However, this was only evident in the "grey" case "V3:AHT*". PredAHT had the greatest impact in this vignette, where there is a consistent history of a fall in a mobile child, but several concerning clinical features that do not fit with this history and which are strongly associated with AHT; PredAHT scores highly on this vignette (98.4% at baseline). The next largest effect was demonstrated in the confirmed AHT case "V1:AHT", where the clinical features and therefore the PredAHT score are identical to those in "V3:AHT*". The child in "V1:AHT" presented with no history of trauma whatsoever, a feature with a high specificity and positive predictive value for abuse.[9] Although all clinicians estimated the probability of AHT in "V1:AHT" and "V3:AHT*" to be 50% or greater at Time 1, on average they underestimated the likelihood ratio and predicted probability of AHT for this combination of clinical features compared to PredAHT, and increased their estimates after seeing the score. This suggests that PredAHT may act to increase clinicians' suspicions when there are a number of clinical features indicative of AHT, and that it may help clinicians to remain objective during their assessment of suspected AHT. PredAHT had the least impact in the "grey" case "V5:ICI-only", where the history and presentation is concerning, but due to the absence of any additional clinical features, the PredAHT score is low (3.7% at baseline). Reassuringly, this suggests that clinicians were not simply following PredAHT, but were considering factors that it cannot account for and considering the score alongside their own clinical judgment. Similarly, a number of clinicians reported disregarding the low PredAHT score (14.2% at baseline) for "V2:nAHT", the confirmed nAHT case, due to concerns about the cheek bruising, which is a recognized indicator of physical abuse.[941] Even those who felt reassured by the score would have requested further information about this feature.

It was also hypothesized that PredAHT would have a greater impact in "V3:AHT*" compared to "V1:AHT" and a greater impact in "V4:nAHT*" compared to "V2:nAHT". These hypotheses were supported by the data. On average clinicians' prior and Time 1 probabilities of AHT were higher in "V1:AHT" than "V3:AHT*" due to the lack of history of trauma. Therefore their initial estimates in "V3:AHT*" were further from the PredAHT score than in "V1:AHT", resulting in a greater difference between their Time 1 and Time 2 estimates. Similarly, on average clinicians prior and Time 1 probabilities of AHT were higher in "V4:nAHT*" than "V2:nAHT" due to the inconsistent history, delay in presentation and concerning social history. Therefore their initial estimates in "V4:nAHT*" were further from the PredAHT score than in "V2:nAHT" again resulting in a greater difference between their Time 1 and Time 2 estimates.

The impact of missing information was explored in "V6:missing", where an ophthalmology examination and skeletal survey had not been performed. Despite being aware of the imputation strategy built into PredAHT to account for missing information, the tool only had a small impact on clinicians' probability estimates of AHT in this vignette. This highlights the importance of obtaining an ophthalmology examination and skeletal survey wherever possible in suspected AHT cases, in line with international recommendations.[512, 521] Arguably, in the absence of a witnessed accidental injury, all children in this age group with ICI should undergo investigations for fractures and RH. However, only 68% of clinicians would have ordered an ophthalmology exam in "V6:missing", and only 68%–82% of clinicians would have ordered a skeletal survey in the vignettes where it hadn't already been performed. Very few clinicians would have ordered repeat skeletal surveys. Qualitative analysis of the "think-aloud" data and free text comments suggested that PredAHT may help to standardise investigations in suspected AHT by highlighting the clinical significance of fractures and RH in the context of suspected AHT, and the influence these investigation results would have on the PredAHT score, reiterating the findings from the earlier qualitative interview study reported in Chapter 5.[747]

Although PredAHT significantly influenced clinicians' probability estimates of AHT, there were only 11/174 instances where clinicians changed their proposed CP action after seeing the score. The tool would have prompted one clinician to make a referral to children's services in the confirmed AHT case "V1:AHT" and three clinicians to make a referral to children's services in "V3:AHT*". In "V5:ICI-only", two clinicians (one neuroradiologist and one consultant general paediatrician) reported that if they had used PredAHT, they would have taken no further CP action in this case, as they were reassured by the absence of additional clinical features and the low PredAHT score (3.7% at baseline). This is potentially a cause for concern, however in reality the assessment of AHT involves the input of multiple clinicians rather than the opinion of one individual. For statistical analysis purposes, the categories of CP action were collapsed, however some clinicians would have taken additional actions within the "investigate further" category after seeing the PredAHT score. Most notably, PredAHT would have prompted clinicians to discuss the case with a CP colleague alongside ordering further investigations or referring to children's services.

In "V1:AHT" and "V3:AHT*", where PredAHT consistently gave a high probability of AHT, even when clinicians prior probabilities were lower, most clinicians did not change their proposed CP actions as they had already elected to refer to children's services at Time 1. That is, they were aware that the clinical features described in these vignettes are highly concerning for AHT and warrant onward referral. In the other four vignettes, where PredAHT gave a comparatively low probability of AHT, clinicians may not have downgraded their initial proposed CP actions as they were concerned about other features in the vignettes that PredAHT does not account for. Similarly, qualitative data suggested that in "V4:nAHT*", some clinicians would have referred the child to children's services due to other concerns within the family rather than because of a suspicion of AHT. Additionally, probabilities are interpreted differently by different people, and the clinicians in this study had very different probability thresholds on which to act. There is little professional agreement as to what equates to a "reasonable suspicion" of abuse, varying in one study from a probability of 10%–35%, 40%–50% or 60%–70% and for a smaller group to >75%.[529] In another study, 51% of participants defined the term "reasonable medical certainty" in the context of child abuse as ≥90% probability, 30% defined it as ≤50% probability and 2% used a definition of ≤25% probability.[530] Furthermore, even when clinicians perceive the likelihood of abuse to be high, they may still not refer the child to children's services. Flaherty et al.,[490] found that clinicians in the United States only reported 73% of the children that they thought were likely or very likely abused to child protective services, and only 24% of children that they thought were possibly abused, despite the fact that child abuse reporting is mandatory in all 50 states. In the current study, some clinicians' probability thresholds for referral were as low as 10%, while some clinicians would not have referred even when they estimated the probability of AHT to be 95%. Other studies have found that improving clinicians' judgments of disease probability does not necessarily change or improve their treatment decisions and may have an unpredictable effect on clinicians' behaviour.[942-944] One possibility is that clinicians' proposed CP actions in this study were not based on probabilistic thresholds at all.[944] This is consistent with the observation that some clinicians changed their CP actions after seeing the PredAHT score, but not their probability estimate of AHT. Alternatively, this finding could suggest that PredAHT may help to reduce the uncertainty around clinicians' point estimates of the probability of AHT, and give them more confidence in their decisions; this was confirmed by the "think-aloud" data, where many clinicians stated that they felt reassured by the tool even if they did not change their proposed CP action.

This study found that clinicians' probability estimates of AHT for each vignette varied somewhat. This finding is consistent with other vignette studies evaluating the likelihood of AHT amongst clinicians. One such study asked US paediatricians to rate 16 cases of paediatric traumatic brain injury on a seven point scale ranging from definitive unintentional injury to definitive inflicted injury, and found they were unable to agree on the cause of the injuries in half of the scenarios.[533] Lindberg et al.[532] found extensive variability between experienced CP paediatricians when estimating the likelihood of abuse in video vignettes of cases referred to a hospital child abuse team, using three rating scales and a percentage probability. Neither of these studies estimated inter-rater reliability statistics. A further vignette study simulated eight case reports of femoral fracture, asking participants to categorize the cases into abuse and non-abuse, and to indicate appropriate investigations; 39% gave an incorrect diagnosis, and only 30% ordered the appropriate investigations.[535] In the current study, inter-rater reliability of clinicians' probability estimates of AHT across vignettes was "poor to good" based on the 95% confidence interval of the ICC statistic, while inter-rater reliability of their CP actions across vignettes was "fair". These findings are in contrast with two other recent studies estimating inter-rater reliability of clinicians' perceptions of the likelihood of abuse and their decisions to report cases to child protective services.[945, 946] In one study,[945] a panel of child abuse experts demonstrated good reliability when scoring the likelihood of abuse on a visual analogue scale from 0-100 (ICC 0.82, 95% CI 0.80–0.84). In the other, a panel of experts in paediatric injury demonstrated nearly perfect inter-rater reliability in their assessments of the likelihood of abuse across a range of different scales and classifications, and almost perfect inter-rater reliability in their decisions to report cases to child protective services.[946] However, it should be noted that the clinicians in both of these studies all had extensive expertise in child abuse, in contrast to those in the current study, where the community paediatricians are likely to have had more experience evaluating children with child abuse than the other participants. One study examining inter-rater reliability of physical abuse determinations in young children with fractures found that experience influenced the level of agreement between subsets of clinicians.[947] It is possible that inter-rater reliability would have differed among subsets of clinicians from different specialties in the current study. The vignettes designed in the current study could be used for training purposes to highlight variability in clinicians' probability estimates of AHT and their proposed CP actions.

Encouragingly, after reading the vignettes and seeing the PredAHT score, many clinicians reported that they felt reassured by the tool, and that it gave them greater confidence in their opinions and decisions. This reinforces the findings from the qualitative study reported in Chapter 5, where many participants suggested that PredAHT would be useful to support their professional opinion and may provide them with more confidence in their decisions if they were to use it in a real case.[747] Interestingly, some clinicians reported that they would find the PredAHT score helpful even if it would not prompt them to change their initial CP action; only 9/29 clinicians changed their CP action in one or more vignettes, however the majority (27/29) declared that they would find it useful in their practice.

The PredAHT score with clinicians' priors was higher than the baseline score for "V4:nAHT*", "V5:ICI-only" and "V6:missing". On average, at Time 1 clinicians overestimated the probability of AHT in these vignettes compared to PredAHT, which could be due in part to the fact that clinicians' likelihood ratios for the evidence, as calculated from their prior and Time 1 probabilities, were larger than those provided by PredAHT, and therefore their prior probability estimates were not reduced enough. There are a number of possible reasons why clinicians' likelihood ratios for the clinical evidence were larger than those calculated by PredAHT. Firstly, it is possible that clinicians found it difficult to estimate the diagnostic value of combinations of multiple clinical features.[580, 618, 619] Secondly, it is possible that clinicians did not explicitly "calculate" their Time 1 probability from their prior probability and the clinical information, but used a different process to make their Time 1 probability estimate. One possibility is that clinicians relied on their previous clinical experiences when estimating their Time 1 probability.[948] Thirdly, when clinicians' prior probabilities were very low, they may have been reluctant to decrease their probability as much as was warranted by their estimated likelihood ratio, avoiding the extreme lower end of the probability scale. Finally, it is possible that the clinicians did not make sufficient use of their own estimates of the prior probability and likelihood ratio when estimating their Time 1 probabilities.[949, 950] In a vignette study exploring clinicians' judgments about the post-test probability of colorectal cancer, Rottman[950] asked clinicians to estimate prior probabilities, likelihood ratios and post-test probabilities, and compared their given post-test probabilities to the normative post-test probabilities calculated from their prior probabilities and likelihood ratios. To analyse whether clinicians' post-test probabilities were sufficiently sensitive to their prior probabilities and likelihood ratios, they used the log-odds form of Bayes' rule,[951] and found that neither clinicians' prior probabilities nor their likelihood ratios were used as much as they should have been.[950] Such an analysis was not possible in the current study as clinicians were not asked to explicitly estimate a likelihood ratio for the clinical evidence.

When compared to the PredAHT score calculated using the baseline prior of 0.34, at Time 1 clinicians overestimated the probability of AHT in "V4:nAHT*", "V5:ICI-only" and "V6:missing" to an even greater extent. It seems reasonable to postulate that if clinicians' prior probabilities were not incorporated in the PredAHT score, PredAHT may have had a greater impact on clinicians' probability estimates of AHT in these vignettes, that is, they may have reduced their probability estimates further at Time 2. Allowing clinicians to incorporate their prior probabilities of AHT enables them to take into account factors that PredAHT does not. Although higher prior probabilities may lead to higher PredAHT scores in some cases, this should prompt further investigation and may help to circumvent the possibility of false reassurance provided by a low score. However, it is important that clinicians' prior probabilities are evidence-based, to minimise the possibility of false accusations of abuse. Findings from the qualitative study reported in Chapter 5[747] suggested that clinicians may be uncomfortable estimating a prior probability due to racial and socioeconomic biases that have been demonstrated in the literature.[201] A recent study examining implicit racial and ethnic biases in the evaluation and reporting of AHT found that minority race/ethnicity children were more frequently evaluated and reported for suspected AHT than white/non-Hispanic children and that these disparities occurred predominantly in lower-risk children ultimately categorized as having suffered non-AHT or with an estimated probability of AHT of 25% or less.[202] In the current study, information on race/ethnicity and socioeconomic status was deliberately excluded from the vignettes so as not to introduce bias, but in reality it is possible that these factors may implicitly influence clinicians' prior probability estimates of AHT and their subsequent CP actions.

Some clinicians were alarmed by the impact their prior probability of AHT had on the PredAHT score. Clinicians' prior probabilities may have been derived from statistical estimates in the scientific literature, or from their personal experiences in assessing the likelihood of AHT in young children, but they varied considerably in all vignettes. Clearly, clinicians would benefit from an explanation or graphical depiction of the effect that different prior probabilities have on the PredAHT score. This could be conveyed using Fagan's nomogram, which illustrates the calculation of the post-test probability given the prior probability and likelihood ratio.[952]

The impact of PredAHT differed between clinicians in different hospital types, specialties, and age groups, and between clinicians with different levels of seniority, training, and experience, however these differences were not statistically significant. A larger study would be required to further examine these observed trends. Given that four of the vignettes were fictional, it was not possible to obtain outcome data regarding AHT and nAHT in each case, thus it was not possible to calculate the sensitivity and specificity of either PredAHT or the clinicians' probability estimates of AHT, or to assess the appropriateness of the clinicians' proposed CP actions. However, the aim of the study was to explore the impact of PredAHT in specific controlled scenarios, and not to determine the accuracy of PredAHT or clinicians in predicting AHT.

The actual impact of PredAHT on clinicians' probability estimates of AHT and subsequent CP actions is likely to differ in clinical practice.[299] It is not yet known whether clinicians will use PredAHT, whether they will use it accurately, or what actions they may take in practice based on specific probability scores, particularly as it does not recommend a direct course of action based on the results. In order to determine whether PredAHT can change clinician behaviour for the better, and to determine its impact on relevant outcomes, a formal impact analysis study is required.

6.5.1 Strengths and limitations

6.5.1.1 Study design

Strengths of this study include the use of mixed methods, which harnesses the strengths and minimises the weaknesses of both quantitative and qualitative approaches.[807] Asking clinicians to provide a rationale for their responses allowed for a more meaningful interpretation of the quantitative data. In addition, since all clinicians completed all six vignettes, it was possible to assess both the individual-level and the group-level impact of PredAHT.[878] Finally, only one study to date has explored the inter-rater reliability of clinicians' perceived likelihood of abuse in possible AHT cases specifically, and this study did not estimate inter-rater reliability statistics.[533]

One possible limitation is that the order of the information presented in the vignettes may not have reflected clinical reality. For example, in practice it is likely that clinicians would have the information regarding apnoea, seizures, and head/neck bruising prior to a child undergoing neuroimaging to look for possible intracranial injury, and they may not gather information regarding the social history until later on in the assessment process. The information was presented as such because clinicians' estimated prior probability of AHT should not be based on the clinical features included in PredAHT but on the other features of a case that PredAHT cannot account for. The qualitative results revealed that in reality, it may be difficult for clinicians to estimate a prior probability of AHT excluding the features included in PredAHT once the presence or absence of these are already known. This reinforces the finding from the qualitative study reported in Chapter 5,[747] that any training on PredAHT would need to incorporate guidance on estimating a prior probability of AHT.

A factorial survey design could have been used to explore a larger number of experimental factors, however such a design is very complex and could have resulted in unrealistic vignettes.[875, 877] The primary aim of this study was to investigate the potential impact of PredAHT in a small number of carefully designed cases that reflected the nuances of clinical practice rather than to explore multiple different factors on clinicians' judgments, decision-making and use of PredAHT.

Finally, one limitation of the study design is that 24 of the 29 clinicians who participated in the study had also previously taken part in the qualitative study reported in Chapter 5. These participants were therefore already familiar with PredAHT, and some of them were also known to two members of the supervisory team. Thus, these participants may have responded more favourably to the PredAHT score when completing the vignettes, compared to clinicians with no prior knowledge or awareness of PredAHT or clinicians with no prior relationship with the wider research group.

6.5.1.2 Validity and reliability

A strength of this study is that the experimental control afforded by vignette studies permits researchers to assess the vignette factors' causal effect on the dependent variable. This enhances internal validity compared to traditional surveys,[875-878] defined by Evans et al.[877] as "the degree to which changes in the dependent variable can be accurately attributed to changes in the independent variable" (p. 163). The "think-aloud" method provided additional evidence of internal validity in this study, because clinicians confirmed that their probability estimates differed as a result of the factors manipulated in the vignettes. Conversely, vignette studies are often criticised due to potential limitations in construct, external and criterion validity.[877, 885] Construct validity is defined by Evans et al.[877] as the degree to which a vignette simulates a real-world scenario, while external validity is defined as the generalisability of the results "to real-world situations encountered by the participants and others like them".[877(p.163)] Criterion validity refers to the extent to which participants' responses in a vignette study predict their responses in actual practice.[953, 954] Researchers and participants alike have voiced concerns about the artificiality of vignettes,[881, 882, 955-958] that is, that the vignettes and the responses to them may not accurately represent what happens in real life, thus calling into question the validity of research utilising this methodology. Each of the forms of validity will be addressed in turn.

Firstly, regarding construct validity, it should be acknowledged that vignettes are not intended to represent reality exactly, but to simulate and manipulate specific features of the topic under consideration,[885] and to promote and encourage reasoning and problem-solving.[880] Put another way, vignettes are selective by nature.[885] The lack of detail and how this is interpreted by participants can generate valuable data in itself.[885, 959] This was demonstrated in the current study, where clinicians' comments about clinical investigations, elements of the history, or differential diagnoses not detailed in the vignettes revealed insights about the factors influencing their judgments and decision-making in suspected AHT cases. Similarly, the lack of information or the form of presentation of information provided in vignettes can reflect the reality of practice; in a study of nursing and social work referrals, Wilson and While[957] found that many real life referrals, similar to their vignettes, contained incomplete or insufficient information. In the current study, construct validity was maximised by using the supervisory team, two of whom have expertise in evaluating children with suspected AHT, to assess the relevance and realism of the vignettes, and by piloting the vignettes prior to their use. In addition, two of the vignettes were based on real cases, and none of the participants remarked that they would be unlikely to encounter the situations described in the vignettes in clinical practice. However as previously discussed, the order in which the information was presented in the vignettes may have restricted construct validity.

Likewise, regarding external validity and the generalisability of the findings, Evans et al.[877] argue that participants' responses to the vignettes are not intended to be taken as representative of their responses in clinical practice but as approximations of their responses. Some researchers argue that the use of realistic scenarios that approximate clinical practice increases experimental realism, thus enhancing external validity compared to traditional experiments.[875, 877, 878] A different view is put forward by Steiner et al.,[876] who propose that the selectivity of vignettes restricts external validity, as the choice of variables to be manipulated and the selection of participants does not warrant inferences to other scenarios or samples. In the current study, only six vignettes were used, yet there are numerous scenarios involving different potential combinations of clinical, historical and social features in which PredAHT could be applied. Most participants were consultants, and half were community paediatricians with considerable CP experience; results may have been different amongst trainee doctors or other specialties involved in the assessment of suspected AHT e.g. neurologists, ophthalmologists or neurosurgeons. To maximise external validity, participants were randomly sampled from a larger pool of potential participants, which extends external validity at least to the target population of clinicians involved in suspected AHT cases.[876] In addition, when computing the inter-rater reliability of clinicians' CP actions, jackknifing was used to estimate the variance due to the sampling of clinicians, in order to ensure the results were generalisable to a larger population of clinicians.

Concerning criterion validity, there is some evidence to suggest that clinicians' responses to vignettes are predictive of their responses to real life scenarios.[960-962] In addition, research comparing vignettes with other research methodologies, including standardised patients, the accepted gold standard for assessing quality of care, suggests that vignettes are a valid tool for measuring the quality of clinical practice, and that clinicians responses accurately reflect and capture actual clinical practice.[953, 963, 964] In one study, clinicians' responses to vignettes resembled their responses to actual clinical practice even in complex medical situations.[965] However, Mohan et al.[954] point out that the majority of studies comparing clinicians' responses to vignettes with their responses to actual clinical encounters explore correlations at the group level only. In an attempt to overcome this limitation, the authors used the Spearman correlation coefficient to analyse the individual-level correlation between trauma triage decisions for vignette patients and trauma triage decisions for real patients, and found no correlation, thus concluding that individual clinician performance on vignettes did not predict their performance in practice.[954] In another study investigating regional variations in consultation referrals, a group of clinicians with the lowest hypothetical referral rates also had the lowest actual referral rates, however there was only a small correlation between individual clinicians hypothetical and actual referral rates.[966] Moreover, Shah et al.[967] found that vignettes overestimated clinical performance compared to standardised patient encounters, which was likely due to a social desirability bias in vignette responses.[953] Indeed, practitioners are motivated to justify their clinical decisions and provide "socially desirable" responses,[968] and they may modify their behaviour due to their awareness of being observed or evaluated.[883] However, Gould[969] argues that compared to observational studies, vignette studies actually help to minimise so-called observer effects because of the "psychological distance" created by a hypothetical scenario.[959] In the current study, the hypothetical nature of the vignettes may have helped to distance or desensitise participants from the sensitive issues being investigated.[884] Consequently however, clinicians' responses may have been decontextualized from the types of responses they may have made in highly pressured or difficult real life situations, where decision-making does not just depend on a rational analysis of the features of a case.[970]

Finally, if a vignette study is to be reliable, it must control for measurement error, experimental error and sampling error.[971] Concerning measurement error, the vignettes were piloted and given to the participants in a random order to control for order effects, both of which increase reliability.[876] Experimental error was minimised by using a within-subjects design rather than a between-subjects design, which can give rise to serious measurement error as responses within individuals cannot be compared and are not contextually grounded.[878, 972] However a power analysis was not undertaken as the study was exploratory. Regarding sampling error, purposive and snowball sampling are appropriate for qualitative, "information rich" data, but can lead to sampling errors with quantitative data as they will tend to underestimate the variability in a population. As previously mentioned, participants were randomly sampled from a larger list of possible participants, however such a sample is not as representative of the population as a probability random sample.[808] Therefore there may be some degree of underestimation of standard errors and overstatement of statistical significance. The p values are not interpreted literally but are treated as a guide for further exploration.

6.5.1.3 Think-aloud method

Critiques of the "think-aloud" method have suggested that thinking aloud affects participants' performance on cognitive tasks, thereby producing inaccurate data, and that participants' verbalisations are not representative of their underlying cognitive processes.[973, 974] However a number of studies have demonstrated that thinking aloud does not interfere with participants' cognition and that "think-aloud" data accurately reflects participants' thought processes.[870, 975, 976] The "think-aloud" data was analysed in a systematic fashion, using a purposeful approach to thematic analysis,[940] adding credibility to the study findings.[880] Although only one researcher coded the data, the analysis and findings were regularly discussed and debated at research team meetings. The "think-aloud" data was not audio recorded due to cost and time constraints, however as the researcher was not required to ask questions or interact with the participants during the exercise, the participants verbalisations were able to be transcribed in real time. While some researchers advocate the use of a follow-up interview to validate the researchers' interpretations of the "think-aloud" data,[977] this was not possible in the current study due to the busy schedules of the clinicians. This study has demonstrated that the concurrent "think-aloud" technique is a feasible technique not only for the validation of survey or questionnaire items but for exploring clinicians' decision-making using vignettes, and that it may help to provide evidence of internal and construct validity in vignette studies.

6.6 Conclusions

This study has demonstrated that PredAHT had a significant impact on clinicians' AHT probability estimates, showing that clinicians are willing to alter their own probability estimate of AHT when exposed to a validated CPR. However, clinicians' proposed CP actions were only influenced by PredAHT in a minority of cases. Additional research is required to assess the actual impact of PredAHT in clinical practice. Chapter 7 presents findings from a novel multisite feasibility study of the evaluation of PredAHT in clinical practice.

7 Evaluating the impact of the Predicting Abusive Head Trauma clinical prediction tool in clinical practice: A feasibility study

7.1 Chapter overview

This chapter presents a novel feasibility study of the evaluation of PredAHT in clinical practice. A multisite, non-randomised, before—after study within the same clinicians was conducted, with a substantial qualitative element. Eighteen consecutive children less than three years of age admitted to two UK teaching hospitals with ICI were prospectively enrolled in the study. Children's case notes were reviewed to extract clinical data, and data regarding the explanation of the injury and social history. Several clinicians involved in the admission and/or care of each child participated in an interview where they applied PredAHT to the case and discussed the probability of AHT estimated by PredAHT in relation to the case. Cases were followed up to determine the outcome (AHT vs. nAHT). A range of component outcome measures were used to assess the feasibility of evaluating the impact of PredAHT in clinical practice, and the findings were formally assessed using an established analytic framework. Qualitative data were analysed using thematic analysis facilitated by the Framework Method. The study yielded rich qualitative data regarding the impact of PredAHT on clinicians' AHT probability estimates and CP actions in each case. The findings are summarised and discussed in relation to the current literature, and the strengths and limitations of the study are considered. Finally, the implications of the study for research and practice are discussed.

7.2 Introduction

As outlined in Chapter 3, an important aim of a CPR is to demonstrate a positive impact on clinical decision-making and patient outcomes. It cannot be guaranteed or assumed that a well-validated CPR that demonstrates good predictive performance will enhance medical decision-making when used in clinical practice, let alone improve patient outcomes.[316] The impact of CPRs on process and patient outcomes can be quantified in comparative studies such as randomised clinical trials, which require considerable time and resources.[315] In order to maximise the possibility that the use of a CPR will indeed confer positive effects on decision-making and health outcomes, an important first step is to assess the feasibility of conducting an impact study.[300] The MRC framework emphasizes the importance of feasibility and piloting work prior to assessing the effectiveness of an intervention, and specifically recommends that the study and recruitment procedures are tested.[322, 323] Crucially, Reilly and Evans[299] note that "a prediction rule rises to the level of a decision rule only if clinicians use its predictions to help make decisions for patients" (p. 201). Since PredAHT does not recommend specific actions based on its predicted probabilities, it is unclear how the PredAHT probabilities will affect clinicians' decisions. Therefore a feasibility study was conducted to determine whether a full-scale impact study can and should be conducted, to explore how probabilities relate to clinicians' CP actions, and to explore potentially appropriate outcome measures for use in a definitive study.

7.2.1 Aims and objectives

The primary aim of this study was to establish the feasibility of evaluating the impact of PredAHT in clinical practice, to determine whether a full-scale impact analysis of PredAHT is warranted. To address this, the different components of the methods and processes used to conduct the study were assessed. The study objectives were to:

1. Assess the processes of patient identification, patient follow-up and clinician recruitment
2. Explore how different probability predictions relate to clinicians' CP decision-making
3. Test the feasibility of collecting the proposed outcome measures, including optimal time points for data collection
4. Assess the appropriateness of the proposed outcome measures
5. Qualitatively explore clinicians' experiences of using PredAHT in clinical practice

To decide whether a full scale impact analysis study is indicated, the feasibility study findings were assessed against an established analytic framework.[978, 979] The secondary aims were to assess the performance of PredAHT in the study population, at baseline and when incorporating clinicians' prior probabilities of AHT, and to assess the performance of clinicians at predicting AHT, both before and after using PredAHT.

7.3 Methods

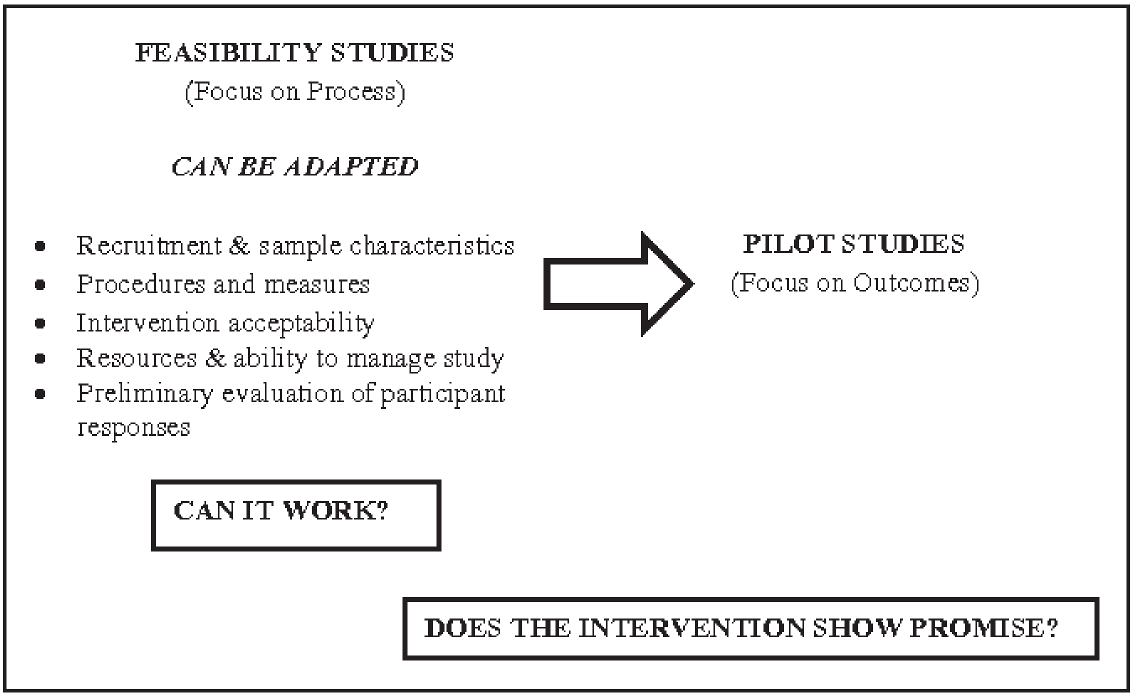
This feasibility study was a multisite, non-randomised, before—after study within the same clinicians, with a substantial qualitative component. The study adopted a mixed methods approach and was conducted within a pragmatic research paradigm,[871] which has been described in Chapter 6, Section 6.3.1.1.

7.3.1 Rationale

Feasibility studies aim to estimate important parameters that are needed to design a full-scale study.[980] They are used to test the different aspects of the study design, methods, and processes used, in preparation for the main study,[322, 980, 981] and to determine whether an intervention is appropriate for further testing.[982] A feasibility study asks whether the future study can be done, should be done, and if so, how.[983, 984] Crucially, the aims and objectives of a feasibility study are not the same as for a definitive study; a feasibility study does not test the efficacy of an intervention, and it need not be randomised or have a primary outcome measure.[980, 981] Feasibility studies can also be considered as distinct from pilot studies, in that feasibility studies focus on the *process* of the development and evaluation of an intervention and a preliminary exploration of participants' responses to the intervention, while pilot studies focus on *outcomes* rather than process, and include a more controlled analysis of participants' responses to the intervention.[985] The distinctive features of a feasibility study are outlined in Figure 7.1.

As described in Chapter 3, the optimal study design for an impact analysis is a cluster randomised trial with centres as clusters.[296] In the index arm, clinicians use the CPR in clinical practice, while in the control arm the CPR is not used i.e. the clinicians are not exposed to its predicted probabilities. The actions of the index group that are guided by the CPR are subsequently compared to the actions of the group providing usual care. The impact on patient outcomes can also be compared between the two groups. However, such a study is expensive to undertake and poses significant practical and logistical challenges, particularly when investigating a condition with a low population prevalence.[295, 296, 305] Therefore, a simpler before—after study design within the same clinicians was chosen for the feasibility study,[295] whereby clinicians indicated their perceived probability estimate of AHT and CP decision for the same patient both before and after seeing the PredAHT predicted probability of AHT. While the possible inclusion of a control group and randomisation process was considered, to the researcher's knowledge there have been no completed impact studies of a CPR designed to assist in the identification of AHT, and it is unknown how a CPR in this field may be received or used in clinical practice, therefore it was considered most important to first establish the feasibility of conducting an impact study of PredAHT. Due to the low prevalence of AHT, two study sites were chosen in order to maximise the number of cases. All clinicians recruited to the study took part in a qualitative interview regarding their experiences of applying PredAHT to specific cases in clinical practice. In-depth qualitative data can help to provide a detailed understanding of the feasibility of an intervention, assist in optimising the design and evaluation of an intervention, and enables the examination of mechanisms of impact of an intervention by capturing participants' experiences of, responses to, and interactions with the intervention.[318]

Figure 7.1 Distinctive features of a feasibility study



Reproduced with permission from Orsmond and Cohn, 2015

7.3.2 Setting and study population

All consecutive children less than three years of age admitted to the Children's Hospital at the University Hospital of Wales (UHW) or the Bristol Royal Hospital for Children (BRHC) with ICI confirmed on neuroimaging, between 31st March and 31st August 2016, were prospectively enrolled to the study. Both sites are large teaching children's hospitals in the UK. The BRHC became the paediatric major trauma centre for the South West of England in May 2014.

7.3.2.1 Inclusion and exclusion criteria

ICI was defined as the presence of any combination of: extra-axial haemorrhage, diffuse or focal parenchymal injury, cerebral oedema, cerebral contusion, hypoxic ischaemic injury or diffuse axonal injury, as per the derivation and validation studies.[59, 60] Children with intracranial findings secondary to confirmed birth injury, a history of a motor vehicle collision, or an underlying structural abnormality or pre-existing disease (hydrocephalus, cystic lesion or tumour, malformation, abnormal brain development) were excluded. In these clinical scenarios AHT is unlikely to be considered as a differential diagnosis, and therefore CP interventions would not be indicated. Children with gunshot wounds, stab wounds and penetrating trauma were excluded as per the Centers for Disease Control definition of AHT.[58]

7.3.3 Site recruitment and set-up

The BRHC was recruited as a study site following a GW4 Alliance two day collaboration event that took place in February 2015 with the objective of setting up a research community around the topic of child head injuries. During this event, the researcher gave a presentation on the development and validation of PredAHT, and the proposed feasibility study, to a group of invited international researchers and clinicians. Dr Giles Haythornthwaite, an emergency medicine consultant, expressed an interest in undertaking the feasibility study at the BRHC and agreed to act as Principal Investigator at this site. Two paediatric major trauma nurse coordinators at BRHC, Mrs Aimee White and Ms Jenni Fryer, agreed to assist in identifying eligible children and clinicians for inclusion in the study. Dr Malcolm Gajraj, a consultant in paediatric intensive care, agreed to act as a clinical supervisor at the UHW.

7.3.3.1 Case identification

A number of different methods were used at each study site to ensure that all eligible children were identified and included in the study.

7.3.3.1.1 University Hospital of Wales

Clinicians working on inpatient wards and in the PICU were asked to notify the researcher by email or telephone of all children eligible for inclusion, as soon as possible following their admission to hospital. The researcher attended weekly safeguarding and radiology peer review meetings, and kept in regular telephone contact with the children's ED, the Children's Admissions Unit, the PICU and the Radiology department. Study posters were placed in prominent areas to remind clinicians to notify the researcher of all eligible children admitted. Despite this, there was a concern that some cases may be missed, as not all clinicians were aware that the study was taking place. These methods were therefore supplemented with regular searches of the hospital radiology database (IMPAX) for head CT scans in children less than three years of age, and eligible children included. The database was retrospectively searched from the beginning of the study period, to ensure that no cases had been missed.

7.3.3.1.2 Bristol Royal Hospital for Children

Eligible cases were identified at BRHC by the two paediatric major trauma nurse coordinators, using a variety of methods. Patients were primarily identified from the ED electronic clerking in system, and hospital databases containing live data on current inpatients were checked. The nurse coordinators received notifications from ward staff via a “tracking email”, and from the clinical site team via a “bed request” email from the ED. Verbal and email referrals were received from the clinical site team, the hospital safeguarding team, advanced nurse practitioners and the neurosurgical team. Cases were tracked and discussed at the weekly multidisciplinary paediatric major trauma team meeting, which the researcher attended whenever possible. Patients admitted on a weekend or out-of-hours were tracked the following working day.

7.3.3.2 Sampling and recruitment of clinicians

The researcher approached several clinicians who were involved in the admission and/or care of each child included in the study, in order to obtain a range of opinions on the likelihood of AHT and the utility and impact of PredAHT. The researcher targeted clinicians from a variety of specialities and with different levels of CP experience and seniority. This included nurses, general paediatricians, paediatric intensivists, emergency medicine paediatricians, community paediatricians, radiologists, neurologists, neurosurgeons, and ophthalmologists. Eligible clinicians were identified from discussions with ward staff, and from entries in the child’s case notes. Clinicians were approached on the ward, or contacted via email, and asked if they could take part in an interview about PredAHT, with regard to the child they had recently cared for. Individuals were excluded if they had not been involved in the child’s admission or care.

7.3.3.3 Study promotion

The researcher, supported by the clinical supervisors and wider study team, visited PICU and paediatric inpatient wards at both sites prior to study commencement, to promote the study to clinicians. The researcher delivered six presentations about PredAHT and the study at PICU, radiology, safeguarding, and paediatric multidisciplinary major trauma meetings at both sites, to ensure that as many clinicians as possible were aware that the study was taking place, and that they may be approached to take part and assist in the identification of eligible patients.

7.3.4 Intervention

The intervention was the use of PredAHT to guide the clinicians' estimate of the probability of AHT and their subsequent CP actions. Clinicians were asked to apply PredAHT to the child that they had cared for. PredAHT provided an LR and percentage probability estimate of AHT to the clinician but no other recommendations regarding their next CP actions. There were no other differences in patient care. All clinicians then took part in an interview where they discussed PredAHT in relation to the case.

7.3.5 Outcome measures

7.3.5.1 Primary outcome

The overall primary outcome of the feasibility study was whether a full-scale impact analysis of PredAHT is warranted and feasible. This was formally assessed using an established analytic framework based on the work of Bugge et al.[978] and Shanyinde and colleagues.[979] A range of component outcome measures were used to assess the feasibility of evaluating the impact of PredAHT in clinical practice, in line with the study objectives (Table 7.1).

Table 7.1 Outcome measures used in the feasibility study, aligned in accordance with the study objectives

Feasibility objective	Outcome measure	Assessment
1. Assess the processes of patient identification, patient follow-up and clinician recruitment	Recruitment rate of children < 3 years old admitted to hospital with ICI identified on neuroimaging	Number of cases identified and enrolled in the study
	Method of recruitment of children < 3 years old admitted to hospital with ICI identified on neuroimaging	Number of cases eligible for inclusion over the study period
	Method of collecting follow-up patient data (AHT vs. nAHT)	Availability and quality of outcome data collected at different stages in the CP process
	Recruitment rate of clinicians involved in the care/admission of the children	Number of clinicians recruited for each case Number of clinicians who declined to participate
	Method of recruitment of clinicians involved in the care/admission of the children	Number of clinicians recruited on the wards

		Number of clinicians recruited by email
2. Explore how different probability predictions relate to clinicians' CP decision-making	Clinicians' probability thresholds for CP investigations and social services referral	Qualitative interviews
3. Test the feasibility of collecting the proposed outcome measures, including optimal time points for data collection	Time taken to identify and enrol eligible patients, extract clinical data, and recruit and interview clinicians	Mean and standard deviation of time taken for case identification from the time of hospital admission Mean and standard deviation of time taken to extract clinical data from the time of case identification Mean and standard deviation of time taken to conduct the first clinician interview from the time of clinical data extraction
	Determine the optimal time-point to interview clinicians about PredAHT	Qualitative interviews
	Test the feasibility of obtaining clinicians' predicted probabilities of AHT	Qualitative interviews
	Completeness of data including: Patient clinical data Clinician data	Availability of clinical variables required for PredAHT in children's case notes Amount of missing data on clinicians' predicted probabilities of AHT or CP actions
4. Assess the appropriateness of the proposed outcome measures	Assess the actions taken by clinicians, and the impact of PredAHT on clinicians' own probability estimates of AHT and subsequent CP actions in each case	Qualitative interviews
5. Qualitatively explore clinicians' experiences of using PredAHT in clinical practice	Explore the factors influencing clinicians' AHT probability estimates and CP actions in each case Explore the reasons why PredAHT did or did not influence clinicians' AHT probability estimates or CP actions in each case	Qualitative interviews

7.3.5.2 Secondary outcomes

The secondary outcomes assessed were the performance of PredAHT in the study population, at baseline and when incorporating clinicians' prior probabilities of AHT, and the performance of clinicians in predicting AHT, both before and after using PredAHT.

7.3.6 Data collection procedure

As soon as each eligible child was identified, the child's case notes were located and reviewed. Each child was assigned a unique study ID number. A list of study ID numbers was stored against the children's NHS numbers, separately to other research data and in a locked filing cabinet, for follow-up purposes. The researcher invited the clinicians involved in the admission or care of each child to participate in an interview about PredAHT with regards to the case. All children were followed up to determine the CP outcome for each case (AHT vs. nAHT). The full data collection procedure is outlined in Figure 7.2. Each of the main stages are discussed in detail below.

7.3.6.1 Case note review

A data collection form was completed for each case (Appendix 33). Data were collected on the six clinical features included in PredAHT (head/neck bruising, seizures, apnoea, rib fractures, long-bone fractures, retinal haemorrhages), the child's gender and age, details of the presenting history and clinical investigations, the explanation for the injury given by the care-giver, the social history, the presence of skull fractures, spinal injury, abdominal injury, additional bruising and additional retinal features, and specific details of the ICI and RH. Anonymised data were entered onto a password-protected database.

7.3.6.2 Clinician interviews

All participating clinicians were given the participant information sheet to read and keep (Appendix 34). Informed consent was obtained, including permission for audio recording (Appendix 35). Participant demographic data were collected in order to describe the characteristics of the sample (Appendix 36). Each interview began by asking the clinician to broadly describe what they knew about the case under discussion, and how they were involved. The researcher then informed the clinician about the six clinical features included in PredAHT, and whether each was present, absent, or missing in the child.

As in the vignette study reported in Chapter 6,[748] clinicians first estimated their "prior" probability of AHT, based on factors pertinent to the case (other than the six clinical features included in PredAHT), e.g. purported history, clinical presentation or psychosocial features. They then estimated their Time 1 probability of AHT and indicated their Time 1 proposed CP action, based on the overall features of the case. If the clinician had already initiated CP procedures prior to the interview, they were asked to record what they had done. If CP procedures were initiated by someone else, they were asked whether they agreed with the actions taken, and if not what they would have done differently. As before, CP actions were aligned with three categories of concern as per National Institute for Health and Care Excellence child maltreatment guidelines[586]; clinicians could choose multiple options.

The PredAHT probability score was then calculated using the clinicians' prior probability, and the child's clinical details. After seeing the PredAHT probability score, participants estimated their Time 2 probability of AHT and indicated their Time 2 proposed CP action. Clinicians were asked to provide reasons for their probability estimates and CP decisions, and to discuss at length why PredAHT did or did not influence their AHT probability estimate or next CP action. If CP investigations were already completed at the time of the interview, the clinician was asked whether PredAHT would have influenced their CP decisions in retrospect. All clinicians used Version 2 of the computerised PredAHT tool, as described in Chapter 4.

7.3.6.3 Case follow-up

The criteria used to define AHT and nAHT are described in Box 7.1. In the absence of a gold-standard diagnostic reference test for AHT, it is important to ensure that circularity is minimised, i.e. that decisions regarding abuse are not based solely on the child's clinical features. In line with the systematic review,[24] derivation study,[59] and validation study[60] described in Chapter 4, injuries were defined as abusive or non-abusive based on the decision made following a multidisciplinary assessment by clinicians, social workers, the police and other relevant agencies at a strategy meeting, case conference, or child death case review meeting. If the child had a multidisciplinary assessment the clinical supervisors liaised with an identified contact from the hospital-based safeguarding team to determine the decision. Children in whom nAHT was diagnosed by the medical team or the hospital-based safeguarding team/community paediatrician, were followed up six months after the study period to determine if they had re-presented to hospital with CP concerns following their initial admission. This was done in order to further confirm the absence of any safeguarding concerns and enhance confidence in the security of the diagnosis of nAHT.

Box 7.1 Criteria used to define AHT and nAHT

Criteria for AHT: Strategy meeting, case conference, child death case review meeting, family proceedings, or criminal proceedings concluded a high probability of AHT Perpetrator admitted AHT AHT was witnessed
Criteria for nAHT: Strategy meeting, case conference, child death case review meeting, family proceedings, or criminal proceedings concluded probable nAHT nAHT was independently witnessed Clinical/hospital safeguarding team presumed nAHT AND the child did not return to hospital with any CP concerns in the six months following their initial admission

Figure 7.2 Data collection procedure

7.3.7 Ethical and governance issues

7.3.7.1 Study sponsorship

Cardiff University agreed to act as Sponsor for the study (reference SPON 1471-15), on 28th October 2015 (Appendix 37).

7.3.7.2 Ethical approval

This study received ethical approval from the NHS Health Research Authority, Wales Research Ethics Committee 3 (reference 16/WA/0003) on January 21st 2016 (Appendix 38). An amendment was submitted on 12th July to request to search the hospital radiology database (IMPAX) at the University Hospital of Wales. This was approved on 29th July 2016 (Appendix 39).

7.3.7.3 Confidentiality Advisory Group approval

Confidentiality Advisory Group (CAG) approval was sought to allow the collection of clinical data from the children's case notes without consent from their parents or care-givers. This was sought as to seek individual consent would reduce and bias case ascertainment and jeopardise the study population. In many research studies that explore child maltreatment concerns, it is often the population of interest whose parents do not consent to participate, yet it is important to study this vulnerable group of children in order to improve the identification of maltreatment. CAG approval (reference 16/CAG/0022) was granted on 21st March 2016 (Appendix 40).

7.3.7.4 NHS Research and Development approval

Approval for the study from the Cardiff and Vale University Health Board Research and Development (R&D) Office was granted (reference 15/RPM/6359) on 29th March 2016 (Appendix 41). A letter of access to conduct research through the health board was issued on 24th March 2016 (Appendix 42). Approval from University Hospitals Bristol NHS Foundation Trust was granted (reference CH/2015/5028) on 31st March 2016 (Appendix 43); a letter of access was issued on 31st March 2016 (Appendix 44).

7.3.7.5 Study management

Three study meetings were held at the BRHC, attended by all members of the study team, to discuss study progress and interim results, any issues relating to the study including recruitment and data collection issues, and any adverse events. The same issues were discussed at regular supervision meetings between the researcher and academic supervisors.

7.3.7.6 Data management

Data were stored and managed in accordance with the Data Protection Act 1998, NHS Caldicott Guardian, The Research Governance Framework for Health and Social Care, and NHS Research Ethics Committee and Confidentiality and Advisory Group approvals. Audio recordings, transcripts, completed surveys and clinical data were securely stored on a password-protected, confidential Cardiff University server. All audio recordings, transcripts, completed surveys, clinical data, consent forms and demographic data will be held securely for 15 years, in line with Cardiff University research data policies. After this, all data will be destroyed. Identifiable data, stored separately to the other research data, were destroyed three months after the study ended.

7.3.7.7 Ethical considerations

This study did not require direct patient contact, and did not require clinicians to perform any clinical or child protection-related investigations that, in their own clinical judgment, fell outside the scope of their usual patient care. Nevertheless, the interviews with clinicians explored their opinions about the probability that AHT had occurred in each case, and thus were highly sensitive, particularly as many of the cases under discussion were distressing. Some clinicians became conflicted, defensive, or upset during the interview. Participants were informed that PredAHT is not a diagnostic tool, but an assistive tool to be used as an adjunct to clinical decision-making, and is intended to complement rather than replace clinical judgment. They were told that the recorder could be stopped at any time, were reminded of their right to withdraw from the study at any point during the interview or beyond, and were assured that they did not have to answer any questions they did not want to. If the researcher had any concerns about a case, these were to be brought to the attention of the academic supervisors, who were to discuss these with the responsible clinicians. The ethical principles of Cardiff University were upheld at all times and the study was carried out according to the principles of the Declaration of Helsinki[816] and the Caldicott principles.[986] All members of the study team completed Good Clinical Practice training.

7.3.7.8 Impact on the researcher

The impact of conducting sensitive and emotive research on the researcher themselves is increasingly being recognized in the literature.[987-989] It was important to consider the emotional wellbeing of the researcher during the study, particularly as they were solely responsible for conducting all of the interviews, reviewing the case notes and extracting clinical data on children with possible AHT. At times, the researcher felt isolated, angry, sad, and powerless when reviewing the case notes of children who had been physically abused or who had died of their injuries, effects which have been documented in other researchers studying child maltreatment.[990] Support mechanisms included regular PhD supervisory team meetings and a counselling session at the Cardiff University student support and wellbeing centre.

7.3.7.9 Reflexivity

The researcher adopted a reflexive approach to the collection and analysis of the qualitative data collected in this study. The potential influence of the researcher's background and positionality, and the relationship between the researcher and the participant on the research process and interpretation of the findings have been discussed in Chapter 5, section 5.3.5.4. During the study the researcher struggled to remain objective and neutral, due to the emotional and upsetting nature of some of the cases being discussed. Visiting the wards and interviewing clinicians about real cases underscored the harsh realities of AHT. The researcher quickly became aware that, when considering the diagnosis of AHT and using the PredAHT tool, the clinician's narratives offered different perspectives on each case, highlighting the convoluted nature of decision-making in suspected AHT cases, and the difficulty of coming to conclusions.

7.3.8 Data analysis

Analysis focused on determining whether or not and how a full-scale impact study should proceed. This included identifying key methodological issues involved in moving from a feasibility study to a full-scale study, and establishing potential solutions to these issues.[978, 979]

7.3.8.1 Feasibility outcomes

Data were summarised using descriptive statistics where appropriate. The analytic framework proposed by Bugge et al.[978] was used to summarise the results and identify potential solutions necessary for a full-scale study to be conducted. This framework is based on the work of Shanyinde et al.[979] who identified 14 methodological components that should be evaluated in feasibility studies. Nine of these were relevant to the current study.

7.3.8.1.1 Qualitative analysis

Data were analysed using thematic analysis[799] within a critical realist paradigm,[835] facilitated by the Framework Method[829] and the constant comparative method.[830] These approaches have all been described in Chapter 5, section 5.3.6. Data were analysed according to the feasibility objectives outlined in Table 7.1. Findings related to clinicians' experiences of applying PredAHT in clinical practice (objective 5) were organised by case. This allowed for an in-depth analysis of the impact of PredAHT on clinicians' probability estimates of AHT and CP actions in each case, and facilitated comparisons both within and across cases. Eighteen interviews (one interview per case), were independently analysed by a second researcher with experience in qualitative methods. Discrepancies were resolved by discussion and consensus. All interviews were transcribed and imported into the data analysis software package NVivo, which was used to organise and manage the data and assist with data analysis.[842] Findings from the descriptive and qualitative analyses are presented concurrently where appropriate. Findings related to the impact of PredAHT on clinicians' AHT probability estimates and CP actions are presented by case, in section 7.4.1.5.

7.3.8.2 Secondary outcomes

The sensitivity, specificity, PPV, NPV, positive likelihood ratio and negative likelihood ratio of PredAHT were calculated 1) using the baseline prior probability of AHT and 2) incorporating clinicians' prior probabilities of AHT. The same performance measures for clinicians' probability estimates of AHT were calculated 1) before they used PredAHT (at Time 1) and 2) after they used PredAHT (at Time 2). As multiple clinicians were interviewed about each case, the data were clustered, and therefore logistic mixed-effects models were used to calculate tool/clinician performance measures wherever possible, to adjust for the potential correlation between observations within each patient.[991]

7.3.9 Reporting

The CONSORT extension for randomised pilot and feasibility trials and the CONSORT-EHEALTH guidelines for the reporting of web-based interventions were used to guide the reporting of this study.[983, 992] Although the CONSORT extension for randomised pilot and feasibility trials does not directly apply to non-randomised feasibility studies, many of the principles of reporting remain applicable to non-randomised study designs.[983] The qualitative components of this study are reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines.[844]

7.4 Results

7.4.1 Feasibility outcomes

7.4.1.1 Objective 1: Assess the processes of patient identification, patient follow-up and clinician recruitment

7.4.1.1.1 Patient identification

Eighteen consecutive children less than three years of age with ICI confirmed on neuroimaging were prospectively enrolled in the study. Twelve (67%) were admitted to BRHC, ten (56%) were aged less than six months, ten (56%) were male, nine (50%) were admitted to the PICU, and two (11%) died (Table 7.2). Eight (44%) children were out-of-area children, seven of whom were transferred to UHW or BRHC from a different regional hospital on the same day or the day following their initial presentation, and had undergone initial investigations (e.g. head CT scan) at the previous hospital. Case follow-up revealed that six children (33%) were deemed to have suffered AHT. The investigations that were performed are detailed in Table 7.3. Of note, 4/6 (66%) children with AHT and 3/12 (25%) children with nAHT had a skeletal survey while all six children with AHT and 4/12 (33%) children with nAHT had an ophthalmology exam. The history and social history are detailed in Table 7.4. In 5/6 (83%) AHT cases, no history of trauma was given, while in contrast a history of trauma was provided for 11/12 nAHT cases. Table 7.5 and Table 7.6 detail the specific history, presentation and clinical findings for each child with AHT and each child with nAHT, respectively. Among AHT cases, seizures, apnoea and RHs were the most common PredAHT predictor variables present, whereas head/neck bruising was more common among nAHT cases (Table 7.7). The presence of PredAHT predictor variables, predicted probabilities of AHT and LRs of AHT for each child with AHT and each child with nAHT are given in Table 7.9 and Table 7.8, respectively.

Regarding the methods used to identify eligible patients with ICI for inclusion in the study, recruitment worked well at BRHC, where cases were tracked by two dedicated paediatric major trauma nurse coordinators as part of their day-to-day role in coordinating the care and rehabilitation of paediatric major trauma patients.[993] However at UHW, when the researcher visited the PICU to review the case notes of an included child, another eligible child was present on the ward, having been admitted the previous day, and thus it became apparent that clinicians were failing to notify the researcher of eligible cases in a timely manner. Thus, 3/8 cases were subsequently identified via regular searches of IMPAX, the hospital radiology database. Nevertheless, a retrospective search of IMPAX found that all eligible cases had been identified and included in the study. Therefore, the researcher was confident that all eligible cases across the two sites had been included.

Table 7.2 Demographics and epidemiology

	Total n=18	**AHT** n=6	**nAHT** n=12
Site, n (%)			
UHW, Cardiff	8 (44.4)	4 (66.7)	4 (33.3)
BRHC, Bristol	10 (55.6)	2 (33.3)	8 (66.7)
Age (in months), n (%)			
< 6	10 (55.6)	5 (83.3)	5 (41.7)
6 – <12	4 (22.2)	0 (0)	4 (33.3)
12 – <24	3 (16.7)	1 (16.7)	2 (16.7)
24 – <36	1 (5.6)	0 (0)	1 (8.3)
Gender, n (%)			
Male	10 (55.6)	3 (50)	7 (58.3)
Female	8 (44.4)	3 (50)	5 (41.7)
PICU admission, n (%)	9 (50)	6 (100)	3 (25)
Neurosurgery, n (%)	1 (5.6)	0 (0)	1 (8.3)
Intubation, n (%)	8 (44.4)	5 (83.3)	3 (25)
ciTBI, n (%)	15 (83.3)	6 (100)	9 (75)
Mortality, n (%)	2 (11.1)	2 (33.3)	0 (0)
Hospitalization ≥2 nights, n (%)	14 (77.8)	5 (83.3)	9 (75)

Table 7.3 Investigations

Investigations	**Total** n=18	**AHT** n=6	**nAHT** n=12
CT, n (%)			
CT head	18 (100)	6 (100)	12 (100)
CT spine	2 (11.1)	1 (16.7)	1 (8.3)
CT thorax	3 (16.7)	1 (16.7)	2 (16.7)
CT abdomen	3 (16.7)	1 (16.7)	2 (16.7)
CT other	3 (16.7)	1 (16.7)	2 (16.7)
MRI, n (%)			
MRI head	7 (38.9)	5 (83.3)	2 (16.7)
MRI spine	6 (33.3)	4 (66.7)	2 (16.7)
X-rays, n (%)			
Chest X-ray	8 (44.4)	5 (83.3)	3 (25)
Follow-up chest X-ray	5 (27.8)	4 (66.7)	1 (8.3)
Additional X-rays	5 (27.8)	1 (16.7)	4 (33.3)
Skeletal survey	7 (38.9)	4* (66.7)	3 (25)
Follow-up skeletal survey	1 (5.6)	0 (0)	1 (8.3)
Radionuclide bone scan, n (%)	1 (5.6)	1 (16.7)	0 (0)
Ophthalmology, n (%)	10 (55.6)	6 (100)	4 (33.3)

*Two skeletal surveys were performed post-mortem

Table 7.4 History and social history

History and social history	Total n=18	AHT n=6	nAHT n=12
History of trauma given, n (%)			
Yes	12 (66.7)	1 (16.7)	11 (91.7)
No	6 (33.3)	5 (83.3)	1 (8.3)
History consistent between caregivers, n (%)			
Yes	2 (11.1)	0 (0)	2 (16.7)
No	2 (11.1)	1 (16.7)	1 (8.3)
Not recorded	5 (27.8)	1 (16.7)	4 (33.3)
Not applicable	9 (50)	4 (66.7)	5 (41.7)
History consistent over time, n (%)			
Yes	9 (50)	2 (33.3)	7 (58.3)
No	4 (22.2)	2 (33.3)	2 (16.7)
Not recorded	5 (27.8)	2 (33.3)	3 (25)
Injuries consistent with mechanism, n (%)			
Yes	10 (55.6)	0 (0)	10 (83.3)
No	7 (38.9)	6 (100)	1 (8.3)
Not recorded	1 (5.5)	0 (0)	1 (8.3)
Independently witnessed injury, n (%)			
Yes	0 (0)	0 (0)	0 (0)
No	18 (100)	6 (100)	12 (100)
Admitted inflicted injury, n (%)			
Yes	2 (11.1)	1 (16.7)	1 (8.3)*
No	16 (88.9)	5 (83.3)	11 (91.7)
Concerning social history, n (%)			
Yes	6 (33.3)	2 (33.3)	4 (33.3)
No	11 (61.1)	4 (66.7)	7 (58.3)
Not recorded	1 (5.6)	0 (0)	1 (8.3)
Previous involvement with social services, n (%)			
Yes	5 (27.8)	3 (50)	2 (16.7)
No	12 (66.7)	3 (50)	9 (75)
Not recorded	1 (5.5)	0 (0)	1 (8.3)

*Inflicted injury by sibling

Table 7.5 Details of the six children with intracranial injury due to abusive head trauma identified during the study period

Case	Age (months)	History	Presentation	Intracranial Injury	Fractures	Retinal Findings	Other Findings
2	<6	Baby slumped forward, went floppy and stopped breathing, father reported shaking the baby after it had stopped breathing	Cardiac arrest, bruising to right side of face, generalised seizures	Multiple bilateral SDH, HIE	None	Multiple bilateral multi-layered RH, small medium and large, in the posterior pole and periphery, swollen optic disc, periocular bruising No subconjunctival haemorrhages	Cervical spine haemorrhage
7	12 – <24	Found collapsed and "blue" at home	Cardiac arrest, bruise to left forehead, convulsive generalised seizures, subsequently died	Multiple bilateral SDH of different ages, HIE, cerebral swelling	No acute # on post-mortem SS Previous distal tibia #, and previous distal femoral metaphyseal # Rib and skull # found following post-mortem examination[994]*	Multiple bilateral multi-layered RH, small and medium, in the posterior pole and periphery, mixed and deep morphology, perimacular folds in left eye, optic disc swelling in both eyes No subconjunctival RH	Bowel perforation No evidence of spinal subdural blood

12	<6	Hit head on father's chin	Bulging tense fontanelle, lower leg bruising, no evidence of scalp swelling or bruising	Multiple bilateral acute SDH, HIE	Possible skull # No rib # on chest X-Ray, SS not done	Right eye: a few small intraretinal RH in the posterior pole Left eye: multiple small intraretinal RH in the posterior pole and periphery Optic disc surrounded by haemorrhages	Apnoea, seizures No evidence of intraspinal subdural blood
13	<6	Unexplained collapse at home when winding after feeding Blood coming out of mouth and nose, not breathing Parents tried to wash blood away with water	PEA cardiac arrest, low blood sodium, abrasions on perineum and buttocks, skin injuries, no external signs of head injury, subsequently died	HIE	Bilateral healing posterior rib # on chest X-Ray No skull # Post-mortem SS results unknown	Multiple bilateral preretinal and intraretinal RH in the posterior pole and periphery, small medium and large Macular retinoschisis possible in right eye Optic disc haemorrhages No perimacular folds No subconjunctival haemorrhages	Tonic seizures
14	<6	Landlady noticed baby increasingly unwell No clear history of trauma from parents	Pale, vomiting, hypothermic, soft fontanelle, severe acidosis and respiratory distress and reduced conscious level, bruises on chest and back and in mouth, no head/neck bruising	Bilateral SAH, DAI, contusion	Healing clavicle # Acute rib # on CT No skull # SS not done	None	Liver laceration, splenic injury No witnessed seizures No intraspinal haematoma

17	<6	Parents left baby with relative. Relative called parents as child was crying, floppy, unresponsive and twitching	Floppy, reduced consciousness, generalised tonic-clonic seizures, tense fontanelle, bruising over eyelids	Multiple bilateral SDH, HIE, cortical swelling, SAH	None	Multiple bilateral multi-layered RH in the posterior pole and periphery, small medium and large, multiple and white centred morphology Macula haemorrhage Subhyaloid haemorrhage Optic disc haemorrhages Perimacular folds No subconjunctival haemorrhages	None

* The rib and skull fractures were not known at the time of the study, the presence of these features were determined from the family court judgment referenced. SDH = subdural haemorrhage, HIE = hypoxic ischaemic encephalopathy, RH = retinal haemorrhage, # = fracture, SS = skeletal survey, PEA = pulseless electrical activity, SAH = subarachnoid haemorrhage, DAI = diffuse axonal injury, CT = computed tomography

Table 7.6 Details of the 12 children with intracranial injury due to non-abusive head trauma identified during the study period

Case	Age (months)	History	Presentation	Intracranial Injury	Fractures	Retinal Findings	Other Findings
1	6 – <12	Crush injury	Vomiting, apnoeic, soft tissue swelling over parietal occipital bones, bruising to front of head	Extensive bilateral SAH, acute shallow SDH, HIE, cerebral swelling	Bilateral comminuted # of occipital bone, transverse # of right temporal bone No rib # on chest x-Ray, SS not done	Ophthalmoscopy not done	Generalised seizures No visceral injury

3	24 – <36	Fall from first floor window onto head	Bruising on the back of the head	EDH	Left occipital skull # No rib # on CT SS not done	Ophthalmoscopy not done	No apnoea, no seizures
4	6 – <12	Fall from sofa onto carpeted floor whilst mobilising No LOC	Boggy swelling, no obvious bruising/lacerations noted	EDH	Linear parietal skull # SS not done	Ophthalmoscopy not done	No apnoea, no seizures
5	<6	Sibling picked baby up out of Moses basket and banged babies head on door	Bruising to the right side of the head	EDH	Linear undisplaced parietal skull vault # No other # on SS	None	No apnoea, no seizures, no spinal blood
6	<6	Breastfeeding, baby unrolled from breast and hit head on metal frame of attached bedside cot No LOC Fell onto carpeted floor after hitting head	Lump/swelling to the head, acute scalp haematoma	Extra-axial blood and remote SDH	Right complex parietal skull # No chest X-Ray or chest CT done, SS not done	None	No respiratory distress, no seizures
8	12 – <24	Playing football in the back garden, fell over backwards onto a gravelled area No LOC, vomiting or drowsiness	Boggy swelling, bruising behind the left ear	EDH	Left parietal skull # No chest X-Ray or chest CT done, SS not done	Ophthalmoscopy not done	No respiratory distress, no seizures
9	<6	Fall from mothers arms onto wooden floor	Petechial bruises to left side of neck, soft anterior fontanelle, swelling above ear, bruise to the side of the head	SDH	Comminuted # of the left parietal bone Linear # of the right parietal bone No chest X-Ray or chest CT done, SS not done	Ophthalmoscopy not done	No apnoea, no seizures

10	12 – <24	Fall from sofa onto carpet, had craniofacial surgery one day prior due to craniofrontonasal dysplaisa	Tonic-clonic seizure, vomiting, no evidence of bruising	Acute high density SDH	No skull #, no chest X-Ray or chest CT done, SS not done	Ophthalmoscopy not done	No respiratory distress
11	<6	No explanation given, high risk child, previous social services involvement	Blood in mouth nose and larynx with respiratory arrest, vomiting	Cortical laceration/shear containing acute blood	None	None	No spinal haematoma, no head/neck bruising
15	6 – <12	Fall down stairs, hit head on radiator	Large bruise to the right side of head	Blood in the underlying subdural and subarachnoid space, haemorrhagic contusion within the right frontal lobe	Depressed # of the right frontal bone extending into the coronal suture, no chest X-Ray or chest CT done, SS not done	Ophthalmoscopy not done	No seizures or apnoea
16	6 – <12	Initially no explanation given, then father reported dropping the baby	Swelling to right side of head	Small SDH	Right parietal skull #, no other # on SS	None	No seizures, no apnoea, no bruising
18	<6	Fall from mothers arms down stairs onto carpeted floor	Scalp haematoma to left side of head	Coritcal haemorrhage	Right linear parietal skull #, no chest X-Ray or chest CT done, SS not done	Ophthalmoscopy not done	No seizures, no apnoea

SAH = subarachnoid haemorrhage, SDH = subdural haemorrhage, HIE = hypoxic ischaemic encephalopathy, # = fracture, SS = skeletal survey, EDH = extradural haemorrhage, CT = computed tomography, LOC = loss of consciousness

Table 7.7 Presence of predictor variables with odds ratios of abusive head trauma for individual variables

	Total (n=18)			AHT (n=6)			nAHT (n=12)			OR for AHT*	95% CI	*p*
	n	%	95% CI	n	%	95% CI	n	%	95% CI			
Head/neck bruising												
Present	11	61.1	(36.1–81.7)	3	50	(14–86.1)	8	66.7	(35.4–88.7)	0.5	(0.07–3.7)	0.5
Absent	7	38.9	(18.3–63.9)	3	50	(14–86.1)	4	33.3	(11.3–64.6)			
Unknown	0	0	(0–21.9)	0	0	(0–48.3)	0	0	(0–30.1)			
Seizures												
Present	8	44.4	(22.4–68.7)	5	83.3	(36.5–99.1)	3	25	(6.7–57.2)	15	(1.21–185.21)	0.03
Absent	10	55.6	(31.4–77.6)	1	16.7	(0.9–63.5)	9	75	(42.8–93.3)			
Unknown	0	0	(0–21.9)	0	0	(0–48.3)	0	0	(0–30.1)			
Apnoea												
Present	7	38.9	(18.3–63.9)	5	83.3	(36.5–99.1)	2	16.7	(2.9–49.1)	25	(1.8–346.71)	0.02
Absent	11	61.1	(36.1–81.7)	1	16.7	(0.9–63.5)	10	83.3	(50.9–97.1)			
Unknown	0	0	(0–21.9)	0	0	(0–48.3)	0	0	(0–30.1)			
Rib fracture												
Present	3	16.7	(4.4–42.3)	3	50	(14–86.1)	0	0	(0–30.1)	6.1	(0.23–162.7)	0.28
Absent	8	44.4	(22.4–68.7)	3	50	(14–86.1)	5	41.7	(16.5–71.4)			
Unknown	7	38.9	(18.3–63.9)	0	0	(0–48.3)	7	58.3	(28.6–83.5)			
Long-bone fracture												
Present	1	5.6	(0.3–29.4)	1	16.7	(0.9–63.5)	0	0	(0–30.1)	4.2	(0.12–152)	0.43
Absent	5	27.8	(10.7–53.6)	2	33.3	(6–75.9)	3	25	(6.7–57.2)			
Unknown	12	66.7	(41.2–85.7)	3	50	(14–86.1)	9	75	(42.8–93.3)			

Retinal haemorrhage												
Present	5	27.8	(10.7–53.6)	5	83.3	(36.5–99.1)	0	0	(0–30.1)	33	(1.06–1023.62)	0.04
Absent	5	27.8	(10.7–53.6)	1	16.7	(0.9–63.5)	4	33.3	(11.3–64.6)			
Unknown	8	44.4	(22.4–68.7)	0	0	(0–48.3)	8	66.7	(35.4–88.7)			

* Odds ratio calculations exclude unknowns. AHT = abusive head trauma, nAHT = non-abusive head trauma, OR = odds ratio, CI = confidence interval

Table 7.8 Presence of predictor variables, predicted probability of abusive head trauma and likelihood ratio of abusive head trauma for each of the children with abusive head trauma

Case ID	Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT Predicted Probability	PredAHT Likelihood Ratio
14	×	✓	?	×	✓	×	94.9%	34.65
17	✓	×	×	✓	×	✓	96.7%	54.99
12	✓	×	?	×	✓	✓	98.2%	101.29
2	✓	×	×	✓	✓	✓	99.5%	379.25
13*	✓	✓	?	×	✓	✓	100%	5964.16
7*	✓	✓	✓	✓	✓	✓	100%	233092.74

*Deceased

Table 7.9 Presence of predictor variables, predicted probability of abusive head trauma and likelihood ratio of abusive head trauma for each child with non-abusive head trauma

Case ID	Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT Predicted Probability	PredAHT Likelihood Ratio
16	×	×	×	×	×	×	3.9%	0.08
4	?	?	?	×	×	×	10.7%	0.22
5	×	×	×	✓	×	×	14.7%	0.32
6	×	?	?	✓	×	×	26.5%	0.67
3	?	×	?	✓	×	×	43.4%	1.43
8	?	?	?	✓	×	×	44.2%	1.48
9	?	?	?	✓	×	×	44.2%	1.48
15	?	?	?	✓	×	×	44.2%	1.48
18	?	?	?	✓	×	×	44.2%	1.48
10	?	?	?	×	×	✓	45.4%	1.55
11	×	×	×	×	✓	✓	58.5%	2.63
1	?	×	?	✓	✓	✓	97.2%	65.63

7.4.1.1.2 Patient follow-up

Obtaining outcome data regarding AHT vs. nAHT proved challenging. If the child had a multidisciplinary assessment, the decision regarding AHT was determined by communicating with the hospital CP team, whose role is to liaise with clinicians, health visitors, social services and the police, arrange and attend strategy meetings, and follow up results of investigations in order to support clinicians in the smooth running of the CP process. Nine children had a multidisplinary assessment involving a strategy meeting or child death review, two were referred to social services where, although a strategy meeting was not conducted, either the police investigated the scene (Case 1), or social services carried out an initial assessment (Case 6), and seven children did not have a multidisciplinary assessment (Table 7.10). For all nine children who had a multidisciplinary assessment, the outcome regarding AHT was successfully obtained from the strategy meeting, Section 47 enquiry findings, or family/criminal proceedings. However in 2/9 of these cases (Cases 2 and 17), the Section 47 enquiry was ongoing at the time of case follow-up; therefore the decision following the strategy meeting was used. In one further case (Case 14) it was unknown if a section 47 enquiry was initiated as the patient was transferred to a different hospital for specialist care. Of the 7/18 children who did not have a multidisciplinary assessment, none returned to the UHW or BRHC with injuries or symptoms prompting CP concerns within six months of the study period. However, one was an out-of-area child and thus could have presented with CP concerns at their local hospital. Ethical approvals were restricted to medical case note review at UHW and BRHC and so precluded six-month follow-up of children presenting from out-of-area.

In addition, categorizing cases as AHT or nAHT proved difficult in two cases. In Case 11 there was a very concerning social history regarding the father, yet the medical evidence was inconclusive and the child was returned home to her mother by the local authority. On balance this case was categorized as nAHT. In Case 12, the Section 47 enquiry concluded that there was a high likelihood of AHT however the family court found that the child's injuries were not "deliberate" but were the result of "inappropriate handling" and the child was returned home subject to a family assistance order. On balance this case was categorized as AHT.

Table 7.10 Flow of children through the child protection system

Case Identifier and History	Hospital safeguarding team contact	Community Paediatrician involvement	Social services referral	Strategy meeting	Section 47 Investigation	Family/Criminal Proceedings	Study Outcome
1: Crush injury	Yes	Yes	Yes	No	No	No	nAHT Police visited the scene and accepted the history as plausible
2: No history of trauma & subsequent admission of inflicted injury	Yes	Yes	Yes	Yes	Yes	Unknown	AHT Police investigation ongoing
3: Fall from window	Yes	No	No	No	No	No	Medical team and hospital safeguarding team concluded nAHT
4: Fall from sofa	No	Yes	No	No	No	No	Medical team and community paediatrician concluded nAHT
5: Inflicted injury by sibling	Yes	Yes	Yes	Yes	No	No	nAHT head injury deemed accidental but written agreement in place regarding parental supervision
6: Hit head on bed frame	Yes	Yes	Yes	No	No	No	Medical team, safeguarding team, community paediatricians and social services concluded nAHT

7: No history of trauma	Yes	Yes	Yes	PRUDIC meeting	N/A	Family and criminal proceedings	Family and criminal proceedings concluded AHT
8: Fall onto gravel	No	No	No	No	No	No	Medical team concluded nAHT
9: Fall from mother’s arms	No	Yes	No	No	No	No	Medical team and community paediatrician concluded nAHT
10: Fall from sofa	No	No	No	No	No	No	Medical team concluded nAHT
11: No history of trauma	Yes	Yes	Yes	Yes	No	No	Probable trauma however medical evidence inconclusive. Local authority returned the baby to mother's care Categorized as nAHT for study purposes
12: Hit head on father’s chin	Yes	Yes	Yes	Yes	Yes	Family proceedings	Section 47 investigation concluded a high probability of AHT. Family proceedings concluded the injury was not “deliberate” but resulted from “inappropriate handling” and returned child home under a family assistance order Categorized as AHT for study purposes
13: No history of trauma	Yes	Yes	Yes	Child death review	N/A	Criminal proceedings	Criminal proceedings concluded AHT

14: No history of trauma	Yes	Yes	Yes	Yes	Unknown	Unknown	Strategy meeting concluded AHT
15: Fall down stairs	No	Yes	No	No	No	No	Presumed nAHT by medical team and community paediatrician. Plan for health visistor to conduct a home safety check.
16: Father dropped baby	Yes	Yes	Yes	Yes	Yes	No	Social care had initial concerns. Police closed case after father admitted dropping the baby. Social care have done some work with father. Some neglect by Dad to not seek medical help after dropping the baby
17: No history of trauma	Yes	Yes	Yes	Yes	Yes	Unknown	Ongoing police investigation. Very high probability of AHT, most likely by a third person. No unsupervised contact, interim care order
18: Fall from mother's arms	No	No	No	No	No	No	Presumed nAHT by medical team

7.4.1.1.3 Clinician recruitment

Of 120 clinicians invited to take part in the study, 87 clinicians (73%) participated (Figure 7.3). The demographics and characteristics of the participating clinicians are presented in Table 7.11. Data regarding the demographics and characteristics of non-responders and non-attendees were not collected. Nine clinicians were interviewed twice about different children, and one clinician participated in two interviews about the same child after more information was collected, resulting in a total of 97 interviews. The number of interviews conducted per case is detailed in Figure 7.4. The researcher conducted all 97 interviews. For 47/97 interviews conducted, clinicians were approached and recruited on the ward and for 50/97 interviews, clinicians were recruited after the researcher identified their involvement in the case from the child's case notes and emailed them to arrange an appointment. In many instances, at the time of the interview, CP investigations had been completed and/or the child had been discharged home, and therefore PredAHT was tested in retrospect (see also section 7.4.1.3, below). Consultants were often not available on the ward; many of the clinicians approached and interviewed on the ward shortly after case note review were nurses. Only one clinician who was approached on the ward declined to participate as they were too busy, suggesting that this method of recruitment is feasible.

Figure 7.3 Flowchart of clinicians participating in a feasibility study of the impact of the Predicting Abusive Head Trauma clinical prediction tool in clinical practice

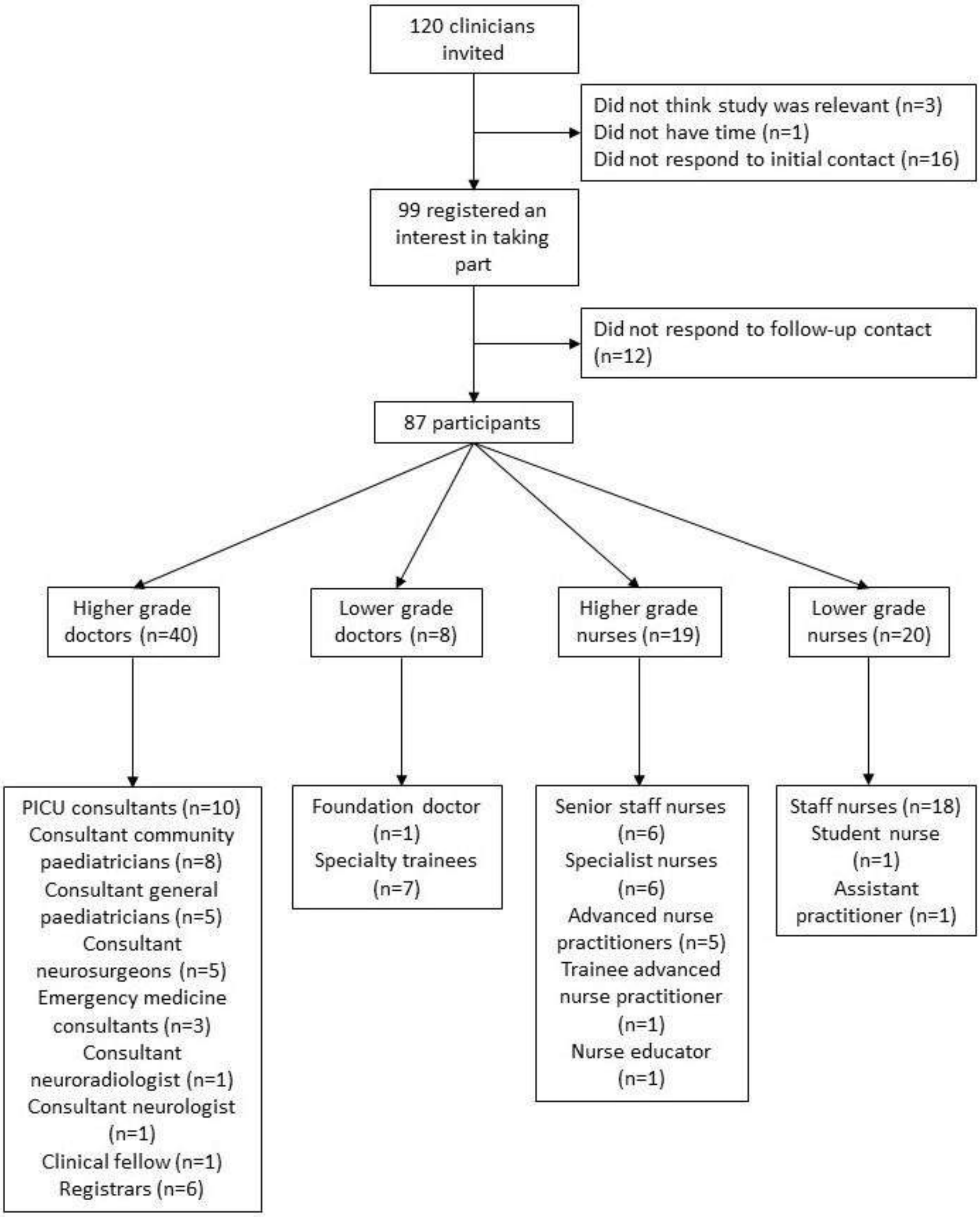
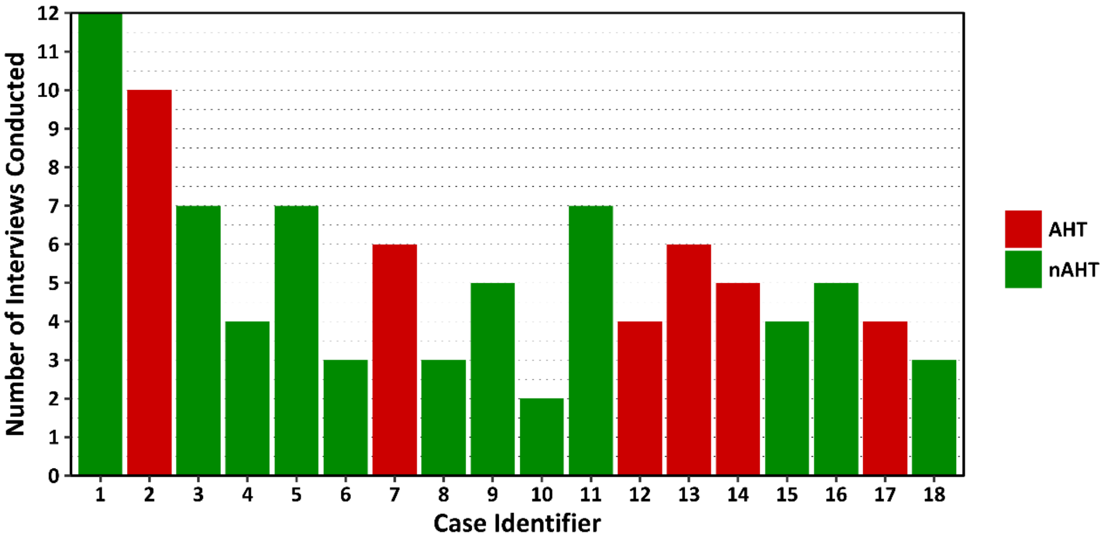


Table 7.11 Demographics and characteristics of clinicians participating in a feasibility study of the impact of the Predicting Abusive Head Trauma clinical prediction tool in clinical practice

Demographics / Characteristics	Doctors (N = 48)		Nurses (N = 39)	
	n	%	n	%
Gender				
Female	24	50	38	97.4
Male	24	50	1	2.6
Age group				
18–24	0	0	6	15.4
25–34	11	23	17	43.6
35–44	16	33.3	7	17.9
45–54	16	33.3	8	20.5
55–64	5	10.4	1	2.6
Ethnicity				
White	37	77.1	39	100
Mixed	3	6.2	0	0
Black or Black British	1	2.1	0	0
Asian or Asian British	7	14.6	0	0
Years since graduation				
<5	4	8.3	8	20.5
5–9	2	4.2	9	23.1
10–20	27	56.2	14	35.9
>20	15	31.3	8	20.5
Site				
Cardiff	26	54.2	15	38.5
Bristol	22	45.8	24	61.5
Specialty				
PICU	15	31.3	15	38.5
Community	9	18.7	2	5.1
General	11	22.9	11	28.2
Emergency	6	12.5	5	12.8
Neuro*	7	14.6	6	15.4

**Neuro includes neurosurgery, neurology, neurosciences, and neuroradiology*

Figure 7.4 Number of interviews conducted per case



7.4.1.2 Objective 2: Explore how different probability predictions relate to clinicians' child protection decision-making

Clinicians reported that their probability thresholds for conducting further investigation were very low, which led some of them to question the relevance and usefulness of a percentage probability in predicting AHT.

> "If my risk was any greater than 5–10% in my head then I am duty bound to do more." ***Clinician 53, Consultant, General paediatrics***

> "Even if there is a 5% chance we have to investigate more anyway." ***Clinician 5, Registrar, PICU***

> "You could argue that if you have anything, if you're suspicious more than nought per cent, you have to investigate. So even if you say there's only a one per cent chance that this is non-accidental injury, you probably have to go with it don't you? So I'm not sure how useful percentages are." ***Clinician 11, Consultant, Neuro***

For this reason, many clinicians said that they were unsure whether PredAHT would change their CP actions in practice.

> "I guess a lot of the ones I see I am just not sure it would change what I do. Because if you had any doubt, if you had any concerns, you would still go

down the whole route. So even if it came out and said the estimated probability of abusive head injury is 20%, that's very low, but still that is 20% you can't ignore, so you would still act you know, so whether it would then change anything…it probably wouldn't do you know what I mean?" ***Clinician 16, Trainee, General paediatrics***

"I'm not entirely sure how necessarily a number would change my practice because it's more a process that we go through for almost every child rather than a 'well if they're between 50 and 90 we do this, and if they're between 30 and 50 you do that', because I think getting to that number is quite difficult." ***Clinician 19, Nurse, higher grade, Neuro***

"Would it change what I did, clinically no, because even if it was a twenty per cent chance, I'd still be going through the investigations." ***Clinician 56, Consultant, PICU***

"If I start out with somewhere around about 5% and it's a likelihood ratio of six, and that bumps it up to the 30% range, fine. That's going to make a difference. I don't think that you're going to see things come out that way. I think that people are going to be in the rough ball park or the numbers are going to be such that even if it's massively multiplied then they're already going to be doing something about it." ***Clinician 28, Consultant, Emergency medicine***

Some clinicians remarked that in some cases they undertake additional safeguarding checks for reasons other than a suspicion of AHT.

"Because of the mechanism of injury I wouldn't have had a high suspicion for abuse anyway, however I wouldn't change all the safeguarding stuff so all the extra checks and things because sometimes it's not abuse, but, not bad parenting, but people just don't think." ***Clinician 24, Nurse, higher grade, Neuro, discussing Case 3***

Clinicians also had a low probability threshold for referring a child to social services. However, one stated that if PredAHT were to provide a specific recommendation to refer a child to social services over a certain probability threshold then they would always follow it.

"Any child who comes in with a significant injury who we don't have a good background on, I think we would follow that through with social services and background checks, even for a child that perhaps you only had a ten percent gut instinct over, because if you have more than zero percent then you have a concern, so therefore it needs to be followed through and excluded." ***Clinician 19, Nurse, higher grade, Neuro***

"What threshold do you need before you refer these…I would have to think carefully about what it would mean for me as an individual decision in terms of thresholds for referral and things like that, but I generally have a low threshold for things like this anyway." ***Clinician 28, Consultant, Emergency medicine***

"I don't know going above 30 per cent maybe, yeah if there is *any* doubt I would just ring [social services] without the scoring, and it wasn't in this case, but as a number you always doubt it, you always question until it's cleared. So I don't know whether percentage scoring would determine…although I suppose if it was down on paper that someone said, you had to do it if it went above this then you would definitely do it wouldn't you." ***Clinician 79, Nurse, lower grade, General paediatrics, discussing case 18***

7.4.1.3 Objective 3: Test the feasibility of collecting the proposed outcome measures, including optimal time points for data collection

The average time taken for the researcher to be notified of a case, extract the clinical data and interview the first clinician across the 18 included cases is displayed in Table 7.12.

Table 7.12 Means and standard deviations of the time taken in hours to identify cases, complete a case note review, and conduct an interview with the first clinician, across the 18 included cases

Summary statistic	Time taken for case identification following child's hospital admission	Time taken to complete case note review following case identification	Time taken to interview first clinician following case note review
Mean	51.25 hours	48.38 hours	12.83 hours
SD	37.47 hours	52.62 hours	33.91 hours

Of the ten children admitted to BRHC, eight were admitted on the weekend or out-of-hours, where cases were tracked the following working day. At UHW, 7/8 children were admitted on the weekend or out-of-hours. The earliest the researcher was notified of a child eligible for inclusion was 13 hours following their admission to hospital with an ICI identified on neuroimaging. At the beginning of the study, it was anticipated that the researcher would be notified of a case immediately following the child's hospitalization, and would be able to extract the clinical data required for PredAHT promptly, and recruit and interview clinicians on the same day, possibly even at multiple subsequent points in the assessment pathway as new information was collected. However, it quickly became apparent that by the time the researcher had identified a patient, extracted the data and approached clinicians to take part, CP procedures had already been initiated, or the child had been discharged. Therefore the majority of interviews were conducted in retrospect. Clinicians suggested that the optimal time-point to assess the impact of PredAHT on their CP actions is as soon as possible following the child's admission.

> "I think it is almost difficult when you've got to our stage of the process because we're almost certain it is non-accidental injury and have gone down the whole route, whereas it would almost make more of a difference if you were to get people at the beginning of it. That would obviously be a much more difficult thing to do but if you could get people at the beginning of the process before they have, after they have literally just seen the CT or something. You almost need to assess this but without hindsight and without knowing what we know now, if that makes sense." ***Clinician 16, Trainee doctor, General paediatrics***

> "As we discuss this case now I have the benefit of a retrospective review on the case and I know a lot of things which at the time I may not have known. So I don't feel that this tool has influenced my thought process or my decision-making. But again with a caveat that by now I already know what happened so I'm biased I'm not blinded to the outcome. I think this tool is much more applicable to the first point of diagnosis, the first point of contact, where the probabilities and the possibilities are being critically analysed and acted upon." ***Clinician 66, Consultant, PICU***

"I think it's different to the people who are the first port of call I guess, you know who saw her first and whereas I was further down the line so everything child protection wise had already happened, I was just following protocol really." ***Clinician 81, Nurse, Lower grade, General paediatrics***

"I can definitely see where it would work. I think our example, it is quite good but actually because everything was just done it's hard isn't it. If you had turned up and no one had suspected it then it is different if then you've got to raise the concern sort of thing." ***Clinician 72, Nurse, Higher grade, General paediatrics***

Regarding data completeness, of 87 clinicians interviewed, two said that they were unable to estimate the probability of AHT, and reported that their role is to simply ensure that CP professionals are involved with the case.

Researcher: So you don't feel comfortable giving me a percentage?

Clinician: Absolutely not

Researcher: Is there any other way you would be able to express it, so do you think it is unlikely?

Clinician: No I would never comment. My practice is by default, if there is an injury, especially an unusual injury, or a severe injury, a severe sort of clinical injury or a situation that is not explained, or uncommon or unusual, my default is to certainly get the social services involved, and the police and the child protection team. ***Clinician 2, Consultant, PICU***

A further two clinicians said that they were unable to provide a *prior* probability estimate of AHT based on the additional factors of the case.

"All the families I see who have got a family history like this one, I'm seeing them because they've got abusive head trauma, so I can't. Yet I know there are many families out there who have got the same problems, but they don't end up bashing their children. So I don't have that information in my head to be honest with you. The likelihood of non-accidental injury, with the history given and the presenting features was high…it was a high percentage of the story, from the very off, not fitting the clinical findings, and that's all I can say." ***Clinician 64, Consultant, PICU, discussing Case 13***

Some clinicians reported that they were uncomfortable estimating percentage probabilities of AHT and reported that they don't usually think in such terms, while others stated that they had no trouble doing so.

"I don't put figures on it. We have to keep an open mind. I just wouldn't do that in clinical practice. You think in terms of risk and safety plans and what is good to do in practice and what the overall conclusion of your report would be but none of that is numeric really." ***Clinician 42, Consultant, Community paediatrics***

"Although I quite like the way it brings up that, I don't use percentages in my report. Because I think it is quite subjective. I don't know, it feels quite subjective. it's not a way I tend to express myself I guess." ***Clinician 32, Registrar, Community paediatrics***

"It's just a different way of presenting what you think isn't it." ***Clinician 27, Nurse, General paediatrics***

Researcher: Did you find it difficult or uncomfortable trying to give probability estimates?

Clinician: No because I've recently been involved in a case, I've looked at the evidence base, and the probability tools are helpful, you know, it supports when you're thinking it is non-accidental injury. ***Clinician 8, Consultant, Community paediatrics***

While clinicians said that they could see the value of incorporating their own prior probabilities of AHT, they suggested that this aspect of PredAHT should be more standardised and evidence-based.

"I think if we are going to do that it'd be good to make it a bit more meaningful rather than say come up with a number. To sort of give some examples of if you are worried about this and this then that might be equal to a certain number. So I think it needs a bit of training to go with it if we are just going to pick numbers out of the air. Because we obviously all have a feeling and a hunch and we have some clinical information and we all compute that in our own way. But you don't think in percentages." ***Clinician 42, Consultant, Community paediatrics***

"We have training and things that we look out for in the history and presentation when we look at children we suspect may have had a non-accidental injury, so delay in presentation and consistent history. All these things are validated and are very useful when you are taking a history and assessing the child to see whether they have had a non-accidental injury. But I am not sure…we probably have in our heads something that we use but I don't think it's necessarily consistent between professionals and it feels a bit unscientific to just pick a number out of thin air." ***Clinician 73, Consultant, General paediatrics***

Clinicians also said that they found it difficult to separate their prior probability estimates of AHT from their overall probability of AHT that involved consideration of the clinical features included in PredAHT.

"If you're not taking any of the clinical factors into consideration, all you've got to go on is the social history and the explanation given for the injuries. But if you're not thinking about the injuries it's hard to think about the explanation that was given, if you see what I mean. I think it's easier to look at the whole thing altogether, rather than just trying to separate it out." ***Clinician 76, Nurse, lower grade, PICU***

"To look at her you can't not take into consideration the fact she is having seizures, like something is obvious. And the bruising so those things were indicative of some sort of head injury and suspicious." ***Clinician 81, Nurse, lower grade, General paediatrics***

"What's gonna happen is that people are gonna put those [clinical features] in anyway, in their prior". ***Clinician 36, Consultant, Emergency medicine***

> "Even in my mind I am cautioning myself because there are co-dependant variables, you know the scoring system is using the same information as I already have." ***Clinician 14, Consultant, PICU***

With regards to patient data, sometimes in the case notes it was unclear whether or not the child had suffered head/neck bruising, apnoea or seizures, as the presence or absence of these features was not always definitively documented, however since the study was prospective this could be clarified with clinicians on the wards. Seven children did not undergo investigations for rib fractures, 12 did not undergo investigations for long-bone fractures, and eight did not have an ophthalmology exam (see Table 7.3 and Table 7.7), emphasizing the value of PredAHT for calculating a probability of AHT when one or more clinical features are unknown.

7.4.1.4 Objective 4: Assess the appropriateness of the proposed outcome measures

As previously detailed, clinicians' probability thresholds for performing further investigations and referring children with ICI to social services are low, implying that these outcome measures may not be relevant for use in a definitive impact study. In all cases where an ophthalmology examination and/or skeletal survey were ordered, the child was referred to social services at the same time, and therefore clinicians reported that if PredAHT was used later on in the CP process once investigations had been completed, it would be unlikely to influence their decisions to refer.

> "By the time you build all this in you might find that people are already at such a level that it wouldn't change what they do and that's really what you're looking for is this kind of change of practice. So by the time you put in retinal haemorrhages people are going to say well this is almost certainly child protection." ***Clinician 28, Consultant, Emergency medicine***

Although clinicians felt that PredAHT would be more useful earlier on in the assessment process, they stressed that it would still not change their clinical practice as their threshold for undertaking further investigations is so low.

> **Clinician:** At the time I had to take the decision to go down the child protection route, I would have had retinal haemorrhages unknown, rib fractures unknown, long-bone fractures unknown, bruising present, seizures absent, and apnoea absent, so I would have had to put unknown in

the first three, because I hadn't had the information. So at that point it might make perhaps a difference, if you don't have the information, it just makes the probability different. It makes it much more likely, because you don't have it, so I was under-estimating it there then. That's interesting to see that, the probability before you have the investigations, and I think you use quite a bit of judgment based on the family history, the circumstances, your own experience, there are so many factors that play in. So I would not have had estimated it as thirty per cent. So that's interesting to see.

Researcher: Do you think that would have been useful then or?

Clinician: It would be interesting to see, but again it wouldn't have changed things, I still would have gone down the route of child protection and done all of the investigations, because you can't take the risk.

In addition, clinicians' highlighted the requirement and importance of a full social services and police investigation, even in many cases where the outcome turns out to be probable nAHT. Therefore, a "correct" referral to social services may not be an appropriate or meaningful outcome measure to assess in a definitive impact study.

"With that history, that sounds plausible that it was inflicted in the way it was, so by being inflicted by a sibling I would say that was therefore accidental, but that still doesn't take away from the fact that there's a huge safeguarding issue there, but the only way we came to that was by going through the process to come to that conclusion." ***Clinician 35, Consultant, Emergency medicine, discussing Case 5***

"What I would have liked to see, if all those [investigations] are negative, is to have a follow-up plan anyway so that the social care would have remained in contact with that family rather than just saying that's the end of it, we've excluded everything. So I think what we've identified both through this tool and for my outstanding risk is that you can't just say okay go home, it's all fine, you've got to have some sort of ongoing monitoring in place bearing in mind that we haven't totally excluded an abusive head injury having taken place. And it is a very young baby and the risks remained there." ***Clinician 42, Consultant, Community paediatrician, discussing Case 5***

Some clinicians felt that PredAHT should not influence clinical practice at all.

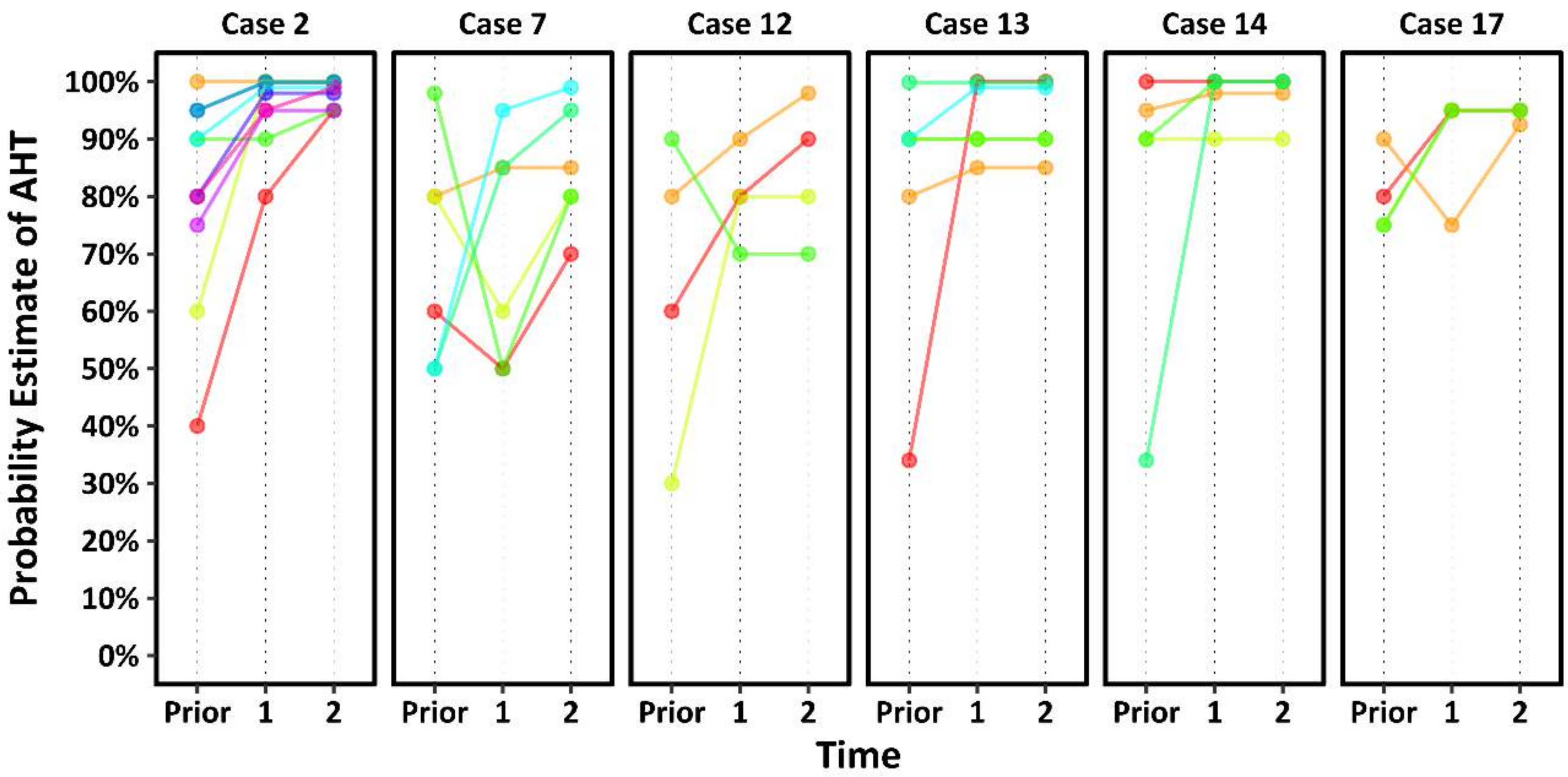
"The only thing worrying about this to me is that what if there is nothing to put in there because it says bruising to the head and neck, what about a bruise on the chest? What if there is no fracture with a bruise on the chest, and then you have put none, none, none, and the probability comes as very low. And someone who is just depending on the tool says, 'oh, that tool gave me a low probability, I will not contact anyone'. What then? So this *shouldn't* change our practice, to be honest. It's just a helping tool."
Clinician 33, Clinical fellow, General paediatrics

Despite clinicians' professed low probability thresholds for performing clinical investigations in children presenting with ICI, only 7/18 (38%) children in this study actually underwent a skeletal survey and only 10/18 (56%) had an ophthalmology review. A detailed analysis of the impact of PredAHT in each of the individual cases, presented below, revealed that PredAHT may prompt clinicians to consider an ophthalmology review, standardise the clinical investigation, encourage discussion and consultation between colleagues, and provide clinicians with confidence and reassurance in their judgments and decision-making, suggesting that these outcomes may be suitable to assess in a future impact study.

7.4.1.5 Objective 5: Qualitatively explore clinicians' experiences of using PredAHT in clinical practice

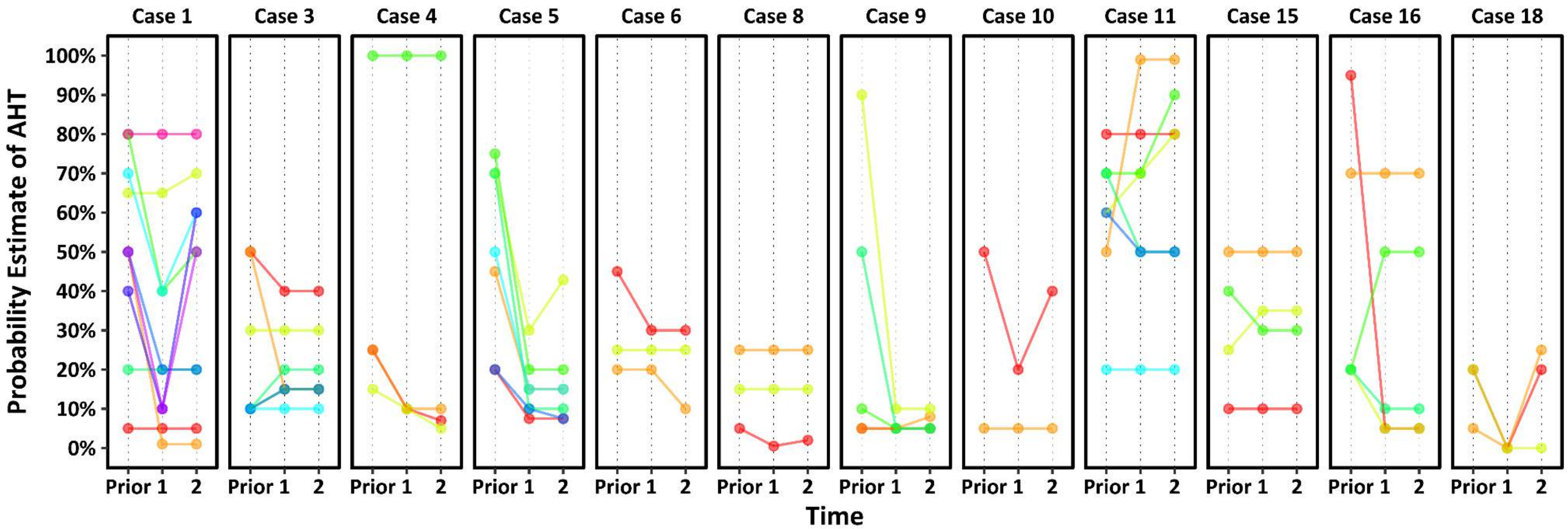
Figure 7.5 and Figure 7.6 show clinicians' prior, Time 1 and Time 2 probability estimates of AHT for each of the AHT cases and for each of the nAHT cases, respectively. The following analysis will focus on the impact of PredAHT on clinicians' probability estimates of AHT and their CP actions, including the reasons why PredAHT did or did not influence their decision-making, and other factors influencing their decision-making, in each of the 18 cases. The six AHT cases will be discussed first, followed by the 12 nAHT cases.

Figure 7.5 Clinicians' prior, Time 1 and Time 2 probability estimates of AHT for each of the abusive head trauma cases



Colours within each case represent different clinicians.

Figure 7.6 Clinicians' prior, Time 1 and Time 2 probability estimates of AHT for each of the non-abusive head trauma cases



Colours within each case represent different clinicians.

7.4.1.5.1 Case 2: Admitted inflicted injury by parent, AHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
✓	×	×	✓	✓	✓	99.5%	379.25

All clinicians interviewed about Case 2 believed that the probability of AHT was very high and that AHT was the most likely diagnosis. Therefore, most clinicians did not alter their probability estimate of AHT, as their Time 1 estimates were already compatible with the PredAHT score (Figure 7.7.) Others were reluctant to estimate higher than 95% even after seeing the PredAHT score, as they could never be 100% certain that AHT had occurred. Clinicians reported that PredAHT did not influence their CP action in this case, due to the number and severity of the clinical features present in the child and their specificity for AHT, alongside a changing history and a subsequent admission of shaking. Despite this, some clinicians stated that PredAHT was useful to support their clinical judgment (Table 7.13). One community paediatrician had already calculated the predicted probability of AHT prior to the interview using the figures provided in the derivation study, and used this to support her opinion throughout the case. One safeguarding nurse thought that PredAHT would be more useful for those with less experience in CP.

Figure 7.7 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 2

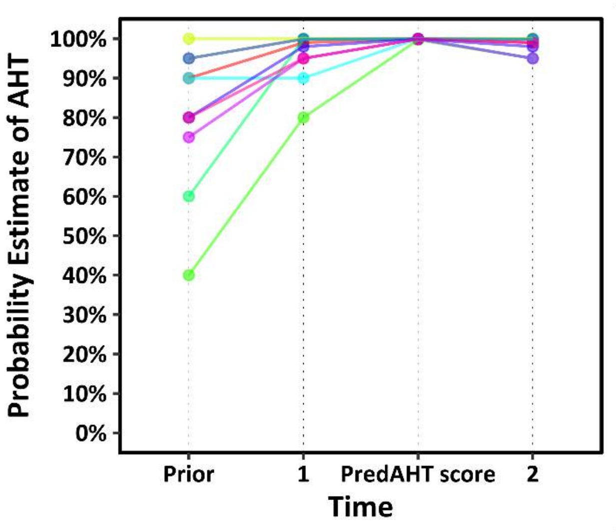


Table 7.13 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 2

Clinician ID	Designation	Specialty	Quote
7	Nurse	Community Paediatrics	"I don't know whether [the PredAHT score] reassured me, and when you look at all the evidence, when a child presents with all those symptoms, it does give a high probability doesn't it of non-accidental head injury. When I was new in role it probably would have been a really useful tool in terms of assisting me to understand the physiology and clinical symptoms, but to determine whether head injury is non-accidental or not, the more experience I have, I think my estimate is more accurate. So moving forward, I don't think it's such a useful tool but definitely, if I had somebody that was new in role I would be using this as a tool for predicting the potential of non-accidental head injury."
10	Doctor	Neuro	"It confirms, it's an additional confirmation for me that it's what I suspect. But you're never 100% sure; you always have that 5% doubt what if I'm wrong. So if you have this tool to say look, you and I think it's 100%, then you kind of get that confirmation thing. I'm thinking 95% when actually at the back of my mind it's probably 100%. But I'm just keeping that 5% in case...so it helps as a support to my clinical judgment to say that I'm right. It helps with that."
12	Doctor	PICU	"To be honest with the history, and knowing what the physical findings were, that would have given me the diagnosis of abusive head trauma. That was my number one diagnosis and really there wasn't much else to look for. Apart from the minor things which we trawl through looking for, metabolic diseases which we never find, but which I know occasionally will be found."
14	Doctor	PICU	"It would be difficult for me to go higher than 95%, unless the child was absolutely covered in bruises. But certainly the fact that a tool came out with such a high score, I have used that as an additional piece of information, not pathognomonic in itself, but another thing that leans me towards this being non-accidental injury."
16	Doctor	PICU	"He has so many features, I guess the more features you have, the more certain you are anyway, and the less this tool probably will change things."

7.4.1.5.2 Case 7: No history of trauma, AHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
✓	✓	✓	✓	✓	✓	100%	233092.74

Clinicians' prior and Time 1 probabilities of AHT varied somewhat in Case 7 (Figure 7.8). Clinicians who were heavily involved with the case reported that they found it extremely difficult and emotionally upsetting, as the child died. Two clinicians who knew the family said that they did not want to believe that AHT had occurred and were conflicted about making a definitive diagnosis of AHT. These clinicians had a high prior probability of AHT, lowered their probability at Time 1 and increased it again after seeing the PredAHT score. Indeed, all clinicians interviewed about Case 7 increased their probability of AHT after seeing the PredAHT score, with the exception of one clinician whose suspicion remained high throughout. One nurse reported that learning of the six clinical features included in PredAHT was useful for her future clinical practice (Table 7.14). The perpertrator was eventually convicted of murder in the criminal court.

Figure 7.8 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 7

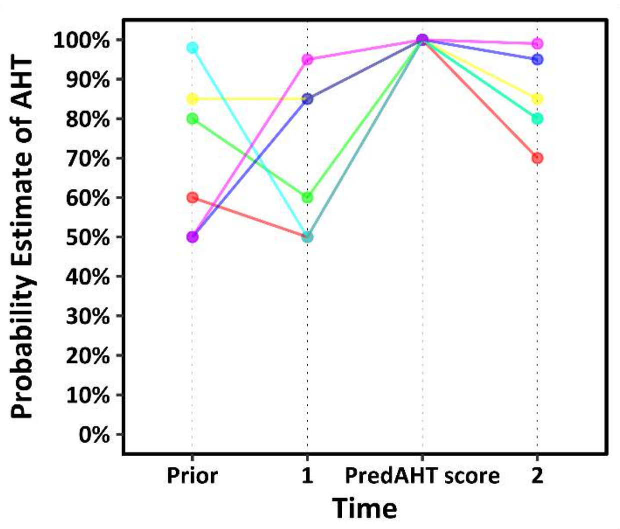


Table 7.14 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 7

Clinician ID	Designation	Specialty	Quote
15	Nurse	PICU	"I had so much involvement with them and because of her background and how she was brought into this family for a better life. Just, you know, I just don't think they could have done it. That's all. I'm very surprised at it and it's a very emotional thing for me because I had her every night. The injury she had is inconsistent with a non-accidental injury, but from what I've noticed with the family and stuff, I was very surprised, I wouldn't say that it was. But obviously the result is pretty much conclusive, saying that it's pretty much conclusive. I don't feel like it could have, but what she's…all the things that she's had done shows it's in line with an abusive case isn't it, so."
34	Nurse	PICU	"I'm going to read up on it and to be more aware of if these are the criteria's here; the retinal haemorrhaging, the bruising, apnoea, and the seizures, then those six criteria would stick in my head for the future."
38	Doctor	Community Paediatrics	"I'm coming round to thinking it is more likely to be an abusive head trauma. I think it is quite difficult when you actually know the parents and know the people, and you get an explanation that may or may not be fairly plausible. I knew this anyway when I was doing it, I knew that tool was going to be really high it was going to be over 98%. And that is helpful, it's reassuring but what this tool doesn't look at is what the actual haemorrhages look like."

7.4.1.5.3 Case 12: Hit head on father's chin, AHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
✓	×	?	×	✓	✓	98.2%	101.29

All clinicians interviewed regarding Case 12 thought that the child's injuries were due to AHT. All estimated the probability of AHT to be high at Time 1, and two increased their probability estimates further after seeing the PredAHT score (Figure 7.9). Interestingly, all four clinicians stated that PredAHT provided them with additional confidence and reassurance that AHT was the most likely diagnosis, with nurses emphasizing that it helped them to remain objective and to put aside any personal biases they had. However, one doctor remarked that PredAHT would be unlikely to impact upon clinicians' CP actions in the PICU as all children with head injuries undergo CP investigations due to the severity of their injuries (Table 7.15).

Figure 7.9 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 12

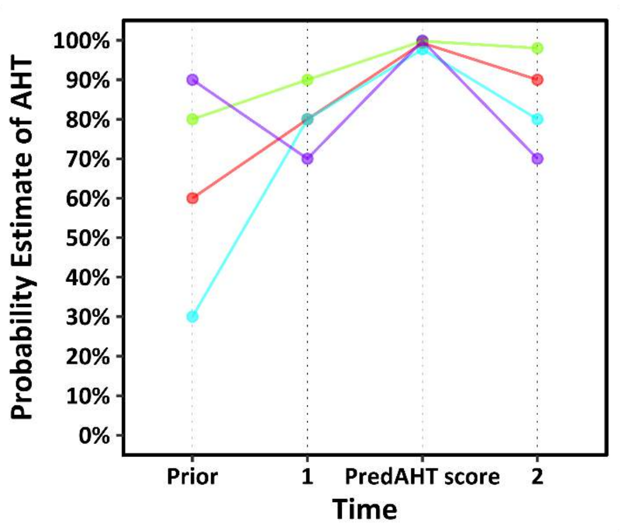


Table 7.15 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 12

Clinician ID	Designation	Specialty	Quote
14	Doctor	PICU	"I like to have as many different sources of information as possible because that helps me complete a jigsaw. So the tool may be potentially useful but it would be unlikely to ever change anything about the management that we do up here, because the management is pretty much the same for any head injury."
61	Doctor	Community Paediatrics	"I think it gives more confidence because I know one of the paediatricians from the first hospital had written a report which had all the facts in it, but concluded in a rather woolly way that abusive head trauma needs to be considered further or something without giving any kind of probability value. So I think they were kind of holding back a little bit and being a little bit soft in giving an opinion, so having a tool like that lends added weight to my opinion that this was a high probability of abuse."
62	Nurse	PICU	"Certainly as a nurse you don't want to think ill of anyone, that somebody is capable of doing that. So it makes it more reflective. It makes you realise well actually, we need to go further with this, we need to investigate it, we need to ensure the siblings are safe."
63	Nurse	PICU	"I think that's actually very useful because we are taught not to go to instincts in these cases, you tend to err on the side of not believing it's a non-accidental injury whereas maybe you should really just completely believe it until you can rule it out. And that makes that quite obvious really."

7.4.1.5.4 Case 13: No history of trauma, AHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
✓	✓	?	×	✓	✓	100%	5964.16

All clinicians interviewed about Case 13 estimated a very high probability of AHT right from the beginning of the case and agreed with the high PredAHT score (Figure 7.10). This child died of their injuries shortly after presentation to the hospital and before any interviews were conducted. Clinicians stated that PredAHT would not have made a difference to their CP actions in this case, as the child's injuries were so severe that it was clear that physical child abuse was the only plausible explanation. The perpetrator was subsequently found to have shaken the child and was convicted of manslaughter in the criminal court. However, clinicians thought that PredAHT may be useful in cases where the diagnosis of AHT is less clear-cut (Table 7.16).

Figure 7.10 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 13

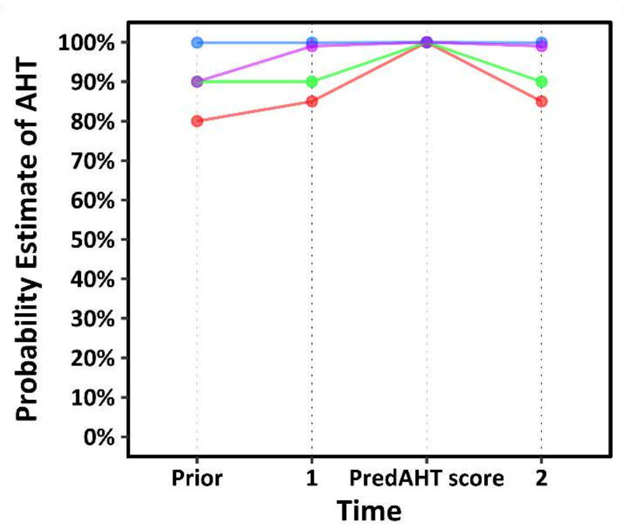


Table 7.16 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 13

Clinician ID	Designation	Specialty	Quote
64	Doctor	PICU	"As a screening tool, would it change my mind, it wouldn't have changed my mind in this case, they were going to get the full NAI work-up, just from the first instant I took the phone call, they were going to scan, social services were going to be involved, etc., etc."
73	Doctor	General paediatrics	"I think in this case there was not much in the way of doubt in terms of the injury, there was nothing else really that would have fitted. As to whether it's useful as a tool I think probably in this situation it wouldn't have made any difference to anything we did to be honest we would have done exactly the same things. I don't think it would have made any difference. But I guess I can see a role for it in the situation…in situations where it's less clear cut so the ones where you are really not sure, that's probably where it would come in."
75	Doctor	PICU	"I think what is the use of this, in I mean a lot of it is about suspicion isn't it and the clinicians just having an index of suspicion. The difficulty is getting that equipoise of not being suspicious of the entire world but not kind of missing a child who has been abused and who survives or there is other siblings and you leave them at risk. So if its purpose is to kind of highlight that and make receiving clinicians suspicious then I think it is potentially useful. If you work in a paediatric environment then that suspicion level is quite high."
86	Doctor	Community paediatrics	"I think in this case it was so clear cut that everything had been triggered before we had investigations. I am sure where this tool is most useful is the grey cases, this case as I said felt probably one of the most clear cut cases I have ever dealt with and just everything was concerning."

7.4.1.5.5 *Case 14: No history of trauma, AHT*

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
×	✓	?	×	✓	×	94.9%	34.65

Again, for Case 14, clinicians estimated a very high probability of AHT throughout (Figure 7.11), due to the severity of the child's injuries, and stated that PredAHT would not have been needed in this case. One nurse stated that it was not her responsibility to make decisions regarding CP, and therefore PredAHT would not have been useful for her. Conversely, the community paediatrician who was preparing a court report for the case at the time of the interview thought that PredAHT would be useful throughout the whole CP process, including when preparing for court. However, this clinician also suggested that the tool be refined to account for more specific patterns of the clinical features (Table 7.17).

Figure 7.11 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 14

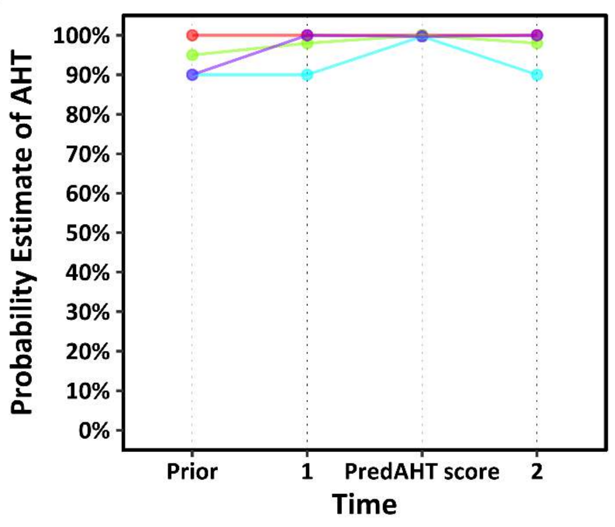


Table 7.17 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 14

Clinician ID	Designation	Specialty	Quote
67	Nurse	PICU	"I mean it is good to obviously check the tool and everything but I think this is one where the tool wouldn't be as useful or because it's such an extreme example of the cases that we do get. There kind of really isn't any other explanation of how she could have got the injuries."
76	Nurse	PICU	"From a nursing point of view I'm not sure whether it would be useful. We don't have decisions to make like that. It's not our responsibility to make a decision as to whether this is accidental or non-accidental. Our responsibility is to treat the child and the family and be an advocate for the child and that doesn't change whether they've been non-accidentally or accidentally injured."
77	Doctor	Community paediatrics	"This is good for somebody working in the A&E. When they are assessing the patient, to get the community paediatric team to see the child. And even when preparing for court. When preparing for court you wait for all your results, you get everything, you go and look at the family history, you think of differentials, you get all the investigations, and at the end, these questions you are asking are the questions I'd want to ask myself, and my conclusions would be based on that." "You can refine your parameters, putting posterior rib fractures, because that has a very high positive predictive value. And then a combination of the retinal haemorrhaging affecting whatever layers, posterior rib fractures, interhemispheric haemorrhage, posterior fossa haemorrhage, thin subdural, and what is the fourth one I always forget, multiple and of different ages, multiple subdurals of different ages. You need to talk about posterior rib fractures, multiple rib fractures, multiple retinal haemorrhages, retinal haemorrhages affecting all of the retina."

7.4.1.5.6 Case 17: No history of trauma, AHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
✓	×	×	✓	×	✓	96.7%	54.99

During interviews with clinicians regarding Case 17, it came to light that CP procedures had already been initiated at the referring hospital where the child first presented. All clinicians agreed with the high PredAHT score and one increased their probability from their Time 1 estimate (Figure 7.12), which they explained was initially lower as they had been wrong about AHT in the past. Clinicians believed that PredAHT would be useful to prompt an ophthalmology examination, but that it would be less useful in the PICU setting due to the severity of the injuries seen, and that it would be unlikely to change practice if used once clinical investigations had been completed (Table 7.18).

Figure 7.12 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 17

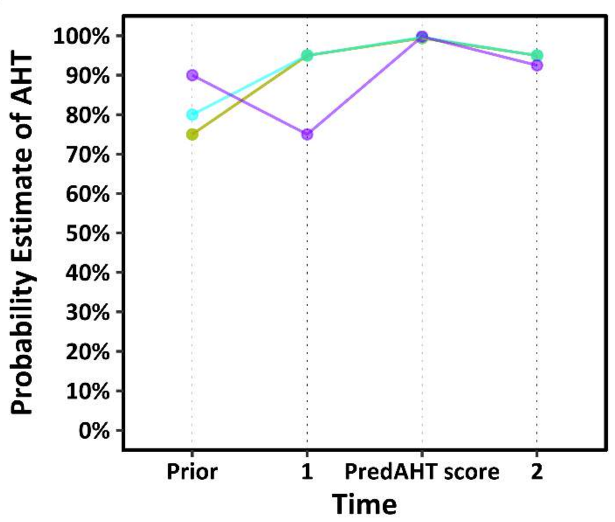


Table 7.18 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 17

Clinician ID	Designation	Specialty	Quote
81	Nurse	General paediatrics	“It probably just confirms what I already think I suppose, there wasn’t really anything else I could have done.”
82	Nurse	PICU	“I don’t really think it changes anything because retinal haemorrhages once they have done the scan is a massive indicator of it anyway so your suspicions are going to be high.” “We see the worst of the worst so generally when they come to us we are already thinking, oh yeah it is, it isn’t, but maybe for clinicians working on the general paediatric population then yeah it probably would be useful.”
83	Nurse	PICU	“It's difficult to know when you'd want us to use it. Because my experience here is when they come in they like to stabilise them so a lot of these things, investigations are left for a few days. So it depends when that was going to be used. I think it would be most useful when you've had more investigations. Could you use this to get our practice to change though? It will encourage people to go and get the eyes looked at. There was big discussions between our consultants about when it should be done.”

7.4.1.5.7 Case 1: Crush injury, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	×	?	✓	✓	✓	97.2%	65.63

The majority (11/12) of clinicians interviewed about Case 1 believed the child's injuries to be accidental, although their prior, Time 1 and Time 2 probabilities of AHT were highly variable (Table 7.19). This was primarily based on a perception that the injuries were consistent with the mechanism described by the care-giver. Other influential factors for clinicians included a lack of additional clinical features concerning for AHT, and a perception that the injuries were the result of neglect rather than physical abuse. Therefore, for most clinicians, the high PredAHT score did not influence their probability estimate of AHT (Figure 7.13), or their CP action. However, after seeing the PredAHT score, one neurosurgeon reported that he was going to discuss the case with a CP colleague, while one clinician reported that he would check whether or not an ophthalmology examination had been performed.

Figure 7.13 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 1

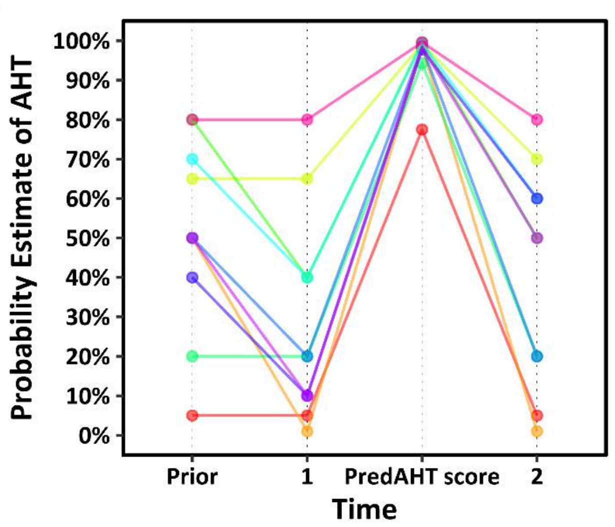


Table 7.19 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 1

Clinician ID	Designation	Specialty	Quote
1	Nurse	PICU	"I don't think anything has changed to be honest, with the story that I had, I think I still need to err on the side of caution but it does sound plausible. You need to look at the story, does it fit the injury."
2	Doctor	PICU	"I did not actually request further investigations. But if there is one investigation I would want, I would want an ophthalmoscopy. I think when I do the ward round this afternoon I will actually check."
3	Doctor	PICU	"If we had any information about the family, any previous incidents or whatever, we could say that our percentage would be much higher, but we don't have any information like this, so we think it is unlikely still."
4	Doctor	PICU	"Abusive means an act of commission so if I am talking about an act of commission it is very low in that way, so it is a negligence, they didn't put a protective mechanism in place, so the parents failed to do that, and it could happen to anyone."
5	Doctor	PICU	"Yes there's a very bad injury on the head but as far as we knew on arrival there's no other signs of injury like big bruises anywhere, so the probability is also based on that fact as well."
8	Doctor	Community Paediatrics	"I would do the same. The tool doesn't take into account the history or the mechanism of injury given. The story tied together nicely in that it sounded plausible and therefore there wasn't immediately any doubt as to the mechanism from the story given. So, the history is quite a significant part of it and plays a greater part in it than those clinical features do."
10	Doctor	Neuro	"So my next plan, my next action will be to discuss with a child protection colleague. When somebody says this is a tool for measuring the probability then you will always get swayed by it a little bit…if somebody comes up who is looking at it at a different view point, then your percentage that is in your head starts going up slowly."

7.4.1.5.8 Case 3: Fall from window, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	×	?	✓	×	×	43.4%	1.43

The seven clinicians interviewed about Case 3 believed the child's injuries to be accidental, and estimated the probability of AHT to be between 10%–40% (Figure 7.14). The PredAHT score was congruent with clinicians' Time 1 probability estimates in 5/7 interviews. Upon seeing the PredAHT score, none of the clinicians changed their probability estimate of AHT, however one nurse reported that she would discuss with the medical team whether an ophthalmology review and skeletal survey was required. Clinicians discussed a range of factors that influenced their decisions, including experience of similar cases, a consistent history, interaction with the family, gut instinct, a lack of additional clinical features suggestive of AHT, and the perception that the injuries were consistent with the mechanism described by the care-giver (Table 7.20). Importantly, although all clinicians agreed that AHT was unlikely, all agreed that further investigation by the hospital safeguarding team was necessary to rule out AHT.

Figure 7.14 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 3

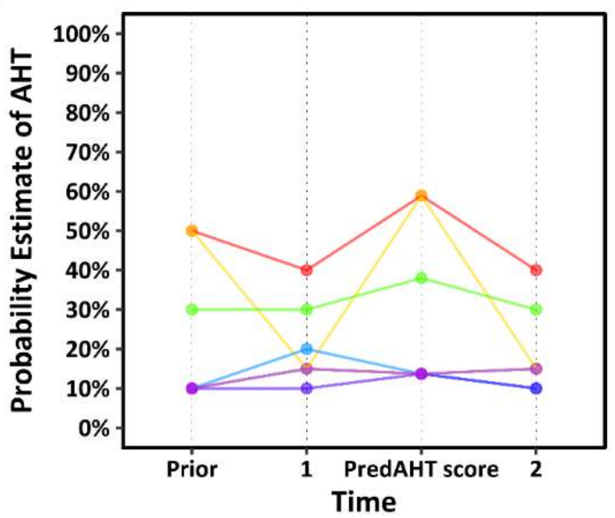


Table 7.20 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 3

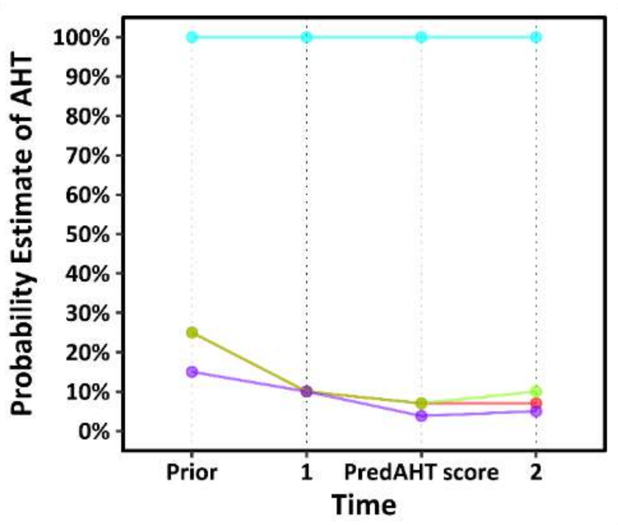
Clinician ID	Designation	Specialty	Quote
18	Nurse	Neuro	"We've seen a few of these injuries before; very similar, falling out of windows and it's never been a case of abuse in this particular injury, it's very specific falling out of a window. Also from observing the family and how they are around the child, those kind of things. From my gut instinct, it doesn't feel like it's something that's been purposely done. With just the facts, it does look suspicious. I think it does look like it could be strange circumstances but I guess as a nurse and working with the family, we may see it in a different light."
19	Nurse	Neuro	"I think in this case we would have had a high suspicion of needing to exclude a safeguarding issue as being the cause of his injury."
26	Nurse	Neuro	"So in this case and largely based on my previous experience of children having had injuries that have fallen from windows, most are caused by an accident and I guess you just have to have an open mind that they could be an abusive head trauma. So I think that would be where I would sit, between 20% and 10% to keep that open mind and to just investigate a bit further."
26	Nurse	General Paediatrics	"I will talk to the team and make sure that those other things happen because I'm quite surprised that they haven't happened to date. I kind of thought with a head injury that was in any way open to suspicion which I guess most are unless there's been a car accident or something very obvious, I would have thought that he would have had a skeletal survey at some point and I would have thought that he'd have an ophthalmology review at some point as well."
27	Nurse	General Paediatrics	"So it was a good story, the story never changed, and there's no other factors, like the rib fractures, or anything to suggest there was previous abuse or anything like that so, the injuries fit the story."

7.4.1.5.9 Case 4: Fall from sofa, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	?	?	×	×	×	10.7%	0.22

Three clinicians involved in Case 4 believed that the probability of AHT was low, in line with the PredAHT score (Figure 7.15). These clinicians' prior probabilities were higher as the initial history taken did not include any detail about exactly how the child had fallen off the sofa, the surface of the floor on which the child had landed, or the child's activity immediately prior to the fall.

Figure 7.15 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 4



"There was a little bit of concern to begin with because if it had been a carpeted floor that is quite a significant injury, and in the initial notes it was a fall from a sofa without any further clarification. So myself and a medical colleague just clarified with mum exactly what happened and it turned out that he hadn't just fallen off a sofa he had been running along the sofa at quite a speed and had hit his head on wooden flooring."
Clinician 22, Nurse, General paediatrics

One neurosurgeon, who at the time of his interview was not aware of the additional detail regarding the history, explained that he assumed the child's injuries were due to AHT until proven otherwise, and would not use PredAHT as further investigations had not been performed, and he did not have all of the information that he would need to come to a decision. All three nurses agreed with the low PredAHT score as they thought that the mechanism of injury fit with the expanded history, however two said that they would have asked their medical colleagues whether an ophthalmology review was required (Table 7.21).

Table 7.21 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 4

Clinician ID	Designation	Specialty	Quote
21	Nurse	PICU	"Looking at it, it is a good indicator and key facts to look at the risk of abuse, and yeah I suppose it could influence my percentage and analysis of something if I was to put the risk factors into a scale and it gave me a percentage yeah. I probably would have looked into the background and made sure I'd read all of the relevant indicators and what had been assessed before I made a prediction, probably."
22	Nurse	General Paediatrics	"I agree with [the PredAHT score] because the mechanism of injury fits with the expanded history, so once we had sat down and talked to mum about exactly how it had happened and the details became finer. I think the only thing looking at the tool is perhaps he should have had an ophthalmology review, or to clarify whether he had an ophthalmology review. Because that would have just put another thing to rest you know, it was a significant injury, it was an extradural and not a subdural but perhaps at least we should have contemplated why an ophthalmology review wasn't requested. If he was still here I would probably query whether we should be thinking about it."
43	Doctor	Neuro	"The problem we find is that the histories aren't very good, which is why we take the default position of just admit them, bung them all under Child Protection scenario, that's my scenario, I'd much rather over-investigate than miss one. We have an incomplete history and therefore I have to assume it is abuse until proven otherwise. Because I can't prove otherwise until I take the history myself or have a sensible history and get the investigations done, so you have to assume it is. I wouldn't trust this tool at all. There's no history and you haven't got half the information. I'd chuck it out. I would not use that tool."
59	Nurse	PICU	"We usually do the retinal haemorrhage ones, but that's for the little, for smaller babies, usually, but it doesn't take long to have an X-ray, so we do usually do skeletal surveys, I mean it wouldn't hurt to, I think everyone should just get a standard, if they've done everything else, why not do a quick X-ray and then you have a look in their eyes, I don't see why that's a problem. They should do it, really. The tool would be good to remind them that they haven't had those done yet."

7.4.1.5.10 Case 5: Inflicted injury by sibling, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
×	×	×	✓	×	×	14.7%	0.32

All clinicians in Case 5 were interviewed once a skeletal survey and ophthalmology examination had been conducted and a referral to social services had been made. Clinicians reported that they were heavily influenced by either the history or the behaviour of the parents in this case. Two clinicians estimated a low prior probability of AHT as they were satisfied with the appropriateness of the parents behaviour and the presentation of the baby (Figure 7.16). However, the majority of clinicians estimated a high prior probability of AHT, as they felt the initial history given did not fit with the purported mechanism of injury. As the case evolved and more information about the history and the mechanism of injury was gathered, clinicians' probabilities of AHT decreased. One clinician was concerned that her "original" high prior probability of AHT was unduly influencing the PredAHT score, raising the question as to whether the prior could be adjusted as new background information comes to light. Exploring the tool, this clinician remarked that it would have been useful to complete before the skeletal survey and ophthalmology had been carried out. One clinician pointed out the value of negative investigation results for helping to bring down the probability of AHT (Table 7.22).

Figure 7.16 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 5

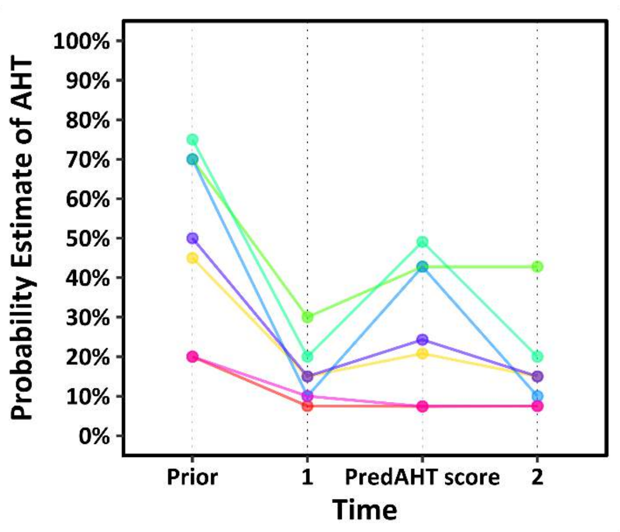


Table 7.22 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 5

Clinician ID	Designation	Specialty	Quote
23	Nurse	Emergency medicine	“I believed her, I believe her, the babies’ behaviour was very normal for such a young baby and mum had done everything she was supposed to, the baby was acting normal. I was very happy with the presentation of the baby.”
25	Nurse	Neuro	“It had been reviewed by a neuro-radiologist who had given their opinion about the, how feasible it was that the injury had been caused by the history that was given. I think a question had been because there were two lumps on the head, two areas of swelling, whether it was possible that the injury could have been caused, with the history, but they went on the scan itself rather than the visible injuries and he felt that it was possible that it was the history given.”
35	Doctor	Emergency medicine	“I think it was all about the history, the history didn’t fit, mum was telling me this story, I was looking at this baby saying I can’t see how a baby could have two separate bruises in separate places, but supposedly from the same incident, I also couldn’t see how, particularly the one that was the more concerning one on the temporoparietal region how that could be so large and the baby just cry for a couple of seconds and then mum say that they were okay, and then when you then coupled that with actually getting the CT back and there’s a big fracture and a bleed. So actually the story now fits, so we’ve got a good reason why that child has now got that injury, there’s no other injuries that would indicate any other abuse or other trauma and that’s why having the negative skeletal survey and the negative ophthalmology is really helpful.”

7.4.1.5.11 Case 6: Hit head on bed frame, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
×	?	?	✓	×	×	26.5%	0.67

The child's injuries in Case 6 were felt to be accidental by all three clinicians interviewed. Clinicians' Time 1 estimated probabilities of AHT were low (Figure 7.17), and they reported that they agreed with the PredAHT score as it was similar to their own predicted probabilities of AHT (Table 7.23). One clinician was reassured by the relatively low PredAHT score (35%) but thought this number was high enough to raise concern and prompt the necessary further investigations. Clinicians stated that they were also reassured by the lack of additional clinical features concerning for AHT.

Figure 7.17 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 6

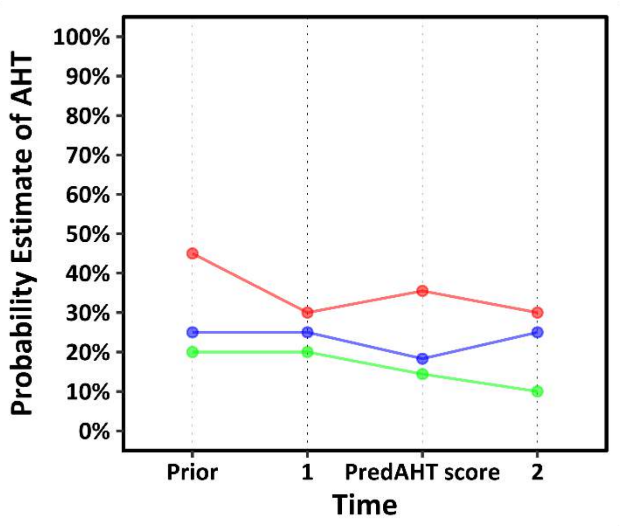


Table 7.23 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 6

Clinician ID	Designation	Specialty	Quote
29	Nurse	Neuro	"I agree with the score, it is enough to raise alarms but shows you that probably it isn't, so the probability is lower for it being non-accidental, but will still raise alarms so that you do all of [the investigations]."
30	Nurse	General paediatrics	"Just thinking about the kind of other warning signs you would see with an abuse case, there's only the single injury that we know of so far so there's not any signs of shaking so no retinal haemorrhages, and there's no other bruises we have not seen any other bruises on any other parts and it's just one injury instead of multiple, there's nothing else to disprove their story."
31	Nurse	Neuro	"A little bit is gut instinct, plus the absence of any other clinical features, that's why I would have thought about 20-25%. It seems to come up with the same thing that I've come out with."

7.4.1.5.12 Case 8: Fall onto gravel, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	?	?	✓	×	×	44.2%	1.48

All three clinicians interviewed believed the child's injuries in Case 8 to be accidental, and did not change their probability estimates of AHT throughout, although of note the clinicians' interpretation of the percentage probability associated with a low risk of AHT varied (Figure 7.18). Although clinicians said that a delay in presentation was an initial concern, all were reasonably confident of the diagnosis of nAHT, despite the fact that a skeletal survey and ophthalmology examination were not performed. Clinicians mentioned a range of factors that influenced their decision, including the mechanism of injury, the pattern of ICI, gut instinct, their interaction with the family, the parent-child interaction, discussions with colleagues, the lack of any other signs of injury suspicious for AHT, and the lack of previous involvement with social services. One clinician estimated a very low probability of AHT at Time 2 (2%) but still felt this level of risk was worryingly high (Table 7.24).

Figure 7.18 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 8

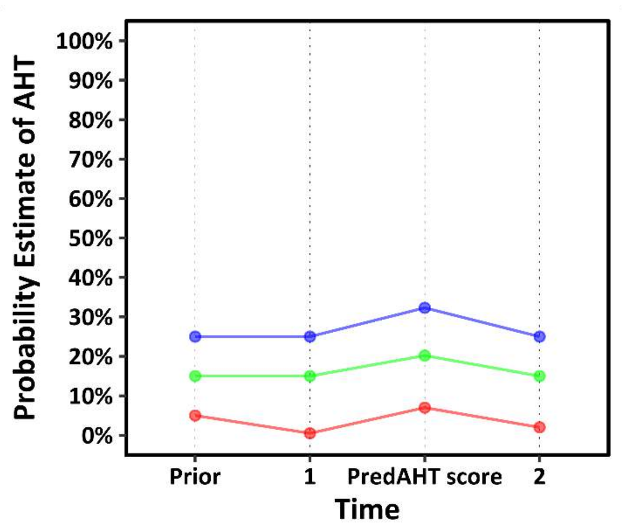


Table 7.24 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 8

Clinician ID	Designation	Specialty	Quote
36	Doctor	Emergency medicine	"There was a delayed presentation so I have to admit there is some risk with the discharge. But some of it is done on gut instinct and the fact that the family were incredibly appropriate and you know you're prepared to take that but uh it does still feel an uncomfortable percentage risk. The fact that it was an extradural meant that it was more likely to be an impact trauma than a shake. Which also reassured me and I discussed that with neurosurgery."
54	Doctor	Emergency medicine	"The mechanism that they gave fitted and otherwise he was really well kempt, didn't have any other bruising, didn't have any other signs of injury that had been caused in a non-accidental way, and he was behaving appropriately with his parents and interactions between them were normal. The concerning things about his presentation was the fact that he had a skull fracture and there was the delay in presentation but in fact all the other bits of his history and his examination were reassuring."
55	Doctor	General paediatrics	"It doesn't change that the devil's in the detail, which is the circumstance, the way they presented, their interaction, their level of concern, their previous history, social care saying 'He's never been known to them', the family set up, who's in the house, the way that the parents are, doesn't change any of the other factors, is the reason why."

7.4.1.5.13 Case 9: Fall from mother's arms, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	?	?	✓	×	×	44.2%	1.48

All clinicians interviewed about Case 9 thought that the likelihood of AHT was low, and they did not alter their probability estimates of AHT after using PredAHT (Figure 7.19). They reported that this was largely due to a consistent history provided by the parents and a perception that the mechanism of injury described was plausible. Two clinicians also reported that they deferred to a senior colleague on the matter, who had elected not to perform further investigations after taking a thorough history. Two higher grade doctors stated that PredAHT would not necessarily have encouraged them to recommend an ophthalmology review, while one trainee doctor reported that they would have asked for an ophthalmology examination and will ensure that this is ordered for future cases (Table 7.25).

Figure 7.19 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 9

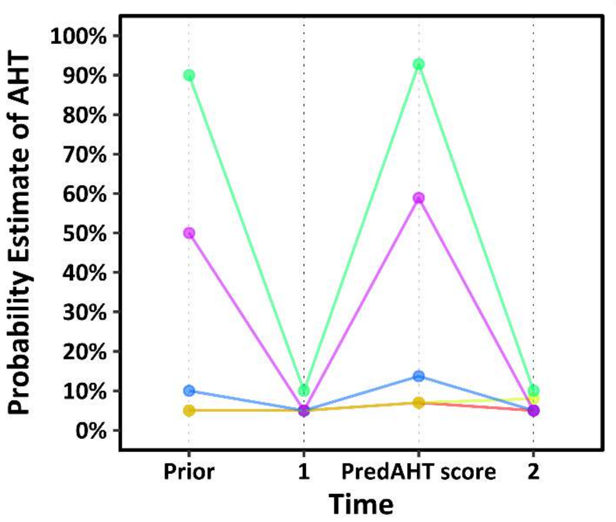


Table 7.25 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 9

Clinician ID	Designation	Specialty	Quote
45	Doctor	General paediatrics	"I would probably have got the ophthalmologist to have a look in his eyes. Knowing that that's on the scale, probably I would have done that. I probably would put that as a standard ophthalmology, if you've got a fracture with a bleed, I would probably use that as a standard. Now I would push for the ophthalmology review just to make sure."
46	Doctor	General paediatrics	"What I felt was that we needed to have done the works, that he should have had the entire works done. Irrespective of how plausible it seemed, I think the right procedure would have been to go down the child protection route. But if for a moment I thought look we were not acting in the best interests of this child or the family then I certainly would have escalated in respect of that. But given that this has been made, I was happy to defer in this instance. Because yes, irrespective of all the data that we have, sometimes you do have to use common sense. A senior consultant colleague has agreed to take the responsibility for those actions."
53	Doctor	General paediatrics	"I think I was totally aware here that I was making a decision based on what I heard and how I heard it, and what other people had heard. What I heard it was so consistent and it was so being relived. So I think I can see that and I think this was a really tricky one and it was a very, I suppose a consultant led decision because I'd taken that responsibility to a certain extent."

7.4.1.5.14 Case 10: Fall from sofa, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	?	?	×	×	✓	45.4%	1.55

Only two interviews were conducted regarding Case 10, as the child was transferred to a specialist hospital shortly after their admission to BRHC. This child had undergone recent facial surgery, however following discussion with the research team they were included in the study as the ICI was felt to be acute. The two clinicians considered the recent surgery in their deliberations, in addition to the behaviour of the family. One thought the risk of AHT was low, while one nurse thought that she may have underestimated the risk of AHT due to her lack of experience with CP (Figure 7.20). Both felt that PredAHT may be useful to initiate safeguarding discussions with senior colleagues (Table 7.26).

Figure 7.20 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 10

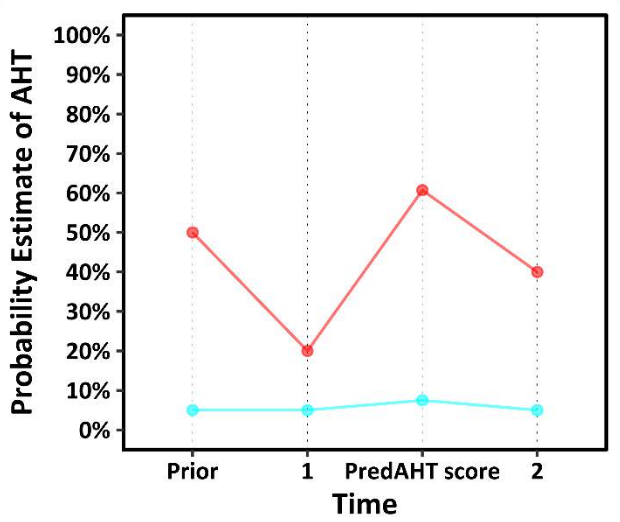


Table 7.26 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 10

Clinician ID	Designation	Specialty	Quote
44	Nurse	Emergency medicine	**Researcher:** If you had had that, would it have changed what you would have done? **Clinician:** Yeah it probably would have made me think a bit more. **Researcher:** Would you have discussed it with your line manager or child protection colleague? **Clinician:** I would probably talk to the consultant who was leading it and just check that she had thought about it and thought it through a bit more seriously, just to make sure that it had happened. **Researcher:** Is there anything else that you would have done differently? **Clinician:** No I wouldn't have done anything differently with the family, and I wouldn't have discussed with child protection because that's what the doctors do. But I would have spoken to people to make sure that that was happening. **Researcher:** Okay, so can I just try and explore with you then what made you go up to 40%? **Clinician:** There probably was more of a risk than I realised…I'm maybe being a bit naïve. I haven't had massive experience with child protection stuff so I generally probably would have been on the naïve side…and if that says it's more likely…
57	Doctor	Emergency medicine	"In this case we knew that she had an underlying vulnerability to head injury as a result of her recent skull surgery which meant that a small impact would by its nature be more likely to cause that kind of injury…the family were well known to the hospital and there hadn't been any concerns. Reading through a few of the letters they have rung immediately and sought help and were very appropriately concerned in resus." "My initial gut reaction is well I wouldn't use it because it's such a low probability in my own head that I don't think I'm going to be wrong. Which is probably arrogant on my part. If it had been massively different so if it was 30% or 35% and I thought it was 5% I think it would, it would make me question my own judgment and think why did I form those opinions. And if I was doing it in live real time I would discuss it further up the chain with someone else and say what do you think?"

7.4.1.5.15 Case 11: No history of trauma, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
×	×	×	×	✓	✓	58.5%	2.63

Clinicians reported that they found it very difficult to come to a decision regarding AHT in Case 11. Their probability estimates of AHT varied, and only two changed their probability estimate of AHT after seeing the PredAHT score, increasing it by 10% and 20%, respectively (Figure 7.21). Nevertheless, the majority thought that the probability of AHT at Time 1 and 2 was at least greater than 50%. Clinicians stated that they were conflicted due to a concerning social history coupled with an unusual and non-specific clinical presentation. Therefore, they said that the higher PredAHT score would not have altered their CP actions in this case, but may have been useful to support their decision-making (Table 7.27).

Figure 7.21 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 11

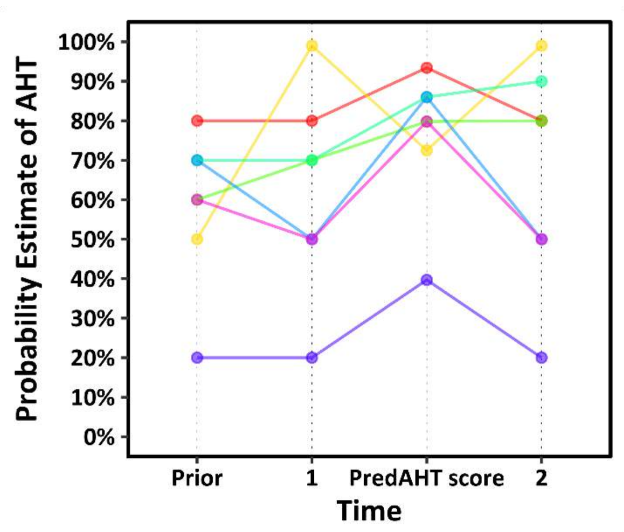


Table 7.27 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 11

Clinician ID	Designation	Specialty	Quote
51	Doctor	General paediatrics	"I guess it might make me think more, yes. That's got good evidence behind it rather than me just kind of picking it up out of my head, I think yes it probably would, you know that's evidence definitely. I think anything that helps you assess how worried you should be will be useful, I think it would be more as a kind of backup tool for what decisions you are making rather than necessarily a frontline, you wouldn't make the decision based on this but this would help you to back up your decisions I think which is useful."
56	Doctor	PICU	"For me, would it make any difference to my clinical practice, knowing that? No. I guess if you had someone who's less experienced, then it might have an influence, but I'm not sure that that figure alone would drive me to manage the patient, or investigate them in a different way." "Given the background and social history, I'm still going to say it's fifty fifty, because it could be that the apnoea, the seizure and the intracranial bit is all part of something else that's an underlying, something that's developmental in her. So because the other bits don't quite fit, if this was a shaking injury, you'd expect perhaps to see something else. Other than the seizure and the apnoea there's not a lot to go on. So it makes it, it's still there as one of the differentials, given the social background, it's still a fairly strong differential, but beyond that I think I'd find it quite hard to be convinced I think, yeah I think it's a really difficult one."
60	Nurse	PICU	"I think the tool is good, I think it's a really cool idea, actually, but I wouldn't like that to sway my mind, I would like to make up my own decisions on investigations, talking to different services that are involved with that child already, and going on results of the CT, the MRI. I like it, but, for me going from 50/50 to being now at 80 percent chance that there's abuse, I wouldn't do care differently, I don't think a percentage would sway you either way I think it would be something that I would do but not rely on."

7.4.1.5.16 Case 15: Fall down stairs, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	?	?	✓	×	×	44.2%	1.48

Clinicians' prior, Time 1 and Time 2 probabilities of AHT in Case 15 varied from 10%–50% (Figure 7.22). None of the clinicians altered their estimated probability of AHT after seeing the PredAHT score, likely because in three cases the PredAHT score was close to their own predicted probabilities. One clinician who thought the likelihood of AHT was very low (10%) said that she was satisfied that the proffered mechanism of injury was feasible and that the parent's behavior was appropriate, however another clinician who estimated a 50% probability of AHT felt that the parent's behavior was inappropriate, and said that he was on the fence regarding whether or not AHT had occurred. Two clinicians mentioned that PredAHT is useful for prompting further investigations, particularly an ophthalmology examination, both to assist in ruling in AHT if positive and ruling out AHT if negative (Table 7.28).

Figure 7.22 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 15

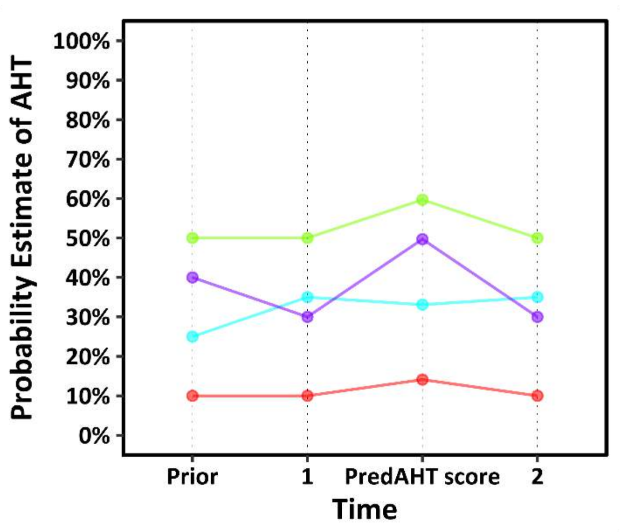


Table 7.28 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 15

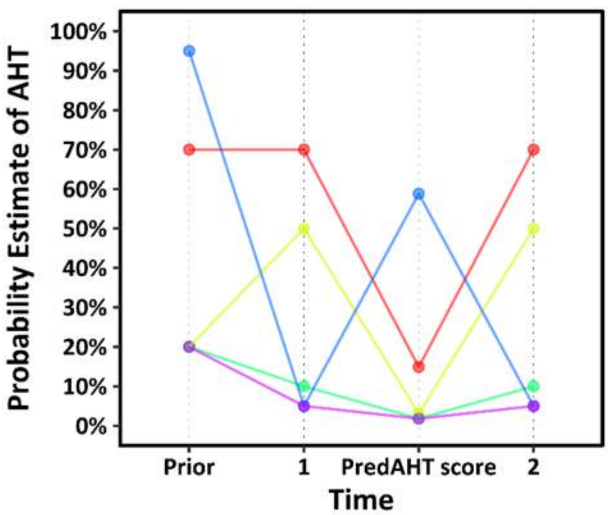
Clinician ID	Designation	Specialty	Quote
69	Nurse	General paediatrics	**Clinician:** If I hadn't thought about it, if I was going along and didn't have any idea that that this may be a non-accidental injury at all, and then I saw this tool, it may make me think a bit more and then ask the doctors has she been seen by ophthalmology, so that through this it may make the nurses suggest to the doctors something they may not have picked up on. **Researcher:** Do you think you might go and do that in this case? **Clinician:** Possibly, I'll certainly find out if they looked in her eyes for retinal haemorrhages. I don't feel I've got enough say on the case to be asking for a skeletal survey for fractures, to find out if those had been done, that would be their decision, not something that I would decide that needs doing. But I'll certainly ask about the other things. Have her eyes been checked and things like that yet.
70	Nurse	General paediatrics	"The tool hasn't changed my reasoning, even if it's gone up 10%, because it is just a tool. I think having a tool probably wouldn't give you the best result, it's about behaviour and how they react that can tell you a bit more than a tool."
71	Doctor	General paediatrics	"I think the tool is useful in terms of flagging up things to ask for and things to be aware of. It's really good. I was looking at whether we have any child protection guidelines and red flags to mention. Even when to say that it is absent, you know what I mean, we don't suspect child protection issues because they didn't have long-bone fracture, didn't have rib fracture and that would be really useful as well. I think I would have asked about these factors. Absent this, absent that, or we haven't investigated that, if you suspect this, follow that."

7.4.1.5.17 Case 16: Father dropped baby, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
×	×	×	×	×	×	3.9%	0.08

Clinicians reported that they were heavily influenced by the history in Case 16. Of four clinicians interviewed, three thought that the child's injuries were accidental while one felt that AHT was more likely than not. One clinician gave a very high probability of AHT as initially there was no history of trauma given whatsoever, however when a further history was obtained along with negative ophthalmology and skeletal survey results, their probability of AHT was reversed. However, one clinician still felt that the delay in proving an adequate history was suspicious for AHT. One clinician (ID 78) was interviewed once before the skeletal survey was conducted, and once afterwards, when the negative results were known. This clinician lowered their Time 1 probability in the second interview, and while PredAHT still did not influence their probability estimate (Figure 7.23), they said that it reassured them that the injury was most likely accidental. This clinician stated that she would only use PredAHT if all investigations were completed. Two clinicians thought that PredAHT would be more useful for those with less experience in CP (Table 7.29).

Figure 7.23 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 16



Clinician 78 is represented by the lime green line (before skeletal survey) and the dark green line (after skeletal survey).

Table 7.29 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 16

Clinician ID	Designation	Specialty	Quote
25	Nurse	Neuro	"I do think there was suspicion around this head injury purely from the history. I think the score is really helpful I just think the history was the thing that muddied the water a little bit and raised my concern in this particular case, with this baby."
78	Nurse	PICU	"The facts are still the same so that has not influenced my opinion because you still don't know what dad said, you still haven't had the skeletal survey and I believe that those things need to be found out and then maybe then you can come back and look at this. So you would have to wait for all the information so that you could put definitely no rib fracture, definitely no long-bone fracture." "There's factors you can't account for. But that's life. You're never going to be able to factor all of those things into a tool because otherwise it will be hundreds of questions long and you have to pick the most important ones don't you, which you've done. The score reassures me. Probability is…everything is based on probability. You could run your whole life by probability. So yeah it is a good test."
85	Doctor	Community paediatrics	"I think the interesting thing of the tool is it really doesn't link in with history and that was the key thing in this, it was because the absence of any history in a baby who couldn't have had that injury without somebody caring for them and knowing how it happened was the main concern. So it wasn't actually the injury itself as such it was the lack of an explanation for it. But it's nice to see that it's still above the 50 per cent that you have really got to think is this an abusive injury."
87	Doctor	Neuro	"It would be useful for the team, the registrars, I think if they are assessing the child, but for me personally, probably not based on the experience I've had but for a new consultant or junior doctor it would be great it sounds very good. You want the sensitivity to be as high as you can make it but it's also got to be specific as well."

7.4.1.5.18 Case 18: Fall from mother's arms, nAHT

Retinal haemorrhage	Rib fracture	Long-bone fracture	Head/neck bruising	Apnoea	Seizure	PredAHT probability	PredAHT Likelihood Ratio
?	?	?	✓	×	×	44.2%	1.48

Interviews regarding Case 18 were conducted after the child had been discharged home with presumed nAHT. At Time 1, all three clinicians estimated the probability of AHT to be 0%, however after seeing PredAHT, two accepted that there may have been a small level of risk for AHT and increased their estimates slightly (Figure 7.24). Clinicians felt that AHT was unlikely due to a consistent history, agreement between the medical team, and the behaviour of the parents, and two stated that PredAHT would not have influenced their CP actions in this case. One thought that PredAHT may help to raise suspicion of AHT with their colleagues (Table 7.30).

Figure 7.24 Clinicians' estimated prior, Time 1 and Time 2 probabilities of AHT, and the PredAHT predicted probabilities, for Case 18

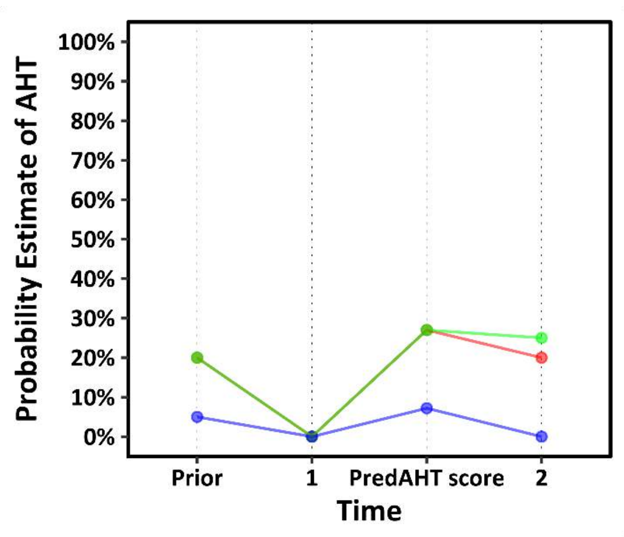


Table 7.30 Impact of PredAHT on clinicians' AHT probability estimates and child protection actions in Case 18

Clinician ID	Designation	Specialty	Quote
79	Nurse	General paediatrics	“It could be useful yeah, it gives you on paper what you are thinking in your head doesn’t it, because if you are thinking all this and then if the next shift comes on and doesn’t think the same as you, you have already done this and somebody may look at it and say, ‘oh right there are concerns then’, although we do hand over. Yeah for this type of patient yeah it could be used definitely.”
80	Nurse	General paediatrics	“No I wouldn’t have done anything differently, there are no strong feelings that I think ‘oh something has been missed’, there isn’t anything I disagree with.”
81	Doctor	General paediatrics	“I'm not going to change my mind because when I further interviewed the mother, she repeated the history and the way of the fall... sometimes when you hear the story you have to take a further detailed history. So it's like you can exclude or include something by taking a further history.”

7.4.1.6 Is a full-scale impact analysis of the Predicting Abusive Head Trauma clinical prediction tool warranted and feasible?

Table 7.31 below sets out the problems identified within this feasibility study, and potential solutions to these problems.

Table 7.31 Methodological issues identified and documented during the feasibility study

Methodological issues	**Findings**	**Evidence**	**Possible solutions**
What factors influenced eligibility and what proportion of those approached were eligible?	All patients with ICI meeting the inclusion criteria were eligible Clinician eligibility was compromised by refusal/non-response however the recruitment rate was encouraging	18/18 (100%) of eligible patients were included 87/120 (72.5%) of clinicians approached participated in the study	Offer incentives for study participation or for the use of the CPR
Was recruitment successful?	Patient recruitment was successful at BRHC with two nurses dedicated to identifying children with ICI as part of their job role. This was less successful at UHW Clinician recruitment of 72.5% was moderately successful Clinician recruitment from direct contact on the wards was more successful than recruitment via email However, a major issue was identified with the timing of recruitment. Clinicians were recruited some time after the case was admitted and clinical assessment was well underway	Clinicians failed to notify the researcher of 3/8 eligible patients at UHW Only one clinician declined to participate after being approached on the ward, whilst 32 who were approached by email did not respond or declined to participate Due to the timing of recruitment of clinicians, the majority of interviews were conducted in retrospect (see Table 7.12)	Increase level of study promotion Improve lines of communication between the researcher and the hospital departments with daily visits to the wards, including the weekend and out-of-hours Employ dedicated research nurses to recruit and interview clinicians as soon as possible after an eligible child is admitted

Did eligible participants consent?	CAG approval was granted to allow enrolment of patients and the collection of their clinical data without consent from their parents or care-givers (*subject to strict case anonymization*) All eligible clinicians consented	All 87 eligible clinicians consented to take part	None required
Did participants adhere to the intervention?	All clinicians calculated the PredAHT score, as this was part of the interview guided by the researcher Some could not estimate their own prior or Time 1 probabilities of AHT	In 93/97 interviews, clinicians provided full data on their probability estimates of AHT and next/retrospective CP actions	Subsequent studies should ensure dedicated research nurses are involved to encourage the use of PredAHT as early as possible in the clinical assessment process
Was the intervention acceptable to the participants?	Yes, overall PredAHT was acceptable to clinicians	Only two clinicians said they would not use PredAHT. Whilst some clinicians said that PredAHT would be unlikely to change their management, some found PredAHT useful to support their judgments and decision-making in spite of this (see analysis of individual cases)	None required
Were outcome assessments completed?	Clinician outcome assessments were completed in 93/97 interviews Patient outcomes regarding AHT vs. nAHT were satisfactory for 16/18 patients, in two cases the outcome was difficult to assign from	Two clinicians could not provide probability estimates of AHT, two could not provide prior probability estimates See table 7.10 for details of patient outcome assessments	Choose different clinician outcome measures to assess i.e. standardisation of clinical investigation, increased clinician confidence in diagnosis

	the available information		Ensure longer time scale for patient follow-up, consider social services or health visitor involvement to obtain more detailed information regarding patient outcomes Consider need to follow up children presenting from out-of-area
Were outcomes measured those that were the most appropriate outcomes?	Outcome measures were not the most appropriate outcomes	Clinicians' probability thresholds for investigating and reporting AHT are low. Where PredAHT influenced clinicians' probability estimates of AHT, these did not cross a threshold of suspicion required to initiate a change in CP action with regards to further investigation or social services referral	Qualitative analysis suggested possible alternative outcome measures i.e. standardisation of clinical investigation, completion of ophthalmology examination, increased clinician confidence in diagnosis, improved communication between team members
Were the logistics of running a multicentre trial assessed?	Yes, greater resources would be needed for a definitive study	One researcher extracted all clinical details and conducted all clinician interviews, interviews were difficult to coordinate in a timely manner at two sites The patient sample size (18) was low	Larger study team required. Buy-in from clinicians would mean impact could be assessed in real-time. Qualitative interviews could be limited to a cross-section of clinicians

			A definitive study would require multiple centres due to low patient numbers, and a much longer study time-frame to maximise patient recruitment
Did all components of the protocol work together?	No, the components had weak synergy	There were considerable difficulties identified in the various study processes and the researcher's ability to implement them. For example the time taken to identify eligible patients had a cascade effect on the time taken to extract clinical data and recruit clinicians for interviews. This meant that PredAHT could not be tested at the beginning of the assessment process. In many cases the outcome regarding AHT/nAHT had been decided by the time of interview and thus interviews were conducted in retrospect. Outcome measures were not appropriate.	Involve dedicated research nurses to ensure PredAHT is tested early in the assessment process Choose different outcome measures to assess

7.4.2 Secondary outcomes

7.4.2.1 Performance of the Predicting Abusive Head Trauma clinical prediction tool

The performance of PredAHT in the feasibility study using a 50% probability cut-off is detailed in Table 7.32 and displayed in Figure 7.25. The performance of PredAHT incorporating clinicians' prior probabilities is detailed in Table 7.33 and displayed in Figure 7.26. When incorporating clinicians' prior probabilities, the sensitivity of PredAHT remained unchanged at 100%, while the specificity decreased from 83.3% to 74.6%.

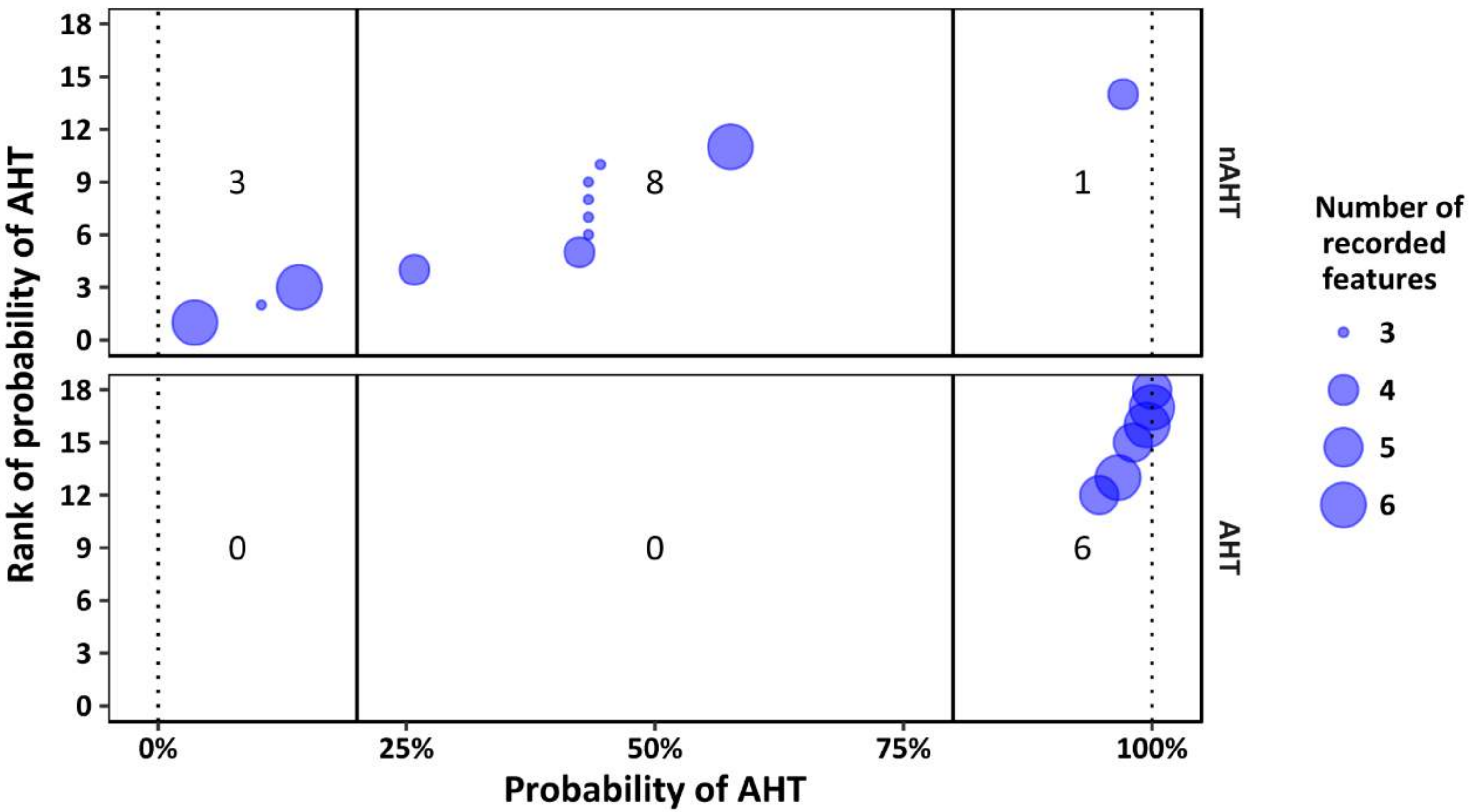
Table 7.32 Performance of the Predicting Abusive Head Trauma tool in the feasibility study

Applying PredAHT	50% cut-off	
	Outcome	
	AHT	nAHT
Higher risk	6	2
Lower risk	0	10
	Value	95% CI
Sensitivity	100%	51.7%–100%
Specificity	83.3%	50.9%–97.1%
Positive predictive value	75%	35.6%–95.5%
Negative predictive value	100%	65.5%–100%
LR +	6	1.69–21.26
LR −	0	0.00–0.48

Table 7.33 Performance of the Predicting Abusive Head Trauma clinical prediction tool in the feasibility study, incorporating clinicians' estimated prior probabilities of abusive head trauma

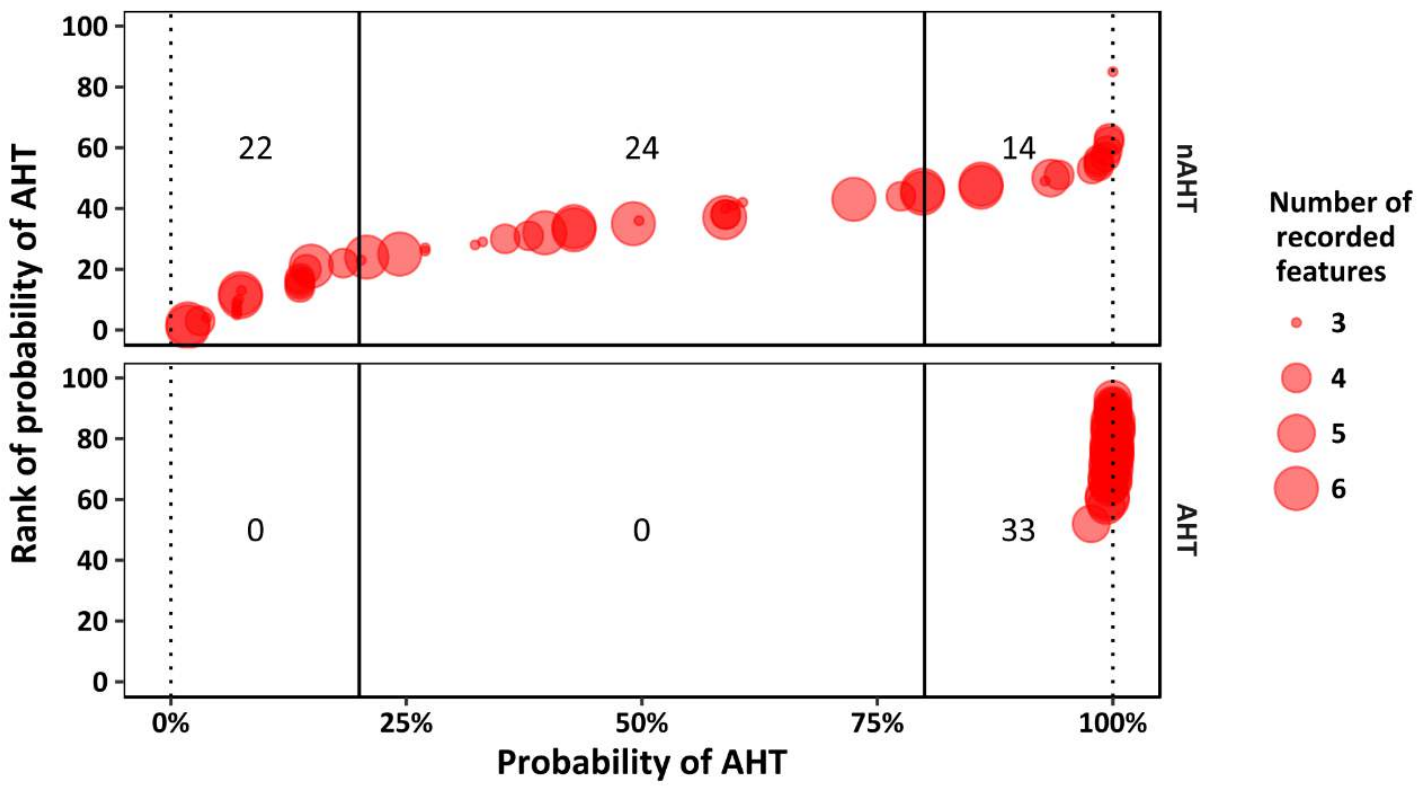
Applying PredAHT	50% cut-off	
	Outcome	
	AHT	nAHT
Higher risk	33	24
Lower risk	0	36
	Value	95% CI
Sensitivity	100%	87%–100%
Specificity	74.6%	40.5%–95.6%
Positive predictive value	60%	50%–75%
Negative predictive value	100%	88%–100%
LR +	3.94	1.68–22.73
LR −	0	0.00–0.17

Figure 7.25 Predicted probability of abusive head trauma assigned by the Predicting Abusive Head Trauma clinical prediction tool for all 18 children with intracranial injury, by outcome and number of recorded features



The circles represent the calculated probability of abusive head trauma for each of the 18 children with intracranial injury included in the study. The circle in the top right hand corner represents Case 1 (crush injury). The size of the circles shows how many of the six features are recorded as present or absent.

Figure 7.26 Predicted probability of abusive head trauma assigned by the Predicting Abusive Head Trauma tool when clinicians' estimated prior probabilities of abusive head trauma are incorporated



The circles represent the probability of abusive head trauma calculated by the Predicting Abusive Head Trauma clinical prediction tool when incorporating clinicians' estimated prior probabilities of abusive head trauma. The size of the circles shows how many of the six features are recorded as present or absent.The tool was less specific in predicting abusive head trauma when clinicians' prior probabilities were included.

7.4.2.2 Performance of clinicians

The performance of clinicians in predicting AHT at Time 1 (before PredAHT) and Time 2 (after PredAHT), using a 50% probability cut-off is detailed in Table 7.34 and Table 7.35 and depicted in Figure 7.27 and Figure 7.28. At Time 1, clinicians' predicted probabilities of AHT were 100% sensitive. Clincians' predictions were more specific than PredAHT, whether or not the PredAHT incorporated clinicians' prior probabilities. At Time 2, the sensitivity and specificity of clinicians in predicting AHT remained largely unchanged.

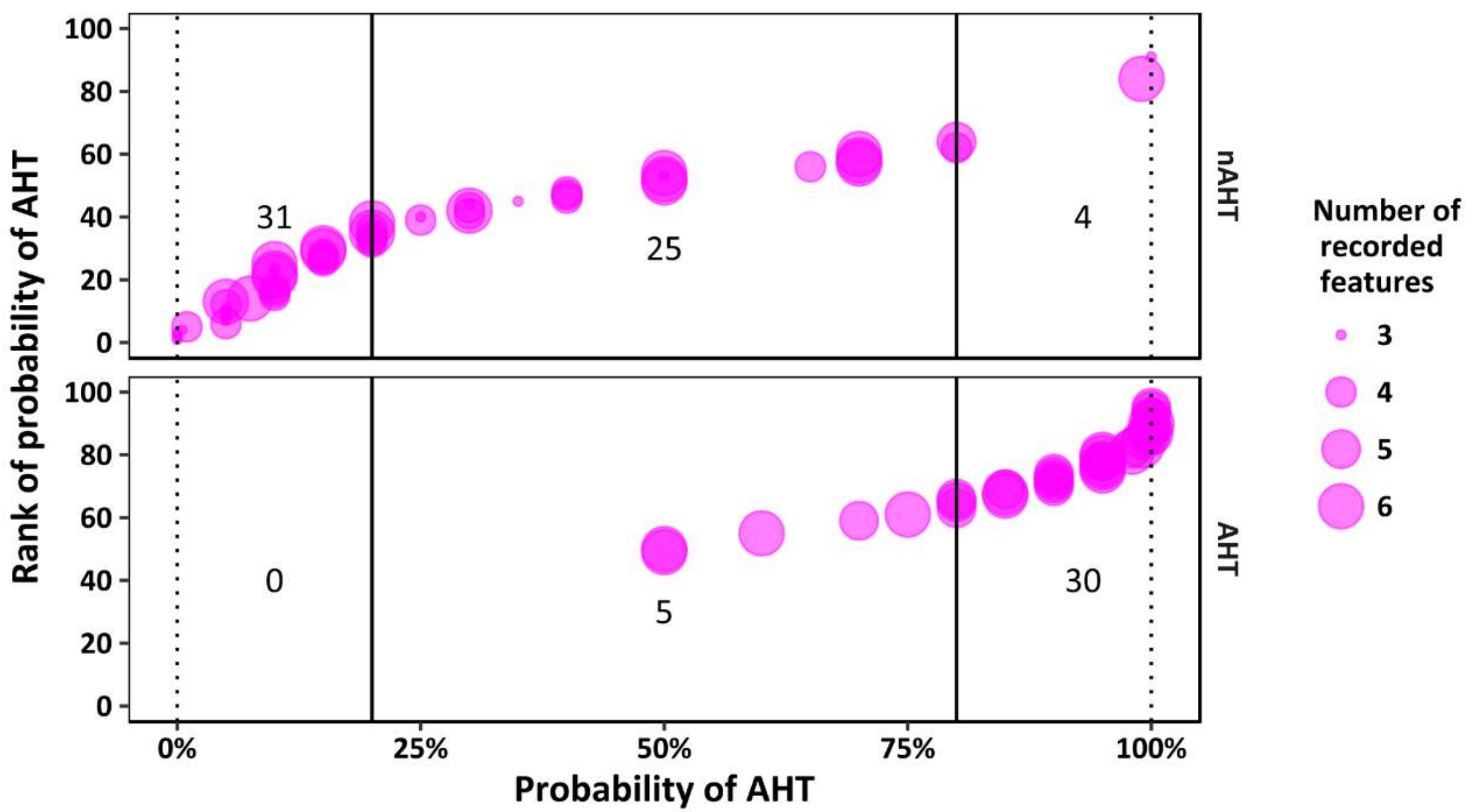
Table 7.34 Performance of clinicians in predicting abusive head trauma at Time 1

Applying PredAHT	**50% cut-off**	
	Outcome	
	AHT	**nAHT**
Higher risk	35	12
Lower risk	0	48
	Value	**95% CI**
Sensitivity	100%	87.7%–100%
Specificity	90.9%	71.7%–99.5%
Positive predictive value	75%	60%–100%
Negative predictive value	100%	90.8%–100%
LR +	11	3.53–200
LR −	0	0.00–0.12

Table 7.35 Performance of clinicians in predicting abusive head trauma at Time 2

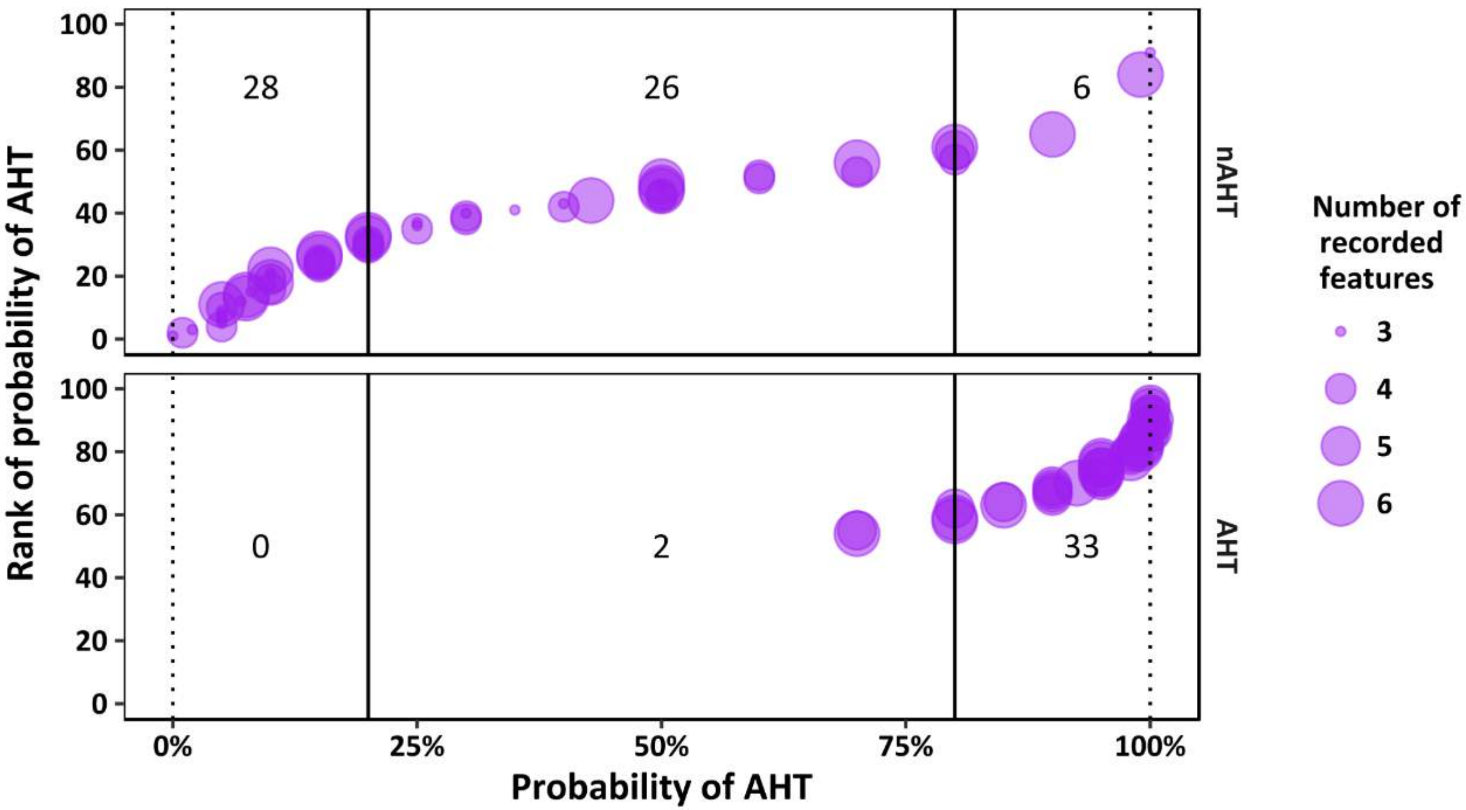
Applying PredAHT	**50% cut-off**	
	Outcome	
	AHT	**nAHT**
Higher risk	35	16
Lower risk	0	44
	Value	**95% CI**
Sensitivity	100%	87.7%–100%
Specificity	90.3%	66.5%–99.7%
Positive predictive value	75%	60%–85.7%
Negative predictive value	100%	90%–100%
LR +	10.3	3–333
LR −	0	0.00–0.13

Figure 7.27 Predicted probability of abusive head trauma assigned by clinicians at Time 1



The circles represent the probability of abusive head trauma estimated by clinicians at Time 1, before seeing the Predicting Abusive Head Trauma clinical prediction tool score. The size of the circles shows how many of the six features are recorded as present or absent. Clinicians were more specific than the tool in predicting abusive head trauma.

Figure 7.28 Predicted probability of abusive head trauma assigned by clinicians at Time 2



The circles represent the probability of abusive head trauma estimated by clinicians at Time 2, after seeing the Predicting Abusive Head Trauma clinical prediction tool score. The size of the circles shows how many of the six features are recorded as present or absent. The tool did not impact on clinicians' sensitivity and specificity in predicting abusive head trauma.

7.5 Discussion

This chapter described a novel feasibility study of the evaluation of the impact of PredAHT in clinical practice. PredAHT is currently the only CPR designed to assist in the identification of AHT that can be used at multiple points in the assessment pathway. The aim of this study was to determine the feasibility of evaluating the impact of PredAHT in the clinical setting, by assessing the different components of the methods and procedures used to conduct the study, including the processes of patient identification and follow-up, clinician recruitment, and the collection, relevance, and appropriateness of the proposed outcome measures. Qualitative methods were used to explore in detail the impact of PredAHT on clinicians' judgments and decision-making in individual children with ICI, and the additional factors influencing clinicians' opinions and practice in each case. The ultimate objective was to decide whether a full-scale, definitive impact analysis of PredAHT is feasible and warranted at this time.

7.5.1 Discussion of the main findings

The findings from this feasibility study indicate that a full-scale impact study of PredAHT is not feasible as designed, and that a definitive impact analysis of PredAHT is not warranted at present. A number of methodological issues were identified and documented during the study, relating to patient identification and follow-up, clinician recruitment, and the appropriateness and relevance of the outcome measures assessed. The findings will be summarised based on a consideration of guiding questions derived from a synthesis of the research methods literature related to feasibility studies.[985]

The first question to ask following a feasibility study, is "Can we recruit appropriate participants?"[985] The 100% recruitment rate of eligible patients less than three years of age admitted to hospital with ICI showed that it is feasible to identify these patients using the methods and procedures employed. Patient identification at BRHC worked particularly well, suggesting that the use of dedicated paediatric major trauma nurses to identify patients is a feasible recruitment strategy for future studies. At UHW, although initially clinicians assisted with patient recruitment, they failed to notify the researcher of 3/8 eligible cases. There are a number of possible reasons for this. Firstly, it is possible that the admitting clinician or the clinicians responsible for the child's care were unaware that the study was taking place. Although the researcher delivered several presentations about the study to the various departments and attended regular departmental safeguarding and multidisciplinary meetings to promote the study as widely as possible and obtain clinician "buy-in", it is likely that some staff members would not have been reached. Secondly, clinicians may have assumed that somebody else had already notified the researcher about the child. Thirdly, clinicians may have been aware of the study but unwilling to assist as there was no incentive to do so, or because they found the PredAHT CPR unacceptable. Possible solutions include improving lines of communication between the researcher and the hospital departments with daily visits to the wards, including the weekend and out-of-hours, and increasing the level of study promotion throughout the data collection period.

The clinician recruitment rate of 72.5% was encouraging, and the researcher was able to recruit a range of clinicians from different specialties to provide their opinions on the likelihood of AHT and the utility and impact of PredAHT in each of the 18 cases. It was feasible and acceptable to recruit clinicians on the wards. However, a major issue was identified with the timing of recruitment of patients and clinicians. As the majority of the children were admitted to hospital at the weekend or out-of-hours, by the time the researcher was notified of a patient, had extracted their clinical data, and approached clinicians for an interview, CP procedures had already been initiated, or the child had been discharged. This meant that PredAHT could not be tested at the beginning of the assessment process, where clinicians felt it would be most useful. In many cases the outcome regarding AHT/nAHT had been decided by the time of the interview and thus interviews were conducted retrospectively. Nevertheless, it was felt to be important to try to obtain as many clinicians' views on the potential impact of PredAHT in real life cases as possible. One possible solution would be to involve dedicated research nurses to be on hand seven days a week to recruit and interview clinicians as soon as possible after an eligible child is admitted, to ensure that PredAHT is tested early on in the clinical assessment process.

A second question to ask is "How appropriate are the data collection procedures and outcome measures for the intended population and purpose of the study?"[985] Selecting outcome measures for an intervention is notoriously difficult.[995] Chosen outcome measures must align with the proposed mechanism of change reflected in an intervention.[985] If inappropriate outcome measures are used in a definitive trial and the trial fails to show any beneficial effect of the intervention, it is possible that the outcome measures are insensitive to change or incongruent with the theoretical causal model of the intervention.[985] Since PredAHT does not recommend specific actions based on its predicted probabilities, one objective of this feasibility study was to explore how different probability predictions related to clinicians' CP decision-making. Although it has been suggested that clinicians may be more likely to use a CPR if its outputs are actionable and presented as decision recommendations,[315, 320] others have expressed concern that recommendations based on the categorization of patients into arbitrary risk groups may not be in their best interests, as it may be assumed that the risks are the same for all patients within each category.[324, 733] Of note, the PediBIRN CPR described in Chapter 4 recommends a thorough work-up for abuse when one or more of four variables are present, yet the calculated probability of AHT corresponding to the individual predictors varies from 11% (skull fracture other than an isolated, unilateral, nondiastatic, linear, parietal skull fracture) to 66% (bilateral or interhemispheric SDH).[705] PredAHT therefore provides patient-specific probabilities of AHT along the full continuum of risk, as recommended in the literature.[326] The threshold approach to decision-making theorises that a CPR that provides predicted probabilities of a condition conceivably acts to influence decision-making by moving a clinicians' pre-test probability above or below their own personal test and treatment threshold probabilities.[580, 585] Clinicians reported that their thresholds for investigating and referring AHT were very low and therefore many questioned the relevance of percentage probabilities for predicting AHT and doubted whether PredAHT would change their practice. In cases where PredAHT did influence clinicians' probability estimates of AHT, their CP actions in terms of further investigation and referral to social services nevertheless remained the same, suggesting that PredAHT did not encourage movement of their probabilities of AHT above or below a threshold of suspicion required to initiate a change in CP action. This suggests that it may not be appropriate or relevant to measure whether PredAHT influences clinicians decisions to undertake further investigations or refer a child to social services. However, in 8/18 cases, ophthalmology examinations were not performed, and in 9/18 cases a skeletal survey was not undertaken. Qualitative analysis of the impact of PredAHT in each of the individual cases suggested that the calculated probability may prompt clinicians to consider an ophthalmology examination, standardise the clinical investigation, encourage discussion and consultation between colleagues, and provide clinicians with confidence and reassurance in their judgments and decision-making, suggesting that these more nuanced outcomes may be suitable to assess in a future impact study.

Regarding patient outcome measures, longer follow-up times and links with social services or health visitors would be required to obtain more detailed information about the likelihood of AHT and to track the flow of children through the CP system. Two cases highlighted the difficulty of coming to conclusions in suspected AHT cases. In one (Case 12) there was a clear mis-match between the opinion of the medical and social care teams and the family court decision, while in another (Case 11), the multidisciplinary team concluded that the medical evidence was inconclusive and that the child had suffered probable trauma but returned the child home based on the consideration that they were safe from the potential perpetrator. One finding of this study was that many children were admitted from out-of-area, and had initially presented to their local regional hospital. This highlights the value of including the hospital CP team in a definitive study, to liaise with local CP teams, and has implications for future applications for ethical approvals.

An important consideration following a feasibility study relates to the logistics of carrying out the study and an evaluation of the resources that would be required to conduct a definitive study.[985] There were several challenges involved in carrying out a multisite study as part of a PhD project. Two neighbouring regional study sites were chosen due to the low population prevalence of AHT, and the researcher reviewed the case notes and conducted all interviews at both sites. One advantage of this was the standardisation of the interviews and data collected, however it was difficult to coordinate clinician interviews in a timely manner when there were children admitted with ICI at both sites at the same time. The interviews conducted with clinicians generated an unprecedented amount of qualitative data. Whilst this was undoubtedly a positive outcome for the feasibility study, a future trial could reasonably limit in-depth qualitative interviews to a cross-section of clinicians and to a narrower focus. Alternatively, focus groups could be conducted with small groups of clinicians; this method was shown to be valuable in assessing clinicians' experiences of using the Burns Risk assessment for Neglect or abuse Tool in clinical practice.[996] Due to the low sample size of 18 cases of children with ICI collected at two sites over a five month period, it was clear from this study that a definitive impact study of PredAHT would need to be conducted in many more sites and over a much longer time-frame to maximise patient numbers.

Several other important findings emerged from the qualitative analysis of the individual cases. Echoing the findings reported in Chapters 5 and 6, it was apparent that clinicians use a wide range of factors in their decision-making in cases of suspected AHT and that each case presents its own individual challenges and nuances. Sadly, all six children with AHT presented with multiple severe injuries and two died. In these cases, it was evident that AHT had occurred and the PredAHT probability calculation would not have changed clinicians' CP actions. However, it did help to reassure clinicians who were conflicted about the diagnosis that AHT was the most likely cause of the child's injuries. Other clinicians interviewed about these children speculated that PredAHT may be more useful in less clear-cut cases. One unexpected finding was the potential utility of PredAHT as an educational and empowerment tool for nurses. Many of whom suggested that PredAHT prompted them to review the clinical indicators of AHT, and would empower them to strike up conversations with doctors, to raise suspicion of AHT and check whether further investigations are required. Recruitment of nurses on the wards was straightforward, suggesting it would be feasible and valuable to include them in any future validation or impact studies of PredAHT.

Analysis of secondary outcomes revealed that PredAHT was 100% sensitive and 83% specific in predicting AHT in this population when applying a 50% probability cut-off. Two nAHT cases were predicted as AHT. The predicted probability of Case 1 (nAHT) was 97.2%. This child had sustained a confirmed accidental crush injury and presented with apnoea, seizures and head/neck bruising. The child was too ill to undergo an ophthalmology exam or skeletal survey and thus information regarding RH and long-bone fractures remained unknown. Due to the severity of this child’s injuries and potential safeguarding concerns around the circumstances of the injury, further investigation was required to rule out AHT. The predicted probability of Case 11 (nAHT) was 58.5%. This child presented with an unexplained cortical laceration, blood in their nose and mouth, apnoea, and seizures, and was classed as a high risk child who had previous involvement with social services and a concerning social history. Further investigations found no fractures or RHs. The medical evidence in this case remained inconclusive for AHT, however following a review of additional factors the case warranted a referral to police and social services which was arguably consistent with the PredAHT probability of 58.5% . These two cases highlight the importance of interpreting PredAHT probabilities in the context of all other available information. In both cases the PredAHT probability exceeded a 50% threshold, which was consistent with the decision to undertake further CP investigations.

When incorporating clinicians’ prior probabilities, the sensitivity of PredAHT in the 18 cases remained unchanged at 100%, while the specificity decreased from 83.3% to 74.6%. This finding, coupled with the finding that clinicians found it difficult to separate their prior probability estimates of AHT from the clinical features included in PredAHT, suggests that the incorporation of clinicians’ estimated prior probabilities into the PredAHT calculation may not be warranted, at least not without further training on and/or standardisation of this feature of the tool. However, the sliding scale feature may still be useful if the prevalence of AHT in a particular setting is known and differs substantially from the PredAHT baseline prior of 34%.

Secondary analysis revealed that clinicians’ predicted probabilities of AHT were 100% sensitive, and were more specific than PredAHT at both Time 1 (90.9%) and Time 2 (90.3%).

Critics have argued that it cannot be assumed that a CPR will be more accurate than clinician judgment and that therefore it is imperative to test a CPR against unaided clinical judgment as early as possible in the process of its development.[648, 680, 997] If it can be demonstrated that a CPRs predictions are more accurate than clinical judgment, clinicians may be more likely to use it.[299] Although earlier studies demonstrated that statistical and mechanical prediction methods consistently outperformed clinical prediction,[619, 637-639] a recent survey and a recent systematic review found that CPRs are rarely compared with subjective clinical judgment, and when a comparison is conducted, CPRs seldom outperform clinicians' unaided judgments.[643, 998] Traditionally it has been argued that a CPR that simply replicates but does not improve on clinical judgment has little added value.[648, 680, 997] In addition, many barriers to the use of CPRs can be explained by Diffusion of Innovations theory, which suggests that the adoption of innovations in healthcare is determined by features of the innovation itself, the adopters, and the wider context.[320, 999] One construct of the Diffusion of Innovations model is "relative advantage", which posits that innovations will only be adopted if they are perceived as more effective than the current way of working.[999]

A recent study found that CPRs are perceived to be of most value when they can be used to guide appropriate treatment and have an unambiguous advantage over usual ways or working, and that many clinicians believed that their unaided judgment was superior to CPRs in predicting patient outcomes.[320] However, this same study found that the use and potential advantages of a CPR may be much more complex than originally thought, and that they may be useful for purposes not previously reported, such as enhancing communication with colleagues and patients, and for medico-legal purposes.[320] These findings are consistent with the findings of this feasibility study, which demonstrated that clinicians proposed that PredAHT would be useful for encouraging communication with colleagues, standardising the clinical investigation of suspected AHT, and providing them with greater confidence and reassurance in their judgments and decision-making. Although clinicians' predictions of AHT were highly sensitive and specific in this study, this may have been because the multidisciplinary team decision regarding AHT had already been decided in some cases and thus the clinicians being interviewed had the benefit of hindsight when providing their predicted probabilities of AHT. One could argue that a tool that is consistent with clinicians' predictions and provides them with additional confidence in their decision-making may be of great value in a field as challenging and contentious as CP.

7.5.2 Comparison with the existing literature

This study has confirmed the association of seizures, apnoea and RHs with AHT, although caution should be taken when interpreting the odds ratios for the individual features due to the small sample size. Rib and long-bone fractures were more associated with AHT than nAHT, however these associations did not reach statistical significance, perhaps due to the amount of missing data for these features. Head/neck bruising was more associated with nAHT than AHT but again not significantly so. Interestingly, in five nAHT cases (3, 8, 9, 5, 18), the children presented with head/neck bruising, a history of an impact to the head, and a skull fracture with an underlying haemorrhage. None of these children had apnoea or seizures, and none underwent a skeletal survey or ophthalmology examination. As described in Chapter 1, this pattern of injury has been consistently associated with nAHT,[356] and the injury pattern was a contributing factor in clinicians' decisions not to perform further investigations. When PredAHT is used with the baseline prior probability of 34%, the predicted probability of AHT for this combination of features is 44.2% and thus arguably may encourage clinicians to investigate further, however, due to the small likelihood ratio, the PredAHT predicted probability for this combination of features will always be very similar to, and dependent on, a clinicians' estimated prior probability (see, for example, Figure 7.18). Clinicians' Time 1 and Time 2 probabilities were generally low in these cases, however in three of the cases (3, 9, 15), at least one clinician was concerned that the child had not had an ophthalmology examination.

These findings suggest that PredAHT could potentially be refined to account for specific patterns of ICI, as recommended by clinicians in the current study and in the qualitative study reported in Chapter 5. Of note, specific patterns of head injury have been found to affect the performance of the PediBIRN CPR. As previously observed, PediBIRN recommends a thorough work-up for AHT for all children presenting to the PICU with a "skull fracture other than an isolated, unilateral, nondiastatic, linear, parietal skull fracture", and the PediBIRN predicted probability of AHT corresponding to this feature is 11%. In the external validation study of PediBIRN in an Australian/New Zealand population, this variable was statistically significantly more common in nAHT cases than AHT cases, and led to false-positive results in 41/94 nAHT cases. When the variable was excluded from the analysis, the specificity of the CPR increased from 43% to 86% in patients admitted to any hospital ward.

To the best of the researcher's knowledge, the current study is the only feasibility study that has been conducted to evaluate whether it is possible to assess the impact of a CPR to identify AHT in clinical practice. A full-scale stratified cluster randomized implementation trial of the PediBIRN CPR for AHT is currently being conducted at eight PICU sites in the US, to assess the CPR's impact on AHT screening accuracy (www.pedibirn.com). The aim of this trial is to compare AHT screening guided by the CPR to AHT screening as usual, and it is hypothesised that application of the CPR will increase AHT detection by encouraging abuse evaluation (skeletal survey's and retinal exams) of high risk patients, and will reduce unnecessary abuse evaluations of patients with nAHT. As far as the researcher is aware, the investigators did not conduct a feasibility or pilot study prior to commencing this trial.

7.5.3 Strengths and limitations

A strength of this study includes the scope of the cases ascertained with regards to injury severity, patterns of clinical presentation, and likelihood of AHT, although it is acknowledged that all six children who were deemed to have suffered AHT were admitted to the PICU with severe injuries and symptoms. Another strength is the number of clinicians recruited, and the range of opinions obtained on the likelihood of AHT and the utility and impact of PredAHT in each of the 18 cases. A variety of clinicians working in different paediatric specialties, with different levels of CP experience and seniority were interviewed regarding each case. The study generated a wealth of rich, detailed qualitative data on the impact of PredAHT on clinicians' judgments and decision-making in each case, as well as the range of different factors influencing their judgments and decision-making.

The main limitation of this study was the timing of the recruitment, which has been discussed in detail above. Although clinicians stated that PredAHT may be valuable for increasing confidence or standardising the clinical investigation, due to the retrospective nature of many of the interviews it was not possible to assess the actual impact of PredAHT on these outcomes in this study. Similarly, although some clinicians said they were going to take specific actions after seeing the PredAHT score, for example check whether an ophthalmology exam had been performed or discuss the case with CP paediatricians or medical colleagues, it is unknown whether clinicians actually took these actions as this was not followed up.

In addition, during the study set-up and promotion, the researcher gave a number of presentations about PredAHT at departmental meetings and informed clinicians that they would be approached by the researcher to take part in an interview about PredAHT if they had been involved in the care and admission of a child with ICI. However, clinicians did not have access to PredAHT prior to the interview, which was sometimes conducted days after the child's admission. While this controlled approach ensured that PredAHT was tested with a number of clinicians, in reality, clinicians would have access to the tool as part of an implementation trial in order that they could complete it without a researcher having to be present. The PediBIRN implementation trial requires clinicians to complete a series of lengthy and detailed forms specifying their estimate of the percentage probability of AHT, details of the predictor variables present in the child, the result of the CPR (high vs. low risk for AHT), and any plans to conduct an abuse evaluation of the patient (see www.pedibirn.com). This approach would clearly require a significant degree of "buy-in" from clinicians prior to study commencement. It is unclear whether this approach is feasible and acceptable to clinicians as a feasibility study of the impact of PediBIRN was not conducted, although the published progress report on the trial website shows a high percentage of complete data for eligible patients. Nevertheless, the approach taken in the current study has demonstrated that it would be beneficial to involve a dedicated research nurse to encourage clinicians to use the tool, as has been shown in previous studies.[301]

Finally, although clinicians' probability thresholds for investigation and referral were explored in the qualitative interviews, on reflection a standardised approach may have been beneficial for understanding the level of certainty that clinicians' ascribe to different percentage probabilities, and the CP actions they would take based on different probabilities, as has been investigated in other studies.[490, 529, 530, 532] Similarly, while clinicians' perceptions of the value of incorporating their prior probabilities of AHT into PredAHT were explored to some extent, it would be useful to explicitly explore clinicians' use and understanding of likelihood ratios and Bayesian/probabilistic reasoning in greater detail in future studies, in order to fully investigate the implications of the application of prior probabilities in the context of decision-making in suspected AHT cases.

7.5.4 Implications for research and practice

Taken together, the results of this feasibility study suggest that in order to assess the impact of PredAHT in clinical practice, a cluster RCT would be required in multiple sites with an extended time scale, and, crucially, dedicated research nurses would be required 24/7 in order to identify cases in a timely manner and encourage clinicians to use PredAHT as early as possible in the assessment process. This strategy could be employed alongside other active, multifaceted implementation strategies in order to promote acceptance and increase uptake of PredAHT. In addition, the findings demonstrate that different outcome measures than those tested in the current study are warranted for a definitive trial. Research questions that could be considered for a future study are "Does the use of PredAHT improve clinicians' level of confidence in their diagnostic decision?" or "Does PredAHT improve the extent to which children less than three years old hospitalised with ICI are fully investigated for AHT?"

However, a full-scale trial would need to be informed by further feasibility and pilot testing before it could be considered viable. The results are further considered below within the context of the PICO format (Population, Intervention, Control, Outcome).

Population. This study showed that recruitment of children with ICI and recruitment of clinicians involved in the admission and care of the children is feasible, however there was an issue with the timing of recruitment. In addition, case numbers were low across the two sites. Thus, a definitive study would require multiple sites over a longer study period. A two-pronged data collection strategy could be utilised involving detailed study proformas as used in the PediBIRN trial, alongside dedicated research nurses to encourage proforma completion and collect additional in-depth qualitative data from a cross-section of clinicians regarding their experiences of using the tool in clinical practice. Alternatively, PredAHT could be integrated into the electronic health record to facilitate point-of-care decision support; this has been shown to be a promising and effective approach in previous studies.[777, 1000]

Intervention. The PredAHT Shiny app could be used on ward computers or clinicians' mobile phones to facilitate calculation of the probability of AHT at the bedside. Possible implementation strategies to encourage acceptance and use of PredAHT include clinician training, onsite visits, site-specific feedback, information sharing sessions, reminder emails, and research nurse support.

Control. Control sites would include sites with a similar population size and demographics to the intervention sites, but employing care-as-usual, without the use of PredAHT.

Outcome. This feasibility study has suggested that appropriate and relevant outcome measures may include increased clinician confidence in their diagnosis of AHT/nAHT, and standardisation of the clinical investigation of children with ICI where AHT may be suspected. Outcomes assessed could include quantitative measures of clinician confidence in their diagnosis and quantitative measures of clinical investigations undertaken in control and intervention sites. Process outcomes to test the feasibility of the study procedures could be measured with descriptive and qualitative data.

7.6 Conclusions

This chapter reported a novel, multisite feasibility study of the evaluation of PredAHT in clinical practice. This study demonstrated that a full-scale impact study of PredAHT is not feasible as designed and that a definitive impact analysis of PredAHT is not warranted at present. However, the feasibility study has been instrumental in informing the re-design of a possible definitive trial. Further feasibility and pilot work is recommended prior to a full-scale impact analysis of PredAHT in clinical practice.

8 General discussion

8.1 Chapter overview

This chapter reiterates the gaps in the evidence-base addressed by this thesis, summarises the main findings, highlights the novel contributions of this research to the field, and discusses the findings in relation to theory. The strengths and limitations of the frameworks used to guide the development and evaluation of PredAHT and the planning and design of the empirical studies presented in this thesis are discussed, and the benefits and challenges of the mixed-methods approach are highlighted. Finally, the implications for future research and practice are presented. Detailed discussions of the development of PredAHT and the empirical studies undertaken have been provided in each of the individual chapters, including comparison with the extant literature and consideration of the strengths and limitations of the study designs and methodological approaches employed.

8.2 Gaps in the evidence-base addressed by this thesis

A general literature review reported in the introductory chapter highlighted that AHT is a serious and potentially incapacitating or fatal condition with a complex clinical picture, and that it may go unrecognized in the clinical setting. The rarity of AHT in a given population means that clinicians likely encounter a limited number of cases in their practice. Doubt sowed around the validity of the diagnosis of AHT in the media, the courts and the medical literature has impacted upon clinicians' confidence to provide a firm opinion regarding the likelihood of AHT in both the clinical and legal setting.[479] Meanwhile, biomechanical, epidemiological and pathophysiological studies have helped to elucidate that AHT and nAHT result in divergent injury patterns; the review thus implied that constellations of clinical findings may be predictive of AHT.[12, 30, 68, 119, 356] Taken together, these findings suggest that evidence-based CPRs that are based upon clinical findings, such as PredAHT, may assist clinicians in identifying AHT.

Chapter 1 acknowledged that the development of a CPR requires several stages, namely, confirming the need for a CPR, derivation, external validation, and finally impact analysis to determine the effect of the CPR on clinical decision-making and patient care.[293, 299, 301] Chapter 2 justified the need for the PredAHT CPR. PredAHT had been derived[59] and externally validated[60] previously, and thus the next stage in the process would be to test its impact in clinical practice. However, the scientific literature regarding the development and evaluation of CPRs and complex interventions recommends that extensive exploratory and preparatory work is undertaken prior to a formal, large-scale definitive impact study, to assess the feasibility of conducting such a study and the acceptability of the CPR to those it is intended for.[300, 322, 323] Guidance also recommends identifying the evidence-base and relevant theories to obtain a theoretical understanding of the processes underlying the behaviour targeted by an intervention.[322, 323] This thesis therefore presented a scoping review of clinical diagnostic decision-making theory and a consideration of the mechanisms by which CPRs are proposed to enhance clinical decision-making (Chapter 3). These findings informed the development of the computerised version of PredAHT that was created to facilitate the conduct of the subsequent studies and the adoption of PredAHT into clinical practice (Chapter 4). These chapters were followed by three novel empirical studies exploring the acceptability[747] (Chapter 5), and potential impact[748] (Chapter 6) of PredAHT, and the feasibility of evaluating its actual impact in clinical practice (Chapter 7).

8.3 Thesis findings and novel contributions

This PhD thesis aimed to build on the knowledge gained from the derivation[24, 59] and validation[60] of PredAHT. The primary aims were to 1) develop a computerised version of PredAHT for use in clinical practice, and 2) to determine the utility of PredAHT in assisting in the identification of AHT, using mixed methods. Three empirical studies were conducted based on existing frameworks and guidance for the development and evaluation of CPRs. [300, 322, 323]

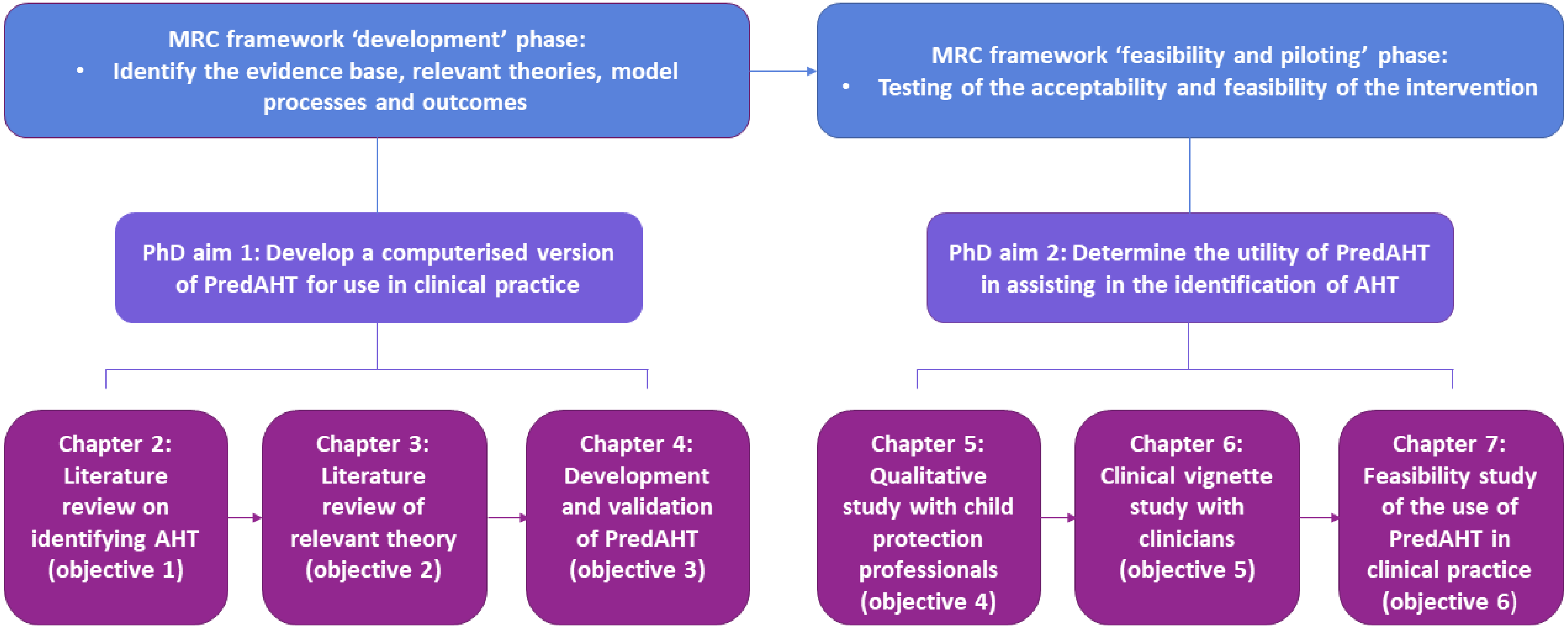
There were six objectives of the PhD:

1. To undertake a review of the literature on the challenges associated with the identification of AHT using systematic search methods.
2. To review relevant theories of clinical decision-making and the logic of CPRs using systematic search methods.
3. To describe the previous derivation and validation of PredAHT, present a critical appraisal of PredAHT and other CPRs for AHT, develop a computerised version of PredAHT using Shiny, a Web application framework for the R language and environment for statistical computing, and to validate the computerised PredAHT in an Australian/New Zealand population.
4. To assess the acceptability of PredAHT with a range of CP professionals, using qualitative methods.
5. To explore the potential impact of PredAHT on clinicians’ judgments and decision-making, using clinical vignettes.

6. To assess the feasibility of evaluating the impact of PredAHT in clinical practice, using mixed methods.

The MRC framework for the development and evaluation of complex interventions was used as an over-arching framework to guide the work, specifically, the “development” and “feasibility and piloting” phases[322, 323] (Figure 8.1). The findings and novel contributions of each chapter of this thesis are discussed below and presented in Table 8.1.

Figure 8.1 A schematic of the work presented in this thesis, according to the aims and objectives of the PhD and relevant phases of the Medical Research Council framework



Reproduced from Chapter 1.

8.3.1 Literature review of the challenges faced by clinicians in identifying abusive head trauma

Previous literature reviews in the field of AHT have focused on reviewing the injury mechanisms, outcomes, clinical features, and differential diagnoses associated with AHT, and the medical evaluation of suspected AHT, however few reviews have been conducted from the perspective of the challenges associated with the identification of AHT in clinical practice. An important step in the development and evaluation of CPRs is to review the evidence base to justify why a CPR is needed and why it is likely to be of value to those it is intended for.[301, 322, 323] The literature review conducted therefore focused on the difficulties faced by clinicians in identifying AHT, with the aim of determining whether a CPR for AHT is needed and to gain an understanding of the likely value of introducing a CPR for AHT into clinical practice. The review highlighted that clinicians face many challenges in the identification of AHT related to the history provided by the care-giver, variability in the clinical manifestations of AHT, potential differential diagnoses, forensic considerations, personal bias, alternative theories of causation, and the evaluation of suspected AHT. Despite this, the review found evidence that certain combinations of clinical features may help discriminate between AHT and nAHT, implying that certain clinical features may be predictive of AHT.[24, 27, 59] Due to the many difficulties involved in identifying AHT, the review concluded that a CPR for AHT that integrates a child's clinical data and calculates an evidence-based patient-specific probability of AHT is needed and would be of value to clinicians working in this field.

8.3.2 Clinical decision-making theories and the logic of clinical prediction rules

A scoping review of relevant theory related to clinical diagnostic decision-making in AHT and the logic underpinning CPRs was conducted. The inherent complexity of a CPR is rarely addressed or captured in studies developing or evaluating prediction rules,[319, 320] yet the use of theory proved to be an invaluable approach to understanding how and why PredAHT may influence clinicians' judgments and decision-making in practice. The findings demonstrated that decision-making functions in two interactive information processing systems, an "intuitive" system and an "analytical" system, that are represented on a cognitive continuum,[544, 596] and that clinicians use a range of decision-making strategies when considering a diagnosis of AHT, including pattern recognition, heuristics, hypothetico-deductive reasoning, causal reasoning, and probabilistic/threshold/Bayesian approaches.[53, 508] However, cognitive and affective biases arising from both "intuitive" and "analytical" reasoning underscore the need for decision support.[560] In addition, the review found that aspects of the diagnostic process in suspected AHT cases are clearly conducive to Bayesian reasoning.[508, 697, 707, 708] Taken together, this suggested that CPRs, which by nature are based on Bayesian reasoning, represent an ideal approach to aid clinicians in the identification of AHT,[627] and facilitated the identification of potential features of a computerised tool that might best support clinicians in their decision-making. Finally, the review highlighted that a CPR may conceivably improve a clinician's decision-making via a number of different mechanisms.[580] A CPR that provides predicted probabilities of disease and/or stratifies patients into risk groups may alter a clinician's decisions via changes to the diagnostic accuracy of their clinical judgment, by shifting their pre-test probability estimate above or below their individual test or treatment threshold probabilities.[545, 580, 585]

8.3.3 Development of the Predicting Abusive Head Trauma clinical prediction tool

The findings of Chapters 2 and 3 motivated and justified the development of a novel computerised version of the PredAHT CPR. Chapter 4 thus reported on the systematic process used to create this. The development of the computerised PredAHT followed an iterative process and included the selection and development of key features, programming, design of the user interface, in-house user testing, and modification based on suggestions and observations from subsequent studies undertaken for this thesis. Key features were selected based on findings reported in Chapter 3, and the acknowledgement that one or more predictor variables required for PredAHT may be unavailable at various stages in the AHT assessment pathway.[738] The computerised PredAHT has several unique features and advantages, including an "unknown" option for each clinical feature, LRs of AHT for each of the possible permutations of clinical features, and a sliding scale providing clinicians with the option to incorporate their estimated prior probability of AHT and facilitating automatic generation of posterior probabilities. The computerised PredAHT thus provides predicted probabilities and LRs for 729 possible permutations of six clinical features depending on whether each is present, unknown, or absent. A systematic review and critical appraisal of four validated CPRs for AHT and an external validation of the computerised PredAHT described in this chapter[706] found that up until the time of publication, PredAHT was the only CPR for AHT designed for use in an in-patient hospital setting that could be used at multiple points in the assessment pathway, and that since PredAHT had been validated in multiple settings it was therefore ready for impact analysis. The computerised PredAHT was subsequently used in the three empirical studies presented. Additional key features incorporated following feedback and observations in these three studies included an option to display confidence intervals around the posterior probabilities and the addition of Fagan's nomogram to illustrate how the user's own prior probability of AHT and the LRs of AHT for the 729 different feature combinations interact to produce the posterior probability of AHT. Strengths of the computerised PredAHT tool include the highly systematic, evidence-based and rigorous approach to its development, however the final version is yet to be formally user-tested.

8.3.4 Acceptability of the Predicting Abusive Head Trauma clinical prediction tool: A qualitative study with child protection professionals

A novel qualitative interview study was conducted with 56 CP professionals involved in suspected AHT cases, including clinicians, CPSWs, police officers, pathologists and legal professionals, including four judges.[187, 747] This study explored the factors influencing these professionals' decision-making and multidisciplinary collaboration in suspected AHT cases,[187] and their attitudes towards PredAHT,[747] and made several novel contributions to the field of AHT diagnosis and the scientific literature. Previously, no study had sought to directly explore how CP professionals make decisions and work together in suspected AHT cases. Six over-arching factors influencing decision-making were identified: "professional", "medical", "circumstantial", "family", "psychological" and "legal" factors. The findings suggested that decision-making in AHT cases is complex and nuanced, and a diagnosis is arrived at only when all potential variables have been carefully explored and considered, including clinical, historical, forensic and social features and potential differential diagnoses. These findings are particularly topical in light of the recent report out of Sweden[251, 252] that questions the validity of AHT/SBS as a medical diagnosis, predicated on the inaccurate premise that the "triad" of SDH, RH and encephalopathy defines AHT and forms the basis of a clinical AHT diagnosis. The findings also suggested that CPSWs and police officers may benefit from additional training in the medical aspects of physical abuse, and that joint training might provide a better understanding of the roles, expectations and limitations of each agency, thereby facilitating more effective collaboration.

The acceptability of PredAHT to CP professionals was also explored .[747] The acceptability of CPRs has been predominantly investigated using survey methodology, however qualitative methods have the potential to provide more in-depth information.[673] The findings demonstrated that PredAHT is acceptable and potentially useful to CP professionals who would be willing to use it as an adjunct to their decision-making. When used in conjunction with a full clinical and social history, PredAHT has the potential to standardise clinical assessment, and minimise subjectivity when weighing up the clinical features in cases of possible AHT. Strengths of this study include the wide range of professionals interviewed, and the richness of the data. The study is one of only two studies[747, 790] exploring the acceptability of a CPR designed for use in the CP field, and the first study to have done so with a range of professionals. There was evidence that data saturation was not reached with the pathologist group, and so further exploration of the acceptability of PredAHT may be warranted with these professionals.

8.3.5 Potential impact of the validated Predicting Abusive Head Trauma clinical prediction tool: A clinical vignette study

A novel clinical vignette study was conducted with 29 clinicians involved in assessing suspected AHT cases.[748] This was a preliminary exploratory study conducted to explore the impact of PredAHT on clinicians' AHT probability estimates and proposed CP actions. This study also made several contributions to the field of AHT diagnosis and the scientific literature. Exploratory studies can provide important information regarding the likelihood that a CPR will influence clinicians' judgments and decision-making but are rarely carried out prior to a formal impact study.[306] Six novel clinical vignettes were carefully designed by the researcher based on the findings from the qualitative study reported in Chapter 5, the clinical experience of the supervisory team, the scientific literature, and actual cases of suspected AHT encountered during previous research. Clinicians estimated the probability of AHT and indicated their CP actions in each vignette, both before and after using PredAHT. Inter-rater agreement of clinicians' judgments was also estimated. The study findings demonstrated that PredAHT significantly influenced clinicians' probability estimates of AHT in all vignettes. Interestingly however, clinicians' proposed CP actions were only influenced by PredAHT in a minority of instances, and PredAHT did not significantly improve the overall agreement between clinicians' AHT probability estimates or their proposed CP actions. Despite this, the "think-aloud" data showed that 27/29 clinicians would find PredAHT useful in their practice, and that it provided them with greater confidence in their opinions in the vignette cases, even if they would not alter their CP actions after seeing the score. However, it was evident that clinicians were influenced by a variety of social, historical and clinical factors in each case, emphasizing the need to consider the PredAHT probabilities in the context of these associated factors. Strengths of this study include enhanced internal validity afforded by the use of carefully manipulated vignettes, and the use of the concurrent "think-aloud" method alongside statistical modelling of survey responses, which allowed for a meaningful interpretation of the quantitative results. Limitations include reduced external validity due to potential “artificiality” of the vignettes, and limited generalisability due to the sampling methods employed.

8.3.6 Evaluating the impact of the Predicting Abusive Head Trauma clinical prediction tool in clinical practice: A feasibility study

Chapter 7 reported a novel multisite feasibility study of the evaluation of PredAHT in clinical practice. This study was conducted to determine whether a full-scale impact analysis of PredAHT is currently warranted. To the best of the researcher’s knowledge, there have been no completed impact studies of a CPR designed to assist in the identification of AHT, and it is unclear how a CPR in this field may be received or used in clinical practice, therefore it was considered paramount to establish the feasibility of conducting an impact study of PredAHT. Eighteen consecutive children less than three years of age admitted to two UK teaching hospitals with ICI were recruited, and several clinicians involved in the admission and/or care of each child participated in an interview where they discussed the probability of AHT estimated by PredAHT in relation to the case. The different components of the methods and procedures used to conduct the study were assessed, and the study yielded rich qualitative data regarding the impact of PredAHT on clinicians’ AHT probability estimates and CP actions in each case. This study demonstrated that a full-scale impact study of PredAHT is not feasible as designed and that a definitive impact analysis of PredAHT is not warranted at present. The results suggested that dedicated research nurses would be required seven days a week in order to identify cases in a timely manner and encourage clinicians to use PredAHT as early as possible in the assessment process. In addition, different outcome measures than those tested in the feasibility study would be warranted for a definitive trial, specifically, improved clinician confidence in their diagnoses, and standardisation or completion rates of the clinical investigation of AHT.

The feasibility study findings were instrumental in informing the re-design of a potential definitive trial. Further feasibility and pilot work is recommended prior to a full-scale impact analysis of PredAHT in clinical practice. Secondary analysis found that PredAHT probability estimates were less specific when clinicians’ prior probabilities were incorporated into the calculation, and that clinicians were more specific in predicting AHT than PredAHT.

The main strengths of this study were the range of clinicians interviewed about each case and the wealth of in-depth qualitative data collected. The main limitation was the timing of recruitment.

Table 8.1 Novel contributions of this thesis

Chapter	Study design	Primary aims	Contribution to research
2	Literature review	To review the literature on the challenges confronting clinicians in the identification of AHT, to identify if a CPR for AHT is needed	This review highlighted that clinicians face many challenges in the identification of AHT related to: • the history provided by the care-giver • variability in the clinical manifestations of AHT • potential differential diagnoses • forensic considerations • personal bias • alternative theories of causation • the evaluation of suspected AHT Despite this, the review found evidence that certain combinations of clinical features may help discriminate between AHT and nAHT. This review concluded that as the identification of AHT is difficult for a myriad of different reasons, a CPR for AHT that integrates a child's clinical data and calculates an evidence-based patient-specific probability of AHT is urgently needed and would be of value to clinicians working in this field.
3	Scoping review	To review clinical decision-making theories and the logic underpinning CPRs, to gain a theoretical understanding of clinical decision-making in suspected AHT cases and the mechanisms by which CPRs may improve clinicians decision-making	This review highlighted that: • clinicians use a range of decision-making strategies when considering a diagnosis of AHT, including Bayesian reasoning, but nevertheless, diagnostic decision-making in AHT is susceptible to a multitude of cognitive and affective biases arising from both "intuitive" and "analytical" reasoning, emphasizing the need for decision support • aspects of the diagnostic process in suspected AHT cases are clearly conducive to Bayesian reasoning Taken together, this suggests that CPRs, which by nature are based on Bayesian reasoning, represent an ideal approach to aid clinicians in the identification of AHT. • a CPR that provides predicted probabilities of disease and/or stratifies patients into risk groups may alter clinician's decisions via changes to the diagnostic

			accuracy of their clinical judgment, by shifting their pre-test probability estimate above or below their individual test or treatment threshold probabilities
4	Development of the computerised PredAHT	To describe the previous derivation of the PredAHT regression model and its subsequent external validation To describe a systematic review of validated CPRs for AHT conducted in collaboration with Australian colleagues and present a critical appraisal of the four validated CPRs for AHT To report the systematic development of the computerised Predicting Abusive Head Trauma clinical prediction tool and its external validation in an Australian/New Zealand population	In order to facilitate its adoption in clinical practice, PredAHT was developed into a simple, novel, web-based calculator. The development of the computerised PredAHT followed an iterative process and included: • selection and development of key features • programming • design of the user interface • in-house user testing • modifications based on suggestions and observations from the subsequent empirical studies undertaken in this thesis Key features were selected based on: • findings from the scoping review of the literature reported in Chapter 3 • the appreciation that if clinicians were to apply PredAHT in practice, they may face the problem of an unknown clinical feature Key features of the computerised PredAHT included: • an “unknown” option for each of the six clinical features • LRs of AHT for each of the possible combinations of clinical features • a sliding scale to enable clinicians to incorporate their own prior probability of AHT and facilitate automatic calculation of the posterior probability of AHT The computerised PredAHT thus provides predicted probabilities and LRs for 729 possible permutations of six clinical features depending on whether each is present, unknown, or absent. Additional key features incorporated following feedback and observations from the empirical studies included: • an option to display confidence intervals around the posterior probabilities • the addition of Fagan’s nomogram to illustrate how the user’s own prior probability of AHT and the LRs of AHT for the different combinations of features interact to produce the posterior probability of AHT Based on a systematic review of four validated CPRs for AHT and an external validation of the computerised PredAHT this chapter concluded that:

			• PredAHT is currently the only CPR for AHT designed for use in an in-patient hospital setting that could be used at multiple points in the assessment pathway • PredAHT has been validated in multiple settings and is thus ready for impact analysis The computerised PredAHT was used in three subsequent empirical studies investigating the acceptability and potential impact of PredAHT and the feasibility of evaluating its impact in clinical practice.
5	Qualitative interview study	To explore the factors influencing decision-making and multidisciplinary collaboration with CP professionals involved in suspected AHT cases, and to determine the acceptability of PredAHT amongst these professionals	Six main factors influencing decision-making in suspected AHT cases were identified: • professional • medical • circumstantial • family • psychological • legal With regards to decision-making and multidisciplinary collaboration in suspected AHT, this study found that AHT is *not* diagnosed based solely on the features of the "triad" alone, that decision-making is complex and nuanced, and a diagnosis is only arrived at when all potential variables are carefully explored and considered. Specifically: • participants diagnose AHT based on clinical features, the history, and the social history, after excluding potential differential diagnoses • participants viewed the diagnosis of AHT as a "jigsaw puzzle" that could only be solved with multiple different pieces of evidence • participants find these cases emotionally challenging but are aware of potential biases in their evaluations and strive to overcome these • barriers to decision-making include lack of experience, uncertainty, the impact on the family, the pressure of making the correct diagnosis, and disagreements between professionals • legal barriers include alternative theories of causation proposed in court • facilitators include support from colleagues, multidisciplinary working, knowledge of the evidence-base, and "gut instinct"

			- participants' experiences with multidisciplinary collaboration are generally positive, however CPSWs and police officers are heavily reliant on clinicians to guide their decision-making, suggesting the need for training on the medical aspects of physical abuse for these professionals and multidisciplinary training that provides knowledge about the roles of each agency With regards to acceptability of PredAHT, this study found that: - clinicians, CPSWs and police thought PredAHT would be beneficial as an objective adjunct to their professional judgment, to give them greater confidence in their decisions - PredAHT was viewed as a piece of the "jigsaw puzzle" of evidence to be used alongside all other information about each case - lawyers and pathologists appreciated its value for prompting multidisciplinary investigations, but were uncertain of its usefulness in court - perceived disadvantages included: possible over-reliance and false reassurance from a low score - interpretations regarding which percentages equate to 'low', 'medium' or 'high' likelihood of AHT varied; participants preferred a precise % probability over these general terms - participants would use PredAHT with provisos: if they received multi-agency training to define accepted risk thresholds for consistent interpretation; with knowledge of its development; if it was accepted by colleagues
6	Clinical vignette study	To explore the impact of PredAHT on clinicians' probability estimates of AHT, and their proposed CP actions, assessing the degree of agreement between clinicians' opinions both before, and after, using PredAHT	In this study six clinical vignettes were designed based on the scientific literature, the findings from the qualitative study reported in Chapter 5, the clinical experience of the supervisory team and actual cases of suspected AHT encountered during previous research. This study found that: - overall, PredAHT significantly influenced clinicians' probability estimates of AHT in all vignettes ($p < 0.001$), although the impact on individual clinicians varied - the influence of PredAHT on clinicians' CP actions was limited; after using PredAHT, 9/29 clinicians changed their CP actions in only 11/174 instances

			• clinicians' AHT probability estimates and CP actions varied somewhat both before and after PredAHT • the "think-aloud" data suggested that PredAHT may increase clinicians' confidence in their decisions when considered alongside other associated clinical, historical and social factors
7	Feasibility study	To establish the feasibility of evaluating the impact of PredAHT in clinical practice To assess the performance of PredAHT in the study population, at baseline and when incorporating clinicians' prior probabilities of AHT, and to assess the performance of clinicians at predicting AHT, both before and after using PredAHT	This study found that a full-scale impact study of PredAHT is not feasible as designed, and that a definitive impact analysis of PredAHT is not warranted at present. The results suggested that: • dedicated research nurses would be required 24/7 in order to identify cases in a timely manner and encourage clinicians to use PredAHT as early as possible in the assessment process • different outcome measures than those tested in the feasibility study are warranted for a definitive trial, specifically, increased clinician confidence in their diagnoses, and standardisation of the clinical investigation of AHT • a definitive impact study of PredAHT would need to be conducted in multiple sites and over a much longer time-frame to maximise patient numbers • further feasibility and pilot work is recommended prior to a full-scale impact analysis of PredAHT in clinical practice Secondary analysis found that: • the sensitivity of PredAHT in predicting AHT was 100%, while specificity was 83% • when clinicians' prior probabilities were incorporated, the specificity of PredAHT decreased to 75% • the sensitivity of clinicians in predicting AHT was 100%, while their specificity was higher than PredAHT, at 90%

8.3.7 Integrating the findings of the novel empirical studies presented in this thesis

Several important findings were echoed throughout all three of the empirical studies conducted. In all three studies, clinicians suggested that PredAHT may help to standardise the clinical evaluation of suspected AHT cases by encouraging them to undertake investigations such as an ophthalmology examination or skeletal survey,[747, 748] in line with international recommendations.[512, 514, 521, 522] This potential use of PredAHT was also appreciated by other CP professionals in the qualitative study reported in Chapter 5,[747] such as pathologists and judges, even when they could not see a use for the tool in their own practice.

Secondly, it was clear that clinicians are influenced by a wide range of factors in their decision-making about suspected AHT.[187, 748] The qualitative study was the first study to directly explore the factors influencing decision-making and multidisciplinary collaboration in suspected AHT cases with clinicians and other CP professionals, finding that decision-making in such cases is multifaceted and that a diagnosis of AHT is made only after a thorough consideration of multiple influential variables.[187] The manipulation of a number of these key variables (clinical features, history, and social history) in the vignette study (Chapter 6)[748] confirmed that clinicians' judgments of the probability of AHT were dependent upon specific details relating to these factors. Further, qualitative data of clinicians' judgments and decision-making in actual cases of suspected AHT in the feasibility study (Chapter 7) revealed that each case presented its own individual challenges and that clinicians considered a long list of factors in their decision-making regarding AHT in each case including clinical features, injury severity, patterns of injury, whether the injuries are consistent with the history and mechanism provided, social factors, and differential diagnoses, to name but a few. The studies presented have thus provided consistent evidence that a diagnosis of AHT is never made based simply on a consideration of a small number of clinical features (i.e. the "triad") in isolation.

Thirdly, a consistent theme emerging from all three studies was that clinicians stated that their probability thresholds for investigating and referring young children with ICI are very low and they maintained that they have a high index of suspicion for AHT in such cases.[187, 747, 748] Despite this, it was evident that in some cases, decisions on whether to perform further investigations or refer a child to social services were influenced by subtle details and circumstances of the case. For example, in the vignette study,[748] some clinicians would have taken no CP action in a scenario where the history was consistent over time and between caregivers, there was no delay in presentation, and the injury was consistent with an impact trauma to the head. Similarly, in five nAHT cases in the feasibility study, an ophthalmology exam and skeletal survey were not performed as the children presented with a skull fracture with an underlying haemorrhage, bruising over the site of the injury, and a lack of other clinical features concerning for AHT. Decisions not to perform an abuse evaluation are likely associated with a need to balance the risk of possible future harm to a child with the harm caused by unnecessary testing and a false accusation of abuse in children where the probability of AHT is low.[290] If the standardisation and completion of the clinical investigation is to be used as an outcome measure in a future definitive study of the impact of PredAHT in clinical practice, further exploration of the probability cut-off at which clinical investigations should be performed is required.

The value of incorporating clinicians' prior probabilities of AHT into the PredAHT calculation was explored in all three studies.[747, 748] Clinicians in the qualitative study felt that this component of PredAHT was subjective and stated that they would need training in order to confidently estimate their prior probabilities of AHT. In the vignette study,[748] some clinicians were surprised at the extent to which their prior probabilities of AHT impacted on the PredAHT score, while in the feasibility study, clinicians found it difficult to separate their prior probability estimates of AHT from their overall probability that incorporated information about the clinical features. The collective findings from this exploration suggest that further investigation of the implications of the use of prior probabilities in decision-making regarding AHT is warranted.

Finally, clinicians and other CP professionals across the three studies thought that PredAHT was valuable for providing them with confidence and reassurance in their judgments, decision-making and diagnoses.[747, 748] Crucially, many clinicians stated that PredAHT would increase their confidence even if it would not have a direct impact on their CP actions. This suggests that increased clinician confidence in their diagnoses of AHT/nAHT could be used as a quantitative outcome measure in a future definitive impact study.

Taken together, the results of the three empirical studies conducted suggested that PredAHT is acceptable to clinicians and other CP professionals,[747] and has the potential to significantly influence clinicians' judgments of the probability of AHT, but may have less impact on their CP actions in terms of deciding whether or not to pursue a child abuse evaluation or refer a child to social services.[748] However, the value of PredAHT may lie in its potential to standardise the clinical investigation of suspected AHT, and increase confidence in clinicians' decisions. At this time, a large-scale impact study of PredAHT is not warranted, however feasibility work has informed the re-design of a potential trial and informed the need for further feasibility and piloting work before a definitive trial would be viable.

8.4 Theoretical underpinning

The review of the selective theories and models of clinical diagnostic decision-making, alongside the logic underpinning CPRs (Chapter 3) found that the revision of disease probability is the principal clinical strategy underlying diagnostic reasoning,[769] and that the use and application of CPRs is based on probabilistic reasoning,[627] thus CPRs represent an ideal strategy for overcoming cognitive errors in clinical diagnostic decision-making at the hypothesis refinement stage. Nevertheless, the review found that clinicians also use a range of other decision-making strategies when considering a diagnosis of AHT. This review was invaluable for assisting in the interpretation of the qualitative data collected in the three empirical studies. It was clear that clinicians used gut instinct, pattern recognition, causal reasoning, and (implicitly and explicitly) Bayesian reasoning in their decision-making regarding suspected AHT, and that their reasoning was not immune to heuristics and biases. Examples of each of these reasoning processes are detailed in Table 8.2. This finding lends support to the dual process model of clinical reasoning[544] and suggests that both “intuitive” System 1 and “analytical” System 2 processes are engaged during diagnostic decision-making. Crucially, as explained in Chapter 3, not all biases, heuristics or cognitive errors originate from System 1 processing. However, when biases do occur, they can only be corrected by engaging System 2 processing.[594] One strategy employed by System 2 is metacognition, meaning “thinking about thinking”.[626] Metacognition describes the ability to step back from the immediate situation, view the problem in a wider context, remind oneself of previous errors, recognise the limits of human cognition and the existence of cognitive biases, and activate cognitive forcing strategies to counter these biases. Cognitive forcing strategies enable clinicians to consider alternative diagnoses, detect potential flaws in their reasoning, and help them to avoid making decisions purely based on pattern-recognition strategies.[626] One such forcing function is the use of CPRs.[625] Clinicians reported that PredAHT would cause them to “step back” and “stop and think” about a diagnosis of suspected AHT (Table 8.2), thereby providing further evidence that CPRs may help to assist clinicians in overcoming cognitive errors associated with faulty heuristics.

Table 8.2 Clinical reasoning strategies reportedly used by clinicians in the empirical studies presented in this thesis

Explicit Bayesian	Clinician 28 Feasibility study, Chapter 7	"If I start out with somewhere around about 5% and it's a likelihood ratio of six, and that bumps it up to the 30% range, fine. That's going to make a difference."
Implicit Bayesian	Clinician 19 Qualitative study, Chapter 5	"That's what we do in our brains, we put all the information together and spit out the probability based on our experience."
Biases	Clinician 9 Qualitative study, Chapter 5	"Well they shouldn't but if it looks like a really nice family that you couldn't imagine doing anything like that and that shouldn't influence you but it makes you think."
Heuristics	Clinician 57 Feasibility study, Chapter 7	"Well I wouldn't use it because it's such a low probability in my own head that I don't think I'm going to be wrong. Which is probably arrogant on my part."
Causal reasoning	Pathologist 1, Qualitative study, Chapter 5	"First of all I see whether there is any injury and decide what sort of injury it is, whether it's a blunt force injury or sharp force injury etcetera and then the distribution of the injuries on the body, and then relate the distribution that I find with what I know about the literature on different patterns of injury for assault or accident, falls."
Pattern recognition	Clinician 10, Qualitative study, Chapter 5	"I would be expecting or might see multiple focal thin layer subdural haemorrhages in different brain compartments."
Gut instinct	Clinician 15, Feasibility study, Chapter 7	"From my gut instinct, it doesn't feel like it's something that's been purposely done. With just the facts, it does look suspicious."
Metacognition	Clinician 10, Qualitative study, Chapter 5	"If there was a mismatch between my clinical opinion and the risk assessment tool that would cause me to stop and think and seriously consider whether I have gone down a bit of a blind alley with this and whether I need to stop and think again…It would be helpful just to reassure us that we are doing the right thing or maybe to cause us to stop and think actually perhaps we shouldn't walk away from this one."

Chapter 3 also established that the mechanism by which a CPR may improve clinicians' diagnostic accuracy is via alterations to their pre-test probability estimate of AHT above or below their individual so-called test and treatment thresholds.[545, 580, 585] The results from the vignette study reported in Chapter 6[748] and the feasibility study reported in Chapter 7 provided evidence that clinicians do revise their probability of AHT in light of new information. However, in both studies, PredAHT had minimal impact on clinicians' CP actions (with the caveat that in the feasibility study many interviews were conducted in retrospect). In the context of the threshold approach to decision-making, this suggests that the shift in probability from their pre-test probability of AHT was not large enough to cross the test or treatment thresholds required to initiate a change in action.[580] This is supported by the finding from the feasibility study that clinicians' probability thresholds for investigating and reporting suspected AHT are very low. However, in the vignette study, three clinicians changed their CP action after seeing the PredAHT score, but not their probability estimate of AHT, suggesting that these clinicians' decisions may not have been based on probabilistic thresholds.[748, 944]

In the vignette study, although clinicians were not asked to estimate a likelihood ratio of AHT in each of the scenarios, it was possible to calculate clinicians' likelihood ratios using their prior and Time 1 probabilities of AHT. This analysis found that clinicians' likelihood ratios were smaller than those calculated by PredAHT in cases where the clinical evidence was strongly suggestive of AHT, and larger than those calculated by PredAHT in cases where the clinical evidence in support of AHT was weaker. This finding is consistent with the findings from other studies in the clinical-decision making literature, which have found that information with high diagnostic value is generally underestimated while information with low diagnostic value is generally overestimated.[616]

Studies investigating the barriers to clinicians use of and adherence to CPRs have traditionally explored clinicians' knowledge, attitudes and behaviours relating to CPRs,[669] but to the best of the researcher's knowledge, have not considered how clinicians' decision-making strategies may influence their use and acceptance of a CPR.

The research findings of have demonstrated that a theoretical understanding of clinical decision-making processes can help to inform researchers of the reasons why clinicians may or may not use or adhere to a CPR in clinical practice or in particular cases. Other theoretical frameworks are useful for interpreting why CPRs may not be used or adopted in clinical practice. The construct of "complexity" within Diffusion of Innovations theory posits that technological innovations that are perceived by clinicians as easy to use are more likely to be adopted.[999] Participants in the qualitative and feasibility studies felt that the computerised PredAHT would be simple to use and integrate into local hospital systems.[747] Other studies have found that this is especially true if CPRs are automated and implemented into the electronic health record,[320] suggesting that this approach could be considered for future impact studies of PredAHT. The Concerns Based Adopted model in Diffusion of Innovations theory[999] suggests that prospective users must be aware of the innovation and possess sufficient knowledge about how it works and how it should be used, which explains why participants in the qualitative study felt that they needed multidisciplinary training on how to use PredAHT and further information on the data behind the tool and how it had been derived and validated.[747] It may therefore be useful to consider the findings from any future impact studies of PredAHT within the theoretical framework of Diffusion of Innovations theory.[999]

8.5 Frameworks to guide the development and evaluation of a clinical prediction rule

This section critiques the frameworks and approaches used to guide the development and evaluation of PredAHT and the planning and design of the empirical studies presented in this thesis.

8.5.1 The Medical Research Council Framework for the development and evaluation of complex interventions

It is increasingly recognized in the literature that decision support interventions and CPRs should be regarded as complex interventions.[315-321] Introducing a CPR into clinical practice with subsequent management actions consists of multiple interacting components, such that the effects of a CPR on downstream patient outcomes are not just the sum of the successive components, and changes in clinician behaviour will not necessarily lead to improvements in patient outcomes.[315, 317] CPRs may be used in different ways by different clinicians, leading to different outcomes.[320] Furthermore, in the context of the theoretical literature on diffusion of innovations, the adoption of a CPR into clinical practice is influenced by numerous contextual factors including the attributes of the CPR itself as perceived by the intended users, clinicians' skills, system factors and external influences such as local policy and incentive structures.[668, 999] However traditionally, decision support interventions are seldom conceptualised in terms of their complexity.[319, 320] The MRC framework[322, 323] for the development and evaluation of complex interventions was considered a suitable overarching framework to use to guide the work conducted for this thesis. This framework advocates a systematic, phased approach to intervention development and evaluation that enabled the researcher to advance the evidence-base and theoretical underpinning for the computerised PredAHT, thus ensuring it is evidence-based and tailored according to the findings. The approach allowed the researcher to determine the acceptability and potential impact of PredAHT, and, crucially, to establish the need for further development and/or feasibility work prior to proceeding to a definitive effectiveness study. The advantage of such an approach is that it prevents premature commissioning of large-scale, expensive evaluation trials which are not feasible and therefore may fail and waste time, effort, funding and resources.[1001, 1002] This is likely to have been the case if a definitive study of the impact of PredAHT in clinical practice had been attempted as designed without first conducting feasibility work. A recent study of 89 feasibility studies funded by the National Institute for Health Research found that 20 were judged as not feasible.[1002] The study estimated that these 20 feasibility studies potentially saved up to £20m of further research funding for studies that would have been unlikely to be successfully completed.[1002]

Although the MRC guidance is presented as a "discussion document", it is often considered to be authoritative, acting as the "gold standard" for intervention development and evaluation.[1003] However, other frameworks and approaches have been proposed to address the complexities in defining, developing and evaluating complex interventions, including intervention mapping,[1004] the behaviour change wheel,[1005] logic models,[318] normalisation process theory,[1006] the multiphase optimisation strategy (MOST) framework,[1007] and the process modelling in implementation research (PRIME) approach.[1008] There is a lack of studies directly comparing the effectiveness of the MRC framework with alternative frameworks,[1009] leading to confusion about which one to follow.[1010] Logic models make explicit the causal assumptions made by researchers regarding how the intervention will produce its intended effects, using a diagram to describe the structures, activities and proposed outcomes of the intervention.[318] Although Chapter 3 of this thesis provided a critical account of the mechanisms by which a CPR may influence decision-making, in hindsight it may have been beneficial to develop a logic model for use in the feasibility study alongside the MRC framework. This approach could be considered if further feasibility testing is undertaken.

In addition, the MRC guidance has been criticised on several fronts.[1011] Firstly, it has been suggested that the recommended phases of development and evaluation resemble too closely the phases involved in developing commercial pharmacological interventions, for which significant resources are allocated.[1003, 1009, 1011] In contrast, complex public health interventions have been historically underfunded and often do not have the resources required to follow the rigorous, iterative development and evaluation process outlined in the MRC framework, which ultimately concludes in a formal RCT.[1009] Secondly, the focus of the MRC guidance on progressing to a full RCT has been criticised as randomized trials often fail to consider how the intervention components interact with each other and the local context, and fail to ask what works, for whom, and under what circumstances.[1012, 1013] Recent guidance on process evaluations for complex interventions thus recommended that qualitative data are collected alongside both feasibility studies and RCTs, to help provide a detailed understanding of intervention functioning, elucidate causal mechanisms, and determine the contextual factors associated with variation in outcomes.[318] Similarly, researchers have argued that the MRC guidance does not acknowledge the complexity and unpredictability of the organisational systems into which interventions may be introduced and that the guidance regards interventions as discrete packages, rather than events in systems.[1014, 1015] There is an assumption that interventions that follow the phases of the MRC framework will be standardised and reach a point of stability.[1011, 1015] While the framework allows for a "specified degree of adaptation to local settings", this conceptualisation of the role of context is arguably too limited, especially when considering CPRs, whereby adjustment and adaptation of a predictive model to a new setting is strongly recommended in order to maximise its validity and predictive performance.[315, 1011] Complex organisational systems are characterised by contextual variation and context is integral to understanding why interventions may work in some individuals or settings and not in others.[1011, 1012, 1014] This has led some researchers to advocate theory-driven approaches to evaluation such as realist methods.[1013, 1016] A discussion of the potential of such approaches was conspicuously absent from the MRC guidance despite the emphasis placed on the importance of theory throughout the document.[1003]

Despite the limitations of the MRC guidance, adhering to the framework has proven invaluable in determining that although PredAHT is acceptable to CP professionals and has potential value, further development and/or feasibility work must be completed before the impact of PredAHT can be tested in clinical practice. This supports the findings of a recent case study exploring the benefits of using the MRC framework to develop a clinical decision support intervention for pain management in patients with dementia.[319] This study highlighted the importance of considering clinical decision support interventions as complex interventions, as they are implemented in a complex environment involving collaboration between different layers of individual and social units, and individuals may interact with the technology in a number of different ways.[319] In particular, the role of theory and a consideration of *how* an intervention may be thought to work in practice was considered to be especially important in this study.[319]

8.5.2 Frameworks for the development and evaluation of clinical prediction rules

The work conducted in this thesis was also guided by methodological guidelines for the development and evaluation of CPRs,[293, 301] and by phases one and two of the four-phased iterative framework for the impact analysis of CPRs proposed by Wallace and colleagues. [300] The computerised PredAHT was externally validated in the original validation dataset and in an Australian/New Zealand population and thus according to Wallace et al.[300] it is ready for impact analysis. In addition, Reilly & Evans[299] categorize such a rule as a Level 3 CPR and suggest that the predictions can be applied to new patients with confidence in their accuracy. However, Kappen et al.[315] point out that as yet no clear guidelines exist regarding the number of external validations that are required before use in clinical practice, raising the question as to when a model has been successfully "broadly" externally validated. One may take the view that as performance of CPRs is heterogeneous,[1017] the consideration of the validity and impact of a CPR should be setting-dependent and a CPR should be customized to a new setting wherever possible.[315]

Phase one of the framework proposed by Wallace et al.[300] involves verifying the sensibility, comprehensibility and appropriateness of the components and predictive abilities of the CPR. This was addressed in Chapter 4, where key features of the computerised PredAHT were selected that reflected the reality of clinical practice, and the validity of the updated PredAHT was re-assessed. For example, since aspects of the diagnostic process in suspected AHT and the application of CPRs are conducive to Bayesian reasoning, LRs were calculated and incorporated into PredAHT to better reflect the probabilistic nature of diagnosis. In addition, the probability of AHT when one or more features were unknown was calculated following consideration that in reality, clinicians may not have access to or knowledge of all of the predictor variables when applying the CPR in clinical practice. As the CPR had been modified since its original derivation and validation, its validity was re-assessed as recommended in the literature,[300, 326] and a further external validation carried out in an Australian/New Zealand dataset. The second phase of the framework involves defining the delivery mode and study design, assessing the acceptability of the CPR, and assessing the feasibility of the impact study which have all been assessed. However, although Wallace et al.[300] suggested the need for a feasibility study, they did not provide any guidance as to how a feasibility study of the impact of a CPR should be conducted. Despite the movement of researchers and funders towards greater consideration of "exploratory" feasibility and pilot studies following the MRC guidance, there is a lack of evidence and guidance in general on how to design and conduct such studies.[1018] A recent systematic review of guidance for exploratory studies of public health interventions found that existing recommendations are often conflicting, and are inconsistent with regards to the aims, designs, and conduct of exploratory studies.[1018] There is also a lack of guidance regarding the evidence required to inform decisions about when to progress to an effectiveness study.[1018] Similar findings were reported by Levati et al.[1010] in their scoping review of strategies for the "optimisation" of complex interventions prior to an RCT, which concluded that it is unclear how to determine when an intervention is ready to be tested in an RCT.[1010] Moore et al.[1001] contend that this poses a challenge for researchers, peer reviewers and funders in assessing the merits of research proposals and findings, and in determining whether further evaluation of an intervention is justified. To this end, the MRC has recently provided funding for the GUEST study to develop guidance for exploratory studies of complex public health interventions.[1001]

Furthermore, guidance for exploratory studies tends to be limited to studies conducted in preparation for RCTs, and therefore discussion regarding the role of randomisation features heavily.[1018] In the feasibility study reported in Chapter 7, the study findings were assessed against an analytic framework derived from the work of Bugge et al.[978] and Shanyinde et al.[979] which was adapted to exclude items relating to randomisation. Researchers exploring the feasibility of evaluating the impact of a prognostic model for management of blunt chest wall trauma patients have devised predetermined success criteria against which to assess their feasibility findings,[1019] however this was in preparation for a randomised trial. The use of predefined success criteria to assess the findings of the PredAHT feasibility study was felt to be premature at this stage. The PredAHT feasibility study included a substantial qualitative element, following recommendations from the MRC guidance for process evaluation of complex interventions.[318] This yielded detailed information regarding the impact of PredAHT on clinicians' AHT probability estimates and CP actions in each case, suggesting that qualitative research is a valuable methodological approach for evaluating the impact of CPRs in clinical practice. Nevertheless, clear guidelines for the design and conduct of pilot and feasibility studies specifically evaluating CPRs would be a welcome addition to the literature, and the absence of such guidance represents a gap in the current evidence-base.

8.6 Benefits and challenges of the mixed-methods approach used in this thesis

The research findings highlight the value of a mixed-methods approach to evaluating CPRs. Historically, "purist" advocates of quantitative and qualitative approaches have argued that the two methods are inherently incompatible due to differences in their paradigmatic assumptions, and thus should not be used together in a single study.[1020] Paradigmatic assumptions are concerned with how quantitative and qualitative methods can be applied to generate valid knowledge of ourselves and the world around us.[1021] Quantitative methods are rooted in a realist/(post-)positivist paradigm which maintains that scientific inquiry should be objective, and that there exists a stable "real" world that is independent of the observer and their error-prone perceptions. In contrast, qualitative methods are rooted in an interpretive/constructionist paradigm, which values research as a means of producing rich understanding of worlds that are inevitably context-bound, and argues for the existence of multiple constructed realities. Johnson and Onwuegbuzie[1020] observed that the quantitative versus qualitative debate has been so divisive that graduate students are often left feeling that they are obligated to pledge allegiance to one school of thought. Indeed, one qualitative purist has asserted that "accommodation between paradigms is impossible…we are led to vastly diverse, disparate, and totally antithetical ends".[1022(p.81)] However, mixed-methods research recognizes that both quantitative and qualitative methods are important and useful and that the use of mixed methods can harness the strengths and minimise the weaknesses of both approaches.[1020] As described in Chapter 6, the pragmatic research paradigm provides an underlying philosophical basis for mixed methods research. Pragmatists contend that the method of data collection and its underlying paradigmatic assumptions are of less importance than the specific research question to be answered, and that mixed methods research is justified if it provides the most appropriate means to address the research question.[1020] In the convergent mixed methods design used in the vignette study (Chapter 6), qualitative methods helped to confirm and explain the quantitative results, and, importantly, suggested that PredAHT may be useful even if it does not directly influence clinicians' decision-making,[748] something that would not have been discovered if quantitative methods alone were used. Qualitative methods also provided evidence of internal validity in this study, as clinicians were able to verify that their probability estimates of AHT differed due to the factors manipulated in the vignettes. The use of qualitative methods in the feasibility study (Chapter 7) allowed for a fascinating and in-depth exploration of the reasons why PredAHT did or did not influence clinicians' AHT probability estimates and CP actions on a case-by-case basis. This led to an understanding of where the value of PredAHT lies, and the specific circumstances in which it is most likely to be useful.

Despite its considerable benefits, mixed methods research is not without its challenges. Firstly, it requires the researcher to be familiar with and skilled in both quantitative and qualitative techniques.[1023] The researcher attended numerous training courses on qualitative interviewing techniques and qualitative and quantitative data analysis methods, sought appropriate guidance where it was needed, and developed a genuine passion for both quantitative and qualitative approaches throughout the PhD. Secondly, a degree of innovation is required to integrate the two different types of data, make meaningful connections between them, and present the results effectively.[1023] Lastly, mixed methods research can be time consuming.[1023] The researcher conducted 97 interviews for the feasibility study (Chapter 7) and an additional 50 interviews for the qualitative study (Chapter 5). This generated a wealth of interesting data, however this meant that data collection and analysis took much longer than anticipated. At times the qualitative interviews were emotionally challenging due to the sensitive subject area. Future impact studies should consider whether all clinicians using a CPR should be interviewed or whether a cross-section would suffice. However, in a field where CPRs have not yet been introduced, obtaining a range of viewpoints regarding each specific case has been instrumental in understanding the potential impact PredAHT may have in clinical practice.

8.7 Implications for future research and practice

The work presented has suggested a number of directions for future research and practice. In terms of future research directions for PredAHT, the feasibility study was invaluable in determining that a full-scale impact study of PredAHT is not feasible as designed and that a definitive impact analysis of PredAHT is not currently warranted. The principal findings suggested a number of modifications to the design of the study, including the use of different outcome measures, and dedicated research nurses to facilitate timely identification of patients and early recruitment of clinicians. A large-scale trial would also require a longer study period and multiple study sites in order to maximise patient numbers, and active implementation strategies to encourage acceptance and use of PredAHT, including possible integration into the electronic health record. However, further feasibility and pilot work is required in order to evaluate the acceptability and feasibility of the modified study procedures and further assess the appropriateness of the proposed new outcome measures.

It is worth bearing in mind that the PredAHT predicted probabilities were originally designed to be calculated once a full clinical work-up had been performed, in order to "assist frontline professionals when deciding whether to refer a child for specialist clinical and multiagency investigation of possible AHT",[59(p.e558)] "assist clinicians in their discussions with social welfare, law enforcement, or other professionals involved in the child protection process"[60(p.291)] and "assist clinicians offering medical testimony in civil or criminal proceedings, in demonstrating why certain combinations of features are more or less predictive of an abusive etiology".[59(p.e558)] However, with the recognition that at different time points in the clinical assessment pathway clinicians may not have access to or knowledge of all six clinical features included in PredAHT, the probability of AHT was estimated when one or more of the six clinical features are unknown, and therefore the computerised PredAHT can be used at multiple points in the clinical investigation process of suspected AHT, including when a child is first admitted. The results of the feasibility study suggested that PredAHT may be best placed at the beginning of the process, when the "probabilities and possibilities are being critically analysed and acted upon" (Clinician 66, Consultant, PICU). Thus, an appropriate outcome measure for a definitive trial may be whether PredAHT encourages standardisation of the clinical investigation when the probability of AHT is high. The ability to explore the predicted probabilities of AHT when RHs or fractures are found to be present or absent may be a valuable asset of PredAHT in this regard. Clinicians in all three studies presented suggested that PredAHT would provide them with confidence and reassurance in their judgments and decision-making. Therefore, when further information is obtained and a diagnostic impression is formed, it may be appropriate to quantitatively assess whether PredAHT increases clinicians' confidence in their diagnoses at a later stage in the assessment process.

A second option would be to go back to the drawing board, and redesign PredAHT altogether. PredAHT was derived based on a systematic review of the clinical features associated with AHT and nAHT in six studies conducted in the 1990's and 2000's.[24] It is recognized that with changes over time, advances in technology and improvements in diagnostic tests may affect the performance of CPRs.[296] Although the six clinical features included in PredAHT continue to be associated with AHT in the medical literature, recent research has also highlighted the strong association of spinal injuries with AHT,[25, 26] and the specificity of certain patterns of ICI[136] and RH[23] for AHT. Clinicians in the qualitative[747] and feasibility study suggested that PredAHT could be refined to account for specific patterns of the clinical features. However, a large-scale prospective collaborative study would need to be conducted to ensure sufficient ascertainment of cases with each possible combination of features and enable valid statistical analysis of every scenario. Of note, such a study was recommended in the original derivation paper published in 2011,[59] but has yet to be conducted. However, consideration needs to be given to the trade-off between the number of variables included in a predictive model and its simplicity and usability. Alternatively, the use of other statistical techniques such as machine learning algorithms may offer a greater degree of flexibility than traditional explanatory statistical models.[1024] In addition, it may be possible to formalise the feature of the computerised PredAHT that allows clinicians to enter their prior probability estimate of AHT as research into the psychosocial risk factors for AHT progresses.[198]

8.7.1 Recommendations arising from this thesis

The findings of this thesis suggest, as argued by others,[315, 319-321] that CPRs should be viewed as complex interventions, and should be thoroughly evaluated using quantitative and qualitative methods within relevant theoretical frameworks in order to acknowledge and capture their inherent complexity. The field of AHT diagnosis would benefit from large-scale international collaborative studies in order to maximise patient numbers and refine, compare, validate, and assess the impact of CPRs to assist in the identification of AHT. There should be a greater focus on evaluating the impact of CPRs, rather than developing new CPRs for the same condition or patient population. In particular, the findings of this thesis have highlighted the benefits of conducting exploratory and feasibility work prior to conducting a full-scale impact study of the effectiveness of a CPR in clinical practice. Guidelines for conducting feasibility studies of the impact of CPRs would be a welcome addition to the literature. Further feasibility and/or development work is required before the impact of the PredAHT CPR in clinical practice can be viably evaluated.

8.8 Conclusion

This thesis describes the development of a computerised version of the PredAHT CPR and explored its clinical utility for assisting in the identification of AHT in three novel empirical studies. The three studies were conducted based on the MRC framework for the development and evaluation of complex interventions[322, 323] and a four-phased iterative framework for the impact analysis of CPRs.[300] The results show that the computerised PredAHT is acceptable to a range of CP professionals, and has the potential to standardise the clinical investigation of AHT and provide clinicians with confidence and reassurance in their diagnostic decisions. Further feasibility and/or development work is recommended before the impact of PredAHT can be tested in a clinical trial. The studies presented in this thesis make important contributions to knowledge in the field of AHT diagnosis.

Hark ye, good parents, to my words true and plain,
When you are shaking your baby, you could be bruising his brain.
So, save the limbs, the brain, even the life of your tot;
By shaking him never; never and not. John Caffey, 1972[42]

9 References

1. Chiesa A & Duhaime A-C (2009). Abusive head trauma. *Pediatric Clinics of North America,* 56(2): 317-331.
2. Adamsbaum C, Grabar S, Mejean N & Rey-Salmon C (2010). Abusive head trauma: Judicial admissions highlight violent and repetitive shaking. *Pediatrics,* 126(3): 546-555.
3. Matschke J, Herrmann B, Sperhake J, Korber F, Bajanowski T & Glatzel M (2009). Shaken baby syndrome: A common variant of non-accidental head injury in infants. *Deutsches Arzteblatt International,* 106(13): 211-217.
4. Duhaime AC, Christian CW, Rorke LB & Zimmerman RA (1998). Nonaccidental head injury in infants--the "shaken-baby syndrome". *New England Journal of Medicine,* 338(25): 1822-1829.
5. Billmire ME & Myers PA (1985). Serious head injury in infants: Accident or abuse? *Pediatrics,* 75(2): 340-342.
6. Gill JR, Goldfeder LB, Armbrustmacher V, Coleman A, Mena H & Hirsch CS (2009). Fatal head injury in children younger than 2 years in New York City and an overview of the shaken baby syndrome. *Archives of Pathology & Laboratory Medicine,* 133(4): 619-627.
7. Bruce DA & Zimmerman RA (1989). Shaken impact syndrome. *Pediatric Annals,* 18(8): 482-494.
8. Bechtel K, Stoessel K, Leventhal JM, Ogle E, Teague B, Lavietes S, Banyas B, et al. (2004). Characteristics that distinguish accidental from abusive injury in hospitalized young children with head trauma. *Pediatrics,* 114(1): 165-168.
9. Hettler J & Greenes DS (2003). Can the initial history predict whether a child with a head injury has been abused? *Pediatrics,* 111(3): 602-607.
10. Reece RM & Sege R (2000). Childhood head injuries: Accidental or inflicted? *Archives of Pediatric & Adolescent Medicine,* 154(1): 11-15.
11. Keenan HT, Runyan DK, Marshall SW, Nocera MA, Merten DF & Sinal SH (2003). A population-based study of inflicted traumatic brain injury in young children. *JAMA,* 290(5): 621-626.
12. King WJ, MacKay M, Sirnick A & with the Canadian Shaken Baby Study Group (2003). Shaken baby syndrome in Canada: Clinical characteristics and outcomes of hospital cases. *Canadian Medical Association Journal,* 168(2): 155-159.
13. Duhaime AC, Gennarelli TA, Thibault LE, Bruce DA, Margulies SS & Wiser R (1987). The shaken baby syndrome. A clinical, pathological, and biomechanical study. *Journal of Neurosurgery,* 66(3): 409-415.
14. Ludwig S & Warman M (1984). Shaken baby syndrome: A review of 20 cases. *Annals of Emergency Medicine,* 13(2): 104-107.
15. Hadley MN, Sonntag VKH, Rekate HL & Murphy A (1989). The infant whiplash-shake injury syndrome: A clinical and pathological study. *Neurosurgery,* 24(4): 536-540.
16. Ettaro L, Berger RP & Songer T (2004). Abusive head trauma in young children: Characteristics and medical charges in a hospitalized population. *Child Abuse & Neglect,* 28(10): 1099-1111.
17. Jayawant S, Rawlinson A, Gibbon F, Price J, Schulte J, Sharples P, Sibert JR, et al. (1998). Subdural haemorrhages in infants: Population based study. *The BMJ,* 317(7172): 1558-1561.
18. Niederkrotenthaler T, Xu L, Parks SE & Sugerman DE (2013). Descriptive factors of abusive head trauma in young children--United States, 2000-2009. *Child Abuse & Neglect,* 37(7): 446-455.
19. Goldstein B, Kelly MM, Bruton D & Cox C (1993). Inflicted versus accidental head injury in critically injured children. *Critical Care Medicine,* 21(9): 1328-1332.

20. Makaroff KL & Putnam FW (2003). Outcomes of infants and children with inflicted traumatic brain injury. *Developmental Medicine & Child Neurology,* 45(7): 497-502.
21. Lind K, Toure H, Brugel D, Meyer P, Laurent-Vannier A & Chevignard M (2016). Extended follow-up of neurological, cognitive, behavioral and academic outcomes after severe abusive head trauma. *Child Abuse & Neglect,* 51: 358-367.
22. Matschke J, Voss J, Obi N, Görndt J, Sperhake J-P, Püschel K & Glatzel M (2009). Nonaccidental head injury is the most common cause of subdural bleeding in infants <1 year of age. *Pediatrics,* 124(6): 1587-1594.
23. Maguire SA, Watts PO, Shaw AD, Holden S, Taylor RH, Watkins WJ, Mann MK, et al. (2013). Retinal haemorrhages and related findings in abusive and non-abusive head trauma: A systematic review. *Eye,* 27(1): 28-36.
24. Maguire S, Pickerd N, Farewell D, Mann M, Tempest V & Kemp AM (2009). Which clinical features distinguish inflicted from non-inflicted brain injury? A systematic review. *Archives of Disease in Childhood,* 94(11): 860-867.
25. Kemp A, Cowley L & Maguire S (2014). Spinal injuries in abusive head trauma: Patterns and recommendations. *Pediatric Radiology,* 44(Suppl 4): S604-S612.
26. Choudhary AK, Ishak R, Zacharia TT & Dias MS (2014). Imaging of spinal injury in abusive head trauma: A retrospective study. *Pediatric Radiology,* 44(9): 1130-1140.
27. Piteau SJ, Ward MGK, Barrowman NJ & Plint AC (2012). Clinical and radiographic characteristics associated with abusive and nonabusive head trauma: A systematic review. *Pediatrics,* 130(2): 315-323.
28. Maguire SA, Upadhyaya M, Evans A, Mann MK, Haroon MM, Tempest V, Lumb RC, et al. (2013). A systematic review of abusive visceral injuries in childhood—Their range and recognition. *Child Abuse & Neglect,* 37(7): 430-445.
29. Sheybani EF, Gonzalez-Araiza G, Kousari YM, Hulett RL & Menias CO (2014). Pediatric nonaccidental abdominal trauma: What the radiologist should know. *RadioGraphics,* 34(1): 139-153.
30. Keenan HT, Runyan DK, Marshall SW, Nocera MA & Merten DF (2004). A population-based comparison of clinical and outcome characteristics of young children with serious inflicted and noninflicted traumatic brain injury. *Pediatrics,* 114(3): 633-639.
31. Jenny C, Hymel KP, Ritzen A, Reinert SE & Hay TC (1999). Analysis of missed cases of abusive head trauma. *JAMA,* 281(7): 621-626.
32. Roche AJ, Fortin G, Labbe J, Brown J & Chadwick D (2005). The work of Ambroise Tardieu: The first definitive description of child abuse. *Child Abuse & Neglect,* 29(4): 325-334.
33. Labbé J (2005). Ambroise Tardieu: The man and his work on child maltreatment a century before Kempe. *Child Abuse & Neglect,* 29(4): 311-324.
34. Tardieu A (1860). Étude médico-légale sur les sévices et mauvais traitements exercés sur des enfants. *Annales d'Hygiène Publique et de Medecine Légale,* 13: 361-398.
35. Tardieu A (1868). *Étude médico-légale sur l'infanticide.* Paris: Librairie JB Baillière et Fils.
36. Al-Holou WN, O'Hara EA, Cohen-Gadol AA & Maher CO (2009). Nonaccidental head injury in children. Historical vignette. *Journal of Neurosurgery Pediatrics,* 3(6): 474-483.
37. Caffey J (1946). Multiple fractures in the long bones of infants suffering from chronic subdural hematoma. *American Journal of Roentgenology & Radium Therapy,* 56(2): 163-173.
38. Caffey J (2011). The classic: Multiple fractures in the long bones of infants suffering from chronic subdural hematoma. *Clinical Orthopaedics & Related Research,* 469(3): 755-758.
39. Kempe CH, Silverman FN, Steele BF, Droegemueller W & Silver HK (1962). The battered-child syndrome. *JAMA,* 181(1): 17-24.

40. Caffey J (1972). The parent-infant traumatic stress syndrome; (Caffey-Kempe syndrome), (battered babe syndrome). *American Journal of Roentgenology, Radium Therapy & Nuclear Medicine,* 114(2): 218-229.
41. Guthkelch AN (1971). Infantile subdural haematoma and its relationship to whiplash injuries. *The BMJ,* 2(5759): 430-431.
42. Caffey J (1972). On the theory and practice of shaking infants. Its potential residual effects of permanent brain damage and mental retardation. *American Journal of Diseases of Children,* 124(2): 161-169.
43. Caffey J (1974). The whiplash shaken infant syndrome: Manual shaking by the extremities with whiplash-induced intracranial and intraocular bleedings, linked with residual permanent brain damage and mental retardation. *Pediatrics,* 54(4): 396-403.
44. American Academy of Pediatrics: Committee on Child Abuse and Neglect (1993). Shaken baby syndrome: Inflicted cerebral trauma. *Pediatrics,* 92(6): 872-875.
45. American Academy of Pediatrics: Committee on Child Abuse and Neglect (2001). Shaken baby syndrome: Rotational cranial injuries—technical report. *Pediatrics,* 108(1): 206-210.
46. Alexander R, Sato Y, Smith W & Bennett T (1990). Incidence of impact trauma with cranial injuries ascribed to shaking. *American Journal of Diseases of Children,* 144(6): 724-726.
47. Hahn YS, Raimondi AJ, McLone DG & Yamanouchi Y (1983). Traumatic mechanisms of head injury in child abuse. *Child's Brain,* 10(4): 229-241.
48. Gilliland M & Folberg R (1996). Shaken babies--some have no impact injuries. *Journal of Forensic Sciences* 41(1): 114-116.
49. Narang SK, Estrada C, Greenberg S & Lindberg D (2016). Acceptance of shaken baby syndrome and abusive head trauma as medical diagnoses. *The Journal of Pediatrics,* 177: 273-278.
50. Christian CW, Block R, Committee on Child Abuse & Neglect & American Academy of Pediatrics (2009). Abusive head trauma in infants and children. *Pediatrics,* 123(5): 1409-1411.
51. Choudhary AK, Servaes S, Slovis TL, Palusci VJ, Hedlund GL, Narang SK, Moreno JA, et al. (2018). Consensus statement on abusive head trauma in infants and young children. *Pediatric Radiology,* 48(8): 1048-1065.
52. Strouse PJ (2018). Shaken baby syndrome is real. *Pediatric Radiology,* 48(8): 1043-1047.
53. Narang S (2011). A daubert analysis of abusive head trauma/shaken baby syndrome. *Houston Journal of Health, Law & Policy,* 11: 505-636.
54. American Academy of Pediatrics (2015). *Understanding abusive head trauma in infants and children: Answers from America's pediatricians.* Available from: https://www.aap.org/en-us/Documents/cocan_understanding_aht_in_infants_children.pdf. Accessed 18 September 2018.
55. Greeley CS (2014). "Shaken baby syndrome" and forensic pathology. *Forensic Science, Medicine & Pathology,* 10(2): 253-255.
56. Narang SK & Greeley CS (2017). Lynøe et al. – #theRestoftheStory. *Acta Paediatrica,* 106(7): 1047-1049.
57. Crown Prosecution Service (2011). *Non accidental head injury cases (NAHI, formerly referred to as shaken baby syndrome [SBS]): Prosecution approach*. Available from: www.cps.gov.uk/legal/l_to_o/non_accidental_head_injury_cases/#a01. Accessed 4 November 2014.
58. Parks SE, Annest JL, Hill HA & Karch. D.L (2012). *Pediatric abusive head trauma: Recommended definitions for public health surveillance and research*. Atlanta, GA:

Centers for Disease Control and Prevention. Available from: https://www.cdc.gov/violenceprevention/pdf/pedheadtrauma-a.pdf. Accessed 1 October 2014.
59. Maguire SA, Kemp AM, Lumb RC & Farewell DM (2011). Estimating the probability of abusive head trauma: A pooled analysis. *Pediatrics,* 128(3): e550-e564.
60. Cowley LE, Morris CB, Maguire SA, Farewell DM & Kemp AM (2015). Validation of a clinical prediction tool for abusive head trauma. *Pediatrics,* 136(2): 291-298.
61. Christian C (2017). Child abuse: Epidemiology, mechanisms, and types of abusive head trauma in infants and children. *In:* Wiley JF (Ed.) *UpToDate*. Available from: https://www.uptodate.com/contents/child-abuse-epidemiology-mechanisms-and-types-of-abusive-head-trauma-in-infants-and-children. Accessed 16 July 2018.
62. Kelly P, John S, Vincent AL & Reed P (2015). Abusive head trauma and accidental head injury: A 20-year comparative study of referrals to a hospital child protection team. *Archives of Disease in Childhood,* 100(12): 1123-1130.
63. Deye KP, Berger RP & Lindberg DM (2013). Occult abusive injuries in infants with apparently isolated skull fractures. *Journal of Trauma & Acute Care Surgery,* 74(6): 1553-1558.
64. Amagasa S, Matsui H, Tsuji S, Uematsu S, Moriya T & Kinoshita K (2018). Characteristics distinguishing abusive head trauma from accidental head trauma in infants with traumatic intracranial hemorrhage in Japan. *Acute Medicine & Surgery,* 5(3): 265-271.
65. Ibrahim NG, Wood J, Margulies SS & Christian CW (2012). Influence of age and fall type on head injuries in infants and toddlers. *International Journal of Developmental Neuroscience,* 30(3): 201-206.
66. Sieswerda-Hoogendoorn T, Robben SGF, Karst WA, Moesker FM, van Aalderen WM, Lameris JS & van Rijn RR (2014). Abusive head trauma: Differentiation between impact and non-impact cases based on neuroimaging findings and skeletal surveys. *European Journal of Radiology,* 83(3): 584-588.
67. Case ME, Graham MA, Handy TC, Jentzen JM, Monteleone JA & National Association of Medical Examiners Ad Hoc Committee on Shaken Baby Syndrome (2001). Position paper on fatal abusive head injuries in infants and young children. *The American Journal of Forensic Medicine & Pathology,* 22(2): 112-122.
68. Case ME (2007). Abusive head injuries in infants and young children. *Legal Medicine,* 9(2): 83-87.
69. Case ME (2014). Distinguishing accidental from inflicted head trauma at autopsy. *Pediatric Radiology,* 44 (Suppl 4): S632-S640.
70. Duhaime AC, Alario AJ, Lewander WJ, Schut L, Sutton LN, Seidl TS, Nudelman S, et al. (1992). Head injury in very young children: Mechanisms, injury types, and ophthalmologic findings in 100 hospitalized patients younger than 2 years of age. *Pediatrics,* 90(2 Pt 1): 179-185.
71. Raghupathi R, Mehr MF, Helfaer MA & Margulies SS (2004). Traumatic axonal injury is exacerbated following repetitive closed head injury in the neonatal pig. *Journal of Neurotrauma,* 21(3): 307-316.
72. Raghupathi R & Margulies SS (2002). Traumatic axonal injury after closed head injury in the neonatal pig. *Journal of Neurotrauma,* 19(7): 843-853.
73. Adamsbaum C & Rambaud C (2012). Abusive head trauma: Don't overlook bridging vein thrombosis. *Pediatric Radiology,* 42(11): 1298-300.
74. Rambaud C (2015). Bridging veins and autopsy findings in abusive head trauma. *Pediatric Radiology,* 45(8): 1126-1131.
75. Maxeiner H (2001). Demonstration and interpretation of bridging vein ruptures in cases of infantile subdural bleedings. *Journal of Forensic Sciences* 46(1): 85-93.

76. Finnie JW, Blumbergs PC, Manavis J, Turner RJ, Helps S, Vink R, Byard RW, et al. (2012). Neuropathological changes in a lamb model of non-accidental head injury (the shaken baby syndrome). *Journal of Clinical Neuroscience,* 19(8): 1159-1164.
77. Ommaya AK & Gennarelli TA (1974). Cerebral concussion and traumatic unconciousness. *Brain,* 97: 633-654.
78. Gennarelli TA & Thibault LE (1985). Biomechanics of head injury. *In:* Wilkins RH & Rengachary SS (Eds.) *Neurosurgery* (1531-1536). New York: McGraw Hill.
79. Margulies SS & Thibault LE (1989). An analytical model of traumatic diffuse brain injury. *Journal of Biomechanical Engineering,* 111(3): 241-249.
80. Hanigan WC, Peterson RA & Njus G (1987). Tin ear syndrome: Rotational acceleration in pediatric head injuries. *Pediatrics,* 80(5): 618-622.
81. Steinbok P, Singhal A, Poskitt K & Cochrane DD (2007). Early hypodensity on computed tomographic scan of the brain in an accidental pediatric head injury. *Neurosurgery,* 60(4): 689-695.
82. Wright JN (2017). CNS injuries in abusive head trauma. *American Journal of Roentgenology,* 208(5): 991-1001.
83. Vavilala MS & Tasker RC (2018). Severe traumatic brain injury in children: Initial evaluation and management. *In:* Wiley JF (Ed.) *UpToDate*. Available from: https://www.uptodate.com/contents/severe-traumatic-brain-injury-in-children-initial-evaluation-and-management. Accessed 25 September 2018.
84. Ruppel RA, Kochanek PM, Adelson PD, Rose ME, Wisniewski SR, Bell MJ, Clark RS, et al. (2001). Excitatory amino acid concentrations in ventricular cerebrospinal fluid after severe traumatic brain injury in infants and children: The role of child abuse. *Journal of Pediatrics,* 138(1): 18-25.
85. Berger RP, Adelson PD, Richichi R & Kochanek PM (2006). Serum biomarkers after traumatic and hypoxemic brain injuries: Insight into the biochemical response of the pediatric brain to inflicted brain injury. *Developmental Neuroscience,* 28(4-5): 327-335.
86. Ichord RN, Naim M, Pollock AN, Nance ML, Margulies SS & Christian CW (2007). Hypoxic-ischemic injury complicates inflicted and accidental traumatic brain injury in young children: The role of diffusion-weighted imaging. *Journal of Neurotrauma,* 24(1): 106-118.
87. Orru' E, Huisman TAGM & Izbudak I (2018). Prevalence, patterns, and clinical relevance of hypoxic-ischemic injuries in children exposed to abusive head trauma. *Journal of Neuroimaging,* 28(6): 608-614.
88. Geddes JF, Hackshaw AK, Vowles GH, Nickols CD & Whitwell HL (2001). Neuropathology of inflicted head injury in children. I. Patterns of brain damage. *Brain,* 124(Pt 7): 1290-1298.
89. Geddes JF, Vowles GH, Hackshaw AK, Nickols CD, Scott IS & Whitwell HL (2001). Neuropathology of inflicted head injury in children. II. Microscopic brain injury in infants. *Brain,* 124(Pt 7): 1299-1306.
90. Shannon P, Smith CR, Deck J, Ang LC, Ho M & Becker L (1998). Axonal injury and the neuropathology of shaken baby syndrome. *Acta Neuropathologica,* 95(6): 625-631.
91. Kemp AM, Stoodley N, Cobley C, Coles L & Kemp KW (2003). Apnoea and brain swelling in non-accidental head injury. *Archives of Disease in Childhood,* 88(6): 472-476.
92. Johnson DL, Boal D & Baule R (1995). Role of apnea in nonaccidental head injury. *Pediatric Neurosurgery,* 23(6): 305-310.
93. Matschke J, Buttner A, Bergmann M, Hagel C, Puschel K & Glatzel M (2015). Encephalopathy and death in infants with abusive head trauma is due to hypoxic-ischemic injury following local brain trauma to vital brainstem centers. *International Journal of Legal Medicine,* 129(1): 105-114.

94. Kadom N, Khademian Z, Vezina G, Shalaby-Rana E, Rice A & Hinds T (2014). Usefulness of MRI detection of cervical spine and brain injuries in the evaluation of abusive head trauma. *Pediatric Radiology,* 44(7): 839-848.
95. Jacob R, Cox M, Koral K, Greenwell C, Xi Y, Vinson L, Reeder K, et al. (2016). MR imaging of the cervical spine in nonaccidental trauma: A tertiary institution experience. *American Journal of Neuroradiology,* 37(10): 1944-1950.
96. Baerg J, Thirumoorthi A, Hazboun R, Vannix R, Krafft P & Zouros A (2017). Cervical spine injuries in young children: Pattern and outcomes in accidental versus inflicted trauma. *Journal of Surgical Research,* 219(Supplement C): 366-373.
97. Brennan LK, Rubin D, Christian CW, Duhaime A-C, Mirchandani HG & Rorke-Adams LB (2009). Neck injuries in young pediatric homicide victims. *Journal of Neurosurgery Pediatrics,* 3(3): 232-239.
98. Hymel KP, Makoroff KL, Laskey AL, Conaway MR & Blackman JA (2007). Mechanisms, clinical presentations, injuries, and outcomes from inflicted versus noninflicted head trauma during infancy: Results of a prospective, multicentered, comparative study. *Pediatrics,* 119(5): 922-929.
99. Greenwald MJ, Weiss A, Oesterle CS & Friendly DS (1986). Traumatic retinoschisis in battered babies. *Ophthalmology,* 93(5): 618-625.
100. Nadarasa J, Deck C, Meyer F, Willinger R & Raul J-S (2014). Update on injury mechanisms in abusive head trauma--shaken baby syndrome. *Pediatric Radiology,* 44(Suppl 4): S565-S570.
101. Rangarajan N, Kamalakkannan SB, Hasija V, Shams T, Jenny C, Serbanescu I, Ho J, et al. (2009). Finite element model of ocular injury in abusive head trauma. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus,* 13(4): 364-369.
102. Levin AV (2010). Retinal hemorrhage in abusive head trauma. *Pediatrics,* 126(5): 961-970.
103. Ghatan S & Ellenbogen RG (2002). Pediatric spine and spinal cord injury after inflicted trauma. *Neurosurgery Clinics of North America,* 13(2): 227-233.
104. Gilles FH, Bina M & Sotrel A (1979). Infantile atlantooccipital instability: The potential danger of extreme extension. *American Journal of Diseases of Children,* 133(1): 30-37.
105. Leonard J (2018). Spinal cord injury without radiographic abnormality (SCIWORA) in children. *In:* Wiley JF (Ed.) *UpToDate*. Available from: https://www.uptodate.com/contents/spinal-cord-injury-without-radiographic-abnormality-sciwora-in-children. Accessed 25 September 2018.
106. Choudhary AK, Bradford RK, Dias MS, Moore GJ & Boal DKB (2012). Spinal subdural hemorrhage in abusive head trauma: A retrospective study. *Radiology,* 262(1): 216-223.
107. Feldman KW, Weinberger E, Milstein JM & Fligner CL (1997). Cervical spine MRI in abused infants. *Child Abuse & Neglect,* 21(2): 199-205.
108. Rutty GN, Squier WMV & Padfield CJH (2005). Epidural haemorrhage of the cervical spinal cord: A post-mortem artefact? *Neuropathology & Applied Neurobiology,* 31(3): 247-257.
109. Koumellis P, McConachie NS & Jaspan T (2009). Spinal subdural haematomas in children with non-accidental head injury. *Archives of Disease in Childhood,* 94(3): 216-219.
110. Gruber TJ & Rozzelle CJ (2008). Thoracolumbar spine subdural hematoma as a result of nonaccidental trauma in a 4-month-old infant. *Journal of Neurosurgery Pediatrics,* 2(2): 139-142.
111. Geddes JF & Plunkett J (2004). The evidence base for shaken baby syndrome. *The BMJ,* 328(7442): 719-720.

112. Squier W (2008). Shaken baby syndrome: The quest for evidence. *Developmental Medicine & Child Neurology,* 50(1): 10-14.
113. Squier W (2011). The triad of retinal haemorrhage, subdural haemorrhage and encephalopathy in an infant unassociated with evidence of physical injury is not the result of shaking, but is most likely to have been caused by a natural disease: Yes. *Journal of Primary Health Care,* 3(2): 159-161.
114. Moreno JA & Holmgren B (2016). The Supreme Court screws up the science: There is no abusive head trauma/shaken baby syndrome 'scientific' controversy. *Utah Law Review,* 5: 1357-1435.
115. Moreno JA & Holmgren B (2013). Dissent into confusion: The Supreme Court, denialism, and the false 'scientific' controversy over shaken baby syndrome. *Utah Law Review,* 1: 153-217.
116. R v. Harris & others (2005). EWCA Crim 1980. Available from: http://www.bailii.org/ew/cases/EWCA/Crim/2005/1980.html. Accessed 12 June 2016.
117. Cory CZ & Jones BMD (2003). Can shaking alone cause fatal brain injury? A biomechanical assessment of the Duhaime shaken baby syndrome model. *Medicine, Science & the Law,* 43(4): 317-333.
118. Prange MT, Coats B, Duhaime A-C & Margulies SS (2003). Anthropomorphic simulations of falls, shakes, and inflicted impacts in infants. *Journal of Neurosurgery,* 99(1): 143-150.
119. Pierce MC & Bertocci GE (2008). Injury biomechanics and child abuse. *Annual Review of Biomedical Engineering,* 10: 85-106.
120. Dias M, S. (2010). The case for shaking. *In:* Jenny C (Ed.) *Child abuse and neglect: Diagnosis, treatment, and evidence* (364-372). Philadelphia: Elsevier Saunders.
121. Berger RP, Adelson PD, Pierce MC, Dulani T, Cassidy LD & Kochanek PM (2005). Serum neuron-specific enolase, S100B, and myelin basic protein concentrations after inflicted and noninflicted traumatic brain injury in children. *Journal of Neurosurgery: Pediatrics,* 103(1): 61-68.
122. Berger RP, Pierce MC, Wisniewski SR, Adelson PD, Clark RSB, Ruppel RA & Kochanek PM (2002). Neuron-specific enolase and S100B in cerebrospinal fluid after severe traumatic brain injury in infants and children. *Pediatrics,* 109(2): e31.
123. Berger RP, Kochanek PM & Pierce MC (2004). Biochemical markers of brain injury: Could they be used as diagnostic adjuncts in cases of inflicted traumatic brain injury? *Child Abuse & Neglect,* 28(7): 739-754.
124. Ruppel RA, Clark RSB, Bayir H, Satchell MA & Kochanek PM (2002). Critical mechanisms of secondary damage after inflicted head injury in infants and children. *Neurosurgery Clinics of North America,* 13(2): 169-182.
125. Kochanek PM (2006). Pediatric traumatic brain injury: Quo vadis? *Developmental Neuroscience,* 28(4-5): 244-255.
126. Kochanek PM, Clark RSB, Ruppel RA, Adelson PD, Bell MJ, Whalen MJ, Robertson CL, et al. (2000). Biochemical, cellular, and molecular mechanisms in the evolution of secondary damage after severe traumatic brain injury in infants and children: Lessons learned from the bedside. *Pediatric Critical Care Medicine,* 1(1): 4-19.
127. Coats B & Margulies SS (2008). Potential for head injuries in infants from low-height falls. *Journal of Neurosurgery Pediatrics,* 2(5): 321-330.
128. Morison CN (2002). The dynamics of shaken baby syndrome. PhD thesis. University of Birmingham, UK.
129. Roth S, Raul J-S, Ludes B & Willinger R (2007). Finite element analysis of impact and shaking inflicted to a child. *International Journal of Legal Medicine,* 121(3): 223-228.

130. Margulies S & Coats B (2010). Biomechanics of head trauma in infants and young children. *In:* Jenny C (Ed.) *Child abuse and neglect: Diagnosis, treatment, and evidence* (359-363). Philadelphia: Elsevier Saunders.
131. Richards PG, Bertocci GE, Bonshek RE, Giangrande PL, Gregson RM, Jaspan T, Jenny C, et al. (2006). Shaken baby syndrome. *Archives of Disease in Childhood,* 91(3): 205-206.
132. Harding B, Risdon RA & Krous HF (2004). Shaken baby syndrome. *The BMJ,* 328(7442): 720-721.
133. Adams G, Ainsworth J, Butler L, Bonshek R, Clarke M, Doran R, Dutton G, et al. (2004). Update from the ophthalmology child abuse working party: Royal College ophthalmologists. *Eye,* 18(8): 795-798.
134. Hymel KP, Rumack CM, Hay TC, Strain JD & Jenny C (1997). Comparison of intracranial computed tomographic (CT) findings in pediatric abusive and accidental head trauma. *Pediatric Radiology,* 27(9): 743-747.
135. Ewing-Cobbs L, Kramer L, Prasad M, Canales DN, Louis PT, Fletcher JM, Vollero H, et al. (1998). Neuroimaging, physical, and developmental findings after inflicted and noninflicted traumatic brain injury in young children. *Pediatrics,* 102(2 Pt 1): 300-307.
136. Kemp AM, Jaspan T, Griffiths J, Stoodley N, Mann MK, Tempest V & Maguire SA (2011). Neuroimaging: What neuroradiological features distinguish abusive from non-abusive head trauma? A systematic review. *Archives of Disease in Childhood,* 96(12): 1103-1112.
137. Arbogast KB, Margulies SS & Christian CW (2005). Initial neurologic presentation in young children sustaining inflicted and unintentional fatal head injuries. *Pediatrics,* 116(1): 180-184.
138. Starling SP, Patel S, Burke BL, Sirotnak AP, Stronks S & Rosquist P (2004). Analysis of perpetrator admissions to inflicted traumatic brain injury in children. *Archives of Pediatrics & Adolescent Medicine,* 158(5): 454-458.
139. Biron D & Shelton D (2005). Perpetrator accounts in infant abusive head trauma brought about by a shaking event. *Child Abuse & Neglect,* 29(12): 1347-1358.
140. Bell E, Shouldice M & Levin AV (2011). Abusive head trauma: A perpetrator confesses. *Child Abuse & Neglect,* 35(1): 74-77.
141. Feld K, Banaschak S, Remschmidt H & Rothschild MA (2018). Shaken baby syndrome—what convicted perpetrators report. *Rechtsmedizin*, 28(6): 514-517.
142. Leestma JE (2005). Case analysis of brain-injured admittedly shaken infants: 54 cases, 1969-2001. *American Journal of Forensic Medicine & Pathology,* 26(3): 199-212.
143. Vinchon M, Defoort-Dhellemmes S, Desurmont M & Delestret I (2010). Confessed abuse versus witnessed accidents in infants: Comparison of clinical, radiological, and ophthalmological data in corroborated cases. *Child's Nervous System,* 26(5): 637-645.
144. De Leeuw M, Beuls E, Parizel P, Jorens P & Jacobs W (2013). Confessed abusive blunt head trauma. *American Journal of Forensic Medicine & Pathology,* 34(2): 130-132.
145. Frasier LD, Kelly P, Al-Eissa M & Otterman GJ (2014). International issues in abusive head trauma. *Pediatric Radiology,* 44(Suppl 4): S647-S653.
146. Greeley C (2015). Abusive head trauma: A review of the evidence base. *American Journal of Roentgenology,* 204(5): 967-973.
147. Hobbs C, Childs AM, Wynne J, Livingston J & Seal A (2005). Subdural haematoma and effusion in infancy: An epidemiological study. *Archives of Disease in Childhood,* 90(9): 952-955.
148. Kelly P & Farrant B (2008). Shaken baby syndrome in New Zealand, 2000-2002. *Journal of Paediatrics & Child Health,* 44(3): 99-107.
149. Kesler H, Dias MS, Shaffer M, Rottmund C, Cappos K & Thomas NJ (2008). Demographics of abusive head trauma in the Commonwealth of Pennsylvania. *Journal of Neurosurgery: Pediatrics,* 1(5): 351-356.

150. Parks S, Sugerman D, Xu L & Coronado V (2012). Characteristics of non-fatal abusive head trauma among children in the USA, 2003-2008: Application of the CDC operational case definition to national hospital inpatient data. *Injury Prevention,* 18(6): 392-398.
151. Kaltner M, Kenardy J, Le Brocque R & Page A (2013). Infant abusive head trauma incidence in Queensland, Australia. *Injury Prevention,* 19(2): 139-142.
152. Parrish J, Baldwin-Johnson C, Volz M & Goldsmith Y (2013). Abusive head trauma among children in Alaska: A population-based assessment. *International Journal of Circumpolar Health,* 72: 21216.
153. Wirtz SJ & Trent RB (2008). Passive surveillance of shaken baby syndrome using hospital inpatient data. *American Journal of Preventive Medicine,* 34(4 Suppl): S134-S139.
154. Minns RA, Jones PA & Mok JY (2008). Incidence and demography of nonaccidental head injury in Southeast Scotland from a national database. *American Journal of Preventive Medicine,* 34(4S): S126-S133.
155. Barlow KM & Minns RA (2000). Annual incidence of shaken impact syndrome in young children. *The Lancet,* 356(9241): 1571-1572.
156. Ellingson KD, Leventhal JM & Weiss HB (2008). Using hospital discharge data to track inflicted traumatic brain injury. *American Journal of Preventive Medicine,* 34(4 Suppl): S157-S162.
157. Bennett S, Ward M, Moreau K, Fortin G, King J, Mackay M & Plint A (2011). Head injury secondary to suspected child maltreatment: Results of a prospective Canadian national surveillance program. *Child Abuse & Neglect,* 35(11): 930-936.
158. Fujiwara T, Barr RG, Brant RF, Rajabali F & Pike I (2012). Using International Classification of Diseases, 10th edition, codes to estimate abusive head trauma in children. *American Journal of Preventive Medicine,* 43(2): 215-220.
159. Shanahan ME, Zolotor AJ, Parrish JW, Barr RG & Runyan DK (2013). National, regional, and state abusive head trauma: Application of the CDC algorithm. *Pediatrics,* 132(5): 1546-1553.
160. Selassie AW, Borg K, Busch C & Russell WS (2013). Abusive head trauma in young children: A population-based study. *Pediatric Emergency Care* 29(3): 283-291.
161. Talvik I, Metsvaht T, Leito K, Põder H, Kool P, Väli M, Lintrop M, et al. (2006). Inflicted traumatic brain injury (ITBI) or shaken baby syndrome (SBS) in Estonia. *Acta Paediatrica,* 95(7): 799-804.
162. Dias MS, Rottmund CM, Cappos KM, Reed ME, Wang M, Stetter C, Shaffer ML, et al. (2017). Association of a postnatal parent education program for abusive head trauma with subsequent pediatric abusive head trauma hospitalization rates. *JAMA Pediatrics,* 171(3): 223-229.
163. Barr RG, Barr M, Rajabali F, Humphreys C, Pike I, Brant R, Hlady J, et al. (2018). Eight-year outcome of implementation of abusive head trauma prevention. *Child Abuse & Neglect,* 84: 106-114.
164. Zolotor AJ, Runyan DK, Shanahan M, Durrance CP, Nocera M, Sullivan K, Klevens J, et al. (2015). Effectiveness of a statewide abusive head trauma prevention program in North Carolina. *JAMA Pediatrics,* 169(12): 1126-1131.
165. Yu YR, DeMello AS, Greeley CS, Cox CS, Naik-Mathuria BJ & Wesson DE (2018). Injury patterns of child abuse: Experience of two Level 1 pediatric trauma centers. *Journal of Pediatric Surgery,* 53(5): 1028-1032.
166. Salehi-Had H, Brandt JD, Rosas AJ & Rogers KK (2006). Findings in older children with abusive head injury: Does shaken-child syndrome exist? *Pediatrics,* 117(5): e1039-e1044.

167. Boop S, Axente M, Weatherford B & Klimo P, Jr. (2016). Abusive head trauma: An epidemiological and cost analysis. *Journal of Neurosurgery Pediatrics,* 18(5): 542-549.
168. Dias MS, Smith K, deGuehery K, Mazur P, Li V & Shaffer ML (2005). Preventing abusive head trauma among infants and young children: A hospital-based, parent education program. *Pediatrics,* 115(4): e470-e477.
169. Fanconi M & Lips U (2010). Shaken baby syndrome in Switzerland: Results of a prospective follow-up study, 2002-2007. *European Journal of Pediatrics,* 169(8): 1023-1028.
170. Wood JN, French B, Fromkin J, Fakeye O, Scribano PV, Letson MM, Makoroff KL, et al. (2016). Association of pediatric abusive head trauma rates with macroeconomic indicators. *Academic Pediatrics,* 16(3): 224-232.
171. Xiang J, Shi J, Wheeler KK, Yeates KO, Taylor HG & Smith GA (2013). Paediatric patients with abusive head trauma treated in US Emergency Departments, 2006-2009. *Brain Injury,* 27(13-14): 1555-1561.
172. Sinal SH, Petree AR, Herman-Giddens M, Rogers MK, Enand C & Durant RH (2000). Is race or ethnicity a predictive factor in Shaken Baby Syndrome? *Child Abuse & Neglect,* 24(9): 1241-1246.
173. Sun DT, Zhu XL & Poon WS (2006). Non-accidental subdural haemorrhage in Hong Kong: Incidence, clinical features, management and outcome. *Child's Nervous System,* 22(6): 593-598.
174. Sills MR, Libby AM & Orton HD (2005). Prehospital and in-hospital mortality: A comparison of intentional and unintentional traumatic brain injuries in Colorado children. *Archives of Pediatric & Adolescent Medicine,* 159(7): 665-670.
175. Parks SE, Kegler SR, Annest JL & Mercy JA (2012). Characteristics of fatal abusive head trauma among children in the USA: 2003-2007: An application of the CDC operational case definition to national vital statistics data. *Injury Prevention,* 18(3): 193-199.
176. Tursz A & Cook JM (2014). Epidemiological data on shaken baby syndrome in France using judicial sources. *Pediatric Radiology,* 44(Suppl 4): S641-S646.
177. Reece RM (2008). What are we trying to measure? The problems of case ascertainment. *American Journal of Preventive Medicine,* 34(4 Suppl): S116-S119.
178. Runyan DK (2008). The challenges of assessing the incidence of inflicted traumatic brain injury: A world perspective. *American Journal of Preventive Medicine,* 34(4 Suppl): S112-S115.
179. Barr RG & Runyan DK (2008). Inflicted childhood neurotrauma: The problem set and challenges to measuring incidence. *American Journal of Preventive Medicine,* 34(4 Suppl): S106-S111.
180. Oral R, Yagmur F, Nashelsky M, Turkmen M & Kirby P (2008). Fatal abusive head trauma cases: Consequence of medical staff missing milder forms of physical abuse. *Pediatric Emergency Care,* 24(12): 816-821.
181. Rubin DM, Christian CW, Bilaniuk LT, Zazyczny KA & Durbin DR (2003). Occult head injury in high-risk abused children. *Pediatrics,* 111(6 Pt 1): 1382-1386.
182. Letson MM, Cooper JN, Deans KJ, Scribano PV, Makoroff KL, Feldman KW & Berger RP (2016). Prior opportunities to identify abuse in children with abusive head trauma. *Child Abuse & Neglect,* 60: 36-45.
183. Sanders T, Cobley C, Coles L & Kemp A (2003). Factors affecting clinical referral of young children with a subdural haemorrhage to child protection agencies. *Child Abuse Review,* 12(6): 358-373.
184. Zolotor AJ, Theodore AD, Chang JJ, Berkoff MC & Runyan DK (2008). Speak softly—and forget the stick: Corporal punishment and child physical abuse. *American Journal of Preventive Medicine,* 35(4): 364-369.

185. Theodore AD, Chang JJ, Runyan DK, Hunter WM, Bangdiwala SI & Agans R (2005). Epidemiologic features of the physical and sexual maltreatment of children in the Carolinas. *Pediatrics,* 115(3): e331-e337.
186. Reijneveld SA, van der Wal MF, Brugman E, Hira Sing RA & Verloove-Vanhorick SP (2004). Infant crying and abuse. *The Lancet,* 364(9442): 1340-1342.
187. Cowley LE, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO & Kemp AM (2018). Factors influencing child protection professionals' decision-making and multidisciplinary collaboration in suspected abusive head trauma cases: A qualitative study. *Child Abuse & Neglect,* 82: 178-191.
188. Narang S & Clarke J (2014). Abusive head trauma: Past, present, and future. *Journal of Child Neurology,* 29(12): 1747-1756.
189. Moles RL & Asnes AG (2014). Has this child been abused? Exploring uncertainty in the diagnosis of maltreatment. *Pediatric Clinics of North America,* 61(5): 1023-1036.
190. Starling SP, Holden JR & Jenny C (1995). Abusive head trauma: The relationship of perpetrators to their victims. *Pediatrics,* 95(2): 259-262.
191. Gumbs GR, Keenan HT, Sevick CJ, Conlin AMS, Lloyd DW, Runyan DK, Ryan MAK, et al. (2013). Infant abusive head trauma in a military cohort. *Pediatrics,* 132(4): 668-676.
192. Barr RG, Trent RB & Cross J (2006). Age-related incidence curve of hospitalized Shaken Baby Syndrome cases: Convergent evidence for crying as a trigger to shaking. *Child Abuse & Neglect,* 30(1): 7-16.
193. Lee C, Barr RG, Catherine N & Wicks A (2007). Age-related incidence of publicly reported shaken baby syndrome cases: Is crying a trigger for shaking? *Journal of Developmental & Behavioral Pediatrics,* 28(4): 288-293.
194. Barr RG (2014). Crying as a trigger for abusive head trauma: A key to prevention. *Pediatric Radiology,* 44(Suppl 4): S559-S564.
195. Vinchon M, Defoort-Dhellemmes S, Desurmont M & Dhellemmes P (2005). Accidental and nonaccidental head injuries in infants: A prospective study. *Journal of Neurosurgery,* 102(Suppl 4): 380-384.
196. Kelly P, Thompson JMD, Koh J, Ameratunga S, Jelleyman T, Percival TM, Elder H, et al. (2017). Perinatal risk and protective factors for pediatric abusive head trauma: A multicenter case-control study. *Journal of Pediatrics,* 187: 240-246.e4.
197. Ricci L, Giantris A, Merriam P, Hodge S & Doyle T (2003). Abusive head trauma in Maine infants: Medical, child protective, and law enforcement analysis. *Child Abuse & Neglect,* 27(3): 271-283.
198. Pierce MC, Kaczor K, Acker D, Webb T, Brenzel A, Lorenz DJ, Young A, et al. (2017). History, injury, and psychosocial risk factor commonalities among cases of fatal and near-fatal physical child abuse. *Child Abuse & Neglect,* 69: 263-277.
199. Davies WH & Garwood MM (2001). Who are the perpetrators and why do they do it? *In:* Lazoritz S & Palusci VJ (Eds.) *The shaken baby syndrome: A multidisciplinary approach* (41-54). New York: Harwood Press.
200. Keenan HT (2010). Epidemiology of abusive head trauma. *In:* Jenny C (Ed.) *Child abuse and neglect: Diagnosis, treatment, and evidence* (35-38). Philadelphia: Elsevier Saunders.
201. Wood JN, Hall M, Schilling S, Keren R, Mitra N & Rubin DM (2010). Disparities in the evaluation and diagnosis of abuse among infants with traumatic brain injury. *Pediatrics,* 126(3): 408-414.
202. Hymel KP, Laskey AL, Crowell KR, Wang M, Armijo-Garcia V, Frazier TN, Tieves KS, et al. (2018). Racial and ethnic disparities and bias in the evaluation and reporting of abusive head trauma. *The Journal of Pediatrics,* 198: 137-143.e1.

203. Cobley C, Sanders, T (2003). ‘Shaken Baby Syndrome’: Child protection issues when children sustain a subdural haemorrhage. *Journal of Social Welfare & Family Law,* 25(2): 101-119.
204. Mok JYQ, Jones PA, Myerscough E, Shah ARF & Minns RA (2010). Non-accidental head injury: A consequence of deprivation? *Journal of Epidemiology & Community Health,* 64: 1049-1055.
205. Brewster AL, Nelson JP, Hymel KP, Colby DR, Lucas DR, McCanne TR & Milner JS (1998). Victim, perpetrator, family, and incident characteristics of 32 infant maltreatment deaths in the United States Air Force. *Child Abuse & Neglect,* 22(2): 91-101.
206. Milner JS & Chilamkurti C (1991). Physical child abuse perpetrator characteristics: A review of the literature. *Journal of Interpersonal Violence,* 6(3): 345-366.
207. Hudson CG (2005). Socioeconomic status and mental illness: Tests of the social causation and selection hypotheses. *American Journal of Orthopsychiatry,* 75(1): 3-18.
208. Berger RP, Fromkin JB, Stutz H, Makoroff K, Scribano PV, Feldman K, Tu LC, et al. (2011). Abusive head trauma during a time of increased unemployment: A multicenter analysis. *Pediatrics,* 128(4): 637-643.
209. Huang MI, O'Riordan MA, Fitzenrider E, McDavid L, Cohen AR & Robinson S (2011). Increased incidence of nonaccidental head trauma in infants associated with the economic recession. *Journal of Neurosurgery Pediatrics,* 8(2): 171-176.
210. Keenan HT, Marshall SW, Nocera MA & Runyan DK (2004). Increased incidence of inflicted traumatic brain injury in children after a natural disaster. *American Journal of Preventive Medicine,* 26(3): 189-193.
211. Johnson DA & Rose D (2004). Prognosis, rehabilitation and outcome after inflicted brain injury in children--a case of professional developmental delay. *Pediatric Rehabilitation,* 7(3): 185-193.
212. Karandikar S, Coles L, Jayawant S & Kemp AM (2004). The neurodevelopmental outcome in infants who have sustained a subdural haemorrhage from non-accidental head injury. *Child Abuse Review,* 13(3): 178-187.
213. Barlow K, Thompson E, Johnson D & Minns RA (2004). The neurological outcome of non-accidental head injury. *Pediatric Rehabilitation,* 7(3): 195-203.
214. Haviland J & Russell RI (1997). Outcome after severe non-accidental head injury. *Archives of Disease in Childhood,* 77(6): 504-507.
215. Bonnier C, Nassogne MC & Evrard P (1995). Outcome and prognosis of whiplash shaken infant syndrome; late consequences after a symptom-free interval. *Developmental Medicine & Child Neurology,* 37(11): 943-956.
216. Barlow KM, Thomson E, Johnson D & Minns RA (2005). Late neurologic and cognitive sequelae of inflicted traumatic brain injury in infancy. *Pediatrics,* 116(2): e174-e185.
217. Jayawant S & Parr J (2007). Outcome following subdural haemorrhages in infancy. *Archives of Disease in Childhood,* 92(4): 343-347.
218. Miller TR, Steinbeigle R, Wicks A, Lawrence BA, Barr M & Barr RG (2014). Disability-adjusted life-year burden of abusive head trauma at ages 0-4. *Pediatrics,* 134(6): e1545-e1550.
219. Nuño M, Pelissier L, Varshneya K, Adamo MA & Drazin D (2015). Outcomes and factors associated with infant abusive head trauma in the US. *Journal of Neurosurgery Pediatrics,* 16(5): 515-522.
220. Bartschat S, Richter C, Stiller D & Banschak S (2016). Long-term outcome in a case of shaken baby syndrome. *Medicine, Science & the Law,* 56(2): 147-149.
221. Chevignard MP & Lind K (2014). Long-term outcome of abusive head trauma. *Pediatric Radiology,* 44(Suppl 4): S548-S558.

222. Nuño M, Ugiliweneza B, Zepeda V, Anderson JE, Coulter K, Magana JN, Drazin D, et al. (2018). Long-term impact of abusive head trauma in young children. *Child Abuse & Neglect,* 85: 39-46.
223. Stipanicic A, Nolin P, Fortin G & Gobeil M-F (2008). Comparative study of the cognitive sequelae of school-aged victims of Shaken Baby Syndrome. *Child Abuse & Neglect,* 32(3): 415-428.
224. Minns RA, Jones PA & Barlow KM (2005). Outcome and prognosis of non-accidental head injury in infants. *In:* Minns R & Brown K (Eds.) *Shaking and other non-accidental head injuries in children* (364-414). Edinburgh: Mac Keith Press.
225. Duhaime AC, Christian C, Moss E & Seidl T (1996). Long-term outcome in infants with the shaking-impact syndrome. *Pediatric Neurosurgery,* 24(6): 292-298.
226. Miller Ferguson N, Sarnaik A, Miles D, Shafi N, Peters MJ, Truemper E, Vavilala MS, et al. (2017). Abusive head trauma and mortality – an analysis from an international comparative effectiveness study of children with severe traumatic brain injury. *Critical Care Medicine,* 45(8): 1398-1407.
227. Dias MS, Backstrom J, Falk M & Li V (1998). Serial radiography in the infant shaken impact syndrome. *Pediatric Neurosurgery,* 29(2): 77-85.
228. Shein SL, Bell MJ, Kochanek PM, Tyler-Kabara EC, Wisniewski SR, Feldman K, Makoroff K, et al. (2012). Risk factors for mortality in children with abusive head trauma. *Journal of Pediatrics,* 161(4): 716-722.e1.
229. Keenan HT, Hooper SR, Wetherington CE, Nocera M & Runyan DK (2007). Neurodevelopmental consequences of early traumatic brain injury in 3-year-old children. *Pediatrics,* 119(3): e616-e623.
230. Keenan HT, Runyan DK & Nocera M (2006). Child outcomes and family characteristics 1 year after severe inflicted or noninflicted traumatic brain injury. *Pediatrics,* 117(2): 317-324.
231. Feldman KW, Bethel R, Shugerman RP, Grossman DC, Grady MS & Ellenbogen RG (2001). The cause of infant and toddler subdural hemorrhage: A prospective study. *Pediatrics,* 108(3): 636-646.
232. Bonnier C, Nassogne M-C, Saint-Martin C, Mesples B, Kadhim H & Sebire G (2003). Neuroimaging of intraparenchymal lesions predicts outcome in shaken baby syndrome. *Pediatrics,* 112(4): 808-814.
233. Keenan HT, Runyan DK & Nocera M (2006). Longitudinal follow-up of families and young children with traumatic brain injury. *Pediatrics,* 117(4): 1291-1297.
234. Rhine T, Wade SL, Makoroff KL, Cassedy A & Michaud LJ (2012). Clinical predictors of outcome following inflicted traumatic brain injury in children. *Journal of Trauma & Acute Care Surgery,* 73(4 Suppl 3): S248-S253.
235. Greiner MV, Lawrence AP, Horn P, Newmeyer AJ & Makoroff KL (2012). Early clinical indicators of developmental outcome in abusive head trauma. *Child's Nervous System,* 28(6): 889-896.
236. Tanoue K, Matsui K, Nozawa K & Aida N (2012). Predictive value of early radiological findings in inflicted traumatic brain injury. *Acta Paediatrica,* 101(6): 614-617.
237. Ilves P, Lintrop M, Talvik I, Sisko A & Talvik T (2010). Predictive value of clinical and radiological findings in inflicted traumatic brain injury. *Acta Paediatrica,* 99(9): 1329-1336.
238. Libby AM, Sills MR, Thurston NK & Orton HD (2003). Costs of childhood physical abuse: comparing inflicted and unintentional traumatic brain injuries. *Pediatrics,* 112(1 Pt 1): 58-65.
239. Friedman J, Reed P, Sharplin P & Kelly P (2012). Primary prevention of pediatric abusive head trauma: A cost audit and cost-utility analysis. *Child Abuse & Neglect,* 36(11-12): 760-770.

240. Peterson C, Xu L, Florence C, Parks SE, Miller TR, Barr RG, Barr M, et al. (2014). The medical cost of abusive head trauma in the United States. *Pediatrics,* 134(1): 91-99.
241. Peterson C, Xu L, Florence C & Parks SE (2015). Annual cost of U.S. hospital visits for pediatric abusive head trauma. *Child Maltreatment,* 20(3): 162-169.
242. Miller TR, Steinbeigle R, Lawrence BA, Peterson C, Florence C, Barr M & Barr RG (2018). Lifetime cost of abusive head trauma at ages 0–4, USA. *Prevention Science,* 19(6): 695-704.
243. Tuerkheimer D (2009). The next innocence project: Shaken baby syndrome and the criminal courts. *Washington University Law Review,* 87(1): 1-58.
244. Findley KA, Barnes PD, Moran DA & Squier W (2012). Shaken baby syndrome, abusive head trauma, and actual innocence: Getting it right. *Houston Journal of Health, Law & Policy,* 12(2): 209-312.
245. Tuerkheimer D (2010). *Anatomy of a misdiagnosis*. The New York Times, Sept 20. Available from: http://www.nytimes.com/2010/09/21/opinion/21tuerkheimer.html. Accessed 24 October 2015.
246. Cenziper D (2015). *Shaken science: A disputed diagnosis imprisons parents*. Washington Post, March 20. Available from: https://www.washingtonpost.com/graphics/investigations/shaken-baby-syndrome/??noredirect=on. Accessed 20 September 2018.
247. Bazelon E (2011). *Shaken-baby syndrome faces new questions in court*. New York Times Magazine, Feb 2. Available from: http://www.nytimes.com/2011/02/06/magazine/06baby-t.html. Accessed 17 December 2015.
248. Cavazos v. Smith (2011). 132 S. Ct. 2. Available from: https://www.courtlistener.com/opinion/616357/cavazos-v-smith/. Accessed 12 June 2016.
249. State v. Edmunds (2008). WI App 33, 746 N.W.2d 590 Available from: https://www.courtlistener.com/opinion/2021651/state-v-edmunds/. Accessed 12 June 2016.
250. Donohoe M (2003). Evidence-based medicine and shaken baby syndrome: Part I: Literature review, 1966-1998. *American Journal of Forensic Medicine & Pathology,* 24(3): 239-242.
251. Lynøe N, Elinder G, Hallberg B, Rosén M, Sundgren P & Eriksson A (2017). Insufficient evidence for 'shaken baby syndrome' – a systematic review. *Acta Paediatrica,* 106(7): 1021-1027.
252. Elinder G, Eriksson A, Hallberg B, Lynöe N, Rosén M, & & Sundgren P (2016). *Traumatic shaking: The role of the triad in medical investigations of suspected traumatic shaking. A systematic review*. Swedish Agency for Health Technology Assessment and Assessment of Social Services. Report 255e. Available from: http://www.sbu.se/en/publications/sbu-assesses/traumatic-shaking--the-role-of-the-triad-in-medical-investigations-of-suspected-traumatic-shaking/. Accessed 13 September 2018.
253. Ceccarelli L (2011). Manufactured scientific controversy: Science, rhetoric, and public debate. *Rhetoric & Public Affairs,* 14(2): 195-228.
254. Shabde N (2006). Child protection training for paediatricians. *Archives of Disease in Childhood,* 91(8): 639-641.
255. R v. Henderson & others (2010). EWCA Crim 1269. Available from: http://www.bailii.org/ew/cases/EWCA/Crim/2010/1269.html. Accessed 6 May 2017.
256. R v. Ben Butler & Jennie Gray (2016). Sentencing remarks of Mr Justice Wilkie. Available from: https://www.judiciary.uk/judgments/r-v-ben-butler-and-jennie-gray/. Accessed 6 May 2017.

257. Davis M (2016). *Child D. A serious case review overview report*. Sutton Local Safeguarding Children Board. Available from: https://drive.google.com/file/d/0B5ILmebheQx3YmhvbjVYOHlpN2M/view. Accessed 6 May 2017.
258. Geddes JF, Tasker RC, Hackshaw AK, Nickols CD, Adams GGW, Whitwell HL & Scheimberg I (2003). Dural haemorrhage in non-traumatic infant deaths: Does it explain the bleeding in 'shaken baby syndrome'? *Neuropathology & Applied Neurobiology,* 29(1): 14-22.
259. A Local Authority v. S (2009). EWHC 2115 (Fam). Available from: http://www.familylawweek.co.uk/site.aspx?i=ed53850. Accessed 8 May 2017.
260. Laurent-Vannier A, Adamsbaum C, Raul J-S, Rey-Salmon C & Rambaud C (2018). Flawed Swedish study on traumatic shaking is already being used by defence lawyers and its findings must be ignored. *Acta Paediatrica,* 107(12): 2048-2050.
261. Offiah AC, Servaes S, Adamsbaum CS, Argyropoulou MI, Halliday KE, Jaspan T, Owens CM, et al. (2017). Initial response of the European Society of Paediatric Radiology and Society for Pediatric Radiology to the Swedish Agency for Health Technology Assessment and Assessment of Social Services' document on the triad of shaken baby syndrome. *Pediatric Radiology,* 47(4): 369-371.
262. Saunders D, Raissaki M, Servaes S, Adamsbaum C, Choudhary AK, Moreno JA, van Rijn RR, et al. (2017). Throwing the baby out with the bath water — response to the Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) report on traumatic shaking. *Pediatric Radiology,* 47(11): 1386-1389.
263. Bilo RAC, Banaschak S, Herrmann B, Karst WA, Kubat B, Nijs HGT, van Rijn RR, et al. (2017). Using the table in the Swedish review on shaken baby syndrome will not help courts deliver justice. *Acta Paediatrica,* 106(7): 1043-1045.
264. Hellgren K, Hellström A, Hård A-L, Jacobson L, Lidén U, Löfgren S, Fahnehjelm KT, et al. (2017). The new Swedish report on Shaken Baby Syndrome is misleading. *Acta Paediatrica,* 106(7): 1040.
265. Levin AV (2017). The SBU report: A different view. *Acta Paediatrica,* 106(7): 1037-1039.
266. Debelle GD, Maguire S, Watts P, Nieto Hernandez R & Kemp AM (2018). Abusive head trauma and the triad: A critique on behalf of RCPCH of 'Traumatic shaking: the role of the triad in medical investigations of suspected traumatic shaking'. *Archives of Disease in Childhood,* 103(6): 606-610.
267. Greeley CS (2010). Infant fatality. *Seminars in Pediatric Neurology,* 17(4): 275-278.
268. Jenny C (2014). Alternate theories of causation in abusive head trauma: What the science tells us. *Pediatric Radiology,* 44(Suppl 4): S543-S547.
269. Leventhal J & Edwards G (2017). Flawed theories to explain child physical abuse: What are the medical-legal consequences? *JAMA,* 318(14): 1317-1318.
270. Strouse PJ (2016). Child abuse: We have problems. *Pediatric Radiology,* 46(5): 587-590.
271. Edwards GA (2015). Mimics of child abuse: Can choking explain abusive head trauma? *Journal of Forensic & Legal Medicine,* 35: 33-37.
272. Stephenson T (2015). Shedding light on non-accidental bruising. *Archives of Disease in Childhood,* 100(5): 419-420.
273. Knox B, Rorke-Adams LB & Luyet FM (2016). Subdural hematoma rebleeding in relation to abusive head trauma. *Journal of Family Violence,* 31(7): 815-821.
274. Hymel KP, Jenny C & Block RW (2002). Intracranial hemorrhage and rebleeding in suspected victims of abusive head trauma: Addressing the forensic controversies. *Child Maltreatment,* 7(4): 329-348.
275. Feldman KW, Sugar NF & Browd SR (2015). Initial clinical presentation of children with acute and chronic versus acute subdural hemorrhage resulting from abusive head trauma. *Journal of Neurosurgery Pediatrics,* 16(2): 177-185.

276. Alexander R, Crabbe L, Sato Y, Smith W & Bennett T (1990). Serial abuse in children who are shaken. *American Journal of Diseases of Children,* 144(1): 58-60.
277. Sheets LK, Leach ME, Koszewski IJ, Lessmeier AM, Nugent M & Simpson P (2013). Sentinel injuries in infants evaluated for child physical abuse. *Pediatrics,* 131(4): 701-707.
278. Petska HW, Sheets LK & Knox BL (2013). Facial bruising as a precursor to abusive head trauma. *Clinical Pediatrics,* 52(1): 86-88.
279. Sieswerda-Hoogendoorn T, Bilo RAC, van Duurling LLBM, Karst WA, Maaskant JM, van Aalderen WMC & van Rijn RR (2013). Abusive head trauma in young children in the Netherlands: Evidence for multiple incidents of abuse. *Acta Paediatrica,* 102(11): e497-e501.
280. Thackeray JD (2007). Frena tears and abusive head injury: A cautionary tale. *Pediatric Emergency Care,* 23(10): 735-737.
281. King WK, Kiesel EL & Simon HK (2006). Child abuse fatalities: Are we missing opportunities for intervention? *Pediatric Emergency Care,* 22(4): 211-214.
282. Thorpe EL, Zuckerbraun NS, Wolford JE & Berger RP (2014). Missed opportunities to diagnose child physical abuse. *Pediatric Emergency Care,* 30(11): 771-776.
283. Tilak GS & Pollock AN (2013). Missed opportunities in fatal child abuse. *Pediatric Emergency Care,* 29(5): 685-687.
284. Puls HT, Anderst JD, Bettenhausen JL, Masonbrink A, Markham JL, Plencner L, Krager M, et al. (2018). Potential opportunities for prevention or earlier diagnosis of child physical abuse in the inpatient setting. *Hospital Pediatrics,* 8(2): 81-88.
285. Laskey AL, Holsti M, Runyan DK & Socolar RR (2004). Occult head trauma in young suspected victims of physical abuse. *Journal of Pediatrics,* 144(6): 719-722.
286. Harper NS, Feldman KW, Sugar NF, Anderst JD & Lindberg DM (2014). Additional injuries in young infants with concern for abuse and apparently isolated bruises. *Journal of Pediatrics,* 165(2): 383-388.e1.
287. Fickenscher KA, Dean JS, Mena DC, Green BA & Lowe LH (2010). Occult cranial injuries found with neuroimaging in clinically asymptomatic young children due to abusive compared to accidental head trauma. *Southern Medical Journal,* 103(2): 121-125.
288. Boehnke M, Mirsky D, Stence N, Stanley RM, Lindberg DM & for the ExStra investigators (2018). Occult head injury is common in children with concern for physical abuse. *Pediatric Radiology,* 48(8): 1123-1129.
289. Leventhal JM, Asnes AG, Pavlovic L & Moles RL (2014). Diagnosing abusive head trauma: The challenges faced by clinicians. *Pediatric Radiology,* 44(Suppl 4): S537-S442.
290. Berger RP & Lindberg DM (2019). Early recognition of physical abuse: Bridging the gap between knowledge and practice. *The Journal of Pediatrics*, 204: 16-23.
291. Slovis TL, Strouse PJ & Strauss KJ (2015). Radiation exposure in imaging of suspected child abuse: Benefits versus risks. *Journal of Pediatrics,* 167(5): 963-938.
292. Laupacis A, Sekar N & Stiell IG (1997). Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA,* 277(6): 488-494.
293. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG & Richardson WS (2000). Users' guides to the medical literature: XXII: How to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA,* 284(1): 79-84.
294. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG & Grobbee DE (2012). Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart,* 98(9): 683-690.
295. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG & Woodward M (2012). Risk prediction models: II. External validation, model updating, and impact assessment. *Heart,* 98(9): 691-698.

296. Moons KG, Altman DG, Vergouwe Y & Royston P (2009). Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *The BMJ,* 338: b606.
297. Royston P, Moons KGM, Altman DG & Vergouwe Y (2009). Prognosis and prognostic research: Developing a prognostic model. *The BMJ,* 338: b604.
298. Altman DG, Vergouwe Y, Royston P & Moons KG (2009). Prognosis and prognostic research: Validating a prognostic model. *The BMJ,* 338: b605.
299. Reilly BM & Evans AT (2006). Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Annals of Internal Medicine,* 144(3): 201-209.
300. Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, Verbakel J, et al. (2011). Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). *BMC Medical Informatics & Decision Making,* 11: 62.
301. Stiell I & Wells G (1999). Methodologic standards for the development of clinical decision rules in emergency medicine. *Annals of Emergency Medicine,* 33(4): 437-447.
302. Lee TH (1990). Evaluating decision aids. *Journal of General Internal Medicine,* 5(6): 528-529.
303. Maguire JL, Kulik DM, Laupacis A, Kuppermann N, Uleryk EM & Parkin PC (2011). Clinical prediction rules for children: A systematic review. *Pediatrics,* 128(3): e666-e677.
304. Keogh C, Wallace E, O'Brien KK, Galvin R, Smith SM, Lewis C, Cummins A, et al. (2014). Developing an international register of clinical prediction rules for use in primary care: A descriptive analysis. *Annals of Family Medicine,* 12(4): 359-366.
305. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, et al. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLoS Medicine,* 10(2): e1001381.
306. Wallace E, Uijen MJM, Clyne B, Zarabzadeh A, Keogh C, Galvin R, Smith SM, et al. (2016). Impact analysis studies of clinical prediction rules relevant to primary care: A systematic review. *BMJ Open,* 6(3): e009957.
307. Sanders SL, Rathbone J, Bell KJL, Glasziou PP & Doust JA (2017). Systematic review of the effects of care provided with and without diagnostic clinical prediction rules. *BMC Diagnostic & Prognostic Research,* 1: 13.
308. Collins GS, Mallett S, Omar O & Yu L-M (2011). Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting. *BMC Medicine,* 9: 103.
309. Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, Mol BW, et al. (2016). Prognostic models in obstetrics: Available, but far from applicable. *American Journal of Obstetrics & Gynecology,* 214(1): 79-90.e36.
310. Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ & Moons KG (2010). Prediction models for prolonged intensive care unit stay after cardiac surgery: Systematic review and validation study. *Circulation,* 122(7): 682-689.
311. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, et al. (2016). Prediction models for cardiovascular disease risk in the general population: Systematic review. *The BMJ,* 353: i2416.
312. Shariat SF, Karakiewicz PI, Margulis V & Kattan MW (2008). Inventory of prostate cancer predictive tools. *Current Opinion in Urology,* 18(3): 279-296.
313. Perel P, Edwards P, Wentz R & Roberts I (2006). Systematic review of prognostic models in traumatic brain injury. *BMC Medical Informatics & Decision Making,* 6: 38.
314. Wessler BS, Lai Yh L, Kramer W, Cangelosi M, Raman G, Lutz JS & Kent DM (2015). Clinical prediction models for cardiovascular disease: Tufts predictive analytics and

comparative effectiveness clinical prediction model database. *Circulation Cardiovascular Quality & Outcomes,* 8(4): 368-375.
315. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y & Moons KGM (2018). Evaluating the impact of prediction models: Lessons learned, challenges, and recommendations. *BMC Diagnostic & Prognostic Research,* 2: 11.
316. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM & Deeks JJ (2012). Assessing the value of diagnostic tests: A framework for designing and evaluating trials. *The BMJ,* 344: e686.
317. White H (2009). Theory-based impact evaluation: Principles and practice. *Journal of Development Effectiveness,* 1(3): 271-284.
318. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, et al. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *The BMJ,* 350: h1258.
319. Dowding D, Lichtner V & Closs SJ (2017). Using the MRC framework for complex interventions to develop clinical decision support: A case study. *In:* Randell R, Cornet R, McCowan C, Peek N & Scott PJ (Eds.) *Studies in health technology and informatics* (544-548). IOS Press.
320. Brown B, Cheraghi-Sohi S, Jaki T, Su T-L, Buchan I & Sperrin M (2016). Understanding clinical prediction models as 'innovations': A mixed methods study in UK family practice. *BMC Medical Informatics & Decision Making,* 16: 106.
321. Noble D, Mathur R, Dent T, Meads C & Greenhalgh T (2011). Risk models and scores for type 2 diabetes: Systematic review. *The BMJ,* 343: d7163.
322. Medical Research Council (2008). *Developing and evaluating complex interventions: New guidance*. London: Medical Research Council. Available from: http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC004871. Accessed 18 September 2014.
323. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I & Petticrew M (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *The BMJ,* 337: a1655.
324. Steyerberg E (2009). *Clinical prediction models: A practical approach to development, validation and updating.* New York: Springer-Verlag.
325. Ban JW, Wallace E, Stevens R & Perera R (2017). Why do authors derive new cardiovascular clinical prediction rules in the presence of existing rules? A mixed methods study. *PLoS One,* 12(6): e0179102.
326. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, et al. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine,* 162(1): W1-W73.
327. Leventhal JM (2000). Thinking clearly about evaluations of suspected child abuse. *Clinical Child Psychology & Psychiatry,* 5(1): 139-147.
328. Higgins J & Green S (2011). *Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011].* The Cochrane Collaboration, 2011. Available from: http://handbook.cochrane.org. Accessed 9 October 2014.
329. Kemp AM, Rajaram S, Mann M, Tempest V, Farewell D, Gawne-Cain ML, Jaspan T, et al. (2009). What neuroimaging should be performed in children in whom inflicted brain injury (iBI) is suspected? A systematic review. *Clinical Radiology,* 64(5): 473-483.
330. Royal College of Paediatrics & Child Health (2015). *Child protection evidence systematic review on neurological injuries*. Available from: https://www.rcpch.ac.uk/resources/child-protection-evidence-neurological-injuries. Accessed 12 September 2018.

331. Jenny C (2010). *Child abuse and neglect. Diagnosis, treatment, and evidence.* Philadelphia: Elsevier Saunders.
332. Minns RA & Brown JK (2005). *Shaking and other non-accidental head injuries in children.* Edinburgh: Mac Keith Press.
333. Christian CW & Committee on Child Abuse and Neglect (2015). The evaluation of suspected child physical abuse. *Pediatrics,* 135(5): e1337-e1354.
334. Canadian Paediatric Society (2007). *Multidisciplinary guidelines on the identification, investigation and management of suspected abusive head trauma*. Ottawa: Canadian Paediatric Society. Available from: https://www.cps.ca/en/documents/position/multidisciplinary-guidelines-abusive-head-trauma. Accessed 20 December 2014.
335. Amagasa S, Matsui H, Tsuji S, Moriya T & Kinoshita K (2016). Accuracy of the history of injury obtained from the caregiver in infantile head trauma. *American Journal of Emergency Medicine,* 34(9): 1863-1867.
336. Fujiwara T, Okuyama M & Miyasaka M (2008). Characteristics that distinguish abusive from nonabusive head trauma among young children who underwent head computed tomography in Japan. *Pediatrics,* 122(4): e841-e847.
337. Altman RL, Brand DA, Forman S & et al. (2003). Abusive head injury as a cause of apparent life-threatening events in infancy. *Archives of Pediatrics & Adolescent Medicine,* 157(10): 1011-1015.
338. Hymel KP & Deye KP (2010). Abusive head trauma. *In:* Jenny C (Ed.) *Child abuse and neglect Diagnosis, treatment, and evidence* (349-359). Philadelphia: Elsevier Saunders.
339. Bechtel K & Berger R (2006). Inflicted traumatic brain injury: Making the diagnosis in the emergency department. *Clinical Pediatric Emergency Medicine*, 7(3): 138-142.
340. Morris MW, Smith S, Cressman J & Ancheta J (2000). Evaluation of infants with subdural hematoma who lack external evidence of abuse. *Pediatrics,* 105(3 Pt 1): 549-553.
341. Minns RA & Busuttil A (2004). Patterns of presentation of the shaken baby syndrome: Four types of inflicted brain injury predominate. *The BMJ,* 328(7442): 766.
342. Altman RL, Forman S & Brand DA (2007). Ophthalmologic findings in infants after an apparent life-threatening event. *European Journal of Ophthalmology,* 17(4): 648-653.
343. Parker K & Pitetti R (2011). Mortality and child abuse in children presenting with apparent life-threatening events. *Pediatric Emergency Care,* 27(7): 591-595.
344. Guenther E, Powers A, Srivastava R & Bonkowsky JL (2010). Abusive head trauma in children presenting with an apparent life-threatening event. *Journal of Pediatrics,* 157(5): 821-825.
345. Pitetti RD, Maffei F, Chang K, Hickey R, Berger R & Pierce MC (2002). Prevalence of retinal hemorrhages and child abuse in children who present with an apparent life-threatening event. *Pediatrics,* 110(3): 557-562.
346. Bonkowsky JL, Guenther E, Filloux FM & Srivastava R (2008). Death, child abuse, and adverse neurological outcome of infants after an apparent life-threatening event. *Pediatrics,* 122(1): 125-131.
347. National Institutes of Health (1987). National Institutes of Health consensus development conference on infantile apnea and home monitoring, Sept 29 to Oct 1, 1986. *Pediatrics,* 79(2): 292-299.
348. McGovern M & Smith M (2004). Causes of apparent life threatening events in infants: A systematic review. *Archives of Disease in Childhood,* 89(11): 1043-1048.
349. Hansen JB, Frazier T, Moffatt M, Zinkus T & Anderst JD (2017). Evaluation of the hypothesis that choking/ALTE may mimic abusive head trauma. *Academic Pediatrics,* 17(4): 362-367.

350. Barnes PD, Galaznik J, Gardner H & Shuman M (2010). Infant acute life-threatening event--dysphagic choking versus nonaccidental injury. *Seminars in Pediatric Neurology,* 17(1): 7-11.
351. Hymel KP, Willson DF, Boos SC, Pullin DA, Homa K, Lorenz DJ, Herman BE, et al. (2013). Derivation of a clinical prediction rule for pediatric abusive head trauma. *Pediatric Critical Care Medicine,* 14(2): 210-220.
352. Thalayasingam M, Veerakumarasivam A, Kulanthayan S, Khairuddin F & Cheah IG (2012). Clinical clues for head injuries amongst Malaysian infants: Accidental or non-accidental? *Injury,* 43(12): 2083-2087.
353. Diaz-Olavarrieta C, Garcia-Pina CA, Loredo-Abdala A, Paz F, Garcia SG & Schilmann A (2011). Abusive head trauma at a tertiary care children's hospital in Mexico City. A preliminary study. *Child Abuse & Neglect,* 35(11): 915-923.
354. Datta S, Stoodley N, Jayawant S, Renowden S & Kemp A (2005). Neuroradiological aspects of subdural haemorrhages. *Archives of Disease in Childhood,* 90(9): 947-951.
355. Vinchon M, Noule N, Tchofo PJ, Soto-Ares G, Fourier C & Dhellemmes P (2004). Imaging of head injuries in infants: Temporal correlates and forensic implications for the diagnosis of child abuse. *Journal of Neurosurgery,* 101(1 Suppl): 44-52.
356. Case ME (2008). Accidental traumatic head Injury in infants and young children. *Brain Pathology,* 18(4): 583-589.
357. Rivas JJ, Lobato RD, Sarabia R, Cordobés F, Cabrera A & Gomez P (1988). Extradural hematoma: Analysis of factors influencing the courses of 161 patients. *Neurosurgery,* 23(1): 44-51.
358. Denton S & Mileusnic D (2003). Delayed sudden death in an infant following an accidental fall: A case report with review of the literature. *The American Journal of Forensic Medicine & Pathology,* 24(4): 371-376.
359. Forbes BJ, Cox M & Christian CW (2008). Retinal hemorrhages in patients with epidural hematomas. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus,* 12(2): 177-180.
360. Palifka LA, Frasier LD, Metzger RR & Hedlund GL (2016). Parenchymal brain laceration as a predictor of abusive head trauma. *American Journal of Neuroradiology,* 37(1): 163-168.
361. Bhardwaj G, Chowdhury V, Jacobs MB, Moran KT, Martin FJ & Coroneo MT (2010). A systematic review of the diagnostic accuracy of ocular signs in pediatric abusive head trauma. *Ophthalmology,* 117(5): 983-992.e17.
362. Levin A (2000). Retinal haemorrhage and child abuse. *In:* David T (Ed.) *Recent advances in paediatrics* (151-219). London: Churchill Livingstone.
363. Binenbaum G, Christian CW, Ichord RN, Ying GS, Simon MA, Romero K, Pollock AN, et al. (2013). Retinal hemorrhage and brain injury patterns on diffusion-weighted magnetic resonance imaging in children with head trauma. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus,* 17(6): 603-608.
364. Morad Y, Kim YM, Armstrong DC, Huyer D, Mian M & Levin AV (2002). Correlation between retinal abnormalities and intracranial abnormalities in the shaken baby syndrome. *American Journal of Ophthalmology,* 134(3): 354-359.
365. Binenbaum G, Mirza-George N, Christian CW & Forbes BJ (2009). Odds of abuse associated with retinal hemorrhages in children suspected of child abuse. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus,* 13(3): 268-272.
366. Levin AV (2002). Ophthalmology of shaken baby syndrome. *Neurosurgery Clinics of North America,* 13(2): 201-211.

367. Betz P, Püschel K, Miltner E, Lignitz E & Eisenmenger W (1996). Morphometrical analysis of retinal hemorrhages in the shaken baby syndrome. *Forensic Science International,* 78(1): 71-80.
368. Levin AL, Luyet FM & Knox BL (2016). Ophthalmologic concerns in abusive head trauma. *Journal of Family Violence,* 31(7): 797-804.
369. Shouldice M, Al-Khattabi F, Thau A, McIntyre S, Ng WK & Levin AV (2018). Traumatic macular retinoschisis in infants and children. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus*, 22(6): 433-437.
370. Breazzano MP, Unkrich KH & Barker-Griffith AE (2014). Clinicopathological findings in abusive head trauma: Analysis of 110 infant autopsy eyes. *American Journal of Ophthalmology,* 158(6): 1146-1154.e2.
371. Wygnanski-Jaffe T, Levin AV, Shafiq A, Smith C, Enzenauer RW, Elder JE, Morin JD, et al. (2006). Postmortem orbital findings in shaken baby syndrome. *American Journal of Ophthalmology,* 142(2): 233-240.
372. Kivlin JD, Simons KB, Lazoritz S & Ruttum MS (2000). Shaken baby syndrome. *Ophthalmology,* 107(7): 1246-1254.
373. Rao N, Smith RE, Choi JH, Xiaohu X & Kornblum RN (1988). Autopsy findings in the eyes of fourteen fatally abused children. *Forensic Science International,* 39(3): 293-299.
374. Kemp AM, Dunstan F, Harrison S, Morris S, Mann M, Rolfe K, Datta S, et al. (2008). Patterns of skeletal fractures in child abuse: Systematic review. *The BMJ,* 337: a1518.
375. Maguire S, Cowley L, Mann M & Kemp A (2013). What does the recent literature add to the identification and investigation of fractures in child abuse: An overview of review updates 2005–2013. *Evidence-Based Child Health: A Cochrane Review Journal,* 8(5): 2044-2057.
376. Lonergan GJ, Baker AM, Morey MK & Boos SC (2003). From the archives of the AFIP. Child abuse: Radiologic-pathologic correlation. *Radiographics,* 23(4): 811-845.
377. Paine CW, Fakeye O, Christian CW & Wood JN (2016). Prevalence of abuse among young children with rib fractures: A systematic review. *Pediatric Emergency Care*. doi: 10.1097/PEC.0000000000000911.
378. Garcia VF, Gotschall CS, Eichelberger MR & Bowman LM (1990). Rib fractures in children: A marker of severe trauma. *The Journal of Trauma,* 30(6): 695-700.
379. Weber MA, Risdon RA, Offiah AC, Malone M & Sebire NJ (2009). Rib fractures identified at post-mortem examination in sudden unexpected deaths in infancy (SUDI). *Forensic Science International,* 189(1-3): 75-81.
380. Darling SE, Done SL, Friedman SD & Feldman KW (2014). Frequency of intrathoracic injuries in children younger than 3 years with rib fractures. *Pediatric Radiology,* 44(10): 1230-1236.
381. Maguire S, Mann M, John N, Ellaway B, Sibert JR & Kemp AM (2006). Does cardiopulmonary resuscitation cause rib fractures in children? A systematic review. *Child Abuse & Neglect,* 30(7): 739-751.
382. Royal College of Paediatrics & Child Health (2017). *Child protection evidence systematic review on fractures*. Available from: https://www.rcpch.ac.uk/resources/child-protection-evidence-fractures. Accessed 20 August 2018.
383. Flaherty EG, Perez-Rossello JM, Levine MA & Hennrikus WL (2014). Evaluating children with fractures for child physical abuse. *Pediatrics,* 133(2): e477-e489.
384. Kleinman PK, Perez-Rossello JM, Newton AW, Feldman HA & Kleinman PL (2011). Prevalence of the classic metaphyseal lesion in infants at low versus high risk for abuse. *American Journal of Roentgenology,* 197(4): 1005-1008.
385. Arkader A, Friedman JE, Warner WCJ & Wells L (2007). Complete distal femoral metaphyseal fractures: A harbinger of child abuse before walking age. *Journal of Pediatric Orthopaedics,* 27(7): 751-753.

386. Kleinman PK, Marks SC, Richmond JM & Blackbourne BD (1995). Inflicted skeletal injury: A postmortem radiologic-histopathologic study in 31 infants. *American Journal of Roentgenology,* 165(3): 647-650.
387. Pandya NK, Baldwin KD, Wolfgruber H, Drummond DS & Hosalkar HS (2010). Humerus fractures in the pediatric population: An algorithm to identify abuse. *Journal of Pediatric Orthopaedics B,* 19(6): 535-541.
388. Capra L, Levin AV, Howard A & Shouldice M (2013). Characteristics of femur fractures in ambulatory young children. *Emergency Medicine Journal,* 30(9): 749-753.
389. Clarke NM, Shelton FR, Taylor CC, Khan T & Needhirajan S (2012). The incidence of fractures in children under the age of 24 months--in relation to non-accidental injury. *Injury,* 43(6): 762-765.
390. Son-Hing JP & Deniz Olgun Z (2018). The frequency of nonaccidental trauma in children under the age of 3 years with femur fractures: Is there a better cutoff point for universal workups? *Journal of Pediatric Orthopaedics B,* 27(4): 366-368.
391. Leventhal JM, Thomas SA, Rosenfield NS & Markowitz RI (1993). Fractures in young children: Distinguishing child abuse from unintentional injuries. *American Journal of Diseases of Children,* 147(1): 87-92.
392. Kowal-Vern A, Paxton TP, Ros SP, Lietz H, Fitzgerald M & Gamelli RL (1992). Fractures in the under-3-year-old age cohort. *Clinical Pediatrics,* 31(11): 653-659.
393. Adamo MA, Drazin D, Smith C & Waldman JB (2009). Comparison of accidental and nonaccidental traumatic brain injuries in infants and toddlers: Demographics, neurosurgical interventions, and outcomes. *Journal of Neurosurgery Pediatrics,* 4(5): 414-419.
394. Meservy CJ, Towbin R, McLaurin RL, Myers PA & Ball W (1987). Radiographic characteristics of skull fractures resulting from child abuse. *American Journal of Roentgenology,* 149(1): 173-175.
395. Worlock P, Stower M & Barbor P (1986). Patterns of fractures in accidental and non-accidental injury in children: A comparative study. *The BMJ,* 293(6539): 100-102.
396. Hobbs CJ (1984). Skull fracture and the diagnosis of abuse. *Archives of Disease in Childhood,* 59(3): 246-252.
397. Wood JN, Christian CW, Adams CM & Rubin DM (2009). Skeletal surveys in infants with isolated skull fractures. *Pediatrics,* 123(2): e247-e252.
398. Ingham AI, Langlois NE & Byard RW (2011). The significance of bruising in infants—a forensic postmortem study. *Archives of Disease in Childhood,* 96(3): 218-220.
399. Valvano TJ, Binns HJ, Flaherty EG & Leonhardt DE (2008). Does bruising help determine which fractures are caused by abuse? *Child Maltreatment,* 14(4): 376-381.
400. Mathew MO, Ramamohan N & Bennet GC (1998). Importance of bruising associated with paediatric fractures: Prospective observational study. *The BMJ,* 317(7166): 1117-1118.
401. Peters ML, Starling SP, Barnes-Eley ML & Heisler KW (2008). The presence of bruising associated with fractures. *Archives of Pediatrics & Adolescent Medicine,* 162(9): 877-881.
402. Henry MK & Wood JN (2018). Advanced cervical spine imaging in abusive head trauma: An update on recent literature and future directions. *Academic Pediatrics,* 18(7): 733-735.
403. Kemp AM, Joshi AH, Mann M, Tempest V, Liu A, Holden S & Maguire S (2010). What are the clinical and radiological characteristics of spinal injuries from physical abuse: A systematic review. *Archives of Disease in Childhood,* 95(5): 355-360.
404. Barber I, Perez-Rossello JM, Wilson CR, Silvera MV & Kleinman PK (2013). Prevalence and relevance of pediatric spinal fractures in suspected child abuse. *Pediatric Radiology,* 43(11): 1507-1515.

405. Trokel M, Discala C, Terrin NC & Sege RD (2006). Patient and injury characteristics in abusive abdominal injuries. *Pediatric Emergency Care,* 22(10): 700-704.
406. Pawlik MC, Kemp A, Maguire S, Nuttall D, Feldman KW & Lindberg DM (2016). Children with burns referred for child abuse evaluation: Burn characteristics and co-existent injuries. *Child Abuse & Neglect,* 55: 52-61.
407. Kemp AM (2011). Abusive head trauma: Recognition and the essential investigation. *Archives of Disease in Childhood Education & Practice Edition,* 96(6): 202-208.
408. Greeley C (2010). Conditions confused with head trauma. *In:* C J (Ed.) *Child abuse and neglect: Diagnosis, treatment, and evidence* (441-450). Philadelphia: Elsevier Saunders.
409. Maguire SA, Lumb RC, Kemp AM, Moynihan S, Bunting HJ, Watts PO & Adams GG (2013). A systematic review of the differential diagnosis of retinal haemorrhages in children with clinical features associated with child abuse. *Child Abuse Review,* 22(1): 29-43.
410. Girard N, Brunel H, Dory-Lautrec P & Chabrol B (2016). Neuroimaging differential diagnoses to abusive head trauma. *Pediatric Radiology,* 46(5): 603-614.
411. Christian CW & States LJ (2017). Medical mimics of child abuse. *American Journal of Roentgenology,* 208(5): 982-990.
412. David TJ (2008). Non-accidental head injury--the evidence. *Pediatric Radiology,* 38 (Suppl 3): S370-S377.
413. Sirotnak A (2006). Medical disorders that mimic abusive head trauma. *In:* Frasier L (Ed.) *Abusive head trauma in infants and children* (191-226). St. Louis, MO: GW Medical Publishing.
414. Hansen JB, Killough EF, Moffatt ME & Knapp JF (2018). Retinal hemorrhages: Abusive head trauma or not? *Pediatric Emergency Care,* 34(9): 665-670.
415. Sieswerda-Hoogendoorn T, Boos S, Spivack B, Bilo RAC & van Rijn RR (2012). Educational paper: Abusive head trauma part I. Clinical aspects. *European Journal of Pediatrics,* 171(3): 415-423.
416. Kemp A (2002). Investigating subdural haemorrhage in infants. *Archives of Disease in Childhood,* 86(2): 98-102.
417. Chamnanvanakij S, Rollins N & Perlman JM (2002). Subdural hematoma in term infants. *Pediatric Neurology,* 26(4): 301-304.
418. Rooks VJ, Eaton JP, Ruess L, Petermann GW, Keck-Wherley J & Pedersen RC (2008). Prevalence and evolution of intracranial hemorrhage in asymptomatic term infants. *American Journal of Neuroradiology,* 29(6): 1082-1089.
419. Whitby EH, Griffiths PD, Rutter S, Smith MF, Sprigg A, Ohadike P, Davies NP, et al. (2004). Frequency and natural history of subdural haemorrhages in babies and relation to obstetric factors. *The Lancet,* 363(9412): 846-851.
420. Looney CB, Smith JK, Merck LH, Wolfe HM, Chescheir NC, Hamer RM & Gilmore JH (2007). Intracranial hemorrhage in asymptomatic neonates: Prevalence on MR images and relationship to obstetric and neonatal risk factors. *Radiology,* 242(2): 535-541.
421. Watts P, Maguire S, Kwok T, Talabani B, Mann M, Wiener J, Lawson Z, et al. (2013). Newborn retinal hemorrhages: A systematic review. *Journal of AAPOS: American Association for Pediatric Ophthalmology and Strabismus,* 17(1): 70-78.
422. Donaldson JW, Edwards-Brown M & Luerssen TG (2000). Arachnoid cyst rupture with concurrent subdural hygroma. *Pediatric Neurosurgery,* 32(3): 137-139.
423. Vapalahti PM, Schugk P, Leena T & Björkesten Ga (1969). Intracranial arterial aneurysm in a three-month-old infant. *Journal of Neurosurgery,* 30(2): 169-171.
424. El-Ghanem M, Kass-Hout T, Kass-Hout O, Alderazi YJ, Amuluru K, Al-Mufti F, Prestigiacomo CJ, et al. (2016). Arteriovenous malformations in the pediatric population: Review of the existing literature. *Interventional Neurology,* 5(3-4): 218-225.

425. Snedeker JD, Kaplan SL, Dodge PR, Holmes SJ & Feigin RD (1990). Subdural effusion and its relationship with neurologic sequelae of bacterial meningitis in infancy: A prospective study. *Pediatrics,* 86(2): 163-170.
426. Dinakaran S, Chan TKJ, Rogers NK & Brosnahan DM (2002). Retinal hemorrhages in meningococcal septicemia. *Journal of American Association for Pediatric Ophthalmology & Strabismus,* 6(4): 221-223.
427. Mashiyama S, Fukawa O, Mitani S, Ito S, Ito K, Asano S & Sai T (2000). Chronic subdural hematoma associated with malignancy: Report of three cases. *No Shinkei Geka,* 28(2): 173-178.
428. Vernant JC, Delaporte JM, Buisson G, Bellance R, Bokor J & Loiseau P (1988). Cerebrovascular complications of sickle-cell anemia. *Revue Neurologique (Paris),* 144(8-9): 465-473.
429. Choo KE, Tan KK, Chuah SP, Ariffin WA & Gururaj A (1994). Haemorrhagic disease in newborn and older infants: A study in hospitalized children in Kelantan, Malaysia. *Annals of Tropical Paediatrics,* 14(3): 231-237.
430. Gelabert González M, García Allut A, Alvez Gonzalez F, Fernández Seara MJ & Bollar Zabala A (1984). Chronic subdural hematoma in a girl with idiopathic thrombopenic purpura. *Anales Espanoles de Pediatria,* 21(3): 260-262.
431. Wetzel RC, Slater AJ & Dover GJ (1995). Fatal intramuscular bleeding misdiagnosed as suspected nonaccidental injury. *Pediatrics,* 95(5): 771-773.
432. Anderst JD, Carpenter SL, Presley R, Berkoff MC, Wheeler AP, Sidonio RF & Soucie JM (2018). Relevance of abusive head trauma to intracranial hemorrhages and bleeding disorders. *Pediatrics,* 141(5): e20173485.
433. Gago LC, Wegner RK, Capone AJ & Williams GA (2003). Intraretinal hemorrhages and chronic subdural effusions: Glutaric aciduria Type 1 can be mistaken for shaken baby syndrome. *Retina,* 23(5): 724-726.
434. Morris AAM, Hoffmann GF, Naughten ER, Monavari AA, Collins JE & Leonard JV (1999). Glutaric aciduria and suspected child abuse. *Archives of Disease in Childhood,* 80(5): 404-405.
435. Vester ME, Bilo RA, Karst WA, Daams JG, Duijst WL & van Rijn RR (2015). Subdural hematomas: Glutaric aciduria type 1 or abusive head trauma? A systematic review. *Forensic Science Medicine & Pathology,* 11(3): 405-415.
436. Nassogne M-C, Sharrard M, Hertz-Pannier L, Armengaud D, Touati G, Delonlay-Debeney P, Zerah M, et al. (2002). Massive subdural haematomas in Menkes disease mimicking shaken baby syndrome. *Child's Nervous System,* 18(12): 729-731.
437. Ganesh A, Jenny C, Geyer J, Shouldice M & Levin AV (2004). Retinal hemorrhages in type I osteogenesis imperfecta after minor trauma. *Ophthalmology,* 111(7): 1428-1431.
438. Carpenter SL, Abshire TC & Anderst JD (2013). Evaluating for suspected child abuse: Conditions that predispose to bleeding. *Pediatrics,* 131(4): e1357.
439. Dyer C (2017). Baby thought to have been shaken has rare syndrome, family court finds. *The BMJ,* 357: j2136.
440. Castori M (2015). Ehlers–Danlos syndrome(s) mimicking child abuse: Is there an impact on clinical practice? *American Journal of Medical Genetics Part C: Seminars in Medical Genetics,* 169(4): 289-292.
441. Shur N & Carey JC (2015). Genetic differentials of child abuse: Is your case rare or real? *American Journal of Medical Genetics Part C: Seminars in Medical Genetics,* 169(4): 281-288.
442. Stern CM, Pepin MJ, Stoler JM, Kramer DE, Spencer SA & Stein CJ (2017). Musculoskeletal conditions in a pediatric population with Ehlers-Danlos syndrome. *The Journal of Pediatrics,* 181: 261-266.

443. Cabral WA, Makareeva E, Letocha AD, Scribanu N, Fertala A, Steplewski A, Keene DR, et al. (2007). Y-position cysteine substitution in type I collagen (α1(I) R888C/p.R1066C) is associated with osteogenesis imperfecta/Ehlers-Danlos syndrome phenotype. *Human Mutation,* 28(4): 396-405.
444. Malfait F, Symoens S, Goemans N, Gyftodimou Y, Holmberg E, López-González V, Mortier G, et al. (2013). Helical mutations in type I collagen that affect the processing of the amino-propeptide result in an Osteogenesis Imperfecta/Ehlers-Danlos Syndrome overlap syndrome. *Orphanet Journal of Rare Diseases,* 8(1): 78.
445. Handy TC, Hanzlick R, Shields LB, Reichard R & Goudy S (1999). Hypernatremia and subdural hematoma in the pediatric age group: Is there a causal relationship? *Journal of Forensic Sciences,* 44(6): 1114-1118.
446. Ravid S & Maytal J (2003). External hydrocephalus: A probable cause for subdural hematoma in infancy. *Pediatric Neurology,* 28(2): 139-141.
447. Raul J-S, Roth S, Ludes B & Willinger R (2008). Influence of the benign enlargement of the subarachnoid space on the bridging veins strain during a shaking event: A finite element study. *International Journal of Legal Medicine,* 122(4): 337-340.
448. Alper G, Ekinci G, Yilmaz Y, Ankan Ç, Telyar G & Erzen C (1999). Magnetic resonance imaging characteristics of benign macrocephaly in children. *Journal of Child Neurology,* 14(10): 678-682.
449. Haws ME, Linscott L, Thomas C, Orscheln E, Radhakrishnan R & Kline-Fath B (2017). A retrospective analysis of the utility of head computed tomography and/or magnetic resonance imaging in the management of benign macrocrania. *The Journal of Pediatrics,* 182: 283-289.e1.
450. Greiner MV, Richards TJ, Care MM & Leach JL (2013). Prevalence of subdural collections in children with macrocrania. *American Journal of Neuroradiology,* 34(12): 2373-2378.
451. Tucker J, Choudhary AK & Piatt J (2016). Macrocephaly in infancy: Benign enlargement of the subarachnoid spaces and subdural collections. *Journal of Neurosurgery: Pediatrics,* 18(1): 16-20.
452. McKeag H, Christian CW, Rubin D, Daymont C, Pollock AN & Wood J (2013). Subdural hemorrhage in pediatric patients with enlargement of the subarachnoid spaces. *Journal of Neurosurgery: Pediatrics,* 11(4): 438-444.
453. Wilms G, Vanderschueren G, Demaerel PH, Smet MH, Van Calenbergh F, Plets C, Goffin J, et al. (1993). CT and MR in infants with pericerebral collections and macrocephaly: Benign enlargement of the subarachnoid spaces versus subdural collections. *American Journal of Neuroradiology,* 14(4): 855-860.
454. Hansen JB, Frazier T, Moffatt M, Zinkus T & Anderst JD (2018). Evaluations for abuse in young children with subdural hemorrhages: Findings based on symptom severity and benign enlargement of the subarachnoid spaces. *Journal of Neurosurgery: Pediatrics,* 21(1): 31-37.
455. Thomas AG, Hegde SV, Dineen RA & Jaspan T (2013). Patterns of accidental craniocerebral injury occurring in early childhood. *Archives of Disease in Childhood,* 98(10): 787-792.
456. Burrows P, Trefan L, Houston R, Hughes J, Pearson G, Edwards RJ, Hyde P, et al. (2015). Head injury from falls in children younger than 6 years of age. *Archives of Disease in Childhood,* 100(11): 1032-1037.
457. Crowe LM, Catroppa C, Anderson V & Babl FE (2012). Head injuries in children under 3 years. *Injury,* 43(12): 2141-2145.
458. Helfer RE, Slovis TL & Black M (1977). Injuries resulting when small children fall out of bed. *Pediatrics,* 60(4): 533-535.

459. Nimityongskul P & Anderson L (1987). The likelihood of injuries when children fall out of bed. *Journal of Pediatric Orthopedics,* 7: 184-186.
460. Warrington SA, Wright CM & Team AS (2001). Accidents and resulting injuries in premobile infants: Data from the ALSPAC study. *Archives of Disease in Childhood,* 85(2): 104-107.
461. Lyons TJ & Oates RK (1993). Falling out of bed: A relatively benign occurrence. *Pediatrics,* 92(1): 125-127.
462. Tarantino CA, Dowd MD & Murdock TC (1999). Short vertical falls in infants. *Pediatric Emergency Care* 15(1): 5-8.
463. Williams RA (1991). Injuries in infants and small children resulting from witnessed and corroborated free falls. *Journal of Trauma,* 31: 1350-1352.
464. Samuel N, Jacob R, Eilon Y, Mashiach T & Shavit I (2015). Falls in young children with minor head injury: A prospective analysis of injury mechanisms. *Brain Injury,* 29(7-8): 946-950.
465. Chadwick DL, Bertocci G, Castillo E, Frasier L, Guenther E, Hansen K, Herman B, et al. (2008). Annual risk of death resulting from short falls among young children: Less than 1 in 1 million. *Pediatrics,* 121(6): 1213-1224.
466. Tzioumi D & Oates RK (1998). Subdural hematomas in children under 2 years. Accidental or inflicted? A 10-year experience. *Child Abuse & Neglect,* 22(11): 1105-1112.
467. Wells RG, Vetter C & Laud P (2002). Intracranial hemorrhage in children younger than 3 years: Prediction of intent. *Archives of Pediatric & Adolescent Medicine,* 156(3): 252-257.
468. Anderst JD, Carpenter SL & Abshire TC (2013). Evaluation for bleeding disorders in suspected child abuse. *Pediatrics,* 131(4): e1314-e1322.
469. Jackson J, Carpenter S & Anderst J (2012). Challenges in the evaluation for possible abuse: Presentations of congenital bleeding disorders in childhood. *Child Abuse & Neglect,* 36(2): 127-134.
470. Hymel KP, Abshire TC, Luckey DW & Jenny C (1997). Coagulopathy in pediatric abusive head trauma. *Pediatrics,* 99(3): 371-375.
471. Kivlin JD, Currie ML, Greenbaum V, Simons KB & Jentzen J (2008). Retinal hemorrhages in children following fatal motor vehicle crashes: A case series. *Archives of Ophthalmology,* 126(6): 800-804.
472. Gnanaraj L, Gilliland MGF, Yahya RR, Rutka JT, Drake J, Dirks P & Levin AV (2005). Ocular manifestations of crush head injury in children. *Eye,* 21(1): 5-10.
473. Lantz PE, Sinal SH, Stanton CA & Weaver RGJ (2004). Perimacular retinal folds from childhood head trauma. *The BMJ,* 328(7442): 754-756.
474. Lueder GT, Turner JW & Paschall R (2006). Perimacular retinal folds simulating nonaccidental injury in an infant. *Archives of Ophthalmology,* 124(12): 1782-1783.
475. Watts P & Obi E (2008). Retinal folds and retinoschisis in accidental and non-accidental head injury. *Eye,* 22(12): 1514-1516.
476. Binenbaum G, Rogers DL, Forbes BJ, Levin AV, Clark SA, Christian CW, Liu GT, et al. (2013). Patterns of retinal hemorrhage associated with increased intracranial pressure in children. *Pediatrics,* 132(2): e430-e434.
477. Shiau T & Levin AV (2012). Retinal hemorrhages in children: The role of intracranial pressure. *Archives of Pediatrics & Adolescent Medicine,* 166(7): 623-628.
478. Narang SK & Melville JD (2014). Legal issues in child maltreatment. *Pediatric Clinics of North America,* 61(5): 1049-1058.
479. Lane WG & Dubowitz D (2009). Primary care pediatricians' experience, comfort and competence in the evaluation and management of child maltreatment: Do we need child abuse experts? *Child Abuse & Neglect,* 33(2): 76–83.

480. Flaherty EG, Jones R & Sege R (2004). Telling their stories: Primary care practitioners' experience evaluating and reporting injuries caused by child abuse. *Child Abuse & Neglect,* 28(9): 939-945.
481. Beresford EB (1991). Uncertainty and the shaping of medical decisions. *Hastings Center Report,* 21(4): 6-11.
482. Willman KY, Bank DE, Senac M & Chadwick DL (1997). Restricting the time of injury in fatal inflicted head injuries. *Child Abuse & Neglect,* 21(10): 929-940.
483. Biron DL & Shelton D (2007). Functional time limit and onset of symptoms in infant abusive head trauma. *Journal of Paediatrics & Child Health,* 43(1-2): 60-65.
484. Adamsbaum C, Morel B, Ducot B, Antoni G & Rey-Salmon C (2014). Dating the abusive head trauma episode and perpetrator statements: Key points for imaging. *Pediatric Radiology,* 44 (Suppl 4): S578-S588.
485. Sieswerda-Hoogendoorn T, Postema FAM, Verbaan D, Majoie CB & van Rijn RR (2014). Age determination of subdural hematomas with CT and MRI: A systematic review. *European Journal of Radiology,* 83(7): 1257-1268.
486. Postema FAM, Sieswerda-Hoogendoorn T, Majoie CBLM & van Rijn RR (2014). Age determination of subdural hematomas: Survey among radiologists. *Emergency Radiology,* 21(4): 349-358.
487. Binenbaum G, Chen W, Huang J, Ying G-S & Forbes BJ (2016). The natural history of retinal hemorrhage in pediatric head trauma. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus,* 20(2): 131-135.
488. Levin AV (2016). Retinal hemorrhage: Science versus speculation. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus,* 20(2): 93-95.
489. Maguire S, Mann MK, Sibert J & Kemp A (2005). Can you age bruises accurately in children? A systematic review. *Archives of Disease in Childhood,* 90(2): 187-189.
490. Flaherty EG, Sege RD, Griffith J, Price LL, Wasserman R, Slora E, Dhepyasuwan N, et al. (2008). From suspicion of physical child abuse to reporting: Primary care clinician decision-making. *Pediatrics,* 122(3): 611-619.
491. Jones R, Flaherty EG, Binns HJ, Price LL, Slora E, Abney D, Harris DL, et al. (2008). Clinicians' description of factors influencing their reporting of suspected child abuse: Report of the Child Abuse Reporting Experience Study Research Group. *Pediatrics,* 122(2): 259-266.
492. Rangel EL, Cook BS, Bennett BL, Shebesta K, Ying J & Falcone RA (2009). Eliminating disparity in evaluation for abuse in infants with head injury: Use of a screening guideline. *Journal of Pediatric Surgery,* 44(6): 1229-1234.
493. Lane WG, Rubin DM, Monteith R & Christian CW (2002). Racial differences in the evaluation of pediatric fractures for physical abuse. *JAMA,* 288(13): 1603-1609.
494. Lane WG & Dubowitz H (2007). What factors affect the identification and reporting of child abuse-related fractures? *Clinical Orthopaedics & Related Research,* 461: 219-225.
495. Benson DE, Swann A, O'Toole R & Turbett JP (1991). Physicians' recognition of and response to child abuse: Northern Ireland and the U.S.A. *Child Abuse & Neglect,* 15(1): 57-67.
496. Hampton RL & Newberger EH (1985). Child abuse incidence and reporting by hospitals: Significance of severity, class, and race. *American Journal of Public Health,* 75(1): 56-60.
497. Laskey AL, Stump TE, Perkins SM, Zimet GD, Sherman SJ & Downs SM (2012). Influence of race and socioeconomic status on the diagnosis of child abuse: A randomized study. *The Journal of Pediatrics,* 160(6): 1003-1008.e1.
498. Keenan HT, Campbell KA, Page K, Cook LJ, Bardsley T & Olson LM (2017). Perceived social risk in medical decision-making for physical child abuse: A mixed-methods study. *BMC Pediatrics,* 17: 214.

499. Najdowski CJ & Bernstein KM (2018). Race, social class, and child abuse: Content and strength of medical professionals' stereotypes. *Child Abuse & Neglect,* 86: 217-222.
500. Marcovitch H & Mughal MZ (2010). Cases do not support temporary brittle bone disease. *Acta Paediatrica,* 99(4): 485-486.
501. Jaspan T (2008). Current controversies in the interpretation of non-accidental head injury. *Pediatric Radiology,* 38 (Suppl 3): S378-S387.
502. Squier W (2017). Retinal haemorrhage: A red flag for raised intracranial pressure. *Developmental Medicine & Child Neurology,* 59(6): 565.
503. Plunkett J (2001). Fatal pediatric head injuries caused by short-distance falls. *The American Journal of Forensic Medicine & Pathology,* 22(1): 1-12.
504. Gabaeff SC (2013). Investigating the possibility and probability of perinatal subdural hematoma progressing to chronic subdural hematoma, with and without complications, in neonates, and its potential relationship to the misdiagnosis of abusive head trauma. *Legal Medicine,* 15(4): 177-192.
505. Buttram HE (2001). Shaken baby syndrome or vaccine-induced encephalitis? *Medical Sentinel,* 6(3): 83-89.
506. Matsuda M, Matsuda I, Sato M & Handa J (1982). Superior sagittal sinus thrombosis followed by subdural hematoma. *Surgical Neurology,* 18(3): 206-211.
507. Fernando S, Obaldo RE, Walsh IR & Lowe LH (2008). Neuroimaging of nonaccidental head trauma: pitfalls and controversies. *Pediatric Radiology,* 38(8): 827-838.
508. Narang SK, Melville JD, Greeley CS, Anderst JD, Carpenter SL & Spivack B (2013). A daubert analysis of abusive head trauma/shaken baby syndrome — part II: An examination of the differential diagnosis. *Houston Journal of Health, Law & Policy,* 13: 203-327.
509. Campbell KA, Olson LM & Keenan HT (2015). Critical elements in the medical evaluation of suspected child physical abuse. *Pediatrics,* 136(1): 35-43.
510. Burrell T, Moffatt M, Toy S, Nielsen-Parker M & Anderst J (2016). Preliminary development of a performance assessment tool for documentation of history taking in child physical abuse. *Pediatric Emergency Care,* 32(10): 675-681.
511. Pierce MC, Kaczor K & Thompson R (2014). Bringing back the social history. *Pediatric Clinics of North America,* 61(5): 889-905.
512. Society and College of Radiographers & The Royal College of Radiologists (2017). *The radiological investigation of suspected physical abuse in children.* London: The Royal College of Radiologists.
513. Section on Radiology; American Academy of Pediatrics (2009). Diagnostic imaging of child abuse. *Pediatrics,* 123(5): 1430-1435.
514. Wootton-Gorges SL, Soares BP, Alazraki AL, Anupindi SA, Blount JP, Booth TN, Dempsey ME, et al. (2017). ACR Appropriateness Criteria® Suspected Physical Abuse—Child. *Journal of the American College of Radiology,* 14(5, Supplement): S338-S349.
515. National Institute for Health & Care Excellence (2014). *Head injury: Assessment and early management. Clinical guideline 176*. London: National Institute for Health & Care Excellence. Available from: https://www.nice.org.uk/guidance/cg176. Accessed 7 November 2017.
516. Hsieh KL-C, Zimmerman RA, Kao HW & Chen C-Y (2015). Revisiting neuroimaging of abusive head trauma in infants and young children. *American Journal of Roentgenology,* 204(5): 944-952.
517. Foerster BR, Petrou M, Lin D, Thurnher MM, Carlson MD, Strouse PJ & Sundgren PC (2009). Neuroimaging evaluation of non-accidental head trauma with correlation to clinical outcomes: A review of 57 cases. *The Journal of Pediatrics,* 154(4): 573-577.

518. Flom L, Fromkin J, Panigrahy A, Tyler-Kabara E & Berger RP (2016). Development of a screening MRI for infants at risk for abusive head trauma. *Pediatric Radiology,* 46(4): 519-526.
519. Royal College of Paediatrics & Child Health (2013). *Child protection companion.* London: Royal College of Paediatrics & Child Health.
520. Barnes PD & Robson CD (2000). CT findings in hyperacute nonaccidental brain injury. *Pediatric Radiology,* 30(2): 74-81.
521. The Royal College of Paediatrics and Child Health & The Royal College of Ophthalmologists (2013). *Abusive head trauma and the eye in infancy.* London: The Royal College of Ophthalmologists.
522. Christian CW, Levin AV, Council on Child Abuse and Neglect, AAP Section on Ophthalmology, American Association of Certified Orthoptists, American Association for Pediatric Ophthalmology and Strabismus & American Academy of Ophthalmology (2018). The eye examination in the evaluation of child abuse. *Pediatrics,* 142(2): e20181411.
523. Christian C (2018). Child abuse: Evaluation and diagnosis of abusive head trauma in infants and children. *In:* Wiley JF (Ed.) *UpToDate*. Available from: https://www.uptodate.com/contents/child-abuse-evaluation-and-diagnosis-of-abusive-head-trauma-in-infants-and-children. Accessed 9 July 2018.
524. Department for Education (2018). *Working together to safeguard children: A guide to inter-agency working to safeguard and promote the welfare of children*. London: Department for Education. Available from: https://www.gov.uk/government/publications/working-together-to-safeguard-children--2. Accessed 12 September 2018.
525. Royal College of Pathologists & Royal College of Paediatrics & Child Health (2016). *Sudden unexpected death in infancy: Multi-agency guidelines for care and investigation*. London: The Royal College of Pathologists. Available from: https://www.rcpath.org/discover-pathology/news/new-guidelines-for-the-investigation-of-sudden-unexpected-death-in-infancy-launched.html. Accessed 27 May 2017.
526. Newman BS & Dannenfelser PL (2005). Children's protective services and law enforcement: Fostering partnerships in investigations of child abuse. *Journal of Child Sexual Abuse,* 14(2): 97-111.
527. Matthews L, Kemp A & Maguire S (2017). Bruising in children: Exploring the attitudes, knowledge and training of child protection social workers and the interface with paediatricians regarding childhood bruising. *Child Abuse Review,* 26(6): 425-438.
528. Kellogg ND (2014). Working with child protective services and law enforcement: What to expect. *Pediatric Clinics of North America,* 61(5): 1037-1047.
529. Levi BH & Brown G (2005). Reasonable suspicion: A study of Pennsylvania pediatricians regarding child abuse. *Pediatrics,* 116(1): e5-e12.
530. Dias MS, Boehmer S, Johnston-Walsh L & Levi BH (2015). Defining 'reasonable medical certainty' in court: What does it mean to medical experts in child abuse cases? *Child Abuse & Neglect,* 50: 218-227.
531. Flaherty EG, Sege R, Price LL, Christoffel KK, Norton DP & O'Connor KG (2006). Pediatrician characteristics associated with child abuse identification and reporting: Results from a national survey of pediatricians. *Child Maltreatment,* 11(4): 361-369.
532. Lindberg DM, Lindsell CJ & Shapiro RA (2008). Variability in expert assessments of child physical abuse likelihood. *Pediatrics,* 121(4): e945-e953.
533. Laskey AL, Sheridan MJ & Hymel KP (2007). Physicians' initial forensic impressions of hypothetical cases of pediatric traumatic brain injury. *Child Abuse & Neglect,* 31(4): 329–342.

534. Wood JN, Feudtner C, Medina SP, Luan X, Localio R & Rubin DM (2012). Variation in occult injury screening for children with suspected abuse in selected US children's hospitals. *Pediatrics,* 130(5): 853-860.
535. Anderst J, Nielsen-Parker M, Moffatt M, Frazier T & Kennedy C (2016). Using simulation to identify sources of medical diagnostic error in child physical abuse. *Child Abuse & Neglect,* 52: 62-69.
536. Trokel M, Waddimba A, Griffith J & Sege R (2006). Variation in the diagnosis of child abuse in severely injured infants. *Pediatrics,* 117(3): 722-728.
537. Lindberg DM, Beaty B, Juarez-Colunga E, Wood JN & Runyan DK (2015). Testing for abuse in children with sentinel injuries. *Pediatrics,* 136(5): 831-838.
538. Cowley L, Tempest V, Maguire S, Mann M, Naughton A, Wain L & Kemp A (2013). Implementing scientific evidence to improve the quality of Child Protection. *BMJ Quality Improvement Reports,* 2(1): u201101.w703.
539. Anderst J, Kellogg N & Jung I (2009). Is the diagnosis of physical abuse changed when Child Protective Services consults a Child Abuse Pediatrics subspecialty group as a second opinion? *Child Abuse & Neglect,* 33(8): 481-489.
540. Butler-Sloss E & Hall A (2002). Expert witnesses, courts and the law. *Journal of the Royal Society of Medicine,* 95(9): 431-434.
541. Glancy GD & Regehr C (2012). From schadenfreude to contemplation: Lessons for forensic experts. *Journal of the American Academy of Psychiatry & the Law Online,* 40(1): 81-88.
542. Dubowitz H, Christian CW, Hymel K & Kellogg ND (2014). Forensic medical evaluations of child maltreatment: A proposed research agenda. *Child Abuse & Neglect,* 38(11): 1734-1746.
543. Osler W (1930). *Aequanimitas with other addresses.* Philadelphia: Blakiston.
544. Croskerry P (2009). A universal model of diagnostic reasoning. *Academic Medicine,* 84(8): 1022-1028.
545. Kassirer JP, Wong JI & Kopelman R (2010). *Learning clinical reasoning.* Baltimore, MD: Lippincott Williams & Wilkins.
546. Knottnerus JA & Buntinx F (2009). *The evidence base of clinical diagnosis: Theory and methods of diagnostic research.* Oxford: Blackwell Publishing Ltd.
547. Sox HC, Higgins MC & Owens DK (2013). *Medical decision making.* Oxford: John Wiley & Sons.
548. Walters G (2013). Teaching and learning clinical decision-making for person-centered medicine: Recommendations from a systematic review of the literature. *European Journal for Person-Centred Healthcare,* 1(1): 112-123.
549. Elstein AS & Schwartz A (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *The BMJ,* 324(7339): 729-732.
550. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M & Moons KG (2012). Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One,* 7(2): e32844.
551. Henifin MS, Kipen HM & Poulter SR (2000). Reference guide on medical testimony. *In:* Federal Judicial Center (Ed.) *Reference manual on scientific evidence* (439-484). Washington, D.C.: Federal Judician Center.
552. Kassirer JP (2010). Teaching clinical reasoning: Case-based and coached. *Academic Medicine,* 85(7): 1118-1124.
553. Moons KG, de Groot JA, Linnet K, Reitsma JB & Bossuyt PM (2012). Quantifying the added value of a diagnostic test or marker. *Clinical Chemistry,* 58(10): 1408-1417.
554. Moons KG, van Es GA, Michel BC, Buller HR, Habbema JD & Grobbee DE (1999). Redundancy of single diagnostic test evaluation. *Epidemiology,* 10(3): 276-281.

555. Moons KG & Grobbee DE (2002). Diagnostic studies as multivariable, prediction research. *Journal of Epidemiology & Community Health,* 56(5): 337-338.
556. Moons KG, Biesheuvel CJ & Grobbee DE (2004). Test research versus diagnostic research. *Clinical Chemistry,* 50(3): 473-476.
557. Hatch S (2017). Uncertainty in medicine. *The BMJ,* 357: j2180.
558. Peabody F (1922). The physician and the laboratory. *Boston Medical & Surgical Journal,* 187: 324-327.
559. Groopman J (2007). *How doctors think.* Boston: Houghton Mifflin.
560. Croskerry P (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine,* 78(8): 775-780.
561. Risinger DM, Saks MJ, Thompson WC & Rosenthal R (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review,* 90(1): 1-56.
562. McClellan MB, McGinnis M, Nabel EG & Olsen LM (2007). *Evidence-based medicine and the changing nature of health care: 2007 IOM annual meeting summary.* Washington, DC: The National Academies Press.
563. Sackett DL, Strauss SE, Richardson WS, Rosenberg W & Haynes RB (2000). *Evidence-Based medicine: How to practice and teach EBM.* Edinburgh: Churchill Livingstone.
564. Post SG. Encyclopaedia of Bioethics. New York, NY: Macmillan; 2004.
565. Patel VL, Arocha JF & Kaufman DR (1999). Expertise and tacit knowledge in medicine. *In:* Sternberg RJ & Horvath JA (Eds.) *Tacit knowledge in professional practice: Researcher and practitioner perspectives* (75-99). Mahwah, NJ Lawrence Erlbaum Associates.
566. Cradock G (2011). Thinking Goudge: Fatal child abuse and the problem of uncertainty. *Current Sociology,* 59(3): 362-378.
567. Elstein AS, Shulman LS & Sprafka SA (1978). *Medical problem solving: An analysis of clinical reasoning.* Cambridge, MA: Harvard University Press.
568. Elstein A, Kagan N, Shulman LS, Jason H & MJ. L (1972). Methods and theory in the study of medical inquiry. *Journal of Medical Education,* 47(2): 85-92.
569. Heneghan C, Glasziou P, Thompson M, Rose P, Balla J, Lasserson D, Scott C, et al. (2009). Diagnostic strategies used in primary care. *The BMJ,* 338: b946.
570. Glick JC & Staley K (2007). Inflicted traumatic brain injury: Advances in evaluation and collaborative diagnosis. *Pediatric Neurosurgery,* 43(5): 436-441.
571. Norman G (2005). Research in clinical reasoning: Past history and current trends. *Medical Education,* 39(4): 418-427.
572. Bordage G (2007). Prototypes and semantic qualifiers: From past to present. *Medical Education,* 41(12): 1117-1121.
573. Schmidt HG, Norman GR & Boshuizen HP (1990). A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine,* 65(10): 611-621.
574. Schmidt HG & Rikers RMJP (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education,* 41(12): 1133-1139.
575. Owens D & Sox HC (2001). Biomedical decision making: Probabilistic clinical reasoning. *In:* Shortliffe EH & Perreault LE (Eds.) *Medical informatics: Computer applications in healthcare and biomedicine* (76-131). London: Springer-Verlag.
576. Sackett DL, Haynes RB, Guyatt GH & Tugwell P (1991). *Clinical epidemiology: A basic science for clinical medicine.* Boston: Little Brown.
577. Hunink M, Glasziou P, Siegel J, Weeks J, Pliskin J, Elstein AS & Weinstein MC (2001). *Decision making in health and medicine: Integrating evidence and values.* Cambridge: Cambridge University Press.
578. Lusted LB (1968). *Introduction to medical decision making.* Springfield, IL: Charles C Thomas.

579. Bayes TR (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions,* 53: 370-418.
580. Fahey T & van der Lei J (2009). Producing and using clinical prediction rules. *In:* Knottnerus JA & Buntinx F (Eds.) *The evidence base of clinical diagnosis: Theory and methods of diagnostic research* (213-236). Oxford: Blackwell Publishing Ltd.
581. Whiting PF, Davenport C, Jameson C, Burke M, Sterne JAC, Hyde C & Ben-Shlomo Y (2015). How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open,* 5: e008155.
582. Simel DL & Rennie D (1997). The clinical examination an agenda to make it more rational. *JAMA,* 277(7): 572-574.
583. Whiting P, Martin RM, Ben-Shlomo Y, Gunnell D & Sterne JAC (2013). How to apply the results of a research paper on diagnosis to your patient. *Journal of the Royal Society of Medicine Short Reports,* 4(1): 7.
584. Knottnerus JA (1995). Diagnostic prediction rules: Principles, requirements and pitfalls. *Primary Care,* 22(2): 341-363.
585. Pauker SG & Kassirer JP (1980). The threshold approach to clinical decision making. *New England Journal of Medicine,* 302(20): 1109-1117.
586. National Institute for Health & Care Excellence (2017). *Child maltreatment: When to suspect maltreatment in under 18s*. NICE Clinical Guideline 89. Available from: http://www.nice.org.uk/guidance/cg89/resources/guidance-when-to-suspect-child-maltreatment-pdf. London: NICE.
587. Hymel KP. Development of a clinical prediction rule for pediatric abusive head trauma. Fifteenth International Conference on Shaken Baby Syndrome/Abusive Head Trauma; 27 September Montreal, Quebec, 25-27 September 2016.
588. Wong JB, Gostin LO & Cabrera OA (2011). Reference guide on medical testimony. *In:* Federal Judicial Center (Ed.) *Reference manual on scientific evidence* (687-745). Washington, D.C.: The National Academies Press.
589. Kuipers B & Kassirer JP (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science,* 8(4): 363-385.
590. Kassirer JP & Cecil JS (2002). Inconsistency in evidentiary standards for medical testimony: Disorder in the courts. *JAMA,* 288(11): 1382-1387.
591. Marcum JA (2012). An integrated model of clinical reasoning: Dual-process theory of cognition and metacognition. *Journal of Evaluation in Clinical Practice,* 18(5): 954-961.
592. Pelaccia T, Tardif J, Triby E & Charlin B (2011). An analysis of clinical reasoning through a recent and comprehensive approach: The dual-process theory. *Medical Education Online,* 16: 10.3402/meo.v16i0.5890.
593. Kahneman D (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American Psychologist,* 58(9): 697-720.
594. Croskerry P, Singhal G & Mamede S (2013). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality & Safety,* 22(Suppl 2): ii58-ii64.
595. Stolper E, Van de Wiel M, Van Royen P, Van Bokhoven M, Van der Weijden T & Dinant GJ (2011). Gut feelings as a third track in general practitioners' diagnostic reasoning. *Journal of General Internal Medicine,* 26(2): 197-203.
596. Hammond K (2000). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* New York: Oxford University Press.
597. Eva KW, Hatala RM, Leblanc VR & Brooks LR (2007). Teaching from the clinical reasoning literature: Combined reasoning strategies help novice diagnosticians overcome misleading information. *Medical Education,* 41(12): 1152-1158.
598. Norman GR & Eva KW (2010). Diagnostic error and clinical reasoning. *Medical Education,* 44(1): 94-100.

599. Kulatunga-Moruzi C, Brooks LR & Norman GR (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching & Learning in Medicine,* 13(2): 110-116.
600. Ark TK, Brooks LR & Eva KW (2006). Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Academic Medicine,* 81(4): 405-409.
601. Balla JI, Heneghan C, Glasziou P, Thompson M & Balla ME (2009). A model for reflection for good clinical practice. *Journal of Evaluation in Clinical Practice,* 15(6): 964-969.
602. Croskerry P (2009). Context is everything or how could I have been that stupid? *Healthcare Quarterly (Toronto, Ont),* 12(Sp): e171-e176.
603. Tversky A & Kahneman D (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185(4157): 1124-1131.
604. Croskerry P (2006). Critical thinking and decisionmaking: Avoiding the perils of thin-slicing. *Annals of Emergency Medicine,* 48(6): 720-722.
605. Klein JG (2005). Five pitfalls in decisions about diagnosis and prescribing. *The BMJ,* 330(7494): 781-783.
606. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG & Mamede S (2017). The causes of errors in clinical reasoning: Cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine,* 92(1): 23-30.
607. Evans JS & Stanovich KE (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science,* 8(3): 223-241.
608. Wilson TD & Schooler JW (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality & Social Psychology,* 60(2): 181-192.
609. Evans JS (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology,* 59: 255-278.
610. Graber ML, Franklin N & Gordon R (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine,* 165(13): 1493-1499.
611. Jenicek M (2011). *Medical error and harm: Understanding, prevention and control.* New York: Productivity Press.
612. Croskerry P, Abbass AA & Wu AW (2008). How doctors feel: Affective issues in patients' safety. *The Lancet,* 372(9645): 1205-1206.
613. Arkes HR (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin,* 110(3): 486-498.
614. Noguchi Y, Matsui K, Imura H, Kiyota M & Fukui T (2002). Quantitative evaluation of the diagnostic thinking process in medical students. *Journal of General Internal Medicine,* 17(11): 848-853.
615. Puhan MA, Steurer J, Bachmann LM & ter Riet G (2005). A randomized trial of ways to describe test accuracy: The effect on physicians' post-test probability estimates. *Annals of Internal Medicine,* 143(3): 184-189.
616. Edwards W (1982). Conservatism in human information processing. *In:* Kahneman D, Slovic P & Tversky A (Eds.) *Judgement under uncertainty: Heuristics and biases* (359-369). Cambridge: Cambridge University Press.
617. Schwartz A & Elstein AS (2009). Clinical problem solving and diagnostic decision making: A selective review of the cognitive literature. *In:* Knottnerus JA & Buntinx F (Eds.) *The evidence base of clinical diagnosis: Theory and methods of diagnostic research* (237-255). Oxford: Blackwell Publishing Ltd.
618. Sanders S (2015). *Clinical prediction rules for assisting diagnosis (doctoral thesis).* Faculty of Heath Sciences & Medicine, Bond University, Australia.

619. Dawes RM, Faust D & Meehl PE (1989). Clinical versus actuarial judgment. *Science,* 243(4899): 1668-1674.

620. Laskey AL (2014). Cognitive errors: Thinking clearly when it could be child maltreatment. *Pediatric Clinics of North America,* 61(5): 997-1005.

621. Skellern C (2015). Minimising bias in the forensic evaluation of suspicious paediatric injury. *Journal of Forensic & Legal Medicine,* 34: 11-16.

622. Keenan HT, Cook LJ, Olson LM, Bardsley T & Campbell KA (2017). Social intuition and social information in physical child abuse evaluation and diagnosis. *Pediatrics,* 140(5): e20171188.

623. Jackson J, Miller M, Moffatt M, Carpenter S, Sherman A & Anderst J (2015). Bruising in children: Practice patterns of pediatric hematologists and child abuse pediatricians. *Clinical Pediatrics,* 54(6): 563-569.

624. Graber ML, Kissam S, Payne VL, Meyer AND, Sorensen A, Lenfestey N, Tant E, et al. (2012). Cognitive interventions to reduce diagnostic error: A narrative review. *BMJ Quality & Safety,* 21(7): 535-557.

625. Croskerry P, Singhal G & Mamede S (2013). Cognitive debiasing 2: Impediments to and strategies for change. *BMJ Quality & Safety,* 22(Suppl 2): ii65-ii72.

626. Croskerry P (2003). Cognitive forcing strategies in clinical decision making. *Annals of Emergency Medicine,* 41(1): 110-120.

627. Falk G & Fahey T (2009). Clinical prediction rules. *The BMJ,* 339: b2899.

628. Grimes DA & Schulz KF (2005). Refining clinical diagnosis with likelihood ratios. *The Lancet,* 365(9469): 1500-1555.

629. Adams ST & Leveson SH (2012). Clinical prediction rules. *The BMJ,* 344: d8312.

630. Beattie P & Nelson R (2006). Clinical prediction rules: What are they and what do they tell us? *The Australian Journal of Physiotherapy,* 52(3): 157-163.

631. Hendriksen JM, Geersing GJ, Moons KG & de Groot JA (2013). Diagnostic and prognostic prediction models. *Journal of Thrombosis & Haemostasis,* 11(Suppl 1): 129-141.

632. Pace N, Eberhart L & Kranke P (2012). Quantifying prognosis with risk predictions. *European Journal of Anaesthesiology,* 29(1): 7-16.

633. Steyerberg EW & Vergouwe Y (2014). Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal,* 35(29): 1925-1931.

634. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, et al. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology,* 21(1): 128-138.

635. Upshur R & Chin-Yee B (2017). Clinical judgment. *In:* Solomon M, Simon JR & Kincaid H (Eds.) *The Routledge companion to philosophy of medicine* (363-371). New York: Routledge.

636. Wasson JH, Sox HC, Neff RK & Goldman L (1985). Clinical prediction rules. Applications and methodological standards. *The New England Journal of Medicine,* 313(13): 793-798.

637. Ægisdóttir S, White MJ, Spengler PM, Maugherman AS, Anderson LA, Cook RS, Nichols CN, et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counselling Psychologist,* 34(3): 341-382

638. Grove WM, Zald DH, Lebow BS, Snitz BE & Nelson C (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment,* 12(1): 19-30.

639. Meehl P (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

640. van der Put CE, Assink M & Boekhout van Solinge NF (2017). Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. *Child Abuse & Neglect,* 73: 71-88.
641. D'Andrade A, Austin MJ & Benton A (2008). Risk and safety assessment in child welfare: Instrument comparisons. *Journal of Evidence-Based Social Work,* 5(1-2): 31-56.
642. Gigerenzer G (2008). *Rationality for mortals: How people cope with uncertainty.* New York: Oxford University Press.
643. Sanders S, Doust J & Glasziou P (2015). A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. *PLoS One,* 10(6): e0128233.
644. Bossuyt PMM & McCaffery K (2009). Additional patient outcomes and pathways in evaluations of testing. *Medical Decision Making,* 29(5): E30-E38.
645. Parikh R, Mathai A, Parikh S, Chandra Sekhar G & Thomas R (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology,* 56(1): 45-50.
646. Søreide K (2009). Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. *Journal of Clinical Pathology,* 62(1): 1-5.
647. Ebell MH, Locatelli I & Senn N (2015). A novel approach to the determination of clinical decision thresholds. *BMJ Evidence-Based Medicine,* 20(2): 41-47.
648. Green SM, Schriger DL & Yealy DM (2014). Methodologic standards for interpreting clinical decision rules in emergency medicine: 2014 update. *Annals of Emergency Medicine,* 64(3): 286-291.
649. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW & Moons KG (2015). A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of Clinical Epidemiology,* 68(3): 279-289.
650. McGinn T (2016). Putting meaning into meaningful use: A roadmap to successful integration of evidence at the point of care. *JMIR Medical Informatics,* 4(2): e16.
651. Kappen TH, Vergouwe Y, van Wolfswinkel L, Kalkman CJ, Moons KG & van Klei WA (2015). Impact of adding therapeutic recommendations to risk assessments from a prediction model for postoperative nausea and vomiting. *British Journal of Anaesthesia,* 114(2): 252-260.
652. Michie S & Johnston M (2004). Changing clinical behaviour by making guidelines specific. *The BMJ,* 328(7435): 343-345.
653. Campbell MK, Elbourne DR & Altman DG (2004). CONSORT statement: Extension to cluster randomised trials. *The BMJ,* 328(7441): 702-708.
654. Hemming K, Haines TP, Chilton PJ, Girling AJ & Lilford RJ (2015). The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting. *The BMJ,* 350: h391.
655. Poldervaart JM, Reitsma JB, Koffijberg H, Backus BE, Six AJ, Doevendans PA & Hoes AW (2013). The impact of the HEART risk score in the early assessment of patients with acute chest pain: Design of a stepped wedge, cluster randomised trial. *BMC Cardiovascular Disorders,* 13: 77.
656. Reilly BM, Evans AT, Schaider JJ, Das K, Calvin JE, Moran LA, Roberts RR, et al. (2002a). Impact of a clinical decision rule on hospital triage of patients with suspected acute cardiac ischemia in the emergency department. *JAMA,* 288(3): 342-350.
657. Pearson SD, Goldman L, Garcia TB, Cook EF & Lee TH (1994). Physician response to a prediction rule for the triage of emergency department patients with chest pain. *Journal of General Internal Medicine,* 9(5): 241-247.
658. Graham ID, Stiell IG, Laupacis A, O'Connor AM & Wells GA (1998). Emergency physicians' attitudes toward and use of clinical decision rules for radiography. *Academic Emergency Medicine,* 5(2): 134-140.

659. Brehaut JC, Stiell IG, Visentin L & Graham ID (2005). Clinical decision rules "in the real world": How a widely disseminated rule is used in everyday practice. *Academic Emergency Medicine,* 12(10): 948-956.
660. Brehaut JC, Stiell IG & Graham ID (2006). Will a new clinical decision rule be widely used? The case of the Canadian C-spine rule. *Academic Emergency Medicine,* 13(4): 413-420.
661. Eichler K, Zoller M, Tschudi P & Steurer J (2007). Barriers to apply cardiovascular prediction rules in primary care: A postal survey. *BMC Family Practice,* 8: 1.
662. Runyon MS, Richman PB & Kline JA (2007). Emergency medicine practitioner knowledge and use of decision rules for the evaluation of patients with suspected pulmonary embolism: Variations by practice setting and training level. *Academic Emergency Medicine,* 14(1): 53-57.
663. Boutis K, Constantine E, Schuh S, Pecaric M, Stephens D & Narayanan UG (2010). Pediatric emergency physician opinions on ankle radiograph clinical decision rules. *Academic Emergency Medicine,* 17(7): 709-717.
664. Keogh C & Fahey T (2010). Clinical prediction rules in primary care: What can be done to maximise their implementation? *Clinical Evidence,* Available from: http://clinicalevidence.bmj.com/downloads/05-10-10.pdf.
665. van der Steen JT, Albers G, Licht-Strunk E, Muller MT & Ribbe MW (2011). A validated risk score to estimate mortality risk in patients with dementia and pneumonia: Barriers to clinical impact. *International Psychogeriatrics,* 23(1): 31-43.
666. Beutel BG, Trehan SK, Shalvoy RM & Mello MJ (2012). The Ottawa knee rule: Examining use in an academic emergency department. *The Western Journal of Emergency Medicine,* 13(4): 366-372.
667. Ballard DW, Rauchwerger AS, Reed ME, Vinson DR, Mark DG, Offerman SR, Chettipally UK, et al. (2013). Emergency physicians' knowledge and attitudes of clinical decision support in the electronic health record: A survey-based study. *Academic Emergency Medicine,* 20(4): 352-360.
668. Sheehan B, Nigrovic LE, Dayan PS, Kuppermann N, Ballard DW, Alessandrini E, Bajaj L, et al. (2013). Informing the design of clinical decision support services for evaluation of children with minor blunt head trauma in the emergency department: A sociotechnical analysis. *Journal of Biomedical Informatics,* 46(5): 905-913.
669. Haskins R, Osmotherly PG, Southgate E & Rivett DA (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy,* 19(2): 142-151.
670. Haskins R, Osmotherly PG, Southgate E & Rivett DA (2015). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: A qualitative study. *Physiotherapy,* 101(1): 44-49.
671. Pluddemann A, Wallace E, Bankhead C, Keogh C, Van der Windt D, Lasserson D, Galvin R, et al. (2014). Clinical prediction rules in practice: Review of clinical guidelines and survey of GPs. *British Journal of General Practice,* 64(621): e233-e242.
672. Kappen TH, van Loon K, Kappen MA, van Wolfswinkel L, Vergouwe Y, van Klei WA, Moons KG, et al. (2016). Barriers and facilitators perceived by physicians when using prediction models in practice. *Journal of Clinical Epidemiology,* 70: 136-145.
673. Kelly J, Sterling M, Rebbeck T, Bandong AN, Leaver A, Mackey M & Ritchie C (2017). Health practitioners' perceptions of adopting clinical prediction rules in the management of musculoskeletal pain: A qualitative study in Australia. *BMJ Open,* 7(8): e015916.
674. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA & Rubin HR (1999). Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA,* 282(15): 1458-1465.

675. Collins GS & Moons KG (2012). Comparing risk prediction models. *The BMJ,* 344: e3186.
676. Atabaki SM, Hoyle JDJ, Schunk JE, Monroe DJ, Alpern ER, Quayle KS, Glass TF, et al. (2016). Comparison of prediction rules and clinician suspicion for identifying children with clinically important brain injuries after blunt head trauma. *Academic Emergency Medicine,* 23(5): 566-575.
677. Mahajan P, Kuppermann N, Tunik M, Yen K, Atabaki SM, Lee LK, Ellison AM, et al. (2015). Comparison of clinician suspicion versus a clinical prediction rule in identifying children at risk for intra-abdominal injuries after blunt torso trauma. *Academic Emergency Medicine,* 22(9): 1034-1041.
678. Reilly BM, Evans AT, Schaider JJ & Wang Y (2002). Triage of patients with chest pain in the emergency department: A comparative study of physicians' decisions. *The American Journal of Medicine,* 112(2): 95-103.
679. Broekhuizen BD, Sachs A, Janssen K, Geersing GJ, Moons K, Hoes A & Verheij T (2011). Does a decision aid help physicians to detect chronic obstructive pulmonary disease? *The British Journal of General Practice,* 61(591): e674-e679.
680. Schriger DL & Newman DH (2012). Medical decisionmaking: Let's not forget the physician. *Annals of Emergency Medicine,* 59(3): 219-220.
681. Finnerty N, Rodriguez R, Carpenter C, Sun B, Theyyunni N, Ohle R, Dodd K, et al. (2015). Clinical decision rules for diagnostic imaging in the emergency department: A research agenda. *Academic Emergency Medicine,* 22(12): 1406-1416.
682. Benger JR & Pearce AV (2002). Simple intervention to improve detection of child abuse in emergency departments. *The BMJ,* 324(7340): 780-782.
683. Ezpeleta L, Perez-Robles R, Fanti KA, Karveli V, Katsimicha E, Nikolaidis G, Hadjicharalambous MZ, et al. (2017). Development of a screening tool enabling identification of infants and toddlers at risk of family abuse and neglect: A feasibility study from three South European countries. *Child Care Health & Development,* 43(1): 75-80.
684. Milani GP, Vianello FA, Cantoni B, Agostoni C & Fossali EF (2016). New program for identification of child maltreatment in emergency department: Preliminary data. *Italian Journal of Pediatrics,* 42: 66.
685. Gonzalez DO & Deans KJ (2017). Hospital-based screening tools in the identification of non-accidental trauma. *Seminars in Pediatric Surgery,* 26(1): 43-46.
686. Escobar MA, Jr., Pflugeisen BM, Duralde Y, Morris CJ, Haferbecker D, Amoroso PJ, Lemley H, et al. (2016). Development of a systematic protocol to identify victims of non-accidental trauma. *Pediatric Surgery International,* 32(4): 377-386.
687. Louwers E, Korfage IJ, Affourtit MJ, Scheewe DJH, van de Merwe MH, Vooijs-Moulaert F, Woltering CMC, et al. (2011). Detection of child abuse in emergency departments: A multi-centre study. *Archives of Disease in Childhood,* 96(5): 422-425.
688. Schouten MC, van Stel HF, Verheij TJ, Nieuwenhuis EE & van de Putte EM (2016). A screening protocol for child abuse at out-of-hours primary care locations: A descriptive study. *BMC Family Practice,* 17(1): 155.
689. Hoytema van Konijnenburg EMM, Teeuw AH, Sieswerda-Hoogendoorn T, Leenders AGE & van der Lee JH (2013). Insufficient evidence for the use of a physical examination to detect maltreatment in children without prior suspicion: A systematic review. *Systematic Reviews,* 2: 109.
690. Carter YH, Bannon MJ, Limbert C, Docherty A & Barlow J (2006). Improving child protection: A systematic review of training and procedural interventions. *Archives of Disease in Childhood,* 91(9): 740-743.

691. Woodman J, Lecky F, Hodes D, Pitt M, Taylor B & Gilbert R (2010). Screening injured children for physical abuse or neglect in emergency departments: A systematic review. *Child Care Health & Development,* 36(2): 153-164.
692. Woodman J, Pitt M, Wentz R, Taylor B, Hodes D & Gilbert RE (2008). Performance of screening tests for child physical abuse in accident and emergency departments. *Health Technology Assessment,* 12(33): 1-95.
693. Bailhache M, Leroy V, Pillet P & Salmi LR (2013). Is early detection of abused children possible? A systematic review of the diagnostic accuracy of the identification of abused children. *BMC Pediatrics,* 13: 202.
694. Louwers EC, Affourtit MJ, Moll HA, de Koning HJ & Korfage IJ (2010). Screening for child abuse at emergency departments: A systematic review. *Archives of Disease in Childhood,* 95(3): 214-218.
695. Hoft M & Haddad L (2017). Screening children for abuse and neglect: A review of the literature. *Journal of Forensic Nursing,* 13(1): 26-34.
696. Schouten MC, van Stel HF, Verheij TJ, Houben ML, Russel IM, Nieuwenhuis EE & van de Putte EM (2017). The value of a checklist for child abuse in out-of-hours primary care: To screen or not to screen. *PLoS One,* 12(1): e0165641.
697. Dunstan FD, Guildea ZE, Kontos K, Kemp AM & Sibert JR (2002). A scoring system for bruise patterns: A tool for identifying abuse. *Archives of Disease in Childhood,* 86(5): 330-333.
698. Pierce MC, Kaczor K, Aldridge S, O'Flynn J & Lorenz DJ (2010). Bruising characteristics discriminating physical child abuse from accidental trauma. *Pediatrics,* 125(1): 67-74.
699. Kemp AM, Hollén L, Emond AM, Nuttall D, Rea D & Maguire S (2017). Raising suspicion of maltreatment from burns: Derivation and validation of the BuRN-Tool. *Burns,* 44(2): 335-343.
700. Berger R, Pak BJ, Kolesnikova MD, Fromkin J, Saladino R, Herman BE, Pierce MC, et al. (2017). Derivation and validation of a serum biomarker panel to identify infants with acute intracranial hemorrhage. *JAMA Pediatrics,* 171(6): e170429.
701. Berger RP, Fromkin J, Herman B, Pierce MC, Saladino RA, Flom L, Tyler-Kabara EC, et al. (2016). Validation of the Pittsburgh infant brain injury score for abusive head trauma. *Pediatrics,* 138(1): e20153756.
702. Hymel KP, Armijo-Garcia V, Foster R, Frazier TN, Stoiko M, Christie LM, Harper NS, et al. (2014). Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics,* 134(6): e1537-e1544.
703. Minns RA, Jones PA, Tandon A, Fleck BW, Mulvihill AO & Elton RA (2012). Prediction of inflicted brain injury in infants and children using retinal imaging. *Pediatrics,* 130(5): e1227-e1234.
704. Scavarda D, Gabaudan C, Ughetto F, Lamy F, Imada V, Lena G & Paut O (2010). Initial predictive factors of outcome in severe non-accidental head trauma in children. *Child's Nervous System,* 26(11): 1555-1561.
705. Hymel KP, Herman BE, Narang SK, Graf JM, Frazier TN, Stoiko M, Christie LM, et al. (2015). Potential impact of a validated screening tool for pediatric abusive head trauma. *The Journal of Pediatrics,* 167(6): 1375-1381.
706. Pfeiffer H, Crowe L, Kemp AM, Cowley LE, Smith AS & Babl FE (2018). Clinical prediction rules for abusive head trauma: A systematic review. *Archives of Disease in Childhood,* 103(8): 776-783.
707. Best N, Ashby D, Dunstan F, Foreman D & McIntosh N (2013). A Bayesian approach to complex clinical diagnoses: A case-study in child abuse. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 176(1): 53-96.

708. Kemp AM, Kemp KW, Evans R, Murray L, Guildea ZES, Dunstan FDJ & Sibert JR (1998). Diagnosing physical abuse using Bayes' theorem: A preliminary study. *Child Abuse Review,* 7(3): 178-188.
709. O'Connor GT & Sox HC (1991). Bayesian reasoning in medicine: The contributions of Lee B. Lusted, MD. *Medical Decision Making,* 11(2): 107-111.
710. Colbourne M (2016). Abusive head trauma: Evolution of a diagnosis. *BC Medical Journal,* 57(8): 331-335.
711. Hoskote A, Richards P, Anslow P & McShane T (2002). Subdural haematoma and non-accidental head injury in children. *Child's Nervous System,* 18(6-7): 311-317.
712. Ewing-Cobbs L, Prasad M, Kramer L, Louis PT, Baumgartner J, Fletcher JM & Alpert B (2000). Acute neuroradiologic findings in young children with inflicted or noninflicted traumatic brain injury. *Child's Nervous System,* 16(1): 25-33.
713. Hilbe J (2009). *Logistic regression models.* Boca Raton, FL: Chapman & Hall/CRC.
714. Vergouwe Y, Royston P, Moons KG & Altman DG (2010). Development and validation of a prediction model with missing predictor data: A practical approach. *Journal of Clinical Epidemiology,* 63(2): 205-214.
715. Little RJA & Rubin DB (2002). *Statistical analysis with missing data.* New York: John Wiley & Sons.
716. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *The BMJ,* 338: b2393.
717. Moons KGM, Donders RART, Stijnen T & Harrell FE (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology,* 59(10): 1092-1101.
718. Donders ART, van der Heijden GJMG, Stijnen T & Moons KGM (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology,* 59(10): 1087-1091.
719. Rubin DB (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
720. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L & Petersen I (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology,* 9: 157-166.
721. van Buuren S & Groothuis-Oudshoorn K (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software,* 45(3): 1-67.
722. White IR, Royston P & Wood AM (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine,* 30(4): 377-399.
723. Harel O, Mitchell EM, Perkins NJ, Cole SR, Tchetgen Tchetgen EJ, Sun B & Schisterman EF (2018). Multiple imputation for incomplete data in epidemiologic studies. *American Journal of Epidemiology,* 187(3): 576-584.
724. Murray JS (2018). Multiple imputation: A review of practical and theoretical findings. *Statistical Science,* 33(2): 142-159.
725. Andridge RR & Little RJA (2010). A review of hot deck imputation for survey non-response. *International Statistical Review,* 78(1): 40-64.
726. Myers TA (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods & Measures,* 5(4): 297-310.
727. Vittinghoff E (2005). *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models.* New York: Springer.
728. Vickers AJ, Cronin AM, Elkin EB & Gonen M (2008). Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Medical Informatics & Decision Making,* 8: 53.

729. Toll DB, Janssen KJ, Vergouwe Y & Moons KG (2008). Validation, updating and impact of clinical prediction rules: A review. *Journal of Clinical Epidemiology,* 61(11): 1085-1094.
730. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE & Moons KG (2003). External validation is necessary in prediction research: A clinical example. *Journal of Clinical Epidemiology,* 56(9): 826-832.
731. Altman DG & Royston P (2000). What do we mean by validating a prognostic model? *Statistics in Medicine,* 19(4): 453-473.
732. Hanley JA & McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology,* 143(1): 29-36.
733. Harrell F (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis.* New York: Springer.
734. Cohrs AC, Agbese E, Leslie DL & Hymel KP (2017). A cost analysis of a validated screening tool for pediatric abusive head trauma. *Journal of Child & Family Studies,* 27(3): 750-755.
735. Pfeiffer H, Smith A, Kemp AM, Cowley LE, Cheek JA, Dalziel SR, Borland ML, et al. (2018). External validation of the PediBIRN clinical prediction rule for abusive head trauma. *Pediatrics,* 141(5): e20173674.
736. Berger R & McGinn T (2013). Deciding whether to screen for abusive head trauma: Do we need a clinical decision rule? *Pediatric Critical Care Medicine,* 14(2): 230-231.
737. Maguire S, Cowley L, Farewell D & Kemp A (2016). Theoretical re-analysis of two previously published datasets. *The Journal of Pediatrics,* 171: 321.
738. Janssen KJ, Vergouwe Y, Donders AR, Harrell FE, Jr., Chen Q, Grobbee DE & Moons KG (2009). Dealing with missing predictor values when applying clinical prediction models. *Clinical Chemistry,* 55(5): 994-1001.
739. Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ & Moons KG (2012). Adaptation of clinical prediction models for application in local settings. *Medical Decision Making,* 32(3): E1-E10.
740. Sullivan LM, Massaro JM & D'Agostino RB (2004). Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Statistics in Medicine,* 23(10): 1631-1660.
741. Jaja BNR, Saposnik G, Lingsma HF, Macdonald E, Thorpe KE, Mamdani M, Steyerberg EW, et al. (2018). Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: The SAHIT multinational cohort study. *The BMJ,* 360: j5745
742. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
743. Chang W, Cheng J, Allaire JJ, Xie Y & McPherson J (2015). Shiny: Web application framework for R. R package version 0.11.1. http://CRAN.R-project.org/package=shiny.
744. RStudio Team. RStudio: Integrated Development for R. 2015.
745. Wojciechowski J, Hopkins A & Upton R (2015). Interactive pharmacometric applications using R and the shiny package. *CPT: Pharmacometrics & Systems Pharmacology,* 4(3): 146-159.
746. Ji X & Kattan MW (2018). Tutorial: Development of an online risk calculator platform. *Annals of Translational Medicine,* 6(3): 46.
747. Cowley LE, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO & Kemp AM (2018). Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals. *Child Abuse & Neglect,* 81: 192-205.

748. Cowley LE, Farewell DM & Kemp AM (2018). Potential impact of the validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A clinical vignette study. *Child Abuse & Neglect,* 86: 184-196.
749. Babl FE, Lyttle MD, Bressan S, Borland M, Phillips N, Kochar A, Dalziel SR, et al. (2014). A prospective observational study to assess the diagnostic accuracy of clinical decision rules for children presenting to emergency departments after head injuries (protocol): The Australasian Paediatric Head Injury Rules Study (APHIRST). *BMC Pediatrics,* 14: 148.
750. Babl FE, Borland ML, Phillips N, Kochar A, Dalton S, McCaskill M, Cheek JA, et al. (2017). Accuracy of PECARN, CATCH, and CHALICE head injury decision rules in children: a prospective cohort study. *The Lancet,* 389(10087): 2393-2402.
751. Collins LM, Schafer JL & Kam CM (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods,* 6(4): 330-351.
752. Graham JW (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology,* 60: 549-576.
753. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L & Bossuyt PM (2013). Variation of a test's sensitivity and specificity with disease prevalence. *Canadian Medical Association Journal,* 185(11): E537-E544.
754. Sotiriadis A, Papatheodorou SI & Martins WP (2015). Synthesizing evidence from diagnostic accuracy tests: The SEDATE guideline. *Ultrasound in Obstetrics & Gynecology,* 47(3): 386-395.
755. Guyatt G, Sackett DL & Haynes RB (2006). Evaluating diagnostic tests. *In:* Haynes RB, Sackett D, Guyatt GH & Tugwell P (Eds.) *Clinical epidemiology: How to do clinical practice research* (273-322). New York: Lippincott William and Wilkins.
756. Linden A (2006). Measuring diagnostic and predictive accuracy in disease management: An introduction to receiver operating characteristic (ROC) analysis. *Journal of Evaluation in Clinical Practice,* 12: 132-129.
757. Scales CDJ, Dahm P, Sultan S, Campbell-Scherer D & Devereaux PJ (2008). How to use an article about a diagnostic test. *Journal of Urology,* 180(2): 469-476.
758. Leeflang MM, Bossuyt PM & Irwig L (2009). Diagnostic test accuracy may vary with prevalence: Implications for evidence-based diagnosis. *Journal of Clinical Epidemiology,* 62(1): 5-12.
759. Mulherin SA & Miller WC (2002). Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Annals of Internal Medicine,* 137(7): 598-602.
760. Brenner H & Gefeller O (1998). Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Statistics in Medicine,* 16(9): 981-991.
761. Ransohoff DF & Feinstein AR (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New England Journal of Medicine,* 299(17): 926-930.
762. Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR & Schwartz JS (1992). Spectrum bias in the evaluation of diagnostic tests: Lessons from the rapid dipstick test for urinary tract infection. *Annals of Internal Medicine,* 117(2): 135-140.
763. Moons KGM, G.A. vE, Deckers JW, Habbema JD & Grobbee DE (1997). Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: A clinical example. *Epidemiology,* 8(1): 12-17.
764. Feinstein AR (2002). Misguided efforts and future challenges for research on “diagnostic tests”. *Journal of Epidemiology & Community Health,* 56(5): 330-332.
765. Usher-Smith JA, Sharp SJ & Griffin SJ (2016). The spectrum effect in tests for risk prediction, screening, and diagnosis. *The BMJ,* 353: i3139
766. Bossuyt PMM (1997). No burial for Bayes’ rule. *Epidemiology,* 8(1): 4-5.
767. Justice AC, Covinsky KE & Berlin JA (1999). Assessing the generalizability of prognostic information. *Annals of Internal Medicine,* 130(6): 515-524.

768. Knapp K (2017). *Validation of the Pittsburgh infant brain injury score for abusive head trauma*. CORE EM, Nov 30. Available from: https://coreem.net/journal-reviews/validation-of-the-pittsburgh-infant-brain-injury-score-for-abusive-head-trauma/ Accessed 12 May 2018.
769. Gallagher EJ (2003). The problem with sensitivity and specificity... *Annals of Emergency Medicine,* 42(2): 298–303.
770. Wells PS, Anderson DR, Rodger M, Ginsberg JS, Kearon C, Gent M, Turpie AGG, et al. (2000). Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: Increasing the models utility with the SimpliRED D-dimer. *Thrombosis & Haemostasis,* 83: 416-420.
771. Wicki J, Perneger TV, Junod AF, Bounameaux H & Perrier A (2001). Assessing clinical probability of pulmonary embolism in the emergency ward: A simple score. *Archives of Internal Medicine,* 161(1): 92-97.
772. Leclercq MGL, Kruip MJHA, Mac Gillavry MR, Van Marwijk Kooy M & Büller HR (2004). Observer variability in the assessment of clinical probability in patients with suspected pulmonary embolism. *Journal of Thrombosis & Haemostasis,* 2(7): 1204-1206.
773. Rodger MA, Maser E, Stiell I, Howley HEA & Wells PS (2005). The interobserver reliability of pretest probability assessment in patients with suspected pulmonary embolism. *Thrombosis Research,* 116(2): 101-107.
774. Klok FA, Zidane M, Djurabi RK, Nijkeuter M & Huisman MV (2008). The physician's estimation 'alternative diagnosis is less likely than pulmonary embolism' in the Wells rule is dependent on the presence of other required items. *Thrombosis & Haemostasis,* 99(1): 244-245.
775. Wolf SJ, McCubbin TR, Feldhaus KM, Faragher JP & Adcock DM (2004). Prospective validation of wells criteria in the evaluation of patients with suspected pulmonary embolism. *Annals of Emergency Medicine,* 44(5): 503-510.
776. Klok FA, Karami Djurabi R, Nijkeuter M & Huisman MV (2007). Alternative diagnosis other than pulmonary embolism as a subjective variable in the Wells clinical decision rule: Not so bad after all. *Journal of Thrombosis & Haemostasis,* 5(5): 1079-1080.
777. Kawamoto K, Houlihan CA, Balas EA & Lobach DF (2005). Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *The BMJ,* 330(7494): 765.
778. Brown W, Yen P-Y, Rojas M & Schnall R (2013). Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *Journal of Biomedical Informatics,* 46(6): 1080-1087.
779. Yardley L, Morrison L, Bradbury K & Muller I (2015). The person-based approach to intervention development: Application to digital health-related behavior change interventions. *Journal of Medical Internet Research,* 17(1): e30.
780. Davis FD (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly,* 13(3): 319-340.
781. Richardson S, Mishuris R, O'Connell A, Feldstein D, Hess R, Smith P, McCullagh L, et al. (2017). "Think aloud" and "Near live" usability testing of two complex clinical decision support tools. *International Journal of Medical Informatics,* 106: 1-8.
782. Cowley LE, Maguire SA, Farewell DM & Kemp AM (2018). Letter to Editor. *Law, Probability & Risk,* 17(3): 275-277.
783. Hymel KP, Herman BE, Narang SK, Graf JM, Frazier TN, Stoiko M, Christie LM, et al. (2016). Reply. *The Journal of Pediatrics,* 171: 321-322.
784. Brehaut JC, Graham ID, Wood TJ, Taljaard M, Eagles D, Lott A, Clement C, et al. (2010). Measuring acceptability of clinical decision rules: Validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. *Medical Decision Making,* 30(3): 398-408.

785. David TJ (2004). Avoidable pitfalls when writing medical reports for court proceedings in cases of suspected child abuse. *Archives of Disease in Childhood,* 89(9): 799-804.
786. Inkila J, Flinck A, Luukkaala T, Astedt-Kurki P & Paavilainen E (2013). Interprofessional collaboration in the detection of and early intervention in child maltreatment: Employees' experiences. *Nursing Research & Practice,* 186414.
787. Paradis M, Stiell I, Atkinson KM, Guerinet J, Sequeira Y, Salter L, Forster AJ, et al. (2018). Acceptability of a mobile clinical decision tool among emergency department clinicians: Development and evaluation of The Ottawa Rules App. *JMIR Mhealth & Uhealth,* 6(6): e10263.
788. Stiell IG, McKnight R, Greenberg GH, McDowell I, Nair RC, Wells GA, Johns C, et al. (1994). Implementation of the Ottawa ankle rules. *JAMA,* 271(11): 827-832.
789. Stiell IG, Wells GA, Hoag RH, Sivilotti ML, Cacciotti TF, Verbeek PR, Greenway KT, et al. (1997). Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries. *JAMA,* 278(23): 2075-2079.
790. Johnson EL, Hollen LI, Kemp AM & Maguire S (2016). Exploring the acceptability of a clinical decision rule to identify paediatric burns due to child abuse or neglect. *Emergency Medicine Journal,* 33(7): 465-470.
791. Sedlak AJ, Schultz D, Wells SJ, Lyons P, Doueck HJ & Gragg F (2006). Child protection and justice systems processing of serious child abuse and neglect cases. *Child Abuse & Neglect,* 30(6): 657-677.
792. Gunn VL, Hickson GB & Cooper WO (2005). Factors affecting pediatricians' reporting of suspected child maltreatment. *Ambulatory Pediatrics,* 5(2): 96-101.
793. Huston P & Rowan M (1998). Qualitative studies. Their role in medical research. *Canadian Family Physician,* 44: 2453-2458.
794. Pope C & Mays N (1995). Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *The BMJ,* 311(6996): 42-45.
795. Ploeg J (1999). Identifying the best research design to fit the question. Part 2: Qualitative designs. *BMJ Evidence-Based Nursing,* 2(2): 36-37.
796. Brown C & Lloyd K (2001). Qualitative methods in psychiatric research. *Advances in Psychiatric Treatment,* 7(5): 350-356.
797. Malterud K (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet,* 358(9280): 483-488.
798. Popay J & Williams G (1998). Qualitative research and evidence-based healthcare. *Journal of the Royal Society of Medicine,* 91(Suppl 35): 32-37.
799. Braun V & Clarke V (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3(2): 77-101.
800. Azzara CV (2010). The focus group vs. in-depth interview debate. *Quirks Marketing Research Review,* June: 16. Available from: http://www.quirks.com/pdf/201006_quirks.pdf. Accessed 15 January 2015.
801. Fontana A & Frey JH (1994). Interviewing: The art of science. *In:* Denzin NK & Lincoln YS (Eds.) *Handbook of qualitative research* (361-376). Thousand Oaks: Sage Publications.
802. Ziebland S & McPherson A (2006). Making sense of qualitative data analysis: An introduction with illustrations from DIPEx (personal experiences of health and illness). *Medical Education,* 40(5): 405-414.
803. Gaskell G (2000). Individual and group interviewing. *In:* Bauer M & Gaskell G (Eds.) *Qualitative researching with text, image and sound* (38-56). London: Sage Publications Ltd.
804. Fossey E, Harvey C, Mcdermott F & Davidson L (2002). Understanding and evaluating qualitative research. *Australian & New Zealand Journal of Psychiatry,* 36(6): 717-732.

805. Morse JM, Barrett M, Mayan M, Olson K & Spiers J (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods,* 1(2): 13-22.
806. Battaglia MP (2008). Nonprobability sampling. *In:* Lavrakas PJ (Ed.) *Encyclopaedia of survey research methods* (524-527). Thousand Oaks, CA: Sage Publications, Inc.
807. Creswell J & Plano Clark V (2011). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage Publications, Inc.
808. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N & Hoagwood K (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration & Policy in Mental Health,* 42(5): 533-544.
809. Guest G, Bunce A & Johnson L (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods,* 18(1): 59-82.
810. Bowen GA (2008). Naturalistic inquiry and the saturation concept: A research note. *Qualitative Research,* 8(1): 137-152.
811. O'Reilly M & Parker N (2012). 'Unsatisfactory Saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research,* 13(2): 190-197.
812. Kuzel A (1992). Sampling in qualitative inquiry. *In:* Crabtree B & Miller W (Eds.) *Doing qualitative research* (31-33). Newbury Park, CA: Sage Publications.
813. Whiting LS (2008). Semi-structured interviews: Guidance for novice researchers. *Nursing Standard,* 22(23): 35-40.
814. Britten N (1995). Qualitative interviews in medical research. *The BMJ,* 311(6999): 251-253.
815. King N & Horrocks C (2010). *Interviews in qualitative research.* Thousand Oaks, CA: Sage Publications Ltd.
816. World Medical Association (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA,* 310(20): 2191-2194.
817. Hand H (2003). The mentor's tale: A reflexive account of semi-structured interviews. *Nurse Researcher,* 10(3): 15-27.
818. Spencer L, Ritchie J, Lewis J & Dillon L (2003). *Quality in qualitative evaluation: A framework for assessing evidence.* London: Cabinet Office.
819. Bradbury-Jones C (2007). Enhancing rigour in qualitative health research: Exploring subjectivity through Peshkin's I's. *Journal of Advanced Nursing,* 59(3): 290-298.
820. Guillemin M & Gillam L (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative Inquiry,* 10(2): 261-280.
821. Pillow W (2003). Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *International Journal of Qualitative Studies in Education,* 16(2): 175-196.
822. Stronach I, Garratt D, Pearce C & Piper H (2007). Reflexivity, the picturing of selves, the forging of method. *Qualitative Inquiry,* 13(2): 179-203.
823. Horsburgh D (2003). Evaluation of qualitative research. *Journal of Clinical Nursing,* 12(2): 307-312.
824. Cutcliffe J (2003). Reconsidering reflexivity: Introducing the case for intellectual entrepreneurship. *Qualitative Health Research,* 13(1): 136-148.
825. Schoenberger E (1991). The corporate interview as a research method in economic geography. *The Professional Geographer,* 43(2): 180-189.
826. Ortlipp M (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report,* 13(4): 695-705.
827. Harvey W (2010). Methodological approaches for interviewing elites. *Geography Compass,* 4(3): 193-205.

828. Valentine C (2007). Methodological reflections: Attending and tending to the role of the researcher in the construction of bereavement narratives. *Qualitative Social Work,* 6(2): 159-176.
829. Gale NK, Heath G, Cameron E, Rashid S & Redwood S (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology,* 13: 117.
830. Boeije H (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality & Quantity,* 36(4): 391-409.
831. Crotty M (1998). *The foundations of social research: Meaning and perspective in the research process.* London: Sage Publications Ltd.
832. Holloway I & Todres L (2003). The status of method: Flexibility, consistency and coherence. *Qualitative Research,* 3(3): 345-357.
833. Bryman A & Burgess R (1994). *Analyzing qualitative data*. London: Routledge.
834. Dey I (1993). *Qualitative data analysis.* London: Routledge.
835. Sims-Schouten W, Riley SCE & Willig C (2007). Critical realism in discourse analysis: A presentation of a systematic method of analysis using women's talk of motherhood, childcare and female employment as an example. *Theory & Psychology,* 17(1): 101-124.
836. Korbin J (1991). Cross-cultural perspectives and research directions for the 21st century. *Child Abuse & Neglect,* 15(Suppl 1): 67-77.
837. Cohen S (2001). *States of denial: Knowing about atrocities and suffering.* Cambridge: Polity Press.
838. Morris M, Rivaux S & Faulkner M (2017). Provider ambivalence about using forensic medical evaluation to respond to child abuse: A content and discourse analysis. *Child Abuse & Neglect,* 65:140-151.
839. D'Cruz H (2004). The social construction of child maltreatment. *Journal of Social Work,* 4(1): 99-123.
840. Ritchie J & Lewis J (2003). *Qualitative research practice: A guide for social science students and researchers.* London: Sage Publications.
841. Saldana J (2013). *The coding manual for qualitative researchers.* London: Sage Publications Ltd.
842. QSR International Pty Ltd. NVivo qualitative data analysis software. Version 10. 2014.
843. Glaser BG & Straus AL (1967). *The discovery of grounded theory: Strategies for qualitative research.* New York: Aldine de Gruyter.
844. Tong A, Sainsbury P & Craig J (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care,* 19(6): 349-357.
845. Regnaut O, Jeu-Steenhouwer M, Manaouil C & Gignon M (2015). Risk factors for child abuse: Levels of knowledge and difficulties in family medicine. A mixed method study. *BMC Research Notes,* 8: 620.
846. Lucas S, Bärtås A, Bonamy A-KE, Törnudd L, Wide P & Otterman G (2017). The way forward in addressing abusive head trauma in infants – current perspectives from Sweden. *Acta Paediatrica,* 106(7): 1033-1035.
847. Ludvigsson JF (2017). Extensive shaken baby syndrome review provides a clear signal that more research is needed. *Acta Paediatrica,* 106(7): 1028-1030.
848. Cross TP, Finkelhor D & Ormrod R (2005). Police involvement in child protective services investigations: Literature review and secondary data analysis. *Child Maltreatment,* 10(3): 224-244.
849. Joughin V (2003). Working together for child protection in A&E. *Emergency Nurse,* 11(7): 30-37.

850. Davies C & Ward H (2012). *Safeguarding children across services: Messages from research.* London and Philadelphia: Jessica Kingsley. Available from: https://www.gov.uk/government/publications/safeguarding-children-across-services-messages-from-research. Accessed 14 November 2014.
851. Schols MWA, de Ruiter C & Öry FG (2013). How do public child healthcare professionals and primary school teachers identify and handle child abuse cases? A qualitative study. *BMC Public Health,* 13: 807.
852. Foster RH, Olson-Dorff D, Reiland HM & Budzak-Garza A (2017). Commitment, confidence, and concerns: Assessing health care professionals' child maltreatment reporting attitudes. *Child Abuse & Neglect,* 67: 54-63.
853. Stolper E, van Bokhoven M, Houben P, Van Royen P, van de Wiel M, van der Weijden T & Jan Dinant G (2009). The diagnostic role of gut feelings in general practice. A focus group study of the concept and its determinants. *BMC Family Practice,* 10: 17.
854. Van den Bruel A, Thompson M, Buntinx F & Mant D (2012). Clinicians' gut feeling about serious infections in children: Observational study. *The BMJ,* 345: e6144
855. Van den Bruel A, Aertgeerts B, Bruyninckx R, Aerts M & Buntinx F (2007). Signs and symptoms for diagnosis of serious infections in children: A prospective study in primary care. *The British Journal of General Practice,* 57(540): 538-546.
856. Dhaliwal G (2011). Going with your gut. *Journal of General Internal Medicine,* 26(2): 107-109.
857. Association of Chief Police Officers (2014). *A guide to investigating child deaths.* London: Association of Chief Police Officers. Available from: https://www.app.college.police.uk/app-content/major-investigation-and-public-protection/homicide/. Accessed 8 March 2015.
858. R v Clark (2000). EWCA Crim 54, Oct. 2nd. Available from: http://netk.net.au/UK/SallyClark4.asp. Accessed 20 July 2015.
859. R v Cannings (2004). EWCA Crim 01, Jan. 19th. Available from: http://www.familylawweek.co.uk/site.aspx?i=ed1654. Accessed 20 July 2015.
860. Aitken CGG, Roberts P & Jackson G (2010). *Fundamentals of probability and statistical evidence in criminal proceedings: Guidance for judges, lawyers, forensic scientists and expert witnesses.* London: Royal Statistical Society. Available from: http://www.rss.org.uk/RSS/Influencing_Change/Statistics_and_the_law/Practitioner_guides/RSS/Influencing_Change/Current_projects_sub/Statistics_and_the_law_sub/Practitioner_guides.aspx?hkey=2cfdf562-361e-432e-851b-ef6ff5254145. Accessed 14 June 2015.
861. Escobar MAJ, Auerbach M, Flynn-O'Brien K, Tiyyagura G, Borgman MA, Duffy SJ, Falcone K, et al. (2017). The association of non-accidental trauma with historical factors, exam findings and diagnostic testing during the initial trauma evaluation. *Journal of Trauma & Acute Care Surgery,* 82(6): 1147-1157.
862. Platt D & Turney D (2014). Making threshold decisions in child protection: A conceptual analysis. *The British Journal of Social Work,* 44(6): 1472-1490.
863. Polit DF & Beck CT (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies,* 47(11): 1451-1458.
864. Popay J, Rogers A & Williams G (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research,* 8(3): 341-351.
865. Asch S, Connor SE, Hamilton EG & Fox SA (2000). Problems in recruiting community-based physicians for health services research. *Journal of General Internal Medicine,* 15(8): 591-599.
866. Nir E (2018). Approaching the bench: Accessing elites on the judiciary for qualitative interviews. *International Journal of Social Research Methodology,* 21(1): 77-89.

867. Barlow J, Fisher JD & Jones D (2012). *Systematic review of models of analysing significant harm*. Oxford: Department for Education. Available from: https://www.gov.uk/government/publications/systematic-review-of-models-of-analysing-significant-harm. Accessed 4 February 2015.
868. Department for Education (2014). *Child protection, social work reform and intervention: Research priorities and questions.* London: Department for Education. Available from: https://www.bl.uk/collection-items/child-protection-social-work-reform-and-intervention-research-priorities-and-questions. Accessed 20 May 2015.
869. Learman K, Showalter C & Cook C (2012). Does the use of a prescriptive clinical prediction rule increase the likelihood of applying inappropriate treatments? A survey using clinical vignettes. *Manual Therapy,* 17(6): 538-543.
870. Ericsson KA & Simon HA (1999). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.
871. Creswell JW (2013). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage Publications, Inc.
872. Feilzer MY (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research,* 4(1): 6-16.
873. Tashakkori A & Teddlie C (2003). *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage Publications, Inc.
874. Onwuegbuzie AJ & Leech NL (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology,* 8(5): 375-387.
875. Atzmüller C & Steiner PM (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral & Social Sciences,* 6(3): 128-138.
876. Steiner PM, Atzmüller C & Su D (2016). Designing valid and reliable vignette experiments for survey research: A case study on the fair gender income gap. *Journal of Methods & Measurement in the Social Sciences,* 7(2): 52-94.
877. Evans SC, Roberts MC, Keeley JW, Blossom JB, Amaro CM, Garcia AM, Odar Stough C, et al. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical & Health Psychology,* 15(2): 160-170.
878. Aguinis A & Bradley KJ (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods,* 17(4): 351-371.
879. Bachmann LM, Mühleisen A, Bock A, ter Riet G, Held U & Kessels A (2008). Vignette studies of medical choice and judgement to study caregivers' medical decision behaviour: Systematic review. *BMC Medical Research Methodology,* 8: 50.
880. Fonteyn ME, Kuipers B & Grobe SJ (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research,* 3(4): 430-441.
881. Ludwick R, Wright ME, Zeller RA, Dowding DW, Lauder W & Winchell J (2004). An improved methodology for advancing nursing research: Factorial surveys. *Advances in Nursing Science,* 27(3): 224-238.
882. Wallander L (2012). Measuring social workers' judgements: Why and how to use the factorial survey approach in the study of professional judgements. *Journal of Social Work,* 12(4): 364-384.
883. Veloski J, Tai S, Evans AS & Nash DB (2005). Clinical vignette-based surveys: A tool for assessing physician practice variation. *American Journal of Medical Quality,* 20(3): 151-157.

884. Hughes R & Huby M (2002). The application of vignettes in social and nursing research. *Journal of Advanced Nursing,* 37(4): 382-386.
885. Hughes H & Huby M (2004). The construction and interpretation of vignettes in social research. *Social Work & Social Sciences Review,* 11(1): 36-51.
886. Backlund L, Skånér Y, Montgomery H, Bring J & Strender L-E (2003). Doctors' decision processes in a drug-prescription task: The validity of rating scales and think-aloud reports. *Organizational Behavior & Human Decision Processes,* 91(1): 108-117.
887. Denig P & Haaijer-Ruskamp FM (1994). 'Thinking aloud' as a method of analysing the treatment decisions of physicians. *European Journal of Public Health,* 4(1): 55-59.
888. Lundgrén-Laine H & Salanterä S (2009). Think-aloud technique and protocol analysis in clinical decision-making research. *Qualitative Health Research,* 20(4): 565-575.
889. Van Someren MW, Barnard YF & Sandberg JAC (1994). *Think aloud method: A practical guide to modelling cognitive processes.* London: Academic Press.
890. Davison GC, Vogel RS & Coffman SG (1997). Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. *Journal of Consulting & Clinical Psychology,* 65(6): 950-958.
891. Charters E (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education Journal,* 12(2): 68-82.
892. Willis G (2015). *Analysis of the cognitive interview in questionnaire design. Understanding qualitative research.* New York, NY: Oxford University Press.
893. Willis G (2005). *Cognitive interviewing. A tool for improving questionnaire design.* California: Sage Publications.
894. Willis G (1999). Cognitive interviewing: A "how to" guide. Reducing survey error through research on the cognitive and decision processes in surveys. *Short course presented at the Meeting of the American Statistical Association*.
895. Heist BS, Gonzalo JD, Durning S, Torre D & Elnicki DM (2014). Exploring clinical reasoning strategies and test-taking behaviors during clinical vignette style multiple-choice examinations: A mixed methods study. *Journal of Graduate Medical Education,* 6(4): 709-714.
896. Denig P, Witteman CLM & Schouten HW (2002). Scope and nature of prescribing decisions made by general practitioners. *Quality & Safety in Health Care,* 11(2): 137-143.
897. Skånér Y, Backlund L, Montgomery H, Bring J & Strender LE (2005). General practitioners' reasoning when considering the diagnosis heart failure: a think-aloud study. *BMC Family Practice,* 6:4.
898. Thackray D & Roberts L (2017). Exploring the clinical decision-making used by experienced cardiorespiratory physiotherapists: A mixed method qualitative design of simulation, video recording and think aloud techniques. *Nurse Education Today,* 49:96-105.
899. Chaiyachati BH, Asnes AG, Moles RL, Schaeffer P & Leventhal JM (2016). Gray cases of child abuse: Investigating factors associated with uncertainty. *Child Abuse & Neglect,* 51: 87-92.
900. Sturm V, Knecht PB, Landau K & Menke MN (2009). Rare retinal haemorrhages in translational accidental head trauma in children. *Eye,* 23(7): 1535-1541.
901. Trefan L, Houston R, Pearson G, Edwards R, Hyde P, Maconochie I, Parslow RC, et al. (2016). Epidemiology of children with head injury: A national overview. *Archives of Disease in Childhood,* 101(6): 527-532.
902. NHS Choices (2016). *Birth-to-five development timeline*. Available from: https://www.nhs.uk/Tools/Pages/birthtofive.aspx. Accessed 7th July 2018.
903. Eisenhauer JG (2003). Regression through the origin. *Teaching Statistics: An International Journal for Teachers,* 25(3): 76-80.

904. Bates D, Maechler M, Bolker B & Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software,* 67(1): 1-48.
905. Hothorn T, Bretz F & Westfall P (2008). Simultaneous inference in general parametric models. *Biometrical Journal,* 50(3): 346-363.
906. Lenth RV (2016). Least-Squares means: The R package lsmeans. *Journal of Statistical Software,* 69(1): 1-33.
907. Cochran WG (1954). Some methods for strengthening the common χ^2 tests. *Biometrics,* 10(4): 417-451.
908. Cohen J (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement,* 20(1): 37-46.
909. Bland MJ & Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet,* 327(8476): 307-310.
910. Feng GC (2015). Mistakes and how to avoid mistakes in using intercoder reliability indices. *Methodology,* 11(1): 13-22.
911. LeBreton JM & Senter JL (2007). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods,* 11(4): 815-852.
912. Koo TK & Li MY (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine,* 15(2): 155-163.
913. McGraw KO & Wong SP (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods,* 1(1): 30-46.
914. Hallgren KA (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology,* 8(1): 23-34.
915. Trevethan R (2017). Intraclass correlation coefficients: Clearing the air, extending some cautions, and making some requests. *Health Services & Outcomes Research Methodology,* 17(2): 127-143.
916. Revelle W (2017). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.6.12.
917. Cicchetti DV & Sparrow SA (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency,* 86(2): 127-137.
918. Fleiss JL (1981). *Statistical methods for rates and proportions.* New York, NY: Wiley.
919. Altman DG (1991). *Practical statistics for medical research.* London: Chapman & Hall.
920. Portney LG & Watkins MP (2009). *Foundations of clinical research: Applications to practice.* Upper Saddle River: Pearson Education.
921. Landis JR & Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33(1): 159-174.
922. Fleiss JL (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin,* 76(5): 378-382.
923. Scott WA (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly,* 19(3): 321-325.
924. Gwet KL (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical & Statistical Psychology,* 61(1): 29-48.
925. Gwet KL (2012a). The kappa coefficient: A review. *In:* Gwet KL (Ed.) *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters* (15-46). Gaithersburg, MD: Advanced Analytics, LLC.
926. Feinstein AR & Cicchetti DV (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology,* 43(6): 543-549.
927. Zec S, Soriani N, Comoretto R & Baldi I (2017). High agreement and high prevalence: The paradox of Cohen's kappa. *The Open Nursing Journal,* 11:211-218.

928. Shankar V & Bangdiwala SI (2014). Observer agreement paradoxes in 2x2 tables: Comparison of agreement measures. *BMC Medical Research Methodology,* 14: 100.
929. Warrens MJ (2010). A formal proof of a paradox associated with Cohen's kappa. *Journal of Classification,* 27(3): 322-332.
930. Wongpakaran N, Wongpakaran T, Wedding D & Gwet KL (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology,* 13: 61.
931. Gwet KL (2008). Variance estimation of nominal-scale inter-rater reliability with random selection of raters. *Psychometrika,* 73(3): 407-430.
932. Gwet KL (2012b). Agreement coefficients and statistical inference. *In:* Gwet KL (Ed.) *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters* (93-119). Gaithersburg, MD: Advanced Analytics, LLC.
933. Gwet KL (2002a). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-Rater Reliability Assessment,* 1(6): 1-6.
934. Gwet K (2002b). Inter-rater reliability: Dependency on trait prevalence and marginal homogeneity. *Statistical Methods for Inter-Rater Reliability Assessment,* 2: 1-9.
935. Walsh P, Thornton J, Asato J, Walker N, McCoy G, Baal J, Mendoza N, et al. (2014). Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. *PeerJ,* 2: e651.
936. McCray G (2013). Assessing inter-rater agreement for nominal judgement variables. Paper presented at the Language Testing Forum. Nottingham: November 15-17.
937. Friedl H & Stampfer E (2014). Jackknife resampling. *In:* Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F & Teugels JL (Eds.) *Wiley StatsRef: Statistics Reference Online*.
938. Gwet KL (2012c). Benchmarking the agreement coefficient. *In:* Gwet KL (Ed.) *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters* (121-147). Gaithersburg, MD: Advanced Analytics, LLC.
939. Daemers DOA, van Limbeek EBM, Wijnen HAA, Nieuwenhuijze MJ & de Vries RG (2017). Factors influencing the clinical decision-making of midwives: A qualitative study. *BMC Pregnancy & Childbirth,* 17: 345.
940. Nowell LS, Norris JM, White DE & Moules NJ (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods,* 161-13.
941. Kemp AM, Maguire SA, Nuttall D, Collins P & Dunstan F (2014). Bruising in children who are assessed for suspected physical abuse. *Archives of Disease in Childhood,* 99(2): 108-113.
942. Elstein AS, Holzman GB, Ravitch MM, Metheny WA, Holmes MM, Hoppe RB, Rothert ML, et al. (1986). Comparison of physicians' decisions regarding estrogen replacement therapy for menopausal women and decisions derived from a decision analytic model. *American Journal of Medicine,* 80(2): 246-258.
943. Lee TH, Pearson SD, Johnson PA, Garcia TB, Weisberg MC, Guadagnoli E, Cook EF, et al. (1995). Failure of information as an intervention to modify clinical management. A time-series trial in patients with acute chest pain. *Annals of Internal Medicine,* 122(6): 434-437.
944. Poses RM, Cebul RD & Wigton RS (1995). You can lead a horse to water−improving physicians' knowledge of probabilities may not affect their decisions. *Medical Decision Making,* 15(1): 65-75.

945. Sittig JS, Uiterwaal CSPM, Moons KGM, Russel IMB, Nievelstein RAJ, Nieuwenhuis EES & van de Putte EM (2016). Value of systematic detection of physical child abuse at emergency rooms: A cross-sectional diagnostic accuracy study. *BMJ Open,* 6: e010788.
946. Lorenz DJ, Pierce MC, Kaczor K, Berger RP, Bertocci G, Herman BE, Herr S, et al. (2018). Classifying injuries in young children as abusive or accidental: Reliability and accuracy of an expert panel approach. *Journal of Pediatrics,* 198: 144-150.e4.
947. Buesser KE, Leventhal JM, Gaither JR, Tate V, Cooperman DR, Moles RL, Silva CT, et al. (2017). Inter-rater reliability of physical abuse determinations in young children with fractures. *Child Abuse & Neglect,* 72: 140-146.
948. Kleiter GD (1994). Natural sampling: Rationality without base rates. *In:* Fischer GH & Laming D (Eds.) *Contributions to mathematical psychology, psychometrics, and methodology* (377-388). New York, NY: Springer-Verlag.
949. Evans JS, Handley SJ, Over DE & Perham N (2002). Background beliefs in Bayesian inference. *Memory & Cognition,* 30(2): 179-190.
950. Rottman BM (2017). Physician Bayesian updating from personal beliefs about the base rate and likelihood ratio. *Memory & Cognition,* 45(2): 270-280.
951. Keren G & Thujs LJ (1996). The base rate controversy: Is the glass half-full or half-empty? *Behavioural & Brain Sciences,* 19(1): 26.
952. Fagan TJ (1975). Nomogram for Bayes's Theorem. *New England Journal of Medicine,* 293(5): 257-257.
953. Peabody JW, Luck J, Glassman P, Dresselhaus TR & Lee M (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *JAMA,* 283(13): 1715-1722.
954. Mohan D, Fischhoff B, Farris C, Switzer GE, Rosengart MR, Yealy DM, Saul M, et al. (2014). Validating a vignette-based instrument to study physician decision making in trauma triage. *Medical Decision Making,* 34(2): 242-252.
955. Hughes R (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health & Illness,* 20(3): 381-400.
956. Cullen S (2010). Survey-driven romanticism. *Review of Philosophy & Psychology,* 1(2): 275-296.
957. Wilson J & While AE (1998). Methodological issues surrounding the use of vignettes in qualitative research. *Journal of Interprofessional Care,* 12: 79-87.
958. Faia MA (1980). The vagaries of the vignette world: A comment on Alves and Rossi. *American Journal of Sociology,* 85(4): 951-954.
959. Wilks T (2004). The use of vignettes in qualitative research into social work values. *Qualitative Social Work,* 3(1): 78-87.
960. Lunza ML (1990). A methodological approach to enhance external validity in simulation based research. *Issues in Mental Health Nursing,* 11(4): 407-422.
961. Kirwan JR, Chaput de Saintonge DM, Joyce CR & Currey HL (1983). Clinical judgment in rheumatoid arthritis. I. Rheumatologists' opinions and the development of 'paper patients'. *Annals of the Rheumatic Diseases,* 42(6): 644-647.
962. Lucet JC, Nicolas-Chanoine MH, Lefort A, Roy C, Diamantis S, Papy E, Riveros-Palacios O, et al. (2011). Do case vignettes accurately reflect antibiotic prescription? *Infectious Control & Hospital Epidemiology,* 32(10): 1003-1009.
963. Peabody JW & Liu A (2007). A cross-national comparison of the quality of clinical care using vignettes. *Health Policy Plan,* 22(5): 294-302.
964. Peabody JW, Luck J, Glassman P, Jain S, Hansen J, Spell M & Lee M (2004). Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Annals of Internal Medicine,* 141(10): 771-780.

965. Rousseau A, Rozenberg P & Ravaud P (2015). Assessing complex emergency management with clinical case-vignettes: A validation study. *PLoS One,* 10(9): e0138663.
966. Langley GR, Tritchler DL, Llewellyn-Thomas HA & Till J (1991). Use of written cases to study factors associated with regional variations in referral rates. *Journal of Clinical Epidemiology,* 44(4): 391-402.
967. Shah R, Edgar DF & Evans BJ (2010). A comparison of standardised patients, record abstraction and clinical vignettes for the purpose of measuring clinical practice. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists),* 30(3): 209-224.
968. Taylor BJ (2006). Factorial surveys: Using vignettes to study professional judgement. *The British Journal of Social Work,* 36(7): 1187-1207.
969. Gould D (1996). Using vignettes to collect data for nursing research studies: How valid are the findings? *Journal of Clinical Nursing,* 5(4): 207-212.
970. Spratt T (2001). The Influence of child protection orientation on child welfare practice. *British Journal of Social Work,* 31(6): 933-954.
971. Hinkelmann K & Kempthorne O (1994). *Design and analysis of experiments. Volume 1: Introduction to experimental design.* New York, NY: John Wiley & Sons, Inc.
972. Birnbaum MH (1999). How to show that 9>221: Collect judgments in a between-subjects design. *Psychological Methods,* 4(3): 243-249.
973. Drefyus HL & Dreyfus SE (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* New York, NY: Free Press.
974. Nisbett RE & Wilson TD (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* 84(3): 231-259.
975. Henry SB, Lebreck DB & Holzemer WL (1989). The effect of verbalization of cognitive processes on clinical decision making. *Research in Nursing & Health,* 12(3): 187-193.
976. Ericsson KA & Simon HA (1980). Verbal reports as data. *Psychological Review,* 87(3): 215-251.
977. Qi D (1998). An inquiry into language-switching in second language composing processes. *The Canadian Modern Language Review,* 54(3): 413-435.
978. Bugge C, Williams B, Hagen S, Logan J, Glazener C, Pringle S & Sinclair L (2013). A process for Decision-making after Pilot and feasibility Trials (ADePT): Development following a feasibility study of a complex intervention for pelvic organ prolapse. *Trials,* 14: 353.
979. Shanyinde M, Pickering RM & Weatherall M (2011). Questions asked and answered in pilot and feasibility randomized controlled trials. *BMC Medical Research Methodology,* 11: 117.
980. National Institute for Health Research (2018). *Feasibility studies*. Available from: https://www.nihr.ac.uk/glossary.
981. Arain M, Campbell MJ, Cooper CL & Lancaster GA (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology,* 10: 67.
982. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, Bakken S, et al. (2009). How we design feasibility studies. *American Journal of Preventive Medicine,* 36(5): 452-457.
983. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L & Lancaster GA (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *The BMJ,* 355: i5239.
984. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL & Bond CM (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLoS One,* 11(3): e0150205.

985. Orsmond GI & Cohn ES (2015). The distinctive features of a feasibility study: Objectives and guiding questions. *OTJR: Occupation, Participation & Health,* 35(3): 169-177.
986. Crook MA (2003). The Caldicott report and patient confidentiality. *Journal of Clinical Pathology,* 56(6): 426-428.
987. Clark AM & Sousa BJ (2018). The mental health of people doing qualitative research: Getting serious about risks and remedies. *International Journal of Qualitative Methods,* 17(1-3): 1609406918787244.
988. Dickson-Swift V, James EL, Kippen S & Liamputtong P (2007). Doing sensitive research: What challenges do qualitative researchers face? *Qualitative Research,* 7(3): 327-353.
989. Dickson-Swift V, James EL, Kippen S & Liamputtong P (2008). Risk to researchers in qualitative research on sensitive topics: Issues and strategies. *Qualitative Health Research,* 18(1): 133-144.
990. Kinard EM (1996). Conducting research on child maltreatment: Effects on researchers. *Violence & Victims,* 11(1): 65-69.
991. Genders TS, Spronk S, Stijnen T, Steyerberg EW, Lesaffre E & Hunink MG (2012). Methods for calculating sensitivity and specificity of clustered data: A tutorial. *Radiology,* 265(3): 910-916.
992. Eysenbach G (2013). CONSORT-EHEALTH: Implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. *Studies in HealthTechnology & Informatics,* 192: 657-661.
993. Crouch R, McHale H, Palfrey R & Curtis K (2015). The trauma nurse coordinator in England: A survey of demographics, roles and resources. *International Emergency Nursing,* 23(1): 8-12.
994. The County Council of the City and County of Cardiff v. Matthew Scully-Hicks and Others (2016). Judgment. Available from: https://tinyurl.com/ybfl4ugq. Accessed 20 May 2017.
995. Coster WJ (2013). Making the best match: Selecting outcome measures for clinical trials and outcome studies. *American Journal of Occupational Therapy,* 67(2): 162-170.
996. Mullen S, Quinn-Scoggins HD, Nuttall D & Kemp AM (2018). Qualitative analysis of clinician experience in utilising the BuRN Tool (Burns Risk assessment for Neglect or abuse Tool) in clinical practice. *Burns,* 44(7): 1759-1766.
997. Green SM (2013). When do clinical decision rules improve patient care? *Annals of Emergency Medicine,* 62(2): 132-135.
998. Schriger DL, Elder JW & Cooper RJ (2017). Structured clinical decision aids are seldom compared with subjective physician judgment, and are seldom superior. *Annals of Emergency Medicine,* 70(3): 338-344.e3.
999. Greenhalgh T, Robert G, Macfarlane F, Bate P & Kyriakidou O (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *The Milbank Quarterly,* 82(4): 581-629.
1000. Kiatchai T, Colletti AA, Lyons VH, Grant RM, Vavilala MS & Nair BG (2017). Development and feasibility of a real-time clinical decision support system for traumatic brain injury anesthesia care. *Applied Clinical Informatics,* 8(1): 80-96.
1001. Moore L, Hallingberg B, Wight D, Turley R, Segrott J, Craig P, Robling M, et al. (2018). Exploratory studies to inform full-scale evaluations of complex public health interventions: The need for guidance. *Journal of Epidemiology & Community Health,* 72(10): 865-866.
1002. Morgan B, Hejdenberg J, Hinrichs-Krapels S & Armstrong D (2018). Do feasibility studies contribute to, or avoid, waste in research? *PLoS One,* 13(4): e0195951.
1003. Anderson R (2008). New MRC guidance on evaluating complex interventions. *The BMJ,* 337: a1937

1004. Bartholomew LK, Parcel GS & Kok G (1998). Intervention mapping: A process for developing theory and evidence-based health education programs. *Health Education & Behavior,* 25(5): 545-563.
1005. Michie S, van Stralen MM & West R (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science,* 6: 42-42.
1006. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, Gask L, et al. (2007). Understanding the implementation of complex interventions in health care: The normalization process model. *BMC Health Services Research,* 7: 148.
1007. Collins LM, Murphy SA, Nair VN & Strecher VJ (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine,* 30(1): 65-73.
1008. Walker AE, Grimshaw J, Johnston M, Pitts N, Steen N & Eccles M (2003). PRIME – PRocess modelling in ImpleMEntation research: Selecting a theoretical basis for interventions to change clinical practice. *BMC Health Services Research,* 3: 22.
1009. Lakshman R, Griffin S, Hardeman W, Schiff A, Kinmonth AL & Ong KK (2014). Using the Medical Research Council framework for the development and evaluation of complex interventions in a theory-based infant feeding intervention to prevent childhood obesity: The baby milk intervention and trial. *Journal of Obesity,* 646504.
1010. Levati S, Campbell P, Frost R, Dougall N, Wells M, Donaldson C & Hagen S (2016). Optimisation of complex health interventions prior to a randomised controlled trial: A scoping review of strategies used. *Pilot & Feasibility Studies,* 2: 17.
1011. Mackenzie M, O'Donnell C, Halliday E, Sridharan S & Platt S (2010). Do health improvement programmes fit with MRC guidance on evaluating complex interventions? *The BMJ,* 340: c185.
1012. Pawson R & Tilley N (1997). *Realistic evaluation.* Thousand Oaks, CA: Sage Publications, Inc.
1013. Bonell C, Fletcher A, Morton M, Lorenc T & Moore L (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine,* 75(12): 2299-2306.
1014. Hawe P, Shiell A & Riley T (2009). Theorising interventions as events in systems. *American Journal of Community Psychology,* 43(3-4): 267-276.
1015. Hawe P, Shiell A & Riley T (2004). Complex interventions: How "out of control" can a randomised controlled trial be? *The BMJ,* 328(7455): 1561.
1016. De Silva MJ, Breuer E, Lee L, Asher L, Chowdhary N, Lund C & Patel V (2014). Theory of Change: A theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials,* 15(1): 267.
1017. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM & Collins GS (2016). External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. *The BMJ,* 353: i3140.
1018. Hallingberg B, Turley R, Segrott J, Wight D, Craig P, Moore L, Murphy S, et al. (2018). Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: A systematic review of guidance. *Pilot & Feasibility Studies,* 4: 104.
1019. Battle C, Abbott Z, Hutchings HA, Neill C, Groves S, Watkins A, Lecky FE, et al. (2017). Protocol for a multicentre randomised feasibility STUdy evaluating the impact of a prognostic model for Management of BLunt chest wall trauma patients: STUMBL trial. *BMJ Open,* 7: e015972.
1020. Johnson RB & Onwuegbuzie AJ (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher,* 33(7): 14-26.
1021. Yardley L & Bishop FL (2015). Using mixed methods in health research: Benefits and challenges. *British Journal of Health Psychology,* 20(1): 1-4.

1022. Guba EG (1990). The alternative paradigm dialog. *In:* Guba EG (Ed.) *The paradigm dialog* (17-27). Newbury Park, CA: Sage.
1023. Tariq S & Woodman J (2013). Using mixed methods in health research. *Journal of the Royal Society of Medicine Short Reports,* 4(6): 2042533313479197.
1024. Waljee AK, Higgins PDR & Singal AG (2014). A primer on predictive models. *Clinical & Translational Gastroenterology,* 5(1): e44.

10 Appendices

Appendix 1. Search strategy used to identify relevant literature

1. CHILD/	58. spinal cord injur:.af.
2. CHILD, PRESCHOOL/	59. (subdural haematoma or hemotoma).af.
3. (child: or infant: or toddler: or babies or baby).af.	60. (subarachnoid hematoma or subarachnoid haematoma).af.
4. or/1-3	61. (subdural haemorrhage or subdural hemorrhage).af.
5. ((non-accidental or nonaccidental) adj3 (trauma or injur:)).af.	62. (ventricular haemorrhage or ventricular hemorrhage).af.
6. ((non-abusive or nonabusive) adj3 (injur: or trauma)).af.	63. whiplash impact syndrome.af.
7. (non-accidental: and injur:).af.	64. whiplash injur:.af.
8. soft tissue injur:.af.	65. whiplash shaken infant.af.
9. physical abuse.af.	66. infarction.af.
10. ((inflicted or noninflicted or non-inflicted) adj3 (brain injur: or cerebral injur: or head injur:)).af.	67. (hypoxic-ischemic injur: or hypoxic-ischaemic injur:).af.
	68. (contusion: or contusional tear).af.
11. (inflicted traumatic head injur: or inflicted traumatic brain injur:).af.	69. (hematoma or haematoma).af.
	70. laceration:.af.
12. (or/5-11) and 4	71. shearing injur:.af.
13. (child abuse or child maltreatment or child protection).af.	72. traumatic effusion:.af.
	73. subdural hygroma.af.
14. (battered child or shaken baby or battered baby).af.	74. hygroma.af.
	75. interhemispheric.af.
15. (battered infant or shaken infant).af.	76. parafalcine.af.
16. (Shak: Baby Syndrome or shak: impact syndrome).af.	77. (brain or brainstem).af.
	78. cerebral.af.
17. Caffey-Kempe syndrome.af.	79. intraparenchymal.af.
18. *"Child Abuse"/di [Diagnosis]	80. sciwora.mp.
19. infant traumatic stress syndrome.af.	81. spinal cord injury without radiologic abnormality.af.
20. parent-infant traumatic stress syndrome.af.	82. cervical lumbar.af.

21. or/13-20 22. 12 or 21 23. abusive head trauma.af. 24. bleeding into brain.af. 25. blow to the head.af. 26. brain damage.af. 27. (brain haemorrhage: or brain hemorrhage:).af. 28. (brain swelling or cerebral edema).af. 29. cerebral injur:.af. 30. cervical spine injur:.af. 31. cervical spine neuropathology.af. 32. cranial injur:.af. 33. craniocerebral trauma.af. 34. diffuse axonal injur:.af. 35. extracranial CNS injur:.af. 36. extracranial Central Nervous System injur:.af. 37. central nervous system injur:.af. 38. (extradural haematoma or hematoma).af. 39. extradural haemorrhage.af. 40. ha?morrhagic retinopathy.af. 41. (head inur: or head trauma).af. 42. impact injur:.af. 43. intracerebral bleeding.af. 44. (intracerebral haemorrhage or intracerebral hemorrhage).af. 45. (intracranial haemorrhage or intracranial hemorrhage).af. 46. intracranial injur:.af. 47. (intraventricular hematoma or intraventricular haematoma).af.	83. thoracic lumbar sacral.af. 84. leptomeningeal cyst.af. 85. growing skull fracture.af. 86. (Extradural haemorrhag: or extradural hemorrhag: or extradural spinal haemorrhag: or extradural spinal hemorrhag:).af. 87. laminar necrosis.af. 88. encephalomalacia.af. 89. cerebral atrophy.af. 90. (craniocervical or hydrocephalus).af. 91. encephalopathy.af. 92. (intraparenchymal hemorrhag: or intraparenchymal haemorrhag:).af. 93. (Ha?morrhagic retinopathy adj3 retinal ha?emorrhag:).af. 94. cerebral venous thrombosis.mp. 95. diffuse axonal injur*.tw. 96. spinal subdural.tw. 97. or/23-96 98. Computed tomography.af. 99. (CT or CAT scan:).af. 100. diagnostic imaging.af. 101. (magnetic resonance imaging or MRI).af. 102. neuroradiology.af. 103. neuroimaging.af. 104. plain films.af. 105. radiological imaging.af. 106. X-rays.af. 107. neurologic: imaging.af. 108. diffusion weighted imaging.af. 109. neurologic examination.af.

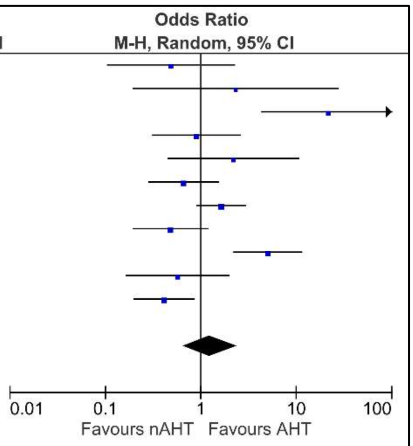
48. (multiple skull fractur: or eggshell fractur:).af. 49. exp Neck Injuries/ 50. neck injur*.af. 51. (neurological injur: adj3 child abuse).af. 52. neuropathology.af. 53. non-accidental head injur:.af. 54. (parenchymal contusion or laceration).af. 55. (retinal hemorrhage or retinal haemorrhage).af. 56. skull fracture:.af. 57. (spinal cord injury adj3 radiologic abnormality).af.	110. ultrasound scan:.af. 111. (Susceptibility Weighted Imaging or SWI).tw. 112. or/98-111 113. 22 and 97 and 112 114. 22 and 97 115. 12 and 97 116. 113 or 114 or 115 117. limit 116 to yr="2013 -Current" 118. management.mp. 119. Review.pt. 120. 118 or 119 121. 117 not 120

Appendix 2. Updated forest plots from Kemp et al., 2011, of the association between neuroradiological features and abusive head trauma. Updated with data from papers identified in the 2014 update of the systematic review (Royal College of Paediatrics and Child Health, 2015)

Forest plot depicting the association between subarachnoid haemorrhage and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Adamo 2009	2	54	12	164	7.6%	0.49 [0.11, 2.25]
Bechtel 2004	1	15	2	67	4.5%	2.32 [0.20, 27.42]
Diaz-Olavarrieta 2011	5	13	3	107	7.3%	21.67 [4.37, 107.51]
Ewing-Cobbs 2000	11	31	11	29	9.7%	0.90 [0.31, 2.57]
Goldstein 1993	4	14	4	26	7.4%	2.20 [0.46, 10.62]
Hettler 2003	9	49	29	114	10.7%	0.66 [0.29, 1.52]
Hymel 2013	37	95	32	114	11.8%	1.63 [0.91, 2.92]
Keenan 2004	9	80	15	72	10.4%	0.48 [0.20, 1.18]
Reece 2000	17	51	13	144	10.8%	5.04 [2.23, 11.38]
Vinchon 2010	5	45	7	39	8.8%	0.57 [0.17, 1.97]
Wells 2002	14	148	22	109	11.2%	0.41 [0.20, 0.85]
Total (95% CI)		**595**		**985**	**100.0%**	**1.24 [0.65, 2.38]**
Total events	114		150			

Heterogeneity: Tau² = 0.85; Chi² = 43.66, df = 10 (P < 0.00001); I² = 77%
Test for overall effect: Z = 0.66 (P = 0.51)

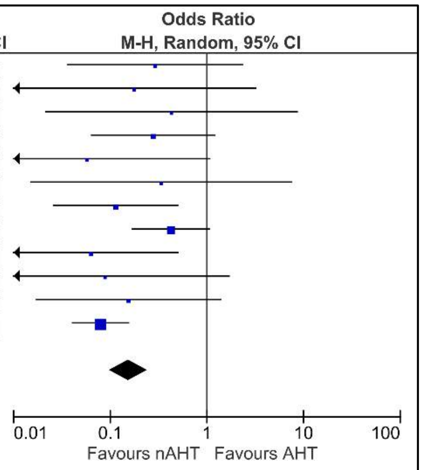


Subarachnoid haemorrhage was not significantly associated with abusive or non-abusive head trauma

Forest plot depicting the association between extradural haemorrhage and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Adamo 2009	1	54	10	164	4.6%	0.29 [0.04, 2.32]
Bechtel 2004	0	15	10	67	2.4%	0.18 [0.01, 3.18]
Duhaime 1992	0	24	3	76	2.3%	0.43 [0.02, 8.59]
Ettaro 2004	2	89	22	288	9.1%	0.28 [0.06, 1.21]
Ewing-Cobbs 2000	0	31	6	29	2.4%	0.06 [0.00, 1.07]
Goldstein 1993	0	14	2	26	2.1%	0.34 [0.02, 7.54]
Hettler 2003	2	49	31	114	9.0%	0.11 [0.03, 0.50]
Hymel 2013	7	95	18	114	21.2%	0.42 [0.17, 1.06]
Keenan 2004	1	80	12	72	4.7%	0.06 [0.01, 0.50]
Thalayasingam 2012	0	43	4	38	2.3%	0.09 [0.00, 1.69]
Vinchon 2010	1	45	5	39	4.2%	0.15 [0.02, 1.39]
Wells 2002	14	148	62	109	35.7%	0.08 [0.04, 0.15]
Total (95% CI)		**687**		**1136**	**100.0%**	**0.15 [0.10, 0.24]**
Total events	28		185			

Heterogeneity: Tau² = 0.03; Chi² = 11.57, df = 11 (P = 0.40); I² = 5%
Test for overall effect: Z = 8.07 (P < 0.00001)

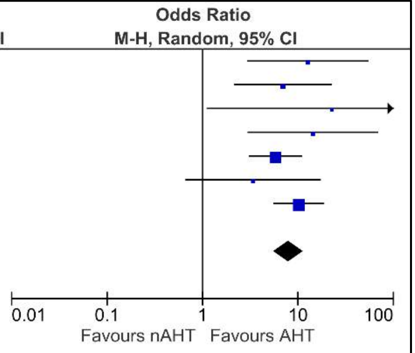


Extradural haemorrhage was significantly associated with non-abusive head trauma

Forest plot depicting the association between interhemispheric subdural haemorrhage and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Datta 2005	38	49	3	14	6.4%	12.67 [2.99, 53.58]
Ewing-Cobbs 2000	20	31	6	29	9.8%	6.97 [2.18, 22.26]
Goldstein 1993	4	14	0	26	1.5%	22.71 [1.12, 459.84]
Hymel 1997	17	39	2	39	5.5%	14.30 [3.01, 67.84]
Hymel 2013	54	95	21	114	34.1%	5.83 [3.13, 10.88]
Kelly 2004	10	41	2	23	5.1%	3.39 [0.67, 17.05]
Wells 2002	105	148	21	109	37.6%	10.23 [5.65, 18.53]
Total (95% CI)		**417**		**354**	**100.0%**	**8.03 [5.58, 11.56]**
Total events	248		55			

Heterogeneity: Tau² = 0.00; Chi² = 4.17, df = 6 (P = 0.65); I² = 0%
Test for overall effect: Z = 11.21 (P < 0.00001)

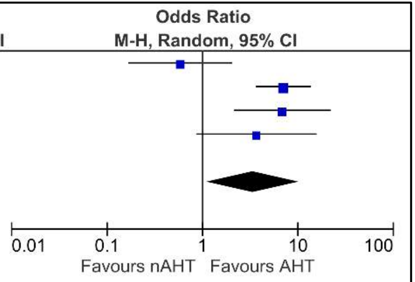


Interhemispheric subdural haemorrhage was significantly associated with abusive head trauma

Forest plot depicting the association between bilateral subdural haemorrhage and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Ewing-Cobbs 2000	15	26	14	20	23.7%	0.58 [0.17, 2.01]
Hymel 2013	54	95	18	114	30.0%	7.02 [3.68, 13.41]
Kelly 2004	29	41	6	23	24.7%	6.85 [2.17, 21.59]
Tzoumi 1998	12	21	4	15	21.5%	3.67 [0.87, 15.38]
Total (95% CI)		**183**		**172**	**100.0%**	**3.36 [1.10, 10.27]**
Total events	110		42			

Heterogeneity: Tau² = 0.97; Chi² = 12.98, df = 3 (P = 0.005); I² = 77%
Test for overall effect: Z = 2.13 (P = 0.03)

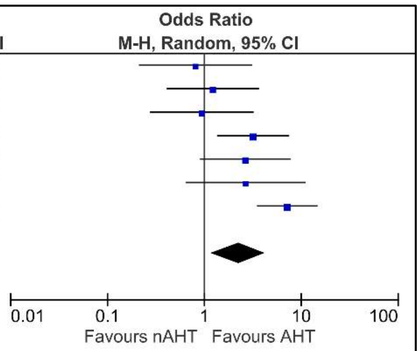


Bilateral subdural haemorrhage was significantly associated with abusive head trauma. This association did not reach significance in the original meta-analysis

Forest plot depicting the association between cerebral oedema and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Datta 2005	12	49	4	14	11.8%	0.81 [0.21, 3.07]
Ewing-Cobbs 2000	11	31	9	29	14.3%	1.22 [0.42, 3.59]
Feldman 2001	15	39	6	15	12.8%	0.94 [0.28, 3.17]
Keenan 2004	25	80	9	72	16.9%	3.18 [1.37, 7.40]
Kelly 2004	24	41	8	23	14.5%	2.65 [0.92, 7.64]
Thalayasingam 2012	8	43	3	38	11.1%	2.67 [0.65, 10.89]
Wells 2002	66	148	11	109	18.6%	7.17 [3.55, 14.48]
Total (95% CI)		**431**		**300**	**100.0%**	**2.24 [1.18, 4.26]**
Total events	161		50			

Heterogeneity: Tau² = 0.45; Chi² = 15.77, df = 6 (P = 0.02); I² = 62%
Test for overall effect: Z = 2.47 (P = 0.01)



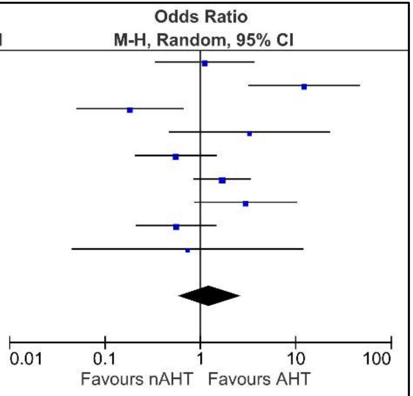
Cerebral oedema was significantly associated with abusive head trauma

Forest plot depicting the association between intraparenchymal haemorrhage and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Adamo 2009	4	54	11	164	11.8%	1.11 [0.34, 3.65]
Diaz-Olavarrieta 2011	6	13	7	107	11.0%	12.24 [3.23, 46.43]
Ewing-Cobbs 2000	4	31	13	29	11.3%	0.18 [0.05, 0.66]
Goldstein 1993	3	14	2	26	8.1%	3.27 [0.48, 22.46]
Hettler 2003	6	49	23	114	13.1%	0.55 [0.21, 1.45]
Hymel 2013	24	95	19	114	14.7%	1.69 [0.86, 3.32]
Ichord 2007	14	30	5	22	11.6%	2.98 [0.87, 10.16]
Keenan 2004	8	80	12	72	13.2%	0.56 [0.21, 1.45]
Wells 2002	1	148	1	109	5.2%	0.73 [0.05, 11.88]
Total (95% CI)		**514**		**757**	**100.0%**	**1.25 [0.58, 2.69]**
Total events	70		93			

Heterogeneity: $Tau^2 = 0.91$; $Chi^2 = 29.35$, df = 8 (P = 0.0003); $I^2 = 73\%$

Test for overall effect: Z = 0.58 (P = 0.56)



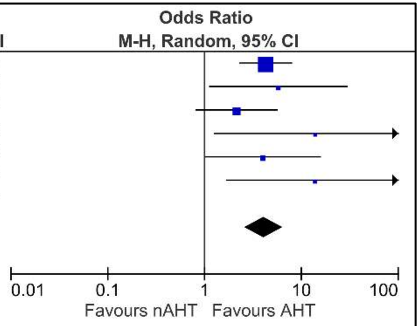
Intraparenchymal haemorrhage was not significantly associated with abusive or non-abusive head trauma. N.B the original meta-analysis included all focal parenchymal injury, which was also not significantly associated with abusive or non-abusive head trauma

Forest plot depicting the association between hypoxic ischaemic injury and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Hymel 2013	48	95	22	114	52.4%	4.27 [2.31, 7.90]
Ichord 2007	11	30	2	22	7.4%	5.79 [1.13, 29.62]
Keenan 2004	15	80	7	72	21.5%	2.14 [0.82, 5.60]
McKinney 2008	3	11	1	38	3.5%	13.88 [1.27, 151.23]
Thalayasingam 2012	11	43	3	38	10.7%	4.01 [1.03, 15.68]
Vinchon 2010	12	45	1	39	4.5%	13.82 [1.70, 112.01]
Total (95% CI)		**304**		**323**	**100.0%**	**4.11 [2.63, 6.42]**
Total events	100		36			

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 4.26$, df = 5 (P = 0.51); $I^2 = 0\%$

Test for overall effect: Z = 6.22 (P < 0.00001)



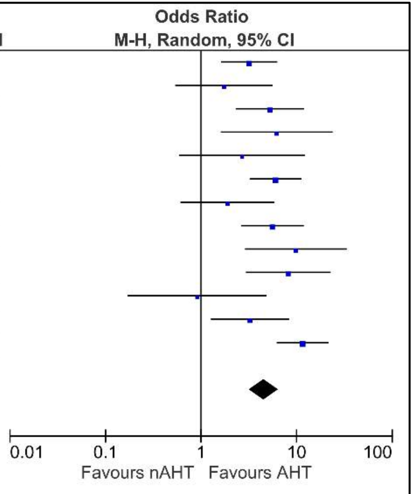
Hypoxic ischaemic injury was significantly associated with abusive head trauma

Forest plot depicting the association between closed head injury and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Adamo 2009	38	54	70	164	10.8%	3.19 [1.65, 6.18]
Diaz-Olavarrieta 2011	7	13	43	107	6.2%	1.74 [0.55, 5.52]
Ettaro 2004	42	58	18	54	9.2%	5.25 [2.34, 11.77]
Feldman 2001	27	39	4	15	5.1%	6.19 [1.63, 23.42]
Goldstein 1993	11	14	15	26	4.3%	2.69 [0.60, 11.99]
Hymel 2013	72	95	39	114	11.5%	6.02 [3.28, 11.06]
Ichord 2007	17	30	9	22	6.5%	1.89 [0.62, 5.76]
Keenan 2004	66	80	33	72	9.9%	5.57 [2.66, 11.68]
Kelly 2004	30	41	5	23	5.9%	9.82 [2.93, 32.86]
Thalayasingam 2012	34	43	12	38	7.3%	8.19 [3.00, 22.34]
Tzoumi 1998	17	21	14	17	3.7%	0.91 [0.17, 4.77]
Vinchon 2010	34	45	19	39	8.0%	3.25 [1.29, 8.21]
Wells 2002	105	148	19	109	11.5%	11.57 [6.29, 21.26]
Total (95% CI)		**681**		**800**	**100.0%**	**4.60 [3.21, 6.60]**
Total events	500		300			

Heterogeneity: Tau² = 0.20; Chi² = 23.43, df = 12 (P = 0.02); I² = 49%

Test for overall effect: Z = 8.30 (P < 0.00001)



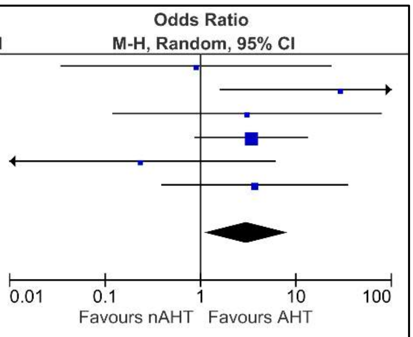
Closed head injury was significantly associated with abusive head trauma

Forest plot depicting the association between diffuse axonal injury/shear injury and abusive head trauma in children

Study or Subgroup	AHT Events	AHT Total	nAHT Events	nAHT Total	Weight	Odds Ratio M-H, Random, 95% CI
Datta 2005	1	49	0	14	9.1%	0.90 [0.03, 23.22]
Ewing-Cobbs 2000	9	29	0	31	11.4%	29.20 [1.61, 529.37]
Hymel 1997	1	39	0	39	9.3%	3.08 [0.12, 77.91]
Hymel 2013	8	95	3	114	42.7%	3.40 [0.88, 13.21]
Ichord 2007	0	30	1	22	9.2%	0.23 [0.01, 6.05]
Vinchon 2010	4	45	1	39	18.4%	3.71 [0.40, 34.66]
Total (95% CI)		**287**		**259**	**100.0%**	**3.03 [1.11, 8.30]**
Total events	23		5			

Heterogeneity: Tau² = 0.14; Chi² = 5.44, df = 5 (P = 0.36); I² = 8%

Test for overall effect: Z = 2.16 (P = 0.03)



Diffuse axonal injury/shear injury was significantly associated with abusive head trauma. This association could not be analysed in the original meta-analysis

Appendix 3. Manuscript accepted for publication in ***BMC Diagnostic & Prognostic Research***

Methodological standards for the development and evaluation of clinical prediction rules: A review of the literature

Laura E. Cowley[1], Daniel M. Farewell[1], Sabine Maguire[1], Alison M. Kemp[1]

Affiliations: [1]Division of Population Medicine, School of Medicine, Neuadd Meirionnydd, Heath Park, Cardiff University, CF14 4YS, Wales, United Kingdom. CowleyLE@cardiff.ac.uk, FarewellD@cardiff.ac.uk, sabinemaguire@gmail.com, KempAM@cardiff.ac.uk

Address correspondence to: Laura E. Cowley, Division of Population Medicine**,** School of Medicine, Neuadd Meirionnydd, Heath Park, Cardiff University, CF14 4YS, Wales, United Kingdom. Telephone: 0044 2920 688688 E-mail: CowleyLE@cardiff.ac.uk

Abstract

Clinical prediction rules (CPRs) that predict the absolute risk of a clinical condition or future outcome for individual patients are abundant in the medical literature, however systematic reviews have demonstrated shortcomings in the methodological quality and reporting of prediction studies. To maximize the potential and clinical usefulness of CPRs, they must be rigorously developed and validated, and their impact on clinical practice and patient outcomes must be evaluated. This review aims to present a comprehensive overview of the stages involved in the development, validation and evaluation of CPRs, and to describe in detail the methodological standards required at each stage, illustrated with examples where appropriate. Important features of the study design, statistical analysis, modelling strategy, data collection, performance assessment, CPR presentation, and reporting are discussed, in addition to other, often overlooked aspects such as the acceptability, cost-effectiveness and longer-term implementation of CPRs, and their comparison with clinical judgment. Although the development and evaluation of a robust, clinically useful CPR is anything but straightforward, adherence to the plethora of methodological standards, recommendations and frameworks at each stage will assist in the development of a rigorous CPR that has the potential to contribute usefully to clinical practice and decision-making and have a positive impact on patient care.

Keywords: Clinical prediction rule, prediction model, risk model, model development, model validation, impact studies, model reporting, implementation, diagnosis, prognosis, study design

Background

The aim of a clinical prediction rule (CPR) is to estimate the probability of a clinical condition or a future outcome by considering a small number of highly valid indicators [1, 2]. CPRs include three or more predictors, from patients' clinical findings, history, or investigation results [3]. Their purpose is to assist clinicians in making decisions under conditions of uncertainty and enhance diagnostic, prognostic or therapeutic accuracy and decision-making, with the ultimate aim of improving the quality of patient care [1, 2, 4]. The predicted probabilities from a CPR allow clinicians to stratify patients into risk groups and help them to decide whether further assessment or treatment is necessary [5]. Some CPRs can help to 'rule in' a condition by identifying patients who are very likely to have a condition and who thus require additional diagnostic testing or treatment, whilst others aim to 'rule out' a condition by identifying patients who are very unlikely to have a condition, thus reducing unnecessary testing without compromising patient care [2, 4]. CPRs that aim to predict the probability of a condition being present are termed *diagnostic* or *screening* rules; those that aim to predict the probability of a future outcome are termed *prognostic* rules; and those that aim to predict the probability that a specific treatment or intervention will be effective are termed *prescriptive* rules [2].

To maximize the predictive accuracy and clinical utility of CPRs, it is vital that they are rigorously developed, validated and evaluated. However, numerous systematic reviews have demonstrated shortcomings in the methodological quality and reporting of prediction studies, which restricts the CPR's usefulness in practice [6-15]. Methodological standards for the development of CPRs were originally outlined by Wasson and colleagues [16]. With the increase in popularity of CPRs inspired by the evidence-based medicine movement, these standards have since been modified and updated by a number of authors over the years [3, 4, 17-19]. Experts have provided thorough and accessible overviews of the principles and methods involved in conducting diagnostic and prognostic research [20-32] and devised frameworks to enhance the conduct and interpretation of prediction studies [33-35]. They have also provided guidance and recommendations for researchers to consider when developing and evaluating CPRs, without aiming to dictate how analyses should be conducted. These recognise that there is no clear consensus on many aspects of model development, that the field is continually evolving, and that methodological standards will therefore require updating accordingly [36]. Guidelines for the *reporting* of clinical prediction research have also been developed, namely, the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [36].

This review aims to outline the stages and methodological standards involved in the development and evaluation of CPRs, illustrated with examples where appropriate.

Terminology used in this review

In the literature the term ‘clinical prediction rule’ is used interchangeably with the terms clinical prediction tool [37], clinical decision rule [17], clinical decision tool [38], clinical prediction algorithm [39], prognostic score [40], prognostic model [21], risk prediction model [23], risk model [30], risk score [41], scoring tool [42], scoring system [43], or risk index [44]. Reilly and Evans [32] distinguish between *assistive prediction* rules that simply provide clinicians with diagnostic or prognostic predicted probabilities without recommending a specific clinical course of action, and *directive decision* rules that explicitly suggest additional diagnostic tests or treatment in line with the obtained score. Decision rules intend to directly influence clinician behaviour, while prediction rules intend to help clinicians predict risk without providing recommendations, with the assumption that accurate predictions will lead to better decisions [32]. Some researchers also distinguish between prediction *models* that provide predicted probabilities along the continuum between certified impossibility (Pi=0) and absolute certainty (Pi=1) [45], and prediction *rules* that classify patients into risk groups, by applying a clinically relevant cut-off that balances the likelihood of benefit with the likelihood of harm [19, 46]. Such cut-offs are known as 'decision thresholds'; a threshold must be applied if a prediction model aims to influence decision-making [19]. In this review the term 'clinical prediction rule' is used to refer to diagnostic, prognostic or prescriptive rules/models derived from multivariable statistical analyses, which predict the probability of a condition or outcome, *with or without* the use of a clinical cut-off or recommendation for further action.

Stages in the development of clinical prediction rules

It is widely acknowledged in the literature that there are three *main* stages in the development of CPRs (Figure 1); derivation; external validation; and impact analysis to determine their impact on patient care [4, 20, 22-25, 32, 33]. Stiell and Wells [17] identified a further three important stages, namely identifying the need for a CPR, determining the cost-effectiveness of a CPR, and long-term dissemination and implementation of a CPR. Therefore all six stages are summarised in Table 1 and discussed in detail below.

Insert Table 1

Detailed methodological and practical recommendations pertaining to the three main stages of development have been published, as each requires a different methodological approach [3, 4, 16-36]. These three stages also correspond to increasing hierarchies of evidence, as outlined in Table 2 [4, 32, 33]. A CPR that has been *derived*, but not externally validated, corresponds to the lowest level of evidence and is not recommended for use in clinical practice, except arguably in rare instances when a CPR is developed for use in only one setting. It has been suggested that a CPR that has been successfully externally *validated* in a setting, or population, similar to the one from which it was derived ('narrow' validation), can be used cautiously in similar future patients [32]. Similarly, it is proposed that a CPR should be consistently successfully externally validated in multiple settings or populations ('broad' validation), before clinicians can use its predictions confidently in future patients [32]. Finally, it is recommended that an *impact analysis* is conducted and that the CPR demonstrates improvements to patient care, before it can be used as a decision rule for the management and treatment of patients [32]. Ideally the impact of a CPR should also be tested in multiple settings. Impact analysis studies correspond to the highest level of evidence [32].

Table 2. Hierarchies of evidence in the development and evaluation of clinical prediction rules

Level of Evidence	**Definitions and Standards of Evaluation**	**Implications for Clinicians**
Level 1: Derivation of CPR	Identification of predictors using multivariable model; blinded assessment of outcomes.	Needs validation and further evaluation before it is used clinically in actual patient care.
Level 2: Narrow validation of CPR	Validation of CPR when tested prospectively in one setting; blinded assessment of outcomes.	Needs validation in varied settings; may use CPR cautiously in patients similar to derivation sample.

Level 3: Broad validation of CPR	Validation of CPR in varied settings with wide spectrum of patients and clinicians.	Needs impact analysis; may use CPR predictions with confidence in their accuracy.
Level 4: Narrow impact analysis of CPR used for decision-making	Prospective demonstration in one setting that use of CPR improves clinicians' decisions (quality or cost-effectiveness of patient care).	May use cautiously to inform decisions in settings similar to that studied.
Level 5: Broad impact analysis of CPR used for decision-making	Prospective demonstration in varied settings that use of CPR improves clinicians' decisions for wide spectrum of patients.	May use in varied settings with confidence that its use will benefit patient care quality or effectiveness.

Adapted from Reilly & Evans, 2016 [32]. CPR: clinical prediction rule

Stage 1: Identifying the need for a clinical prediction rule

Before developing a CPR, researchers need to ensure that there is a clinical need for the rule. CPRs are most valuable when decision-making is challenging, when there is evidence that clinicians are failing to accurately diagnose a condition, and when there are serious consequences associated with an incorrect diagnosis [2, 4]. CPRs are also valuable when there is a need to simplify or speed up the diagnostic or triage process, for example in patients presenting to the emergency department with chest pain and suspected acute cardiac ischaemia [47]. CPRs are most likely to be adopted into clinical practice, and to demonstrate improvements in patient care and reductions in health care costs, when they improve the overall efficiency of clinical practice [17]. For example, ankle injuries are frequently seen in the emergency department. Prior to the implementation of the Ottawa Ankle Rule, clinicians ordered a high proportion of radiographs that were negative for fracture, when the majority of them believed that a fracture was highly unlikely [48]. The rule was found to lead to a reduction in both radiography [49] and health care costs [50], and in one survey 70% of Canadian and UK emergency department clinicians reported frequent use of the rule [51].

Before developing a CPR, researchers should consider whether a new CPR is needed, as many are developed for the same target population or to predict the same outcome [8, 10, 11, 52-55]. The characteristics, performance and level of evidence of existing CPRs should be systematically reviewed using validated search filters for locating prediction studies, and the Critical Appraisal and Data Extraction for Systematic Reviews of prediction modelling studies (CHARMS) checklist [56, 57]. The recently published Prediction model Risk Of Bias ASsessment Tool (PROBAST) can be used to assess the risk of bias and applicability of CPRs [58]. Researchers can also assess the performance of existing CPRs on their own collected data [59]. Existing CPRs with potential should be updated, validated or tested in an impact study before a new CPR is developed [52, 60, 61]. If a new CPR is derived, researchers should clearly justify why it is required, with reference to existing CPRs, to avoid research waste and duplication of efforts [62]. Qualitative research with clinicians can be useful in determining whether a proposed CPR is clinically relevant, and to assess the credibility of the proposed predictor variables [63, 64].

Stage 2: Derivation of a clinical prediction rule according to methodological standards

Once a need for a new CPR is established, and a researcher has an appropriate clinical question, a CPR must be derived according to strict methodological standards [23]. There are various elements to consider, pertaining to the study design, statistical techniques employed, and the assessment, presentation and reporting of the CPR.

Researchers should consider writing and publishing a study protocol and registering the study prior to the derivation of a new CPR, in the interests of transparency [65, 66].

Study design for the derivation of a clinical prediction rule

The first stage in the development of a CPR is the derivation of the rule. This involves an examination of the ability of multiple potential variables from the clinical findings, history, or investigation results to predict the target outcome of interest. Predicted probabilities are derived from the statistical analysis of patients with known outcomes, and the outcome of interest serves as the reference standard by which the performance of the CPR is assessed. The performance of a CPR is dependent upon the quality of the underlying data, and the dataset used to derive the CPR should be representative of the target population it is intended for [17, 30, 67, 68].

The optimal study design for the derivation of a diagnostic CPR is a cross-sectional cohort study, while for prognostic CPRs, the preferred design is a longitudinal cohort study [30]. In general, case-control studies are inappropriate, as they do not allow for the estimation of absolute outcome risk [21, 23, 69], however nested case-control or case-cohort studies can be used [69, 70]. Prospective cohort studies are preferred to retrospective cohort studies, to optimise measurement and documentation of predictive and outcome variables [21, 23]. For prescriptive CPRs, study designs that include a control group, such as randomised controlled trials (RCTs), are essential to ensure that treatment effect modifiers and non-specific prognostic predictors are distinguishable from one another [71, 72]. The study design should be adequately detailed and include the study setting, inclusion and exclusion criteria, and patient demographics and characteristics [17]. To enhance generalisability, multicentre studies are recommended [30].

Statistical analysis

Commonly used statistical methods for the derivation of CPRs include multivariable regression techniques, and recursive partitioning techniques, such as classification and regression tree analysis [73]. Methods based on univariable analysis, where individual risk factors are simply totalled and assigned arbitrary weightings, should be avoided, as they are much less accurate than methods based on multivariable analysis [74]. This is because the final model may include predictors that are potentially related to each other and not independently associated with the outcome of interest [74]. Multivariable methods overcome the limitations of univariable analysis by enabling improved assessment of the association of the predictors with the target outcome [74].

In the case of multivariable regression, logistic regression models are required to predict binary events such as the presence or absence of a condition, while Cox regression models are suitable for time-to-event outcomes. Such models estimate regression coefficients (e.g. log odds or hazard ratios) of each predictor. Regression coefficients are mutually adjusted for the other predictors, and thus represent the contribution of each predictor to the probability of the outcome [23]. The probability of an outcome can be computed for a patient by combining the observed values of the predictors and their corresponding regression coefficients with the model intercept, or estimated baseline hazard [23]. For logistic models, the model intercept and the weighted values applicable to each patient are summed [16]. Specific values are assigned to each predictor, which are multiplied by the corresponding coefficients. In the case of a model with only binary categorical predictors, the predictors are multiplied by 0 or 1, depending on whether they are absent (0) or present (1), as per the model in Box 1 [75]. Exponentiating the final risk score gives the odds, and the probability (absolute risk) is calculated by use of the inverse logistic link function [76]. In this way, the probability of an outcome can be estimated from any combination of the predictor values [36]. The estimated probability for an individual without any of the predictors depends only on the intercept [23]. In this case the value for each of the predictors will be 0; when each of these is multiplied by its relevant coefficient the value of 0 is retained [76]. For Cox regression models the baseline hazard is estimated separately [26, 29].

Recursive partitioning involves repeatedly splitting patients into subpopulations including only individuals with a specific outcome [77], and was the method used to derive the Ottawa Ankle Rule [78]. CPRs can also be derived using discriminant function analysis [3], and machine learning algorithms based on artificial neural networks [1]. Artificial intelligence and machine learning approaches are becoming increasingly more common [79, 80].

Box 1. Clinical prediction rule for postoperative nausea and vomiting (PONV) [75]

Risk of PONV = 1/(1 + exp−[2.28 + 1.27 × female sex + 0.65 × history of PONV or motion sickness + 0.72 × non-smoking + 0.78 × postoperative opioid use])

Missing data

In clinical research, investigators almost always encounter missing observations involving predictor or outcome variables, even in carefully designed studies and in spite of their best efforts to maximize data quality [81]. There are three types of missing-data mechanisms: 1) Missing completely at random (MCAR) 2) Missing at random (MAR), and 3) Missing not at random (MNAR) [82]. When data are MCAR this means that there are no systematic differences between the missing and observed values; for example, laboratory tests may be missing because of a dropped test tube or broken equipment. When data are MAR this means that the probability of a missing value depends on the observed values of other variables (but not the unobserved values); for example, missing blood pressure measurements may be lower than observed measurements because younger people may be more likely to have missing measurements; in this case data can be said to be MAR given age [83]. When data are MNAR this means that the probability of a missing value depends on the unobserved values or other unobserved predictors, conditional on the observed data; for example, people with high blood pressure may be more likely to miss a doctor's appointment due to headaches [83]. Missing values are rarely MCAR, that is, their 'missingness' is usually directly or indirectly related to other subject or disease characteristics, including the outcome [23, 25]. Missing data is frequently addressed with case-wise deletion, which excludes all participants with missing values from the analysis [83]. However, when data are plausibly MAR, this reduces sample size and statistical power and biases the results [83], leading to inaccurate estimates of predictor-outcome relationships and the predictive performance of the model, since the participants with complete data are not a random subsample of the original sample [82, 84, 85].

Multiple imputation is a popular approach to the problem of missing data [81, 83, 84, 86-89], as it quantifies the uncertainty in the imputed values, by generating multiple different plausible imputed datasets, and pooling the results obtained from each of them [83, 89]. Multiple imputation involves three stages [83, 87, 89-91]. First, as the name suggests, multiple imputed datasets are created, based on the distribution of the observed data. This first stage accounts for uncertainty in estimating the missing values by adding variability into the values across the imputed datasets. In the second stage, standard statistical techniques are used to fit the models that are of interest in the substantive analysis to each of the imputed datasets. Estimated associations in each of the imputed datasets will be different, due to the variability introduced in stage one. In the third and final stage, the multiple results are averaged together, and standard errors are calculated using Rubin's combination rules [89], which account for both within-and between-imputation variability and the number of imputed datasets, and therefore the uncertainty of the imputed values. Multiple imputation typically assumes that data are MAR [91]. Importantly, the MAR assumption is just that; an assumption, rather than a property of the data [83]. The MCAR assumption can be tested, but it is not possible to differentiate between MAR and MNAR from the observed data [26, 83]. Most missing data are expected to be at least partly MNAR [83, 92, 93]. Sensitivity analyses can help to determine the effect of different assumptions about the missing data mechanism; work in this area is ongoing [94-98]. Other statistically principled approaches to dealing with missing data have been developed, based on random effects models [99, 100], Bayesian methods or maximum likelihood estimation [101], or, where data are longitudinal, joint models [102, 103]. Guidelines for reporting on the treatment of missing data in clinical and epidemiological research studies have been suggested by Sterne and colleagues [83]. Guidance also exists for handling missing data when deriving and validating CPRs [81, 104, 105]. It has been demonstrated that the outcome should be used for imputation of missing predictor values [85]. It is also becoming increasingly apparent that a real-time strategy to impute missing values is desirable when applying a CPR in clinical practice [106-108]. This is because one or more predictor variables may be unobserved for a particular patient, and thus the CPRs risk prediction cannot be estimated at the time of decision-making [106]. Real-time multiple imputation is not typically straightforward, as it requires access to the derivation dataset via, for example, a website [106, 108]. Of note, although multiple imputation is a widely advocated approach for handling missing data in CPR studies, a recent study showed that implementing simpler imputation methods resulted in similar predictive utility of a CPR to predict undiagnosed diabetes, when compared to multiple imputation [109].

Selection of candidate predictors for inclusion in a multivariable model

Candidate predictors are variables that are preselected for consideration in a multivariable model, and differ from those that are subsequently selected for inclusion in the final model [23]. Candidate predictors should be selected without studying the predictor-outcome relationship in the data; in other words, predictors should not be excluded as candidates solely because they are not statistically significant in univariable analysis [25, 26, 29, 110-112]. Predictor variables do not have to be causally related to the outcome of interest [21, 113]. Effects modelled in studies examining *causality* are expressed with relative risk estimates such as odds ratios, while risk *predictions* are presented as probabilities on an absolute scale between 0 and 1. Relative risk estimates are used in prediction research to calculate an absolute probability of an outcome for a patient, as described above, and can also be reported alongside risk predictions. All variables thought to be related to the target outcome can be selected as candidate predictors for inclusion in a multivariable model, however when the number of outcome events in the dataset is small, there is a risk of overfitting the data when a large number of predictor variables are included. Thus the CPR will perform well on the derivation data, but poorly on new data [29, 67, 111, 114]. CPRs with a smaller number of predictors are also easier to use in practice. To overcome this problem, only the most clinically relevant candidate predictors should be chosen from the larger pool of potential predictor variables, without looking into the data [5, 115]. In addition, sample size recommendations for studies deriving CPRs are often based on the concept of events-per-variable (EVP), whereby the researcher controls the ratio of the number of outcome events to the number of coefficients estimated prior to any data-driven variable selection [31]. A rule-of-thumb of ten EPV has been suggested [29, 31, 112, 116]. Simulation studies examining the effect of this rule-of-thumb have yielded conflicting results [117-121]. One study found that when the EPV was less than ten there were a range of circumstances in which coverage and bias were within acceptable levels [117]. Another found that 20 EPV or more are required when low-prevalence predictors are included in a model [121], while another suggested that problems may arise even when the EPV exceeds ten, as CPR performance may depend on many other factors [118]. Research in this area continues to evolve, as new guidance is clearly needed to support sample size considerations for the derivation of CPRs [119]. Recently, van Smeden and colleagues have suggested that sample size should be guided by three influential parameters: the number of predictors, total sample size and the events fraction [120].

Relevant predictors may be chosen based on a combination of clinical experience, expert opinion surveys, qualitative studies, and formal systematic reviews and meta-analyses of the literature [26, 33, 36, 63, 122]. Strategies for reducing the number of candidate predictors include removing those that are highly correlated with others, and combining similar predictors [29]. Other considerations include selecting predictors that will be readily available for clinicians to observe or measure in the target setting, and selecting predictors that are relatively easy to measure and demonstrate high inter-rater reliability between clinicians [17, 21]. In terms of handling continuous predictors, researchers strongly advise against converting continuous variables into categorical variables, due to information loss and reduced predictive accuracy [123-126]. Similarly, it should not be assumed that continuous variables have a linear relationship [127]. Instead, methods that permit more flexibility in the functional form of the association between the predictors and outcome should be considered [127, 128]; two common approaches are fractional polynomials and restricted cubic splines [129, 130]. However, if sample size is limited, assuming a linear relationship between continuous variables may make a model less sensitive to extreme observations.

Penalised regression can be used to alleviate the problem of overfitting [114]. This approach involves placing a constraint on the values of the estimated regression coefficients in order to shrink them towards zero [114]. This has the effect of yielding less extreme risk predictions, and thus may improve the accuracy of predictions when the CPR is applied in new patients [111, 131]. The two most popular penalised methods are ridge regression [132] and lasso regression [133]. Unlike ridge regression, lasso regression also selects predictors as a consequence of its penalisation [114]. Ridge regression is usually preferred when a set of pre-specified predictors is available, while lasso regression may be preferred if a simpler model with fewer predictors is required [114, 131].

Selection of predictors during multivariable modelling

There is no consensus regarding how predictors should be selected while developing the final model [25]. Two common strategies include the 'full model approach' and the 'predictor selection approach' [23]. An alternative approach, known as 'all possible subsets regression', is less commonly used [28]. In the full model approach, all previously identified candidate predictors are included, and no further analysis is performed. Although this approach precludes selection bias and overfitting, it requires in-depth knowledge about the most relevant candidate predictors [26, 29]. In the predictor selection approach, predictors are chosen either by 'backward elimination' or 'forward selection', based on pre-defined criteria. Backward elimination begins with all predictors in the model and removes predictors, while forward selection begins with an empty model, and predictors are added successively. All possible subsets regression can build models with combinations of predictors not generated by the standard forward or backward procedures, because every conceivable combination of predictors is assessed to find the best fitting model [134]. With all methods, a series of statistical tests are performed to assess the 'goodness of fit' between the different models. Models can be compared by setting a pre-defined significance level and using the log likelihood ratio test, or using other model selection criterion such as the Akaike information criterion, or the Bayesian information criterion [23, 25]. Backward elimination is favoured, as it allows for the assessment of the effects of all predictors concurrently, and can take into account all correlations between predictors [135, 136]. Multiple testing in all possible subsets regression can easily lead to overfitting. However, with all methods, the choice of significance level impacts upon the number of final predictors; the use of smaller significance levels (e.g. $p<0.05$) produces models with fewer predictors at the risk of excluding potentially important predictors, while the use of larger significance levels (e.g. $p<0.25$) may result in the inclusion of less important predictors [25].

Predictor selection by so-called 'automated', data-dependent significance testing may generate overfitted, 'optimistic' models, particularly when the derivation dataset is small [23, 28, 126, 137, 138]. Thus, the Akaike information criterion is preferred, as it discourages overfitting by comparing models based on their fit to the data and penalising for the complexity of the model [25]. In addition, it may be acceptable to retain a non-significant predictor in a model, if there is substantial evidence of its predictive ability in the literature [26].

Definition and assessment of predictor and outcome variables

To ensure that the CPR can be accurately applied in practice, predictor and outcome variables should be clearly defined, and outcome variables should be clinically important [17]. Predictor variables must be reliable to enable their assessment in clinical practice; reliability refers to the reproducibility of the findings by the same clinician (intra-rater reliability) or between different clinicians (inter-rater reliability). Some researchers recommend that the reliability of predictor variables be explicitly evaluated, and that only those demonstrating good agreement beyond that expected by chance alone should be considered for inclusion [17]. A recent study found that measurement error of predictor variables is poorly reported, and that researchers seldom state explicitly when the predictors should be measured, and the CPR applied [139]. Another study demonstrated that predictor measurement heterogeneity across settings can have a detrimental impact on the performance of a CPR at external validation [140]. Ideally, the outcome variable should be assessed independently of the predictor variables to avoid circular reasoning or 'incorporation bias', when the results of the CPR or its predictor variables are used in the determination of the outcome [141]. However, it is acknowledged that this is not always possible, particularly for conditions that require a consensus diagnosis based on all available patient information [142]. It is well known that misclassification in the outcome variable may cause serious problems with prediction accuracy [143, 144].

Internal validation

Prediction models are known to perform better in the dataset from which they are derived, in comparison to applying them in new but plausibly related patients [145, 146]. 'Plausibly related patients' may be defined as those who are suspected of having the same condition or who are at risk of the same outcome examined in the derivation study [147]. This enhanced performance occurs simply because a model is designed to optimally fit the available data [23]. The performance of a model is most likely to be overestimated when the derivation dataset is small, and uses a large number of candidate predictors. Therefore, regardless of the approaches used in the derivation stage of development, internal validation is required to examine and correct the amount of overfitting or 'optimism' in the model, and thus the stability of the model [23].

Internal validation does not validate a model itself, but the process used to fit the model [26, 29]. Optimism is estimated using the original derivation dataset only. A number of methods are available for this purpose, including split-sampling, cross-validation, and bootstrapping. Split-sampling is the simplest method, and is performed by dividing the derivation dataset into a 'training' sample and a 'test' sample prior to modelling. The CPR is then derived using the training sample, and its performance is assessed using the test sample [20]. However, the test sample usually comprises one third of the original derivation dataset and is likely to be relatively small, resulting in imprecise performance estimates [148, 149]. This approach also squanders the test data that could have been used in the derivation of the CPR [23, 149]. In cross-validation, the CPR is derived using the whole derivation dataset, and the whole dataset is then reused to assess performance [20]. It is randomly split into equal samples: five or ten samples are commonly used. In the case of five samples, the model is refitted using four of the five samples and its performance tested using the fifth; this process is repeated five times until each of the five samples has been used as the test data, and an average of the estimated performance is taken. To improve stability, the overall procedure can be replicated several times, using different random subsamples [148]. The preferred internal validation method is bootstrapping, particularly when the derivation dataset is small or a large number of candidate predictors are assessed [23, 29]. The idea is to mimic random sampling from the target population by repeatedly drawing samples of the same size with replacement from the derivation dataset [150]. Sampling with replacement renders bootstrap samples similar, but not identical, to the original derivation sample [23]. Each step of model development is repeated in each bootstrap sample (typically 500), most likely yielding different models with varying performance. Each bootstrap model is then applied to the original derivation sample, yielding a difference in model performance. The average of these differences indicates the optimism in the performance metrics of the model that was initially derived in the derivation dataset [23, 26, 29, 150], and enabling adjustment of the overall performance to better approximate the expected model performance in novel samples [23]. Bootstrapping also estimates a uniform shrinkage factor to enable adjustment of the estimated regression coefficients for over-fitting [26, 29, 150]. However, no internal validation procedures can be a substitute for external validation; internal validation only addresses sampling variability, while external validation considers variation in the patient population [146].

Clinical prediction rule performance measures

CPR predictive performance can be assessed in terms of overall performance, calibration, and discrimination [26]. ‘Overall performance’ can be quantified by calculating the distance between observed and predicted outcomes, using measures such as R^2 or the Brier score [151]. ‘Calibration’ reflects the agreement between the predicted probabilities produced by the model and the observed outcome frequencies [23]. For example, if a model predicts a 20% probability of residual tumour for a testicular cancer patient, residual tumour should be observed in about 20 out of 100 of these patients [46]. ‘Internal calibration’ refers to agreement between predicted probabilities and observed outcome frequencies in the derivation dataset, where poor calibration may indicate lack of model fit or model misspecification [152]. 'External calibration' refers to agreement between predicted probabilities and observed outcome frequencies in novel datasets external to the one from which the model was derived, where poor calibration may indicate an overfitted model [152]. Calibration can be visualised by categorising individuals into quantiles based on their predicted probabilities, and plotting the observed outcome frequencies against the mean predicted probabilities [25]. Such a plot is the graphical equivalent of the Hosmer and Lemeshow goodness-of-fit test [153], which, although frequently used, may lack statistical power to identify overfitting [25, 26]. Alternatively, binary outcomes can be regressed on the predicted probabilities of the fitted model to estimate the observed outcome probabilities using smoothing techniques such as the loess algorithm [29, 152]. A comprehensive overview of calibration is given in Van Calster et al. [154].

Discrimination reflects the ability of a CPR to discriminate between patients with, and without, the outcome of interest. The predicted probabilities for patients *with* the outcome should be higher than the predicted probabilities for those who don't have the outcome [46]. The easiest way to assess discrimination is by calculation of the discrimination slope, which is simply the absolute difference in the average predicted probabilities for patients with and without the outcome [26]. Discrimination can also be visualized with a simple box plot. The most widely used measure to assess discrimination is the concordance index (c-index) [155], or, for logistic models its equivalent, the Area Under the Receiver Operating Characteristic curve (AUROC) [156]. These measures represent the chance that, given one patient with the outcome and one without, the CPR will assign a higher predictive probability to the patient with the outcome compared to the one without. A c-index or AUROC of 0.5 indicates predictions that are no better than random predictions, and a value of 1 represents perfect discrimination between patients with and without the outcome [29]. In theory, a CPR may demonstrate good *discrimination* (classifying patients into the correct risk categories), but poor *calibration* (inaccurately estimating the absolute probability of an outcome), and vice versa [157]. A model that cannot discriminate between patients with and without the outcome has little use as a CPR, however poor calibration can be corrected without compromising discriminatory performance [19, 112]. Van Calster and Vickers [158] found that poorly calibrated models diminish the clinical usefulness of a CPR, and can be harmful for clinical decision-making under certain circumstances, emphasizing the importance of developing well-calibrated CPR's. On the other hand, a CPR with poor calibration but good discrimination at a particular risk threshold may be appropriate if the aim is to prioritise patients for assessment or treatment, by identifying those with a very low risk of the target outcome relative to the rest of the population [159].

Performance measures such as sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios, are used to assess performance following the application of a risk threshold. Choosing a risk threshold can often be arbitrary, and it can therefore be useful to consider a range of thresholds when assessing performance [19]. Ideally a CPR will have both a high sensitivity and a high specificity, and therefore correctly identify the majority of patients who truly have the condition, as well as correctly exclude the majority of patients who do not actually have the condition. However, this scenario rarely occurs in clinical practice. More often than not, the definition of a threshold is based on clinical considerations about the relative consequences of false positive and false negative classifications. Sensitivity and specificity are inversely proportional, so that as sensitivity increases, specificity decreases and vice versa [160]. Defining a high cut-off point will result in good specificity and few false positives, but poor sensitivity and many false negatives. A test with a high specificity is useful for ruling in a disease if a person tests positive. This is because it rarely misdiagnoses those who don't have the condition of interest. Defining a low cut-off point will result in good sensitivity and few false negatives, but poor specificity and many false positives. A test with a high sensitivity is useful for ruling out disease if a person tests negative. This is because it rarely misdiagnoses those who have the condition of interest [160]. Receiver Operating Characteristic (ROC) curves display the sensitivity and specificity of a CPR across the full range of cut-off values, and can be used to choose an optimal cut-off threshold [161]. Other approaches to determining clinical cut-offs have also been proposed [162].

In recent years, some novel model performance measures have been proposed that quantify the clinical usefulness of a CPR, by taking into account the costs and benefits of clinical decisions. These measures include relative utility curves and decision curves [163, 164]. Decision curves in particular are becoming a popular method of evaluating whether clinical decisions based on CPRs would do more good than harm [165]. Decision curve analysis assumes that a given probability threshold is directly related to the cost to benefit ratio, and uses this threshold to weight false positive and false negative predictions. The cost to benefit ratio thus defines the relative weight of false-positive decisions to true-positive decisions [163]. Model performance can subsequently be summarized as a net benefit, by subtracting the proportion of false-positive patients from the proportion of true-positive patients, weighting by the relative costs of a false-positive and a false-negative result. The net benefit of a CPR can be derived across and plotted against the whole range of threshold probabilities, yielding a decision curve, similar to ROC curves that plot the full range of cut-offs for a sensitivity/specificity pair [163].

Presentation of a clinical prediction rule

The final step in the derivation of a CPR is to consider the format in which it should be presented. It is imperative that the regression coefficients and intercept of a final model are presented, and confidence intervals around predicted probabilities can also be provided [23, 26]. If the final regression formula (as in Box 1) is not provided, a CPR could not be applied by future users [36]. A model can be developed into a simple web-based calculator or application to enhance the usability of a CPR. This may be beneficial for complex CPRs, and would facilitate their integration into the electronic health record, allowing them to be used at the point of clinical care [166]. Nomograms, graphical decision trees, and other novel visualization techniques could also be used [26, 167], which may aid in the interpretation and understanding of a CPR [167], however these must be presented alongside the full model formula. Scoring systems are often used to simplify CPRs and facilitate use, where regression coefficients are converted to integer point values that can be easily totalled and related back to the predicted probabilities [168]. However, this transformation leads to a loss of information and therefore reduced predictive accuracy [169].

Reporting the derivation of a clinical prediction rule

Numerous systematic reviews have shown that reporting of the derivation of CPRs is deficient [6-8]. As a result, the TRIPOD guidelines were produced [36], and should be followed by all researchers working in this field.

Stage 3: External validation and refinement of a clinical prediction rule

As previously noted, CPRs perform better in the dataset from which they are derived compared to their application in plausibly related or 'similar but different' individuals, even after internal validation and adjustment [24]. Diminished performance can be due to overfitting, unsatisfactory model derivation, the absence of important predictors, differences in how the predictor variables are interpreted and measured, differences in the patient samples ('case mix'), and differences in the prevalence of the disease [26, 147]. There is no guarantee that even well-developed CPRs will be generalisable to new individuals. In one external validation study a CPR to detect serious bacterial infections in children with fever of unknown source demonstrated considerably worse predictive performance, such that it was rendered useless for clinical care [145]. It is therefore essential to assess the performance of a CPR in individuals outside the derivation dataset; this process is known as external validation [28].

External validation is not simply repeating the steps involved at the derivation stage in a new sample to examine whether the same predictors and regression coefficients are obtained; neither is it refitting the model in a new sample and comparing the performance to that observed in the derivation sample [24, 31]. External validation involves taking the original fully specified model, with its predictors and regression coefficients as estimated from the derivation study; measuring and documenting the predictor and outcome variables in a new patient sample; applying the original model to these data to predict the outcome of interest; and quantifying the predictive performance of the model by comparing the predictions with the observed outcomes [20]. Performance should be assessed using calibration, discrimination, and measures to quantify clinical usefulness such as decision curve analysis [163]. A CPR can also be refined if it demonstrates poor performance in an external validation study.

Regrettably, few CPRs are externally validated [27, 170, 171]. A systematic review of CPRs for children identified 101 CPRs addressing 36 conditions; of these, only 17% had narrow validation and only 8% had broad validation [170].

Study design for the external validation of a clinical prediction rule

Ideally, a validation study should be conducted prospectively, by enrolling new individuals in a specifically predesigned study, and the CPR should be applied to all patients meeting the study inclusion criteria [17, 23]. However, validation studies can be conducted retrospectively, using existing datasets. if adequate data on the predictor and outcome variables is available [23]. Investigators conducting a validation study should receive brief training on the accurate application of the CPR. If possible, all patients should be subjected to the reference standard, to establish their true outcome and enable comparison with the CPR prediction. However, in some cases this may not be feasible or practical, and an appropriate and sensible proxy outcome may be used instead [172]. Stiell and Wells [17] recommend that the inter-rater reliability of the interpretation of the CPR result is assessed, to determine if the CPR is being applied accurately and consistently. In terms of sample size, for a logistic regression model with six predictors, a minimum of 100 patients with the outcome of interest and 100 patients without the outcome of interest has been suggested [173]. Other authors propose that external validation studies require a minimum of 100 events, but ideally 200 events [174]. A minimum of 200 events and 200 non-events has been suggested in order to reliably assess moderate calibration and produce useful calibration plots [154]. The characteristics of patients included in a validation study should be described in detail, and compared with those included in the derivation study. To enhance the interpretation of external validation studies, it is possible to quantify the degree of relatedness between derivation and validation datasets, to determine the extent to which the CPR can be generalised to different populations [34]. Authors have also proposed benchmark values to distinguish between a case-mix effect and incorrect regression coefficients in external validation studies, and therefore assist in the interpretation of a CPR's performance in validation samples [175]. Similarly, a model-based concordance measure has recently been derived that enables quantification of the expected change in a CPR's discriminative ability owing to case-mix heterogeneity [176].

Types of external validation

Many types of external validation are recognized in the literature, but all types consider patients that differ in some respect from the patients included in the derivation study [26]. The greater the differences between the patients in the derivation and validation samples, the stronger the test of generalisability of the CPR [24]. Three types of external validation have received the most attention, namely *temporal* validation, *geographical* validation and *domain* validation [147].

In *temporal* validation studies, the CPR is tested on patients in the same centre(s) but over a different time period [146]. *Geographical* validation studies examine the generalisability of the CPR to other centres, institutes, hospitals or countries [146]. Patient characteristics are likely to vary between locations, and predictor and outcome variables are likely to be interpreted and measured differently in different places, leading to greater differences between the derivation and validation populations than in a temporal validation study [24, 147]. In *domain* validation the CPR is tested in very different patients than those from whom it was derived, for example in patients from a different setting (e.g. primary or secondary care), or in patients of different ages (e.g. adults vs. children). The case mix of patients included in a domain validation study will clearly differ from the derivation population [147]. Differences between the derivation and validation populations are generally smallest in a temporal validation study, and greatest in a domain validation study; therefore good performance of a CPR in a temporal validation study may only provide weak evidence that the CPR can be generalised to new patients, while good performance in a domain validation study can be considered as the strongest evidence of generalisability [147]. Other types of external validation studies include *methodologic* validation which refers to testing using data collected via different methods, *spectrum* validation which refers to testing in patients with different disease severity or prevalence of the outcome of interest, and fully *independent* validation which refers to testing by independent investigators at different sites [26, 146]. A recent study of cardiovascular risk CPRs found that very few were externally validated by independent researchers; to increase the chance of fully independent validation, researchers should report all the information required for risk calculation, to ensure replicability [177].Some authors have found that CPRs demonstrate worse performance in fully independent external validation studies compared to temporal or geographical external validation studies [26, 28], while others have found no difference [178]. When multiple external validations of a CPR have been performed, it is useful to conduct a formal meta-analysis to summarize its overall performance across different settings and to assess the circumstances under which the CPR may need adjusting; a recently published framework provides guidance on how to do this [35].

Refinement of a clinical prediction rule: model updating or adjustment

When researchers encounter an inferior performance of a CPR in an external validation study compared with that found in the derivation study, there is a temptation to reject the CPR and derive an entirely new one in the often considerably smaller validation dataset [147, 179]. This approach leads to a loss of scientific information captured in the derivation study and an abundance of CPRs developed for the same clinical situation, leaving clinicians in a quandary over which one to use [24, 147]. However, a reduction in performance is to be expected in an external validation study [24, 26, 147]. The recommended alternative is to update, adjust or recalibrate the CPR using the validation data, thereby combining information captured in the original CPR with information from new patients and improving generalisability [22, 180, 181]. Several methods for updating CPRs are available. When the outcome prevalence in the validation study is different to that in the derivation study, calibration in the validation sample will be affected, but can be improved by adjusting the baseline risk (intercept) of the original model to the patients in the validation sample [179]. If the CPR is overfitted or underfitted, calibration can be improved by simultaneously adjusting all of the regression coefficients [24]. To improve discrimination, individual regression coefficients can be re-estimated, or additional predictors can be added [24, 179]. Ideally, updated CPRs that are adjusted to validation samples should themselves be externally validated, just like newly derived CPRs [147].

Comparing the performance of clinical prediction rules

Once a CPR has been externally validated, it is useful to compare its performance with the performance of other existing CPRs for the same condition [59]. Improvements in discrimination can be assessed by quantifying the difference in the AUROC or equivalent c-index between two CPRs [182], however this approach is inappropriate in the case of nested models that are fitted in the same data set [183]. Novel metrics have been proposed that quantify the extent to which a new CPR improves the classification of individuals with and without the outcome of interest into predefined risk groups [46]. These include the Net Reclassification Improvement (NRI), and the Integrated Discrimination Improvement (IDI) [184]. Various decision-analytic approaches to model comparison have also been proposed [185]. All of these measures can be used for comparing both nested and non-nested models. However, both the NRI and IDI statistics have come under intense scrutiny in the literature and many researchers caution against their use, as positive values may arise simply due to poorly fitted models [30, 186-190]. Therefore, the NRI and IDI statistics cannot be recommended [191]. Decision-analytic methods are increasingly recommended as they incorporate misclassification costs and therefore indicate the clinical usefulness of CPRs [185]. A systematic review of comparisons of prediction models for cardiovascular disease found that formal and consistent statistical testing of the differences between models was lacking and that appropriate risk reclassification measures were rarely reported [192]. A recent commentary provides a useful and comprehensive overview of the advantages and disadvantages of the various methods available for quantifying the added value of new biomarkers [193].

Reporting the external validation of a clinical prediction rule

External validation studies of CPRs are often poorly reported [9]; researchers should adhere to the TRIPOD checklist and accompanying guidelines [36].

Stage 4: Impact of a clinical prediction rule on clinical practice

Since the ultimate aim of a CPR is to improve the quality of patient care, the effect of a validated CPR on clinician behaviour and patient outcomes should be examined in what are known as impact analysis studies [22, 24]. It is increasingly recognised that CPR's should be regarded as complex interventions, as the introduction of a CPR into clinical practice with subsequent management decisions consists of multiple interacting components [106, 194-200]. The impact of a CPR on clinical practice will depend on several interacting factors, including the accuracy and applicability of the CPR, clinicians' interpretation of probabilities, and clinicians' adherence to and acceptance of the CPR [195]. Evaluating the impact of a CPR has been described as 'the next painful step' in the development process [201]. Impact analysis studies clearly differ from validation studies as they must be comparative, typically requiring a control group of clinicians providing usual care [22, 24, 32]. It is possible to assess the impact of both assistive CPRs that simply provide predicted probabilities, and directive *decision* rules that suggest a specific course of action based on probability categories [32]. Assistive CPRs respect clinicians' individual judgment and leave room for intuition, whereas directive rules may be more likely to influence clinician behaviour [32, 202, 203]. However it is not guaranteed that clinicians will follow CPR, or the recommendations provided by directive rules [32]. Therefore an impact study must demonstrate that clinical behaviour can be altered and patient care improved by the CPR, prior to widespread dissemination and implementation [17].

Unfortunately, even fewer CPRs undergo an impact assessment than undergo external validation. In the systematic review of 101 CPRs for children, none had impact analysis performed [170]. An evaluation of 434 primary care CPRs found that only 12 had undergone impact analysis [171]. A subsequent systematic review of the impact of primary care CPRs found 18 studies relating to 14 CPRs, with 10/18 studies demonstrating an improvement in primary outcome when the CPR was used compared to usual care [204]. This review cautioned that the small number of impact analysis studies found, precluded the possibility of drawing firm conclusions about the overall effectiveness of CPRs in primary care, with the authors pointing out that the methodological quality of the included studies was unclear due to incomplete reporting [204]. Another recent systematic review of the impact of CPRs found that the intermediate consequences of a CPR such as clinical management decisions were the primary outcome in the majority of studies, while few studies aimed to establish the effect of a CPR on patient outcomes [205]. In addition, in many of the included studies, the risk of bias was either high or unclear [205]. Finally, a study describing the distribution of derivation, validation and impact studies in four reviews of leading medical journals since 1981 demonstrated that a minority of studies concerned CPR impact (10/201), with the pattern remaining stable over time [27].

Study design for an impact analysis

Before carrying out a formal impact study, researchers must consider whether the CPR is ready for implementation [106, 206]. If possible, the predictive performance of the CPR should be verified in the new setting, and the CPR tailored to the new setting to enhance performance [106]. The optimal study design for an impact analysis is a cluster randomised trial with centres as clusters [22]. Randomising individual patients is not recommended as clinicians may learn the rule and apply it to patients randomised to the control group [22]. Randomising clinicians is preferable but requires more patients, and may lead to contamination of experience between clinicians in the same centre [24, 207]. An attractive variant of a cluster randomised trial is the stepped-wedge cluster randomised trial. In a stepped-wedge design, all centres apply care-as-usual, and then use the CPR at different, randomly allocated time periods [208]. This design allows for the comparison of outcomes both within and between hospitals, generates a wealth of data regarding potential barriers to implementation, and is particularly beneficial if the

CPR turns out to have a promising effect [209]. When the outcome of interest in an impact study is clinician behaviour or decision-making, a cross-sectional randomised study without patient follow-up is sufficient, with randomisation at either the patient or clinician level. However, to determine the impact of a CPR on patient outcomes or cost-effectiveness, follow-up of patients is essential [22].

Given the significant practical, logistic and economic challenges associated with cluster randomised trials, non-randomised approaches are possible and are often used. Cluster randomised trials can be expensive and time-consuming and it may be difficult to recruit an adequate number of clusters [24, 106]. A suggested rule-of thumb is to regard four clusters per arm as the absolute minimum number required [210], however methods for determining sample size in cluster randomised trials have been proposed by a number of authors [211-213]. A popular design is a before–after study, in which outcomes are assessed in a time period before a CPR is available and compared with outcomes measured in a time period after it is introduced; this design is susceptible to temporal confounding [24]. Finally, a relatively low-cost and simple design is a before–after study within the same clinicians. In this design, clinicians are asked to indicate their treatment or management decision or perceived risk of disease for the same patient both before, and after, receiving the CPR prediction [24]. Single centre impact studies are recommended to inform the planning of multicentre randomised trials [32]. As with derivation and validation studies, a sample size calculation should be performed, with consideration of all relevant impact measures, and where possible assessment of outcome measures should be blinded to the CPR predictions and recommendations [32, 33]. Clinicians must undergo training in order to correctly interpret and use the CPR [17].

The impact of CPRs can also be estimated indirectly using decision analytic modelling, which integrates information on CPR predictions and information about the effectiveness of treatments from therapeutic intervention studies [214, 215]. Such studies cost less, and take less time, than RCTs, however they are limited by the quality of available evidence, and only provide theoretical indications of the impact CPRs may have on patient outcomes. Thus it has been suggested that they should not replace RCTs but rather be performed as an intermediate step prior to an RCT [216].

Measures of impact of a clinical prediction rule

During an impact analysis study the sensitivity and specificity of the CPR should be recalculated to determine its accuracy in the new study population [17]. However, measures of CPR *accuracy* are not synonymous with measures of *impact*, and only represent the *potential* impact of the CPR [32]. This is because clinicians are unlikely to follow the logic of the CPR or its recommendations in every case; they may not use the CPR at all, they may not use it correctly, they may deliberately disregard its predictions or suggestions, or they may be unable to use it for other reasons [32]. Measures that are assessed in traditional RCTs include safety, which refers to any adverse events resulting from the implementation of an intervention, and efficacy, which relates to the extent that an intervention helps to improve patient outcomes, for example by reducing mortality rates [217]. In addition, Reilly and Evans [32] propose that the impact of a CPR is assessed in terms of its ‘safety’ and ‘efficiency’, where safety is defined as the proportion of patients found to have the outcome of interest and who received the appropriate intervention, and efficiency is defined as the proportion of patients *without* the outcome of interest and who *did not* receive the intervention. The sensitivity and specificity of a CPR will only be the same as its safety and efficiency if clinicians follow the logic and recommendations of the CPR exactly [32]. Therefore, in an impact analysis study a CPR may demonstrate more, or less, actual impact than its potential impact. The effect of clinicians' incorrect use of the CPR, or their deviations from its logic or suggestions can provide important insights into its impact under specific circumstances, and may reveal complex interactions between clinicians and the CPR [32]. For example, Reilly and colleagues [218] found that when clinicians did not consult a CPR for suspected acute cardiac ischemia at all, or overruled its recommendations, their decisions were less efficient than if they had followed the CPR in every case.

Acceptability of a clinical prediction rule

If the use of a CPR is warranted but it is not used, the considerable time, money and effort that goes into its development and evaluation is wasted. Assessing the acceptability of a CPR is therefore crucial for successful implementation. Even valid and reliable CPRs may not be accepted or used by clinicians [17]. Impact studies allow researchers to evaluate the acceptability of a CPR to clinicians, patients, or others who may use it, as well as its ease of use and barriers to its uptake [22]. If a CPR proves to be acceptable, its long-term and widespread dissemination and implementation would be justified; if not, the CPR could undergo modification and further evaluation [219]. Acceptability of a CPR and attitudes towards it can be determined via survey, qualitative, simulation or clinical vignette studies [33, 219-222]. The validated Ottawa Acceptability of Decision Rules survey instrument can be used both to measure the overall acceptability of a CPR, and to assess specific barriers to its use, which can inform potential improvements to the CPR as well as the design of dedicated implementation strategies [219]. Qualitative studies can be invaluable for determining the acceptability of a CPR but are relatively rare [199, 220, 222-225].

Comparison of a clinical prediction rule with unstructured clinical judgment

For a CPR to improve the diagnostic accuracy of clinicians, its performance in distinguishing between patients with and without the condition of interest should be superior to that of unstructured clinical judgment alone. Therefore a vital metric is the comparison of the accuracy of the CPR predicted probabilities of disease, or recommended decisions, with the accuracy of clinicians own estimated disease probabilities or management decisions [18]. The sensitivity and specificity of clinicians' predictions or decisions are generally measured under usual practice, and compared to the sensitivity and specificity of the CPR predictions or decisions when applied to the same patients [226, 227]. Some studies have used clinical vignettes [228] while others have used multivariable logistic models to assess the added value of a CPR over and above clinical judgment alone [229]. If it can be demonstrated that the performance of a CPR is superior to unaided clinician judgment, this may aid clinicians' acceptance and use of the CPR [32]. Although comparison of a CPR to clinician suspicion regularly takes place at the impact analysis stage, some researchers have recommended that this is carried out during the derivation or validation stages, arguing that if the CPR does not add anything beyond clinical judgment, then the use of the CPR and an impact study would not be warranted [230]. In addition, Finnerty and colleagues [231] recommend that comparison is undertaken in multiple settings, as the performance of a CPR may be superior to clinical judgment in certain settings, but inferior or no different in other settings. A recent systematic review comparing CPRs with clinical judgment concluded that the differences between the two methods of judgment are likely due to different diagnostic thresholds, and that the preferred judgment method in a given situation would therefore depend on the relative benefits and harms resulting from true positive and false positive diagnoses [232]. Brown and colleagues' [199] found that the use and potential advantages of a CPR may be much more complex than originally thought, and that CPRs may be useful for purposes not previously reported, such as enhancing communication with colleagues and patients, and medico-legal purposes. Recent studies in the child protection field have demonstrated that CPRs may provide clinicians with additional confidence in their decision-making, even if they do not alter their management actions based on the CPRs risk prediction [220, 233].

The four phases of impact analysis for clinical prediction rules

Despite the abundance of methodological guidelines for the derivation and validation of CPRs [26], there is a lack of clear guidance for the design, conduct and reporting of impact analysis studies of CPRs. To this end, Wallace and colleagues [33] formulated an iterative four-phased framework for the impact analysis of CPRs, specifying the importance of substantial preparatory and feasibility work prior to the conduct of a full-scale formal experimental study (Figure 2). Phase 1 involves determining whether the CPR is ready for impact analysis i.e. whether it has been rigorously derived and broadly validated according to pre-defined methodological standards. Phase 2 includes assessing the acceptability of the CPR and identifying potential barriers to its uptake and implementation, as well as assessing the feasibility of conducting an impact study. Evaluating the feasibility of carrying out an impact study involves consideration of multiple factors including the most appropriate study design for measuring relevant outcomes, and how the CPR will be delivered at the point of care or integrated into the clinical workflow. Phase 3 involves formally testing the impact of the CPR using a comparative study design. Phase 4 involves long-term dissemination and implementation of the CPR, which corresponds to Stage 6 in the development of CPRs, discussed below.

Reporting the impact analysis of a clinical prediction rule

There are currently no published reporting guidelines for studies analysing the impact of CPRs. This is a gap in the literature, and a priority for future research. However, researchers assessing the impact of CPRs in an RCT may refer to guidelines on the reporting of clinical trials, such as the Consolidated Standards of Reporting Trials (CONSORT) statement [217].

Stage 5: Cost-effectiveness of the clinical prediction rule

If an impact analysis study shows that a CPR demonstrates safety and efficiency, alters clinician behaviour and improves clinical care, a formal economic evaluation can be carried out to determine the cost-effectiveness of the CPR. The aim is to establish the health care savings associated with routine use of the CPR in clinical practice [17]. Economic evaluation is usually based on decision analytic models [234]. Any economic evaluation must make reasonable assumptions about the accuracy and effectiveness of the CPR and the costs involved [17]. Sensitivity analyses should be performed by re-running models with alternative assumptions, to examine the uncertainty of the model projections [234]. In reality, many economic evaluations are conducted prior to an impact analysis study or even an external validation study, perhaps because they are relatively quick and low cost to perform, and provide a significant part of the justification for the development and implementation of a CPR.

Stage 6: Long-term implementation and dissemination of the clinical prediction rule

The gap between evidence and practice has been consistently demonstrated in health services research [235], and there is no guarantee that a CPR will be widely disseminated or used, even if it is shown to have a positive impact on clinical care and cost benefits. Therefore, in order to maximise the uptake of a CPR, an active dissemination and implementation plan must be in place. Simple passive diffusion of study results via publication in journals or presentations at conferences is unlikely to significantly change clinical practice [236]. Examples of dissemination include actively targeting specific audiences via direct mail or the press, while implementation involves the use of local administrative, educational, organizational and behavioural strategies to put the CPR into effect in clinical practice [236]. Active broad dissemination of the widely accepted Ottawa ankle rule via an educational intervention found no impact of the rule on clinicians' use of ankle radiography [237], leading the authors to recommend implementation strategies at the local level instead. Some implementation strategies have been found to be more effective than others in changing clinician behaviour. A systematic review found the most effective approaches to be reminders in the form of posters, pocket cards, sheets or computer-embedded prompts, face-to-face local clinician education, and the use of multiple interventions simultaneously [238]. Incorporation of CPRs into clinical guidelines may also be of benefit; a recent study found that clinical guidelines and local policies that mandated the use of CPRs were effective in increasing their adoption in clinical practice [199]. In addition, the integration of CPRs into the clinical workflow via electronic health records may promote their use [239]. Since impact in a research study does not ensure impact in real world clinical practice, follow-up of clinicians can be conducted to assess the long-term use and effect of the CPR [17, 33].

Barriers and facilitators to the use of clinical prediction rules

Clearly, identifying the barriers and facilitators to the implementation of CPRs is crucial for the development of targeted implementation strategies that may encourage clinicians to use the CPR. The adoption of CPRs into clinical practice is influenced by various factors including clinician characteristics, patient factors, features of the CPR itself and environmental factors [32, 64, 221, 224, 225, 240-252].

Table 3 provides an overview of the barriers to the adoption of CPRs identified in the literature [253], grouped according to their effect on clinician knowledge, attitudes or behaviours [254]. Barriers relating to knowledge include lack of awareness of the CPR or the burden of the clinical problem it applies to, unfamiliarity with the CPR, and a lack of understanding of the purpose of CPRs in general [225, 240-242]. Clinicians may also be unaware of a CPR due to the increasing volume of CPRs, particularly when they are developed for the same condition [59, 243]. Common barriers relating to clinician attitude include a conviction that clinical judgment is superior to the CPR, and distrust of the accuracy of the CPR [32, 224, 240, 241, 244, 245]. Barriers relating to behaviour include organizational factors [251], the complexity of the CPR and the time it takes to apply; survey studies suggest that clinicians much prefer a CPR that is simple to use, memorable, and saves time [221, 246, 247]. Complex models such as those based on machine and artificial learning algorithms may introduce additional barriers relating to applicability and usability, due to their potential lack of reproducibility and transparency [58, 80]. Other studies have demonstrated that clinicians will be unlikely to use a CPR if there are predictors missing which are deemed to be important, or if the predictor variables are not logically related to the outcome variable [32, 225]. Reilly & Evans [32] offer a number of strategies for overcoming barriers to the use of CPRs. These include emphasizing the discretionary use of the CPR, comparing clinical judgment with the CPR, checking whether any excluded factors affect the CPR predictions, performing a simulated impact analysis and soliciting clinicians input regarding the logic and format of the CPR, among others [32].

Insert Table 3

Summary

For CPRs to be useful in clinical practice, they must be properly planned [65], derived using appropriate statistical techniques [23], and externally validated in multiple settings and by independent investigators to determine their predictive accuracy [147]. In addition, CPRs must undergo impact analysis to determine their effect on clinician behaviour and relevant patient outcomes [22]. There are numerous factors to consider when deriving, validating and assessing the impact of a CPR including the study design, preparatory work, statistical analysis, modelling strategy, performance/impact measures, the presentation of the CPR, and the reporting of the study methodology. New CPRs should only be derived when there is a clear clinical need for them [17]. There is an urgent need to change the focus from the derivation of CPRs, to the validation and impact analysis of existing ones [33]. The CPR must be presented in full, and the study methods reported adequately, to ensure its quality, risk of bias and clinical utility can be evaluated; the TRIPOD guidelines should be followed to ensure completeness of reporting requirements [36]. Feasibility and preparatory work is essential to determine whether a formal impact study of the CPR is warranted [33, 106], and survey and qualitative work should be undertaken to verify whether the CPR is acceptable and relevant to clinicians [63, 219, 220, 222]. If a CPR is found to have a positive impact on patient outcomes, its cost-effectiveness should be evaluated, and a targeted implementation and dissemination strategy devised, with consideration of possible barriers to implementation, to maximize uptake [17].

In summary, the development and evaluation of a robust, clinically useful CPR with high predictive accuracy is challenging, and research in the field concerning derivation, validation and impact evaluation continues to evolve. However, adhering to the existing methodological standards and recommendations in the literature at every step will help to ensure a rigorous CPR that has the potential to contribute usefully to clinical practice and decision-making.

Table 1. Stages in the development and evaluation of clinical prediction rules

Stage of development	Methodological standards
Stage 1. Identifying the need for a CPR	• Consider conducting qualitative research with clinicians to determine clinical relevance and credibility of CPR • Conduct a systematic review of the literature to identify and evaluate existing CPRs developed for the same purpose • Consider updating, validating or testing the impact of existing CPRs
Stage 2. Derivation of a CPR according to methodological standards	**Study design for the derivation of a CPR** • Consider registering the study and publishing a protocol • Ensure the dataset is representative of the population for whom the CPR is intended • Conduct a prospective multicentre cohort study **Statistical analysis** • Conduct multivariable regression analysis (logistic for binary outcomes, Cox for long-term prognostic outcomes) • Identify the model to be used, plus rationale if other methods used **Missing data** • Use multiple imputation

	Selection of candidate predictors for inclusion in a multivariable model • Only include relevant predictors based on evidence in the literature/clinical experience • Aim for a sample size with a minimum of ten events per predictor, preferably more • Avoid selection based on univariable significance testing • Avoid categorizing continuous predictors **Selection of predictors during multivariable modelling** • Backward elimination of predictors is preferred • Avoid data-driven selection and incorporate subject-matter knowledge into the selection process **Definition and assessment of predictor and outcome variables** • Define predictor and outcome variables clearly • Consider inter-rater reliability of predictor measurement and potential measurement error • Aim for blind assessment of predictor and outcome variables **Internal validation** • Use cross-validation or bootstrapping and adjust for optimism • Ensure to repeat each step of model development if using bootstrapping **CPR performance measures** • Assess and report both calibration and discrimination • Consider decision curve analysis to estimate the clinical utility of the CPR **Presentation of a CPR** • Report the regression coefficients of the final model, including the intercept or baseline hazard • Consider a clinical calculator if the CPR is complex **Reporting the derivation of a CPR** • Adhere to the TRIPOD guidelines [36]
Stage 3. External validation and refinement of a CPR	**Study design for the external validation of a CPR** • Conduct a prospective multicentre cohort study • Aim for a sample size with a minimum of 100 outcome events, preferably 200 • Consider using a framework of generalisability to enhance the interpretation of the findings [34] **Types of external validation** • Conduct temporal, geographical and domain validation studies to ensure maximum generalisability • If multiple validations have been performed, conduct a meta-analysis to summarize the overall performance of the CPR, using a published framework [35] **Refinement of a CPR: model updating or adjustment**

	• Consider updating, adjusting or recalibrating the CPR if poor performance is found in an external validation study • Consider further external validation of updated CPRs **Comparing the performance of CPRs** • Compare the CPR with other existing CPRs for the same condition • Ensure the statistical procedures used for comparison are appropriate; consider a decision-analytic approach **Reporting the external validation of a CPR** • Adhere to the TRIPOD guidelines [36]
Stage 4. Impact of a CPR on clinical practice	**Study design for an impact analysis** • Consider whether the CPR is ready for implementation • Conduct a cluster randomised trial with centres as clusters, or a before–after study • Perform appropriate sample size calculations • Consider decision-analytic modelling as an intermediate step prior to a formal impact study **Measures of impact of a CPR** • Report the safety and efficacy of the CPR • Report the impact of the CPR on clinician behaviour if assessed **Acceptability of a CPR** • Evaluate the acceptability of the CPR using the validated OADRI [219], or using qualitative or vignette methods **Comparison of a CPR with unstructured clinical judgment** • Compare the sensitivity and specificity of the CPR with clinicians own predictions/decisions **The four phases of impact analysis for CPRs** • Follow the framework for the impact analysis of CPRs [33] • Ensure extensive preparatory and feasibility work is conducted prior to a formal impact study **Reporting the impact analysis of a CPR** • There are currently no published reporting guidelines for impact studies of CPRs; this is an area for future research
Stage 5. Cost-effectiveness	• Conduct a formal economic evaluation, with sensitivity analyses to examine the uncertainty of the model projections
Stage 6. Long-term implementation and dissemination	• Devise and evaluate targeted implementation strategies to ensure maximum uptake **Barriers and facilitators to the use of CPRs** • Assess barriers to the use of the CPR and devise strategies to overcome these

CPR: clinical prediction rule; TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis; OADRI: Ottawa Acceptability of Decision Rules Instrument

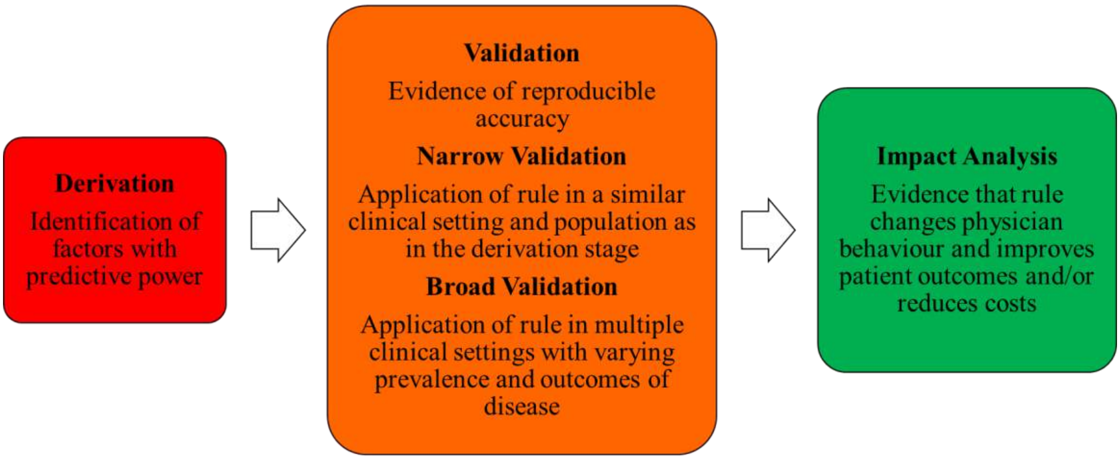
Table 3. Barriers to the use of clinical prediction rules in practice identified in the literature

Theme	Subtheme	Barrier
Knowledge	Awareness	Unaware: • That CPR exists • Of clinical problem or burden of clinical problem to which CPR applies Unable to choose from multiple CPRs
	Familiarity	Unfamiliar with CPR
	Understanding	Lack of knowledge and understanding of the purpose, development and application of CPRs in general
	Forgetting	Clinician forgets to use CPR despite best intentions
Attitudes	Negative beliefs about CPRs	Belief that: • CPRs threaten autonomy • CPRs are too 'cook-book', and oversimplify the clinical assessment process • Clinical judgment is superior to CPRs • Clinical judgment is not error prone • Use of CPRs causes intellectual laziness • The development of the CPR was biased • Patients will deem clinicians less capable if using a CPR • CPRs only apply to the less experienced • Probabilities are not helpful for decision-making Dislike of the term 'rule' Clinician had a false negative result when using a CPR in the past Existing CPRs are not ready for clinical application
	Outcome expectancy	Belief that: • CPRs will not lead to improved patient or process outcomes • The information provided by the CPR is not sufficient to alter clinical decisions Clinician:

		• Fears unintended consequences of use • Is uncertain about using the CPR in patients with an atypical presentation • Worries that improving efficiency threatens patient safety
	Self-efficacy	Belief that the CPR is too difficult to use Clinician uncertain how to interpret or use CPR output
	Motivation	Clinician lacks motivation to use the CPR
Behaviour	Patient factors	Patients expectations are not consistent with the CPR
	Features of the CPR	Clinician: • Finds CPR too complicated • Finds CPR 'too much trouble' to apply Perception that: • The CPR is not an efficient use of time • The CPR does not have face validity or that important predictors are missing • The CPR does not fit in with usual work flow or approach to decision-making • The CPR is not generalisable to the clinician's patient • The CPR is static and does not consider the dynamic nature of clinical practice • Overruling the CPR is often justified Data required for the CPR is difficult to obtain
	Environmental factors	Lack of: • Time • Organisational support • Peer support for use Perceived increased risk of litigation Insufficient incentives or reimbursement for use of the CPR

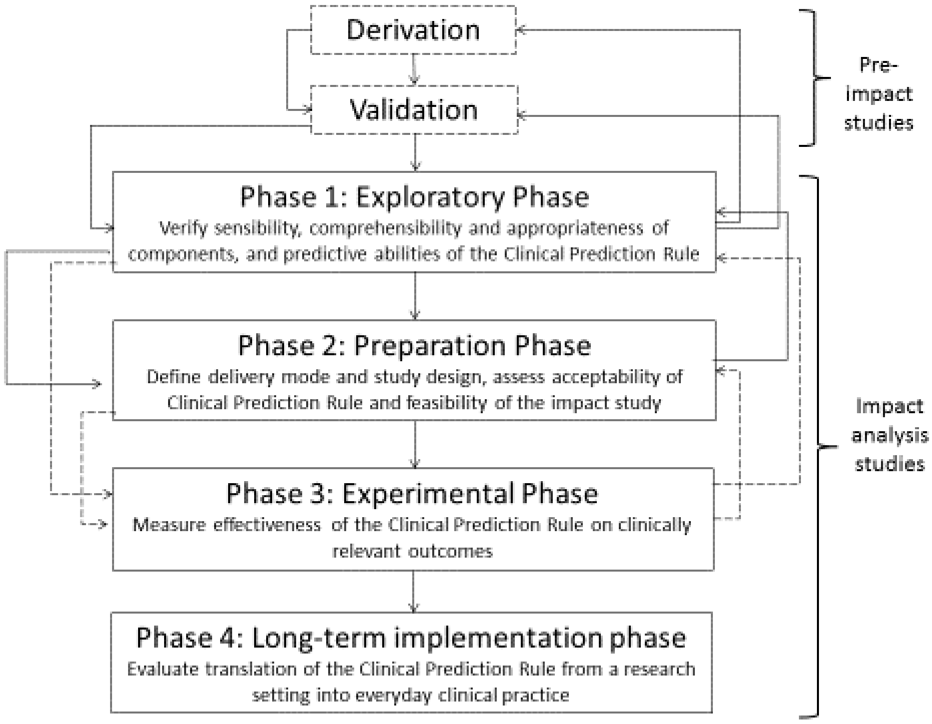
Adapted from Sanders, 2015 [253]. CPR = Clinical prediction rule

Figure 1. The three main stages in the development and evaluation of clinical prediction rules



Adapted from McGinn, 2016 [255]

Figure 2. The four phases of impact analysis for a clinical prediction rule



Reproduced with permission from Wallace et al. 2011 [33]

Abbreviations

AUROC: Area Under the Receiver Operating Characteristic curve; CPR: Clinical prediction rule; IDI: Integrated Discrimination Improvement; MAR: Missing at random; MCAR: Missing completely at random; MNAR: Missing not at random; NRI: Net reclassification Improvement; RCT: Randomised controlled trial; ROC: Receiver operating characteristic curve; TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Health and Care Research Wales (grant number HS-14-24). The funders had no involvement in the study design, the collection, analysis, or interpretation of the data, the writing of the manuscript or the decision to submit the manuscript for publication.

Authors' contributions

LEC conducted the literature search, drafted the manuscript, produced the tables, boxes and figures, and edited the manuscript. DMF, SM and AMK critically revised the manuscript for important intellectual content. All authors approved the final version submitted for publication.

Acknowledgements

We would like to thank Health and Care Research Wales for funding this work.

References

1. Adams ST, Leveson SH. Clinical prediction rules. BMJ. 2012;344**:**d8312.
2. Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Aust J Physiother. 2006;52(3)**:**157-63.
3. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA. 1997;277(6)**:**488-94.
4. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA. 2000;284(1)**:**79-84.
5. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. J Thromb Haemost. 2013;11(Suppl 1)**:**129-41.
6. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9(5)**:**1-12.
7. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med. 2010;8**:**20.
8. Collins GS, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med. 2011;9**:**103.
9. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;14**:**40.
10. Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, et al. Prognostic models in obstetrics: available, but far from applicable. Am J Obstet Gynecol. 2016;214(1)**:**79-90.e36.
11. Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ, Moons KG. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. Circulation. 2010;122(7)**:**682-9.

12. Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol. 2013;66(3):268-77.
13. Nayak S, Edwards DL, Saleh AA, Greenspan SL. Performance of risk assessment instruments for predicting osteoporotic fracture risk: a systematic review. Osteoporos Int. 2014;25(1):23-49.
14. Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. Cancer Invest. 2009;27(3):235-43.
15. Collins GS, Michaelsson K. Fracture risk assessment: state of the art, methodologically unsound, or poorly reported? Curr Osteoporos Rep. 2012;10(3):199-207.
16. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. N Engl J Med. 1985;313(13):793-98.
17. Stiell I, Wells G. Methodologic standards for the development of clinical decision rules in emergency medicine. Ann Emerg Med. 1999;33(4):437-47.
18. Green SM, Schriger DL, Yealy DM. Methodologic standards for interpreting clinical decision rules in emergency medicine: 2014 update. Ann Emerg Med. 2014;64(3):286-91.
19. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-31.
20. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.
21. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
22. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.
23. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683-90.
24. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-98.
25. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ. 2009;338:b604.
26. Steyerberg E. Clinical prediction models: a practical approach to development, validation and updating. New York: Springer-Verlag; 2009.
27. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381.
28. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453-73.
29. Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.
30. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. BJOG. 2017;124(3):423-32.

31. Collins GS, Ma J, Gerry S, Ohuma E, Odondi LO, Trivella M, et al. Risk prediction models in perioperative medicine: methodological considerations. Curr Anesthesiol Rep. 2016;6(3):267-75.
32. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144(3):201-09.
33. Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, et al. Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). BMC Med Inform Decis Mak. 2011;11:62.
34. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015;68(3):279-89.
35. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res. 2018; doi: 10.1177/0962280218785504.
36. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73.
37. Lo BWY, Fukuda H, Nishimura Y, Farrokhyar F, Thabane L, Levine MAH. Systematic review of clinical prediction tools and prognostic factors in aneurysmal subarachnoid hemorrhage. Surg Neurol Int. 2015;6:135.
38. Hopper AD, Cross SS, Hurlstone DP, McAlindon ME, Lobo AJ, Hadjivassiliou M, et al. Pre-endoscopy serological testing for coeliac disease: evaluation of a clinical decision tool. BMJ. 2007;334:729.
39. LaValley MP, Lo GH, Price LL, Driban JB, Eaton CB, McAlindon TE. Development of a clinical prediction algorithm for knee osteoarthritis structural progression in a cohort study: value of adding measurement of subchondral bone density. Arthritis Res Ther. 2017;19:95.
40. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 2008;5(8):e165.
41. Ferro JM, Bacelar-Nicolau H, Rodrigues T, Bacelar-Nicolau L, Canhão P, Crassard I, et al. Risk score to predict the outcome of patients with cerebral vein and dural sinus thrombosis. Cerebrovasc Dis. 2009;28(1):39-44.
42. Woo J, Leung J, Wong S, Kwok T, Lee J, Lynn H. Development of a simple scoring tool in the primary care setting for prediction of recurrent falls in men and women aged 65 years and over living in the community. J Clin Nurs. 2009;18(7):1038-48.
43. Scholz NN, Bäsler KK, Saur PP, Burchardi HH, Felder SS. Outcome prediction in critical care: physicians' prognoses vs. scoring systems. Eur J Anaesthesiol. 2004;21(8):606-11.
44. Kheterpal S, Tremper KK, Heung M, Rosenberg AL, Englesbe M, Shanks AM, Campbell DA. Development and validation of an acute kidney injury risk index for patients undergoing general surgery results from a national data set. Anesthesiology. 2009;110(3):505-15.

45. Pace N, Eberhart L, Kranke P. Quantifying prognosis with risk predictions. Eur J Anaesthesiol. 2012;29(1):7-16.
46. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology. 2010;21(1):128-38.
47. Sarasin FP, Reymond JM, Griffith JL, Beshansky JR, Schifferli JA, Unger PF, et al. Impact of the acute cardiac ischemia time-insensitive predictive instrument (ACI-TIPI) on the speed of triage decision making for emergency department patients presenting with chest pain: a controlled clinical trial. J Gen Intern Med. 1994;9(4):187-94.
48. Stiell IG, McDowell I, Nair RC, Aeta H, Greenberg G, McKnight RD, Ahuja J. Use of radiography in acute ankle injuries: physicians' attitudes and practice. CMAJ. 1992;147(11):1671-78.
49. Stiell IG, McKnight R, Greenberg GH, McDowell I, Nair RC, Wells GA, et al. Implementation of the Ottawa ankle rules. JAMA. 1994;271(11):827-32.
50. Anis AH, Stiell IG, Stewart DG, Laupacis A. Cost-effectiveness analysis of the Ottawa ankle rules. Ann Emerg Med. 1995;26(4):422-28.
51. Graham ID, Stiell IG, Laupacis A, McAuley L, Howell M, Clancy M, et al. Awareness and use of the Ottawa ankle and knee rules in 5 countries: can publication alone be enough to change practice? Ann Emerg Med. 2001;37(3):259-66.
52. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016;353:i2416.
53. Shariat SF, Karakiewicz PI, Margulis V, Kattan MW. Inventory of prostate cancer predictive tools. Curr Opin Urol. 2008;18(3):279-96.
54. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. BMC Med Inform Decis Mak. 2006;6:38.
55. Wessler BS, Lai Yh L, Kramer W, Cangelosi M, Raman G, Lutz JS, Kent DM. Clinical prediction models for cardiovascular disease: Tufts predictive analytics and comparative effectiveness clinical prediction model database. Circ Cardiovasc Qual Outcomes. 2015;8(4):368-75.
56. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One. 2012;7(2):e32844.
57. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.
58. Moons KM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170(1):W1-W33.
59. Collins GS, Moons KG. Comparing risk prediction models. BMJ. 2012;344:e3186.
60. Dekker FW, Ramspek CL, van Diepen M. Con: most clinical risk scores are useless. Nephrol Dial Transplant. 2017;32(5):752-55.
61. Masconi K, Matsha T, Erasmus R, Kengne A. Recalibration in validation studies of diabetes risk prediction models: a systematic review. Int J Stat Med Res. 2015;4(4):347-69.

62. Ban JW, Wallace E, Stevens R, Perera R. Why do authors derive new cardiovascular clinical prediction rules in the presence of existing rules? A mixed methods study. PLoS One. 2017;12(6):e0179102.
63. de Salis I, Whiting P, Sterne JA, Hay AD. Using qualitative research to inform development of a diagnostic algorithm for UTI in children. Fam Pract. 2013;30(3):325-31.
64. Haskins R, Osmotherly PG, Southgate E, Rivett DA. Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. Physiotherapy. 2015;101(1):44-49.
65. Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. PLoS Med. 2014;11(7):e1001671.
66. Altman DG. The time has come to register diagnostic and prognostic research. Clin Chem. 2014;60(4):580-2.
67. Han K, Song K, Choi BW. How to develop, validate, and compare clinical prediction models involving radiological parameters: study design and statistical methods. Korean J Radiol. 2016;17(3):339-50.
68. Lee Y-h, Bang H, Kim DJ. How to establish clinical prediction models. Endocrinol Metab (Seoul). 2016;31(1):38-44.
69. Biesheuvel CJ, Vergouwe Y, Oudega R, Hoes AW, Grobbee DE, Moons KG. Advantages of the nested case-control design in diagnostic research. BMC Med Res Methodol. 2008;8:48.
70. Sanderson J, Thompson SG, White IR, Aspelund T, Pennells L. Derivation and assessment of risk prediction models using case-cohort data. BMC Med Res Methodol. 2013;13:113.
71. Nee RJ, Coppieters MW. Interpreting research on clinical prediction rules for physiotherapy treatments. Man Ther. 2011;16(2):105-08.
72. Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Phys Ther. 2009;89(7):698-704.
73. Labarère J, Renaud B, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med. 2014;40(4):513-27.
74. Grobman WA, Stamilio DM. Methods of clinical prediction. Am J Obstet Gynecol. 2006;194(3):888-94.
75. van den Bosch JE, Kalkman CJ, Vergouwe Y, Van Klei WA, Bonsel GJ, Grobbee DE, Moons KG. Assessing the applicability of scoring systems for predicting postoperative nausea and vomiting. Anaesthesia. 2005;60(4):323-31.
76. Hilbe J. Logistic regression models. Boca Raton, FL: Chapman & Hall/CRC; 2009.
77. Marshall RJ. The use of classification and regression trees in clinical epidemiology. J Clin Epidemiol. 2001;54(6):603-09.
78. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med. 1992;21(4):384-90.
79. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
80. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and AI research for patient benefit: 20 critical questions on

transparency, replicability, ethics and effectiveness. CoRR. 2018;abs/1812.10404.
81. Vergouwe Y, Royston P, Moons KG, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. J Clin Epidemiol. 2010;63(2):205-14.
82. Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley & Sons; 2002.
83. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
84. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59(10):1087-91.
85. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. J Clin Epidemiol. 2006;59(10):1092-101.
86. Janssen KJM, Donders ART, Harrell FE, Vergouwe Y, Chen Q, Grobbee DE, Moons KGM. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol. 2010;63(7):721-27.
87. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, Petersen I. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157-66.
88. van der Heijden GJMG, Donders AR, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J Clin Epidemiol. 2006;59(10):1102-09.
89. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons, Inc.; 1987.
90. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3).
91. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30(4):377-99.
92. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods. 2001;6(4):330-51.
93. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol. 2009;60:549-76.
94. Carpenter JR, Kenward MG, White IR. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. Stat Methods Med Res. 2007;16(3):259-75.
95. Demirtas H, Schafer JL. On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. Stat Med. 2003;22(16):2553-75.
96. Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. Pharmacoeconomics. 2018;36(8):889-901.
97. Leacy FP, Floyd S, Yates TA, White IR. Analyses of sensitivity to the missing-at-random assumption using multiple imputation with delta adjustment: application to a tuberculosis/HIV prevalence survey with incomplete HIV-status data. Am J Epidemiol. 2017;185(4):304-15.

98. Héraud-Bousquet V, Larsen C, Carpenter J, Desenclos J-C, Le Strat Y. Practical considerations for sensitivity analysis after multiple imputation applied to epidemiological studies with incomplete data. BMC Med Res Methodol. 2012;12:73.
99. Carpenter JR, Kenward MG. MAR methods for quantitative data. In: Missing data in randomised controlled trials— a practical guide. Birmingham: National Institute for Health Research; 2008.
100. Goldstein H, Carpenter J, Kenward MG, Levin KA. Multilevel models with multivariate mixed response types. Stat Model. 2009;9(3):173-97.
101. Schafer JL. Analysis of incomplete multivariate data. London: Chapman & Hall; 1997.
102. Dobson A, Diggle P, Henderson R. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000;1(4):465-80.
103. Rizopoulos D. Joint models for longitudinal and time-to-event data with applications in R. New York: Chapman and Hall/CRC; 2012.
104. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol. 2009;9:57.
105. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Med Res Methodol. 2010;10:112.
106. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. BMC Diagn Progn Res. 2018;2:11.
107. Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of clinical prediction models for application in local settings. Med Decis Making. 2012;32(3):E1-E10.
108. Janssen KJM, Vergouwe Y, Donders ART, Harrell FE, Chen Q, Grobbee DE, Moons KGM. Dealing with missing predictor values when applying clinical prediction models. Clin Chem. 2009;55(5):994-1001.
109. Masconi KL, Matsha TE, Erasmus RT, Kengne AP. Effects of different missing data imputation techniques on the performance of undiagnosed diabetes risk prediction models in a mixed-ancestry population of South Africa. PLOS ONE. 2015;10(9):e0139210.
110. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol. 1996;49(8):907-16.
111. Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema JDF. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Making. 2001;21(1):45-56.
112. Harrell FEJ, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.
113. Shmueli G. To explain or to predict? Stat Sci. 2010;25(3):289-310.
114. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, Omar RZ. How to develop a more accurate risk prediction model when there are few events. BMJ. 2015;351:h3868.
115. Heinze G, Dunkler D. Five myths about variable selection. Transpl Int. 2017;30(1):6-10.

116. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-79.
117. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and cox regression. Am J Epidemiol. 2007;165(6):710-18.
118. Courvoisier DS, Combescure C, Agoritsas T, Gayet-Ageron A, Perneger TV. Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. J Clin Epidemiol. 2011;64(9):993-1000.
119. van Smeden M, de Groot JAH, Moons KGM, Collins GS, Altman DG, Eijkemans MJC, Reitsma JB. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol. 2016;16:163.
120. van Smeden M, Moons KGM, de Groot JAH, Collins GS, Altman DG, Eijkemans MJC, Reitsma JB. Sample size for binary logistic prediction models: beyond events per variable criteria. Stat Methods Med Res. 2018; doi: 10.1177/0962280218784726.
121. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2016;76:175-82.
122. Battle CE, Hutchings H, Evans PA. Expert opinion of the risk factors for morbidity and mortality in blunt chest wall trauma: results of a national postal questionnaire survey of Emergency Departments in the United Kingdom. Injury. 2013;44(1):56-59.
123. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Stat Med. 2007;26(30):5512-28.
124. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127-41.
125. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. Stat Med. 2016;35(23):4124-35.
126. Steyerberg EW, Uno H, Ioannidis JPA, van Calster B, Ukaegbu C, Dhingra T, et al. Poor performance of clinical prediction models: the harm of commonly applied methods. J Clin Epidemiol. 2018;98:133-43.
127. Collins GS, Ogundimu EO, Cook JA, Le Manach Y, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. Stat Med. 2016;35(23):4124-35.
128. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester: John Wiley & Sons, Ltd; 2009.
129. Harrell FEJ, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst. 1988;80(15):1198-202.
130. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. J R Stat Soc Ser C Appl Stat. 1994;43(3):429-67.
131. Ambler G, Seaman S, Omar RZ. An evaluation of penalised survival methods for developing prognostic models with rare events. Stat Med. 2012;31(11-12):1150-61.

132. Le Cessie S, Van Houwelingen JC. Ridge estimators in logistic regression. J R Stat Soc Ser C Appl Stat. 1992;41(1):191-201.
133. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996;58(1):267-88.
134. Hosmer DW, Jovanovic B, Lemeshow S. Best subsets logistic regression. Biometrics. 1989;45(4):1265-70.
135. Mantel N. Why stepdown procedures in variable selection. Technometrics. 1970;12(3):621-25.
136. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem. 2004;50(3):473-76.
137. Steyerberg EW, Eijkemans MJC, Habbema JDF. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999;52(10):935-42.
138. Steyerberg EW, Schemper M, Harrell FE. Logistic regression modeling and the number of events per variable: selection bias dominates. Journal of Clinical Epidemiology. 2011;64(12):1464-65.
139. Whittle R, Peat G, Belcher J, Collins GS, Riley RD. Measurement error and timing of predictor values for multivariable risk prediction models are poorly reported. J Clin Epidemiol. 2018; doi: 10.1016/j.jclinepi.2018.05.008.
140. Luijken K, Groenwold RHH, van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on performance of prediction models: a measurement error perspective. arXiv:180610495 [statME]. 2018:arXiv:1806.10495.
141. Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. CJEM. 2008;10(2):174-75.
142. Moons KG, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? J Clin Epidemiol. 2002;55(7):633-36.
143. Wang LE, Shaw PA, Mathelier HM, Kimmel SE, French B. Evaluating risk-prediction models using data from electronic health records. Ann Appl Stat. 2016;10(1):286-304.
144. van Doorn S, Brakenhoff TB, Moons KGM, Rutten FH, Hoes AW, Groenwold RHH, Geersing GJ. The effects of misclassification in routine healthcare databases on the accuracy of prognostic prediction models: a case study of the CHA2DS2-VASc score in atrial fibrillation. BMC Diagn Progn Res. 2017;1:18.
145. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, Moons KG. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol. 2003;56(9):826-32.
146. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130(6):515-24.
147. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol. 2008;61(11):1085-94.
148. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.
149. Steyerberg EW. Validation in prediction research: the waste by data-splitting. J Clin Epidemiol. 2018; doi: 10.1016/j.jclinepi.2018.07.010.
150. Efron B, Tibshirani R. An introduction to the bootstrap. Boca Raton, FL: Chapman & Hall/CRC; 1993.

151. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. Biom J. 2008;50(4):457-79.
152. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med. 2014;33(3):517-35.
153. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 2000.
154. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167-76.
155. Pencina MJ, D'Agostino RBS. Evaluating discrimination of risk prediction models: the C statistic. JAMA. 2015;314(10):1063-64.
156. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
157. Baron JA, Sorensen HT. Clinical epidemiology. In: Olsen J, Saracci R, Trichopoulos D, editors. Teaching epidemiology: a guide for teachers in epidemiology, public health and clinical medicine. New York: Oxford University Press; 2010: 411-28.
158. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making. 2014;35(2):162-69.
159. Meurer WJ, Tolles J. Logistic regression diagnostics: understanding how well a model predicts outcomes. JAMA. 2017;317(10):1068-69.
160. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol. 2008;56(1):45-50.
161. Søreide K. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. J Clin Pathol. 2009;62(1):1-5.
162. Ebell MH, Locatelli I, Senn N. A novel approach to the determination of clinical decision thresholds. BMJ Evid Based Med. 2015;20(2):41-47.
163. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-74.
164. Baker SG, Cook NR, Vickers A, Kramer BS. Using relative utility curves to evaluate risk prediction. J R Stat Soc Ser A Stat Soc. 2009;172(4):729-48.
165. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:i6.
166. Feldstein DA, Hess R, McGinn T, Mishuris RG, McCullagh L, Smith PD, et al. Design and implementation of electronic health record integrated clinical prediction rules (iCPR): a randomized trial in diverse primary care settings. Implement Sci. 2017;12(1):37.
167. Van Belle V, Van Calster B. Visualizing risk prediction models. PLoS One. 2015;10(7):e0132614.
168. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. Stat Med. 2004;23(10):1631-60.
169. Cole TJ. Algorithm AS 281: scaling and rounding regression coefficients to integers. J R Stat Soc Ser C Appl Stat. 1993;42(1):261-68.

170. Maguire JL, Kulik DM, Laupacis A, Kuppermann N, Uleryk EM, Parkin PC. Clinical prediction rules for children: a systematic review. Pediatrics. 2011;128(3):e666-e77.
171. Keogh C, Wallace E, O'Brien KK, Galvin R, Smith SM, Lewis C, et al. Developing an international register of clinical prediction rules for use in primary care: a descriptive analysis. Ann Fam Med. 2014;12(4):359-66.
172. Stiell IG, Greenberg GH, Wells GA, McDowell I, Cwinn AA, Smith NA, et al. Prospective validation of a decision rule for the use of radiography in acute knee injuries. JAMA. 1996;275(8):611-15.
173. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol. 2005;58(5):475-83.
174. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med. 2016;35(2):214-26.
175. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol. 2010;172(8):971-80.
176. van Klaveren D, Gönen M, Steyerberg EW, Vergouwe Y. A new concordance measure for risk prediction models in external validation settings. Stat Med. 2016;35(23):4136-52.
177. Ban J-W, Stevens R, Perera R. Predictors for independent external validation of cardiovascular risk clinical prediction rules: cox proportional hazards regression analyses. BMC Diagn Progn Res. 2018;2:3.
178. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015;68(1):25-34.
179. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol. 2008;61(1):76-86.
180. Ivanov J, Tu JV, Naylor CD. Ready-made, recalibrated, or remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery. Circulation. 1999;99(16):2098-104.
181. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23(16):2567-86.
182. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
183. Demler OV, Pencina MJ, D'Agostino RBS. Misuse of DeLong test to compare AUCs for nested models. Stat Med. 2012;31(23):2577–87.
184. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27(2):157-72.
185. Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW. Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures. Med Decis Making. 2013;33(4):490-501.

186. Leening MJ, Steyerberg EW, Van Calster B, D'Agostino RB, Sr., Pencina MJ. Net reclassification improvement and integrated discrimination improvement require calibrated models: relevance from a marker and model perspective. Stat Med. 2014;33(19):3415-8.
187. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. Ann Intern Med. 2014;160(2):122-31.
188. Pepe MS, Fan J, Feng Z, Gerds T, Hilden J. The net reclassification index (NRI): a misleading measure of prediction improvement even with independent test data sets. Stat Biosci. 2015;7(2):282-95.
189. Burch PM, Glaab WE, Holder DJ, Phillips JA, Sauer JM, Walker EG. Net reclassification index and integrated discrimination index are not appropriate for testing whether a biomarker improves predictive performance. Toxicol Sci. 2017;156(1):11-13.
190. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. Stat Med. 2014;33(19):3405-14.
191. Antolini L, Tassistro E, Valsecchi MG, Bernasconi DP. Graphical representations and summary indicators to assess the performance of risk predictors. Biom J. 2018; doi: 10.1002/bimj.201700186.
192. Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP. Comparisons of established risk prediction models for cardiovascular disease: systematic review. BMJ. 2012;344:e3318.
193. Cook NR. Quantifying the added value of new biomarkers: how and how not. BMC Diagn Progn Res. 2018;2:14.
194. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ. 2012;344:e686.
195. White H. Theory-based impact evaluation: principles and practice. J Dev Effect. 2009;1(3):271-84.
196. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350:h1258.
197. Dowding D, Lichtner V, Closs SJ. Using the MRC framework for complex interventions to develop clinical decision support: a case study. In: Randell R, Cornet R, McCowan C, Peek N, Scott PJ, editors. Studies in health technology and informatics. IOS Press; 2017: 544-48.
198. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ. 2011;343:d7163.
199. Brown B, Cheraghi-Sohi S, Jaki T, Su T-L, Buchan I, Sperrin M. Understanding clinical prediction models as ‘innovations’: a mixed methods study in UK family practice. BMC Med Inform Decis Mak. 2016;16:106.
200. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337:a1655.
201. Lee TH. Evaluating decision aids. J Gen Intern Med. 1990;5(6):528-29.
202. Kappen TH, Vergouwe Y, van Wolfswinkel L, Kalkman CJ, Moons KG, van Klei WA. Impact of adding therapeutic recommendations to risk assessments

from a prediction model for postoperative nausea and vomiting. Br J Anaesth. 2015;114(2):252-60.
203. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. BMJ. 2004;328(7435):343-45.
204. Wallace E, Uijen MJM, Clyne B, Zarabzadeh A, Keogh C, Galvin R, et al. Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. BMJ Open. 2016;6(3):e009957.
205. Sanders SL, Rathbone J, Bell KJL, Glasziou PP, Doust JA. Systematic review of the effects of care provided with and without diagnostic clinical prediction rules. BMC Diagn Progn Res. 2017;1:13.
206. Kappen T, Peelen, LM,. Prediction models: the right tool for the right problem Curr Opin Anesthesiol. 2016;29(6):717-26.
207. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. BMJ. 2004;328(7441):702-08.
208. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015;350:h391.
209. Poldervaart JM, Reitsma JB, Koffijberg H, Backus BE, Six AJ, Doevendans PA, Hoes AW. The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial. BMC Cardiovasc Disord. 2013;13:77.
210. Hayes RJ, Moulton LH. Cluster randomised trials. Boca Raton, FL: CRC Press; 2017.
211. Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: extension to cluster randomised trials. BMJ 2004;328(7441):702-08.
212. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. Int J Epidemiol. 2015;44(3):1051-67.
213. Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. BMJ. 2017;358:j3064.
214. Schaafsma JD, van der Graaf Y, Rinkel GJ, Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. J Clin Epidemiol. 2009;62(12):1248-52.
215. Koffijberg H, van Zaane B, Moons KG. From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. BMC Med Res Methodol. 2013;13:12.
216. Siontis KC, Siontis GC, Contopoulos-Ioannidis DG, Ioannidis JP. Reply to letter by Ferrante di Ruffano et al.: patient outcomes in randomized comparisons of diagnostic tests are still the ultimate judge. J Clin Epidemiol. 2016;69:267-68.
217. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.
218. Reilly BM, Evans AT, Schaider JJ, Das K, Calvin JE, Moran LA, et al. Impact of a clinical decision rule on hospital triage of patients with suspected acute cardiac ischemia in the emergency department. JAMA. 2002;288(3):342-50.
219. Brehaut JC, Graham ID, Wood TJ, Taljaard M, Eagles D, Lott A, et al. Measuring acceptability of clinical decision rules: validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. Med Decis Making. 2010;30(3):398-408.

220. Cowley LE, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO, Kemp AM. Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: a qualitative study with child protection professionals. Child Abuse Negl. 2018;81:192-205.
221. Ballard DW, Rauchwerger AS, Reed ME, Vinson DR, Mark DG, Offerman SR, et al. Emergency physicians' knowledge and attitudes of clinical decision support in the electronic health record: a survey-based study. Acad Emerg Med. 2013;20(4):352-60.
222. Johnson EL, Hollen LI, Kemp AM, Maguire S. Exploring the acceptability of a clinical decision rule to identify paediatric burns due to child abuse or neglect. Emerg Med J. 2016;33(7):465-70.
223. Mullen S, Quinn-Scoggins HD, Nuttall D, Kemp AM. Qualitative analysis of clinician experience in utilising the BuRN Tool (Burns Risk assessment for Neglect or abuse Tool) in clinical practice. Burns. 2018;44(7):1759-66.
224. Haskins R, Osmotherly PG, Southgate E, Rivett DA. Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. Man Ther. 2014;19(2):142-51.
225. Kelly J, Sterling M, Rebbeck T, Bandong AN, Leaver A, Mackey M, Ritchie C. Health practitioners' perceptions of adopting clinical prediction rules in the management of musculoskeletal pain: a qualitative study in Australia. BMJ Open. 2017;7(8):e015916.
226. Atabaki SM, Hoyle JDJ, Schunk JE, Monroe DJ, Alpern ER, Quayle KS, et al. Comparison of prediction rules and clinician suspicion for identifying children with clinically important brain injuries after blunt head trauma. Acad Emerg Med. 2016;23(5):566-75.
227. Mahajan P, Kuppermann N, Tunik M, Yen K, Atabaki SM, Lee LK, et al. Comparison of clinician suspicion versus a clinical prediction rule in identifying children at risk for intra-abdominal injuries after blunt torso trauma. Acad Emerg Med. 2015;22(9):1034-41.
228. Reilly BM, Evans AT, Schaider JJ, Wang Y. Triage of patients with chest pain in the emergency department: a comparative study of physicians' decisions. Am J Med. 2002;112(2):95-103.
229. Broekhuizen BD, Sachs A, Janssen K, Geersing GJ, Moons K, Hoes A, Verheij T. Does a decision aid help physicians to detect chronic obstructive pulmonary disease? Br J Gen Pract. 2011;61(591):e674-e79.
230. Schriger DL, Newman DH. Medical decisionmaking: let's not forget the physician. Ann Emerg Med. 2012;59(3):219-20.
231. Finnerty N, Rodriguez R, Carpenter C, Sun B, Theyyunni N, Ohle R, et al. Clinical decision rules for diagnostic imaging in the emergency department: a research agenda. Acad Emerg Med. 2015;22(12):1406-16.
232. Sanders S, Doust J, Glasziou P. A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. PLoS One. 2015;10(6):e0128233.
233. Cowley LE, Farewell DM, Kemp AM. Potential impact of the validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: a clinical vignette study. Child Abuse Negl. 2018;86:184-96.
234. Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. BMJ. 2011;342:d1766.

235. Grimshaw J, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. Med Care. 2001;39(8 Suppl 2):II2-II45.
236. Stiell IG, Bennett C. Implementation of clinical decision rules in the emergency department. Acad Emerg Med. 2007;14(11):955-59.
237. Cameron C, Naylor CD. No impact from active dissemination of the Ottawa Ankle Rules: further evidence of the need for local implementation of practice guidelines. CMAJ. 1999;160(8):1165-68.
238. Davis DA, Taylor-Vaisey A. Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. CMAJ. 1997;157(4):408-16.
239. Katz MH. Integrating prediction rules into clinical work flow. JAMA Internal Medicine. 2013;173(17):1591-91.
240. Boutis K, Constantine E, Schuh S, Pecaric M, Stephens D, Narayanan UG. Pediatric emergency physician opinions on ankle radiograph clinical decision rules. Acad Emerg Med. 2010;17(7):709-17.
241. Pluddemann A, Wallace E, Bankhead C, Keogh C, Van der Windt D, Lasserson D, et al. Clinical prediction rules in practice: review of clinical guidelines and survey of GPs. Br J Gen Pract. 2014;64(621):e233-e42.
242. Kappen TH, van Loon K, Kappen MA, van Wolfswinkel L, Vergouwe Y, van Klei WA, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. J Clin Epidemiol. 2016;70:136-45.
243. Keogh C, Fahey T. Clinical prediction rules in primary care: what can be done to maximise their implementation? Clinical Evidence. 2010. http://clinicalevidence.bmj.com/downloads/05-10-10.pdf. Accessed 12 Jun 2018.
244. Runyon MS, Richman PB, Kline JA. Emergency medicine practitioner knowledge and use of decision rules for the evaluation of patients with suspected pulmonary embolism: variations by practice setting and training level. Acad Emerg Med. 2007;14(1):53-57.
245. Pearson SD, Goldman L, Garcia TB, Cook EF, Lee TH. Physician response to a prediction rule for the triage of emergency department patients with chest pain. J Gen Intern Med. 1994;9(5):241-47.
246. Brehaut JC, Stiell IG, Visentin L, Graham ID. Clinical decision rules "in the real world": how a widely disseminated rule is used in everyday practice. Acad Emerg Med. 2005;12(10):948-56.
247. Brehaut JC, Stiell IG, Graham ID. Will a new clinical decision rule be widely used? The case of the Canadian C-spine rule. Acad Emerg Med. 2006;13(4):413-20.
248. Graham ID, Stiell IG, Laupacis A, O'Connor AM, Wells GA. Emergency physicians' attitudes toward and use of clinical decision rules for radiography. Acad Emerg Med. 1998;5(2):134-40.
249. Eichler K, Zoller M, Tschudi P, Steurer J. Barriers to apply cardiovascular prediction rules in primary care: a postal survey. BMC Fam Pract. 2007;8:1.
250. Beutel BG, Trehan SK, Shalvoy RM, Mello MJ. The Ottawa knee rule: examining use in an academic emergency department. West J Emerg Med. 2012;13(4):366-72.
251. Sheehan B, Nigrovic LE, Dayan PS, Kuppermann N, Ballard DW, Alessandrini E, et al. Informing the design of clinical decision support services for evaluation

of children with minor blunt head trauma in the emergency department: a sociotechnical analysis. J Biomed Inform. 2013;46(5):905-13.
252. van der Steen JT, Albers G, Licht-Strunk E, Muller MT, Ribbe MW. A validated risk score to estimate mortality risk in patients with dementia and pneumonia: barriers to clinical impact. Int Psychogeriatr. 2011;23(1):31-43.
253. Sanders S. Clinical prediction rules for assisting diagnosis (doctoral thesis). Faculty of Heath Sciences & Medicine, Bond University, Australia. 2015.
254. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, Rubin HR. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA. 1999;282(15):1458-65.
255. McGinn T. Putting meaning into meaningful use: a roadmap to successful integration of evidence at the point of care. JMIR Med Inform. 2016;4(2):e16.

Appendix 4. Published article resulting from a collaboration with Australian colleagues

ADC Online First, published on April 5, 2018 as 10.1136/archdischild-2017-313748
Original article

Clinical prediction rules for abusive head trauma: a systematic review

Helena Pfeiffer,[1,2] Louise Crowe,[2] Alison Mary Kemp,[3] Laura Elizabeth Cowley,[3] Anne S Smith,[4] Franz E Babl,[1,2,5] On behalf of the Paediatric Research in Emergency Departments International Collaborative (PREDICT)

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/archdischild-2017-313748).

[1]Emergency Department, Royal Children's Hospital, Parkville, Victoria, Australia
[2]Emergency Research, Murdoch Children's Research Institute, Parkville, Victoria, Australia
[3]Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, UK
[4]Victorian Forensic Paediatric Medical Service, Royal Children's Hospital, Parkville, Victoria, Australia
[5]Department of Pediatrics, University of Melbourne, Melbourne, Victoria, Australia

Correspondence to
Dr Franz E Babl, Emergency Research, Murdoch Children's Research Institute, Parkville, VIC 3052, Australia; franz.babl@rch.org.au

Received 15 July 2017
Revised 14 February 2018
Accepted 20 February 2018

ABSTRACT

Objective Misdiagnosis of abusive head trauma (AHT) has serious consequences for children and families. This systematic review identifies and compares clinical prediction rules (CPredRs) assisting clinicians in assessing suspected AHT.

Design We searched MEDLINE, Embase, PubMed and Cochrane databases (January 1996 to August 2016). Externally validated CPredRs focusing on the detection of AHT in the clinical setting were included.

Results Of 110 potential articles identified, three studies met the inclusion criteria: the Pediatric Brain Injury Research Network (PediBIRN) 4-Variable AHT CPredR, the Predicting Abusive Head Trauma (PredAHT) tool and the Pittsburgh Infant Brain Injury Score (PIBIS). The CPredRs were designed for different populations and purposes: PediBIRN: intensive care unit admissions (<3 years) with head injury, to inform early decisions to launch or forego an evaluation for abuse (sensitivity 0.96); PredAHT: hospital admissions (<3 years) with intracranial injury, to assist clinicians in discussions with child abuse specialists (sensitivity 0.72); and PIBIS: well-appearing children (<1 year) in the emergency department with no history of trauma, temperature <38.3°C, and ≥1 symptom associated with high risk of AHT, to determine the need for a head CT scan (sensitivity 0.93). There was little overlap between the predictive variables.

Conclusion Three CPredRs for AHT were relevant at different stages in the diagnostic process. None of the CPredRs aimed to diagnose AHT but to act as aids/prompts to clinicians to seek further clinical, social or forensic information. None were widely validated in multiple settings. To assess safety and effectiveness in clinical practice, impact analyses are required and recommended.

What is already known on this topic?

- Abusive head trauma (AHT) has high morbidity and mortality; if AHT is missed, patients are at increased risk of further injury and death.
- Clinical prediction rules (CPredRs) may assist in deciding which head injured children might have sustained injury as a result of AHT.

What this study adds?

- Three recently validated CPredRs for AHT investigate very different populations, focus on different stages of the diagnostic process and use differing predictive and outcome variables.
- The aim of the CPredRs was not to diagnose AHT but to act as aids or prompts to seek further information, investigation and assessment.

INTRODUCTION

Abusive head trauma (AHT) is a leading cause of traumatic death in children less than 1 year of age and the most common cause of fatal child abuse.[1] Children with AHT have an estimated mortality rate of 26%[2] and long-term disability ranging from 44%[3] to 92%[4] among survivors.[5 6] Patients are at increased risk of further injury and death if AHT is missed.[7 8] However, deciding which children should undergo a full evaluation for AHT is difficult, as histories provided by the caregiver might be absent or fabricated, and the clinical findings may be similar to those seen in accidental trauma.[9] Clinicians might hesitate to raise suspicion of AHT, as a wrongful accusation means emotional strain for the families, endangers patient–doctor relationships and leads to unnecessary investigation-related costs and risks for the child.[10]

Clinical prediction rules (CPredRs) are evidence-based tools, which incorporate three or more variables from clinical findings including history, physical examination and results of investigations, to predict aetiology or outcome.[11] They are especially important for conditions where decision-making is difficult, clinical stakes are high and clinical experience and intuition are limited.[12] There are three main phases in the development of CPredRs: derivation, external validation and impact analysis.[13] Each requires a different and rigorous methodological approach. In the initial phase, predicted probabilities are derived from the statistical analysis of patients with known outcomes, typically using multivariable regression techniques or classification and regression tree analysis.[13] Predictor variables should be clinically sensible and clearly defined,[14] and the data set used to derive the rule should be representative of the target population. Reilly and Evans[11] distinguish between assistive prediction rules that simply provide clinicians with predicted probabilities without recommending a specific clinical course of action, and directive decision rules that explicitly suggest additional diagnostic tests or treatment in line with the obtained score.[11]

Check for updates

To cite: Pfeiffer H, Crowe L, Kemp AM, *et al*. *Arch Dis Child* Epub ahead of print: [please include Day Month Year]. doi:10.1136/archdischild-2017-313748

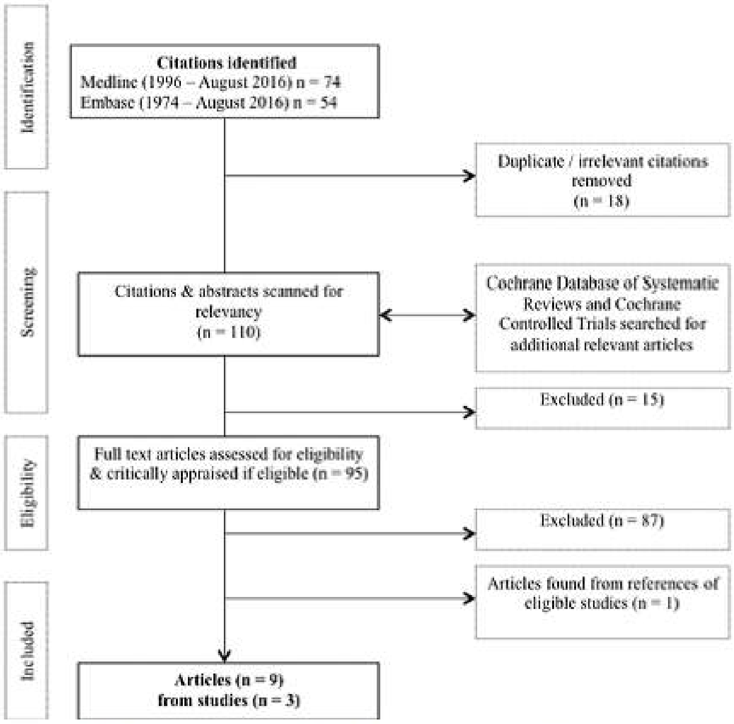


Figure 1 Flow diagram of literature search based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

CPredRs must be validated with a data set external to the one in which it was derived, preferably in multiple settings, and tested in clinical practice to determine their impact on patient care.[15]

There have been few systematic reviews in this field that explore the quality and effectiveness of CPredRs in child abuse. While Louwers *et al* compared several screening techniques for child abuse in emergency departments (ED)[16] we set out to find and critically appraise CPredRs that aim to detect AHT across various medical settings and compare them in terms of their quality and performance.

METHODS

We conducted a systematic review to identify CPredRs for AHT and to compare them in terms of derivation, population, definition of AHT, variables used, external validation and performance. For this purpose, we followed a protocol based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (figure 1; for PRISMA checklist see online supplementary table 1).[17] Our review was registered with the PROSPERO international prospective register of systematic reviews (http://www.crd.york.ac.uk/; record number CRD42017058141).

Inclusion/exclusion criteria

We included externally validated CPredRs with a focus on the detection of AHT in children from 0 to 18 years of age in the clinical setting. Papers that reported associations between one or two variables or markers and AHT were excluded, as well as clinical assessment tools trying to detect child abuse in general.

Search strategy

We searched the electronic databases MEDLINE, the Cochrane Database of Systematic Reviews, Cochrane Controlled Trials (1996 to August 2016) and Embase (1974 to August 2016) using the Ovid and PubMed platforms. No limitations were applied for languages. The search strategy including search terms is presented in the online supplementary appendix.

Study selection

Duplicates were removed and relevant abstracts were reviewed. When eligibility could not be determined from the abstract, a full-text review was performed. When there were uncertainties about inclusion, other lead authors were consulted (LC, FEB).

Assessment of quality

Maguire *et al*[18] proposed 17 quality items for the development and validation of CPredRs for children in their systematic review. We chose this approach to compare the methodological quality of the CPredRs, calculating a score (table 2) according to the number of quality standards achieved.

Data analysis

It became apparent early on that the included CPredRs were heterogeneous and a meta-analysis was unlikely to be possible, therefore a descriptive analysis is provided.

RESULTS

The search in Medline resulted in 74 articles and in Embase 54 articles were found. The search in PubMed and the Cochrane Database of Systematic Reviews and Cochrane Controlled Trials did not add any additional relevant studies. After removing duplicates, reviewing abstracts and excluding irrelevant articles, three recently published, externally validated CPredRs for AHT met the inclusion criteria: the Pediatric Brain Injury Research Network (PediBIRN) 4-Variable AHT CPredR,[19] the Predicting Abusive Head Trauma (PredAHT) tool[20] and the Pittsburgh Infant Brain Injury Score (PIBIS) for AHT[21] (see figure 1 and table 1).

Pediatric Brain Injury Research Network clinical prediction rule

The PediBIRN CPredR[19] was designed to 'inform [pediatric intensivists'] early decisions to launch (or forego) an evaluation for abuse.'[10] It was derived in a prospective study including 209 children published in 2013 (95 AHT cases)[10] and validated in a further prospective study comprising 291 children (125 AHT cases), published in 2014.[19] The validation study took place in 14 hospitals within the USA, of which 10 sites had been part of the derivation study. The population of interest was that of acutely head-injured children aged <3 years admitted to the paediatric intensive care unit (PICU). Following a bivariate analysis of 45 potential factors for discrimination and reliability the authors applied a classification tree to their data using binary recursive partitioning to derive a 4-variable CPredR with maximum sensitivity to determine the risk of AHT. The four variables were respiratory compromise, bruising involving the ears, neck and torso, bilateral/interhemispheric subdural haemorrhages and skull fractures (other than an isolated, unilateral, non-diastatic, linear, parietal skull fracture). If one or more of the variables were present, the child should be thoroughly evaluated for abuse. In the validation study, the CPredR achieved a sensitivity of 0.96 and a specificity of 0.46 to detect AHT cases.

Predicting Abusive Head Trauma tool

The PredAHT CPredR provides an estimated probability of AHT to assist clinicians in discussions with child abuse specialists.[20] It was derived from a pooled analysis of individual patient data from six previously published studies, which included prospectively collected data on 133 children (58 AHT cases) and retrospective data on 920 children (290 AHT cases) with head injury.[22] PredAHT identified the positive predictive values and ORs for AHT given any combination of six possible variables using multilevel logistic regression. The validation study comprised a retrospective data set of 60 children (23 AHT cases) in Cardiff, UK, and prospective data on 138 children (42 AHT cases) from Lille, France.[23] The Welsh data set included children <3 years of age with intracranial injuries confirmed on neuroimaging, who were admitted to hospital, whereas the French cases comprised patients <2 years old with craniocerebral traumatic lesions diagnosed on CT and referred alive to the ED, PICU or neurosurgical department. Missing data were accounted for using multiple imputation by chained equations. PredAHT gave probabilities of AHT that ranged from 4% if none of the variables were present to close to 100% when all six variables were present.[22] Following validation, the probability of AHT was always greater than 81.5% if any three or more of the predictive variables—head/neck bruising, seizures, apnoea, rib fractures, long-bone fractures or retinal haemorrhages—were present. In the validation study the CPredR performed with a sensitivity of 0.72 and a specificity of 0.86 in detecting AHT using a cut-off probability of AHT of 50%.

Pittsburgh Infant Brain Injury Score

The PIBIS for AHT[21] assists in determining which high-risk infants in the ED should undergo head CT to rule out abnormalities including AHT. It was derived based on retrospective data on 187 children (150 without brain injury and 37 with mild AHT), which were not published, and validated using logistic regression in a prospective study carried out in the USA with 1040 infants. Missing data were handled with listwise deletion. The included sample was of well-appearing children between 30 and 364 days of age with a temperature <38.3°C, no history of trauma and at least one symptom associated with high risk of AHT (apparent life-threatening event/apnoea, vomiting without diarrhoea, seizures/seizure-like activity, soft tissue swelling of the scalp, bruising or other non-specific neurological symptoms such as lethargy, fussiness or poor feeding). If children received a score of 2 points or more when adding: abnormality such as bruises observed on dermatologic examination (2 points), age equal to or greater than 3 months (1 point), head circumference >85th percentile (1 point) and haemoglobin <11.3 g/dL (1 point), further neuroimaging should be performed. In the validation study, the CPredR performed with a sensitivity of 0.93 and a specificity of 0.53 to detect abnormal neuroimaging. It is important to stress that the outcome case definition did not exclusively comprise AHT, but included other clinically significant traumatic and non-traumatic abnormalities.

Comparing the rules

All three CPredRs defined their outcome and predictive variables clearly, reported their results adequately and used 95% CIs on rule properties. Using a standardised approach[18] the PediBIRN CPredR[19] received the highest score for methodological quality (table 2). It was the only CPredR that described an independent blinded assessment of predictive and outcome variables and an evaluation of inter-rater reliability in the derivation study, comparing the assessment of blinded duplicate data on 20% of included patients by different investigators.[10] The PIBIS study conducted the only follow-up of cases (6 months after enrolment or up to 1 year of age) to identify further abnormal neuroimaging and assess the progress of symptoms at presentation.[10] This approach was an attempt to verify the true negatives as only 61% of controls had neuroimaging. The PediBIRN and the PIBIS CPredRs both proposed a clear course of action (PediBIRN—thorough evaluations for abuse, PIBIS—CT scan), whereas PredAHT provided a probability of AHT in order to 'assist clinicians in their discussions with child abuse specialists, in addition to facilitating discussions between child abuse specialists and social welfare, law enforcement, or other professionals involved in the child protection process.'[20]

DISCUSSION

We identified three validated CPredRs that met inclusion criteria. These prediction rules are aimed at very different populations and different time points within the clinical assessment (figure 2): PIBIS[21] is targeted at a specific population of well-appearing infants in the ED who might benefit from a head CT

Table 1 Comparison of three clinical prediction rules for abusive head trauma

Name	PediBIRN4-Variable AHT CPredR	PredAHT tool	PIBS for Abusive Head Trauma
Country	USA	UK	USA
Derivation paper	Hymel *et al*, Derivation of a clinical prediction rule for pediatric abusive head trauma, *Pedr Crit Care Med*, 2013[15]	Maguire *et al*, Estimating the probability of NAHI, *Pediatrics*, 2011[22]	Unpublished data
	Prospective: n=209	Prospective: n=133 Retrospective: n=920	Retrospective: n=187
Validation paper	Hymel *et al*, Validation for a clinical prediction rule for pediatric abusive head trauma, *Pediatrics*, 2014[19]	Cowley *et al*, Validation of a prediction tool for abusive head trauma, *Pediatrics*, 2015[23]	Berger *et al*, Validation of the Pittsburgh Infant Brain Injury Score for abusive head trauma, *Pediatrics*, 2016[21]
	Prospective: n=291	Prospective: n=138 Retrospective: n=60	Prospective: n=1040
CPredR	Every acutely head-injured infant/child meeting the inclusion criteria and presenting with ≥1 of these four predictor variables should be thoroughly evaluated for abuse: ▶ Any clinically significant respiratory compromise (infrequent/laboured respirations, apnoea or any need for intubation or assisted ventilation) at the scene of injury, during transport, in the ED or before admission ▶ Any bruising involving the child's ears, neck and torso (including chest, abdomen, genitourinary region, back or buttocks) ▶ Any subdural haemorrhages or fluid collections that are bilateral or involved the interhemispheric space ▶ Any skull fractures other than an isolated, unilateral, non-diastatic, linear, parietal skull fracture	Estimated probability of AHT varies from 4% when none of the features are present to close to 100% when all six features are present and >81.5% (63.3%–91.8%) when ≥3 of these six features are present: ▶ Head or neck bruising ▶ Seizures ▶ Apnoea (documented in initial history or during inpatient stay) ▶ Rib fracture (documented after appropriate radiological imaging) ▶ Long-bone fracture ▶ Retinal haemorrhage (documented after indirect ophthalmological examination by a paediatric ophthalmologist)	Children with a score of ≥2 should undergo neuroimaging to check for abnormal findings: ▶ Abnormality on dermatological examination (2 points) ▶ Age ≥3.0 months (1 point) ▶ Head circumference >85th percentile (1 point) ▶ Haemoglobin <11.2 g/dL (1 point)
Objective	Detection of AHT among acutely head-injured children admitted to PICU	Prediction of the likelihood of AHT in head-injured children	Detection of abnormal neuroimaging in well-appearing children with non-specific symptoms
Inclusion	▶ Children <3 years of age ▶ Admission to PICU ▶ Symptomatic, acute, closed, traumatic, cranial or intracranial injuries confirmed by CT or MRI	Data set 1 (Cardiff, UK): ▶ Children <3 years of age ▶ Hospital admission ▶ ICI (combination of extra-axial haemorrhage, diffuse or focal parenchymal injury, cerebral oedema, cerebral contusion, hypoxic ischaemic injury or diffuse axonal injury) confirmed on neuroimaging Data set 2 (Lille, France): ▶ Children <2 years of age ▶ Craniocerebral traumatic lesions diagnosed based on at least one CT[23] ▶ Referred alive to the neurosurgical department, the PICU or the ED	▶ 30–364 days of age ▶ Well appearing ▶ Temperature <38.3°C ▶ No history of trauma – Seeking medical evaluation for one of the following symptoms: ALTE/apnoea – Vomiting without diarrhoea – Seizures or seizure-like activity – Soft tissue swelling of the scalp – Bruising – Other non-specific neurological symptom not described above, such as lethargy, fussiness or poor feeding
Exclusion	▶ Children ≥3 years of age ▶ HI resulting from a collision involving a motor vehicle ▶ Initial neuroimaging revealed clear evidence of pre-existing brain malformation, disease, infection or hypoxia-ischaemia	▶ Children ≥3 years of age (data set 2: ≥2 years of age) ▶ Normal neuroimaging ▶ Underlying structural abnormality or pre-existing disease (hydrocephalus, cystic lesion/tumour, metabolic cause, malformation, abnormal brain development) ▶ Injuries caused by neglect ▶ Birth injuries	▶ Previous abnormal CT scan of the head

Continued

Table 1 Continued

Name	PediBIRN 4-Variable AHT CPredR	PredAHT tool	PIBS for Abusive Head Trauma
Definition of AHT	► The primary caregiver (PC) admitted abusive acts. ► Abusive acts by the PC were witnessed by an unbiased, independent observer. ► The PC specifically denied that the preambulatory child in his/her care had experienced any head trauma. ► The PC provided an account of the child's HI event that was clearly historically inconsistent with repetition over time. ► The PC provided an account of the child's HI event that was clearly developmentally inconsistent with the child's known (or expected) gross motor skills. ► Further workup confirmed the presence of two or more categories of extracranial injuries considered moderately or highly suspicious for abuse – Classic metaphyseal lesion fracture or epiphyseal separation – Rib fracture, fracture of the scapula or sternum – Fractures of the digits – Vertebral body fractures – Dislocation/fracture of spinous process – Skin bruising/abrasion/laceration in two or more distinct locations other than knee, shins or elbows – Patterned bruising or dry contact burns – Scalding burns with uniform depth, clear lines of demarcation and paucity of splash marks – Confirmed intra-abdominal injuries – Retinoschisis confirmed by an ophthalmologist – Retinal haemorrhages described by an ophthalmologist as dense, extensive, covering a large surface area and/or extending to the ora serrata	Confirmed cases on AHT (ranked 1 or 2 for abuse) ► Rank 1: – Abuse confirmed at case conference or civil, family or criminal court proceedings – Admitted by perpetrator – Independently witnessed ► Rank 2: – Abuse confirmed by stated criteria including multidisciplinary assessment	Brain injury due to definite/probable, but not possible, abuse as assessed by the hospital-based child protection team at each enrolled site (cases=abnormal neuroimaging)
Definition of nAHT	► All remaining patients	► Witnessed accidental mechanisms ► Confirmed organic causes ► Abuse excluded after child protection investigations	► NS
Validation study	n=291	n=198	n=862
Sensitivity*	0.96 (0.90–0.99)	0.72 (0.60–0.82)	0.93 (0.89–96.0)
Specificity	0.43 (0.35–0.50)	0.86 (0.79–0.91)	0.53 (0.49–0.57)
Prevalence	0.43 (0.37–0.49)	0.33 (0.27–0.40)	0.26
PPV	0.55 (0.48–0.62)	0.71 (0.59–0.81)	0.39 (0.35–0.44)
NPV	0.93 (0.85–0.98)	0.86 (0.80–0.91)	0.96 (0.94–0.98)
LR+	1.67 (1.46–1.91)	5.06 (3.25–7.88)	1.98
LR−	0.09 (0.04–0.23)	0.32 (0.22–0.48)	0.13
Area under the curve	0.78	0.88 (0.82–0.93)	0.82
	*Accuracy of detecting AHT cases among children with HI admitted to PICU	*Accuracy of detecting AHT cases among admitted children with HI	*Accuracy of detecting cases with abnormal neuroimaging in well-appearing children with at least one non-specific symptom, which is common in AHT

AHT, abusive head trauma; ALTE, apparent life-threatening event; CPredR, clinical prediction rule; ED, emergency department; HI, head injury; ICI, intracranial injury; LR+, positive likelihood ratio; LR−, negative likelihood ratio.; nAHT, accidental head trauma (non-inflicted); NPV, negative predictive value; NS, not specified; PC, primary caregiver; PIBIS, Pittsburgh Infant Brain Injury Score; PICU, paediatric intensive care unit; PPV, positive predictive value; PredAHT, Predicting Abusive Head Trauma.

Table 2 Assessment of methodological quality as proposed by Maguire *et al*[18]

Quality item	PediBIRN[19]	PredAHT[20]	PIBIS[21]
Prospective validation	Yes	Only DS 2	Yes
Study site well described	Yes	Yes	Yes
Population well described	Yes	Yes	Yes
Rule applied to all patients at risk	>90%	NS	No
Predictive variables			
Clear definition	Yes	Yes	Yes
Blind assessment	Yes	NS	NS
Reproducible	Yes	NS	NS
Outcome variable			
Definition	Yes	Yes	Yes
Blind assessment	Yes	NS	NS
Adequate follow-up	NS	NS	Yes
Sensibility			
Clinically sensible	Yes	Yes	Yes
Easy to use	Yes	Yes	Yes
Course of action	Yes	No	Yes
Statistical analysis			
Mathematical technique reported	Yes	Yes	Yes
Adequate calculated power reported	No	No	No
Adequate reporting of results	Yes	Yes	Yes
95% CIs reported on rule properties	Yes	Yes	Yes
Score	15	9	12

Present=score of 1; not specified/no=score of 0.
AHT, abusive head trauma; DS, data set; NS, not specified; PediBIRN, Pediatric Brain Injury Research Network clinical prediction rule; PIBIS, Pittsburgh Infant Brain Injury Score; PredAHT, Predicting Abusive Head Trauma tool.

scan. PredAHT[20] applies to children <3 years of age admitted to hospital with intracranial injury, where children have been examined and may have had some ophthalmological and radiological investigations, and PediBIRN[19] applies to a narrower population of <3-year-olds admitted to PICU with a cranial or intracranial injury excluding head trauma resulting from motor vehicle collisions. It is notable that PredAHT[20] does not apply to children with cranial injury only, as defined within the Centers for Disease Control and Prevention definition of AHT.

Figure 2 Focus of the clinical prediction rules: the pyramid represents the time point and population where each clinical prediction rule (CPredR) is relevant. ED, emergency department; PediBIRN, Pediatric Brain Injury Research Network; PIBIS, Pittsburgh Infant Brain Injury Score; PICU, paediatric intensive care unit; PredAHT, Predicting Abusive Head Trauma.

Table 3 Variables used in the clinical prediction rules

Variable	Hymel	Cowley	Berger	Availability of item
Abnormality on dermatological examination/bruising	X	X	X	Physical examination
Respiratory compromise/apnoea	X	X		Physical examination/history
Subdural haemorrhages (bilateral, interhemispheric)	X			MRI/CT
Skull fractures (other than isolated unilateral, non-diastatic, linear, parietal skull fracture)	X			MRI/CT/skull X-ray
Rib fractures		X		Chest X-ray
Long-bone fracture		X		Long-bone X-ray
Retinal haemorrhage		X		Ophthalmological funduscopy
Seizures		X		Physical examination/history
Age ≥3.0 months			X	History
Head circumference >85th percentile			X	Physical examination
Haemoglobin <11.2 g/dL			X	FBE

FBE, full blood examination.

Just as the populations are different, so are the predictor variables prima facie. As apparent in table 3, the only overlap among the items used by the different CPredRs are cutaneous injuries and respiratory compromise/apnoea. However, some of the PIBIS high-risk AHT symptoms used for the inclusion of patients (apnoea, seizures, bruising) appeared among the predictor variables of PediBIRN and PredAHT. Beyond that, both of these CPredRs had tested some of the same predictive variables in their derivation studies (eg, PediBIRN: seizures and PredAHT: skull fractures), yet they had not significantly improved their CPredR's performance.[10 22]

In terms of published accuracy of the rules, the PediBIRN CPredR[19] performed best with a sensitivity of 0.96, closely followed by PIBIS[21] with 0.93, whereas the PredAHT CPredR detected 0.72 of AHT cases applying a 50% cut-off; however, the PredAHT tool provides a sliding scale of probability from 4% to nearly 100% depending on the presence or absence of each of the six features.[20] In addition, the sensitivity and specificity of PredAHT depends on the probability cut-off applied. The pretest prevalence in the PediBIRN population[19] was 0.43, which raises the question whether all children in this high-risk group should be screened. In the PredAHT and PIBIS studies the pretest probabilities were lower at 0.33 and 0.26, respectively.[20 21]

The PIBIS[21] variables are all available from physical examination, simple blood test and history, whereas neuroimaging is required for PediBIRN,[19] and further investigations such as ophthalmological examination and rib and long-bone X-rays are required for PredAHT[20] (table 3). These tests would usually only be performed if physical abuse or serious trauma had already been suspected placing PredAHT as a potentially useful tool for assessing the significance of the results of these investigations at a specific stage in the diagnostic process.

Due to the lack of gold standard diagnostic criteria for AHT, different approaches were chosen to minimise circular reasoning where AHT may have been decided based on the presence of the predictor variables within the CPredRs. The PediBIRN CPredR aimed to avoid definitional criteria that involved intracranial injuries, injury severity, any of the predictor variables and child protection team assessment.[19] When challenged on the

 Pfeiffer H, *et al*. *Arch Dis Child* 2018;0:1–8. doi:10.1136/archdischild-2017-313748

issue that bruising was included in their definitional criteria and as a predictor variable[24] the authors stated that of 73 patients with bruising, 61 met other definitional criteria based on the history and the 12 remaining patients were subsequently diagnosed with definite/probable AHT by the treating physicians[25] The PredAHT group applied the outcome of the child protection process including only cases where multidisciplinary or court proceedings had confirmed AHT.[20] Arguably this decision will include a consideration of clinical features, as is the case in any clinical diagnosis. However, within the child protection social care and legal process, there are multiple additional forensic, clinical, social and historical factors that are included in decisions about the balance of probability of abuse and future risks for the child. Similar reasoning applies to PIBIS where the child protection team assessment decision was also used.[21] Regarding PIBIS, the true negatives could not be identified as only 61% of controls had neuroimaging. Case follow-up was undertaken for 6–12 months to determine if any neurological imaging was performed at a later stage and cases potentially missed.

None of the CPredRs have yet been widely validated in multiple settings or undergone an impact analysis to determine their safety and effectiveness in clinical practice. Hymel *et al* have undertaken a theoretical impact analysis of PediBIRN in the combined population of the derivation and validation study.[26] Of note, because the PIBIS rule as originally derived was updated in the validation study, ideally this CPredR should be validated in another external data set before it can be applied to new patients.[27]

As the three CPredRs apply at different time points in the diagnostic process, in different populations and with a different degree of investigations completed, this explains and allows for differences in sensitivity and specificity. At the outset, it is paramount to ensure that cases are not missed and undergo sufficient investigation, therefore high sensitivity is the focus, with specificity of lower importance. When more investigations have been undertaken, a higher specificity would be desirable to ensure that a diagnosis of AHT is not made incorrectly. This suggests that the three CPredRs might complement each other in clinical practice. For instance, if a well-appearing infant between 30 and 364 days of age met the inclusion criteria and predictive values of PIBIS underwent neuroimaging and had an intracranial injury, further investigations should include skeletal survey and ophthalmological exam, providing more of the items required for PredAHT. In the critically ill children admitted to PICU, the PediBIRN rule would apply and would be useful to inform decision-making at a time when the child may be too sick to undergo a skeletal survey.

Limitations

The number of included studies was small and a meta-analysis was not possible. A further CPredR by Wells *et al* on the radiological differentiation of intentional and non-intentional intracranial haemorrhages was excluded, as it had not been externally validated.[28]

CONCLUSION

The three CPredRs for AHT focus on different populations with different inclusion criteria. They use different predictive variables available at different stages in the diagnostic process, and different outcome variables. PediBIRN aims to rule out AHT in the PICU. PredAHT calculates the probability of AHT for hospitalised children. PIBIS aims to detect abnormal neuroimaging in the ED. None of the CPredRs aimed to diagnose AHT but to act as aids or prompts to clinicians to seek further clinical, social or forensic information and move towards a multidisciplinary child protection assessment should more information in support of AHT arise. Wider validation in multiple settings is recommended for each CPredR, in addition to impact analyses to assess their safety and effectiveness in clinical practice.

Acknowledgements We thank Ms Poh Chua, librarian, Library Service, Royal Children's Hospital, Melbourne, Australia, for her assistance in the literature search.

Contributors HP contributed to the design of the study, conducted the systematic review, and drafted and revised the article. LC, AMK and LEC contributed to the design of the study, reviewed the search, made substantial contributions to the interpretation and discussion of findings and critically revised the manuscript for important intellectual content. JASS contributed to the design of the study and critically revised the manuscript for important intellectual content. FEB had the initial study idea, contributed to the design of the study and critically revised the manuscript for important intellectual content. FEB takes responsibility for the paper as a whole.

Funding The study was in part funded by grants from the National Health and Medical Research Council (Centre of Research Excellence for Paediatric Emergency Medicine GNT1058560), Canberra, Australia; the Murdoch Children's Research Institute, Melbourne, Australia; and supported by the Victorian Government's Infrastructure Support Program, Melbourne, Australia. FEB's time was in part funded by a grant from the Royal Children's Hospital Foundation and a Melbourne Campus Clinician Scientist Fellowship, Melbourne, Australia and an NHMRC Practitioner Fellowship.

Competing interests AMK and LEC were involved with the development of one of the clinical prediction rules described in this paper.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The search strategy as well as the PRISMA checklists are available as supplementary files.

REFERENCES

1 Duhaime A-C, Christian CW, Rorke LB, *et al*. Nonaccidental head injury in infants — the "shaken-baby syndrome". *N Engl J Med Overseas Ed* 1998;338:1822–9.
2 Duhaime AC, Christian C, Moss E, *et al*. Long-term outcome in infants with the shaking-impact syndrome. *Pediatr Neurosurg* 1996;24:292–8.
3 Karandikar S, Coles L, Jayawant S, *et al*. The neurodevelopmental outcome in infants who have sustained a subdural haemorrhage from non-accidental head injury. *Child Abuse Review* 2004;13:178–87.
4 Haviland J, Russell RI. Outcome after severe non-accidental head injury. *Arch Dis Child* 1997;77:504–7.
5. Jayawant S, Parr J. Outcome following subdural haemorrhages in infancy. *Arch Dis Child* 2007;92:343–7.
6 Makaroff KL, Putnam FW. Outcomes of infants and children with inflicted traumatic brain injury. *Dev Med Child Neurol* 2003;45:497–502.
7 Jenny C, Hymel KP, Ritzen A, *et al*. Analysis of missed cases of abusive head trauma. *JAMA* 1999;281:621–6.
8 Letson MM, Cooper JN, Deans KJ, *et al*. Prior opportunities to identify abuse in children with abusive head trauma. *Child Abuse Negl* 2016;60:36–45.
9 Hettler J, Greenes DS. Can the initial history predict whether a child with a head injury has been abused? *Pediatrics* 2003;111:602–7.
10 Hymel KP, Willson DF, Boos SC, *et al*. Derivation of a clinical prediction rule for pediatric abusive head trauma. *Pediatr Crit Care Med* 2013;14:210–20.
11 Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144:201–9.
12 McGinn TG, Guyatt GH, Wyer PC, *et al*. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-based medicine working group. *JAMA* 2000;284:79–84.
13 Wasson JH, Sox HC, Neff RK, *et al*. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;313:793–9.
14 Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* 1999;33:437–47.
15 Moons KG, Kengne AP, Grobbee DE, *et al*. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691–8.
16 Louwers EC, Affourtit MJ, Moll HA, *et al*. Screening for child abuse at emergency departments: a systematic review. *Arch Dis Child* 2010;95:214–8.

17 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
18 Maguire JL, Kulik DM, Laupacis A, *et al*. Clinical prediction rules for children: a systematic review. *Pediatrics* 2011;128:e666–77.
19 Hymel KP, Armijo-Garcia V, Foster R, *et al*. Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics* 2014;134:e1537–e1544.
20 Cowley LE, Morris CB, Maguire SA, *et al*. Validation of a prediction tool for abusive head trauma. *Pediatrics* 2015;136:290–8.
21 Berger RP, Fromkin J, Herman B, *et al*. Validation of the Pittsburgh Infant Brain Injury Score for Abusive Head Trauma. *Pediatrics* 2016;138:e20153756.
22 Maguire SA, Kemp AM, Lumb RC, *et al*. Estimating the probability of abusive head trauma: a pooled analysis. *Pediatrics* 2011;128:e550–64.
23 Vinchon M, Defoort-Dhellemmes S, Desurmont M, *et al*. Accidental and nonaccidental head injuries in infants: a prospective study. *J Neurosurg* 2005;102:380–4.
24 Maguire S, Cowley L, Farewell D, *et al*. Theoretical re-analysis of two previously published datasets. *J Pediatr* 2016;171:321.
25 Hymel KP, Herman BE, Narang SK, *et al*. Reply. *J Pediatr* 2016;171:321–2.
26 Hymel KP, Herman BE, Narang SK, *et al*. Potential impact of a validated screening tool for pediatric abusive head trauma. *J Pediatr* 2015;167:1375–81.
27 Moons KG, Altman DG, Reitsma JB, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
28 Wells RG, Vetter C, Laud P. Intracranial hemorrhage in children younger than 3 years: prediction of intent. *Arch Pediatr Adolesc Med* 2002;156:252–7.

Appendix 5. Manuscript submitted to *Archives of Disease in Childhood.*

Title: Validation of the PredAHT Prediction Tool for Abusive Head Trauma

Helena Pfeiffer[a,b], Laura Elizabeth Cowley[c], MSc, PhD, Alison Mary Kemp[c], MRCP, MBBS, Stuart R Dalziel[d,e], PhD, Anne Smith[b,f], MBBS, John A. Cheek[a,b,g], Meredith L Borland[h,i], MBBS, Sharon O'Brien[h], Megan Bonish[d], Jocelyn Neutze[j], MBChB, Ed Oakley[a,b,k], MBBS, Louise Crowe[b], PhD, Stephen J.C. Hearps[b], PGDipBiostat, Mark D. Lyttle[b,l,m], MBChB, Silvia Bressan[b,n], MD, PhD, Franz E Babl[a,b,k], MD on behalf of the Paediatric Research in Emergency Departments International Collaborative (PREDICT)

Affiliations:

[a]Royal Children's Hospital, Melbourne, Australia; [b]Murdoch Children's Research Institute, Melbourne; [c]Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, Wales, United Kingdom; [d]Starship Children's Health, Auckland, New Zealand; [e]Departments of Surgery and Paediatrics: Child and Youth Health, University of Auckland, Auckland, New Zealand; [f]Victorian Forensic Paediatric Medical Service, Royal Children's Hospital, Melbourne, Australia; [g]Monash Medical Centre, Melbourne, Australia; [h]Princess Margaret Hospital for Children, Perth, Australia; [i]Divisions of Paediatrics and Emergency Medicine, School of Medicine, University of Western Australia, Crawley, Australia; [j]Kidzfirst Middlemore Hospital, Auckland, New Zealand; [k]Department of Paediatrics, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia; [l]Bristol Royal Hospital for Children, Bristol, United Kingdom; [m]Academic Department of Emergency Care, University of the West of England, Bristol, United Kingdom; and [n]Department of Women's and Children's Health, University of Padova, Padova, Italy

Address correspondence to: Franz E Babl, MD MPH, Emergency Research, Murdoch Children's Research Institute, 50 Flemington Road, Parkville, Victoria, 3052, Australia, franz.babl@mcri.edu.au, +613993 66748

Table of Contents Summary: This study externally validates a tool to predict the probability of abusive head trauma in young children.

Word count: 2600

Abstract

Objective:

The validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool calculates the probability of abusive head trauma (AHT) in children <3 years of age with intracranial injuries (ICI) identified on neuroimaging, based on combinations of 6 clinical features: head/neck bruising, seizures, apnoea, rib fracture, long-bone fracture, and retinal haemorrhages.. PredAHT version 2 enables a probability calculation when information regarding any of the six features is absent. We aimed to externally validate PredAHT-2 in an Australian/New Zealand population.

Design:

A secondary analysis of a prospective multicentre study of paediatric head injuries.

Setting:

We extracted data on patients with possible AHT at 5 tertiary paediatric centres.

Patients:

We included all children <3 years of age admitted to hospital with ICI identified on neuroimaging. We assigned cases as positive for AHT, negative for AHT or having indeterminate outcome following multidisciplinary review.

Main outcome measures:

The estimated probability of AHT for each case was calculated using PredAHT-2, blinded to outcome. Tool performance measures were calculated with 95% confidence intervals.

Results:

Of 87 cases with ICI, 27 (31%) were positive for AHT; 45 (52%) were negative for AHT, and 15 (17%) had indeterminate outcome. Using a probability cut-off of 50%, excluding indeterminate cases, PredAHT-2 had a sensitivity of 74% (54%–89%) and a specificity of 87% (73%–95%) for AHT.

Conclusion:

PredAHT-2 demonstrated reasonably high point sensitivity and specificity when externally validated in an Australian/New Zealand population. Performance was similar to that in the original validation study.

INTRODUCTION

Abusive head trauma (AHT) continues to be a major cause of traumatic deaths and long term morbidity in infants due to child abuse.(1-3) Ascribing AHT as the cause of an intracranial injury (ICI) is challenging for clinicians as the differential diagnosis may include abuse, accidental trauma, or other childhood diseases.(4) Perpetrators of AHT may deny abuse or offer alternative explanations as to what happened, and the presenting history is frequently inaccurate or incomplete.(4) The consequences of a missed diagnosis of AHT can put the child in increased danger and risk the child's life and future wellbeing;(5, 6) equally a false accusation of abuse can have devastating consequences for the child and family. The validity of AHT as a medical diagnosis is constantly questioned and any evidence regarding which combinations of clinical features are associated with a diagnosis of AHT or accidental trauma can support decision-making and lend weight to the diagnostic process.

Clinical prediction rules (CPRs) are evidence based tools that combine clinical features, history or results of investigations to predict diagnosis, prognosis or response to therapy. (7) They may assist clinicians in making complex decisions, improving their accuracy and decreasing variability between clinicians.(7) In 2011 and 2015, a Welsh team of experts in child protection research derived and validated the Predicting Abusive Head Trauma (PredAHT) tool, which calculates the probability of AHT in children <3 years of age with ICI based upon different combinations of 6 clinical features (head or neck bruising, seizure, apnoea, rib fracture, long bone fracture and retinal haemorrhage, detailed in Supplementary Table 1. (8, 9)

The PredAHT derivation study provided predicted probabilities of AHT for all 64 possible combinations of the presence or absence of these features.(8) In the validation study, using a 50% probability cut-off, PredAHT performed with a sensitivity of 72% and a specificity of 86% in identifying AHT.(9) In order to address one or more missing elements of clinical features the authors used their derivation data set (8) to create PredAHT-version 2 (PredAHT-2).(10)(11) (12) (see Appendix 1 for all possible permutations of present, absent and unknown). While validated at the derivation sites (13) PredAHT-2 requires validation in multiple locations and by independent investigators.(13, 14)

We describe an external validation of PredAHT-2 on an Australian/New Zealand population, of children <3 years of age admitted to hospital with ICI confirmed on neuroimaging.(8, 9)

METHODS

Study design, setting and patients

This study was a planned secondary subgroup analysis of children enrolled into a prospective multi-centre observational study of children (0 to <18 years) with head injuries in 10 Australian and New Zealand paediatric emergency departments (EDs) between April 2011 and November 2014.(15, 16) We obtained ethics approval from 5 participating sites (Australia: Royal Children's Hospital Melbourne, Monash Children's Hospital, Clayton, and Princess Margaret Hospital for Children, Perth; and New Zealand: Starship Children's Hospitals and KidzFirst Children's Hospitals, Auckland) for additional medical record review of possible cases of AHT.

Study procedures

Full details of the primary study protocol are described elsewhere.(16) In short, the parent study aimed to externally validate and compare three clinical decision rules, which assist in determining the need for CT in head injured children. Children and young people with head injury were enrolled by the treating ED clinician who collected clinical data and a research assistant recorded ED and hospital management data following the visit. We collected injury and clinical variables, demographic and epidemiological information as well as information about neuroimaging, admission and neurosurgery. In this study, we analysed data from a subset of children <3 years of age with ICI identified on neuroimaging, indicating that a differential diagnosis of AHT should be considered.

Exclusion criteria

Patients who did not undergo neuroimaging or who had normal neuroimaging results were excluded. Patients with skull fracture with no accompanying ICI and those with an underlying structural abnormality or pre-existing disease (hydrocephalus, cystic lesion or tumour, metabolic cause, malformation, or abnormal brain development), injuries caused by neglect or birth injuries were excluded as in the original PredAHT validation study.(8)

Strategy to identify possible AHT cases

In order to identify all possible AHT cases from the parent study(17) we extracted all children aged <3 years of age admitted to the hospital with head injury and abnormal neuroimaging results. In addition, at The Royal Children's Hospital, Melbourne (5372 (40.2%) of 13,371 patients enrolled at the 5 sites), we also accessed the database of the

Victorian Forensic Paediatric Medical Service, the hospital child protection team, which we searched for all children aged <3 years of age admitted to the hospital with head injury and abnormal neuroimaging results. We then accessed the medical records of all possible AHT cases at 5 sites and abstracted relevant data including predictor variables, outcomes and eligibility criteria for PredAHT.(8)

Study Definitions

We used senior radiologists' reports to determine the results of neuroimaging. ICI was defined as any combination of: any extra-axial haemorrhage, diffuse or focal parenchymal injury, cerebral oedema, cerebral contusion, hypoxic ischemic injury, or diffuse axonal injury visible on head computed tomography or magnetic resonance imaging.(8) AHT was defined as the diagnosis of ICI (confirmed on neuroimaging), which was due to physical child abuse by parents or caregivers, rather than neglect, according to the decision of a multidisciplinary child protection team at the conclusion of their investigation and their consideration of the relevant social, forensic, and clinical features in the context of the presenting history, in accordance with the Australian and New Zealand standard child protection assessment processes. Non-AHT was defined as ICI following a witnessed accidental injury or an accidental injury confirmed by the decision of a multidisciplinary child protection team. Cases were categorized as AHT-positive or AHT-negative (non-AHT) by the study investigators upon retrospective review of the multidisciplinary team records. Cases in which this categorization was not clear were deemed indeterminate.(18) Any uncertainty in terms of category assignment on review of the records was arbitrated by the director of the Victorian Forensic Paediatric Medical Service (A.S.) on the basis of the forensic reports and medical records.

Application of PredAHT-2 to the data set

We applied PredAHT-2 to each child <3 years of age with ICI confirmed on neuroimaging (8) blinded to the clinical outcome categorization. We calculated the specific probability of AHT for each individual patient based upon whether the six clinical features were present, absent or unknown (Appendix 1). As a primary analysis we used a 50% probability cut-off to categorize all patients with a probability of ≥50% as higher risk for AHT and those with a probability of <50% as lower risk for AHT. Individual clinician's interpretation and application of probability thresholds to risk and decision-making differs and therefore in a secondary analysis, we explored the implications of using different probability cut-off points to categorize cases as AHT.(12, 19, 20) To do this, we used a 20% probability cut-off, and an 80% probability cut-off, respectively.

Statistical analysis

Data were entered into Epidata (The Epidata Association, Odense, Denmark), and were later entered into REDCap.(21) Data were analysed using Stata 13 (Statacorp, College Station, TX). Summary statistics were derived to describe total and subgroup characteristics; proportions and frequencies for categorical variables, and the median (interquartile range) for continuous variables. Using the Stata command diagt we calculated the sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio and negative likelihood ratio of PredAHT-2 using the three different probability cut-offs (20%, 50% and 80%), with 95% confidence intervals (CIs), excluding AHT-indeterminate cases.

Reporting

This study is reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines.(22) A checklist is included in Appendix 2.

RESULTS

Of the 20,137 patients at 10 centres in the parent study, 13,371 (66%) patients presented at the 5 centres included in this secondary analysis. Of these patients 5,264 were <3 years old (39%), of which 3,038 (58%) were male. The medical records of 142 cases of children <3 years old admitted with possible physical-abuse related head injuries on neuroimaging were reviewed, and 87 children with ICI were identified (Figure 1).

Sixty-one (70%) were aged <1 year, 51 (59%) were male, 13 (15%) were admitted to PICU, 26 (30%) underwent neurosurgery and 6 (7%) died (Table 1).

Patients were categorized as AHT-positive in 27 (31%), AHT-negative in 45 (52%) and AHT-indeterminate in 15 cases (17%). Head or neck bruising was more strongly associated with AHT-negative cases than AHT-positive cases, while seizures, apnoea, rib fractures, long-bone fractures, and retinal haemorrhages were more strongly associated with AHT-positive cases than AHT-negative cases. Many AHT-negative cases did not have an ophthalmology examination or skeletal radiology (Table 2).

Figure 2 shows the distribution of the PredAHT-2 predicted probabilities against the outcome (AHT-positive, AHT-negative, indeterminate). The presence or absence of all six clinical features were recorded in all but one AHT-positive cases and all indeterminate cases, while only 36% (16/45) AHT-negative cases had complete data recorded. Using a probability cut-off of 50%, PredAHT-2 correctly identified 20/27 AHT-positive cases (sensitivity = 74% (95% CI, 54%–89%)) and correctly identified 39/45 AHT-negative cases (specificity = 87% (73%–95%)) (Table 3). For the AHT-negative cases, the probability of AHT was more likely to be <20% for cases where all six features were known. A total of 7/15 (47%) AHT-indeterminate cases were classified as higher risk by PredAHT-2 while 8/15 (53%) were classified as lower risk.

Applying PredAHT-2 using a 20% probability cut-off increased the sensitivity to 81% (62%–94%) at the expense of a much lower specificity (33% (20%–49%)). Conversely, applying PredAHT-2 using an 80% probability cut-off increased the specificity to 91% (79%–98%), at the expense of a much lower sensitivity (56% (35%–75%)).

DISCUSSION

In this external validation of PredAHT-2 in an Australian and New Zealand dataset, the performance of the tool was very similar to the performance of PredAHT in the original validation study (sensitivity 74% vs. 72% and specificity 87% vs. 86%).(8) With its added capacity to give a probability of AHT for an individual case where one or more of the six features are unknown, PredAHT-2 has the potential to contribute to decision-making at multiple points along the assessment and referral pathway (10)

Exploring the implications for clinical practice

There were 7 AHT-positive cases that were assigned a probability of AHT of <50% by PredAHT-2 (Figure 2). In 4 of these cases the perpetrator confessed or was accused by the other parent, or the child's injuries were severe, and included complex skull fractures and widespread bruising. These were additional factors that strongly increased the likelihood of physical abuse, highlighting the importance of interpreting probability estimates given by PredAHT-2 in combination with all other available information.

The extent of unknown features was considerable for AHT-negative cases (Figure 2) and likely to be related to the clinicians' decisions not to undertake a skeletal survey or ophthalmology examination, based upon their level of confidence that the injury was accidental and consistent with the mechanism of injury described. Independently witnessed mechanisms of injury included falls from a parent's arms, a fall down the stairs or being hit by a falling heavy object, which contrasted with the lack of history or inadequate explanation of trauma given by parents in AHT-positive cases. Where AHT-negative cases were fully investigated, the predicted probability of AHT was low.

Six AHT-negative cases were assigned a probability of >50% (Figure 2). Five of these cases did not have an ophthalmology examination or skeletal survey. Completing the investigation would identify whether retinal haemorrhages, rib or long-bone fractures were present, and refine the probability estimate. For example, in children with ICI and head/neck bruising but no information about retinal haemorrhages or fractures, the calculated probability of AHT is 44.2% (see Appendix 1). If skeletal survey and ophthalmology were normal this would decrease to 14.7%. Conversely, if either long-bone fracture, retinal haemorrhage or rib fracture were identified, the probability would increase to 70.2%, 85.3%, and 88.5%, respectively. This highlights the importance of considering an ophthalmology examination and skeletal survey for those children presenting with ICI in the absence of an independently witnessed accident.

All of the 15 indeterminate cases were fully investigated and PredAHT predicted 7/15 to be high risk (>80%) and 7/15 to be low risk (<20%) for AHT. This finding suggests that the uncertainty in these cases did not arise from a consideration of the clinical features alone, and is consistent with a study conducted by Chaiyachati et al.,(23) who found that there was no single component of the injury, incident or history associated with the uncertainty around clinicians' perceived likelihood of physical abuse. Of 7 indeterminate cases with a probability of abuse of >80%, 2 died and AHT was deemed "likely" in 4 cases. Among the 7 with a probability of AHT of <20%, AHT was deemed "likely" in 1 case; however, in each of these cases AHT could not be definitively confirmed, partly due to differing opinions between members of the multidisciplinary child protection team, most notably between medical clinicians and child protection social workers.

The study findings reinforce those from the original derivation and validation studies,(8, 9) that no set of clinical features is specific for AHT. It is therefore unlikely that any CPR based upon clinical features alone could perfectly predict AHT and emphasizes that PredAHT-2 should be used in combination with a full multidisciplinary assessment and consideration of all of the other clinical, social, historical and forensic elements of each individual case. PredAHT is designed to provide a specific probability estimate for each individual case based upon six key features that should be identified during an assessment of a young child with ICI to inform further investigations or decisions.

This validation strengthens the utility of PredAHT-2, and raises its level of evidence.(24) (20)

Roll-out of a computerized version would enable simple application of PredAHT-2 at the bedside, as new information is collected. PredAHT-2 should now be tested in an impact analysis study to determine its impact on clinician behaviour and relevant patient outcomes.(25)

Comparison with existing literature

The Pediatric Brain Injury Research Network (PediBIRN) 4-variable CPR is the only other CPR for paediatric AHT intended for use in an inpatient setting, and was designed to assist in deciding which children with cranial or ICI admitted to the paediatric intensive care unit (PICU) should be evaluated further for abuse.(26) (27-29) In a recent external validation of PediBIRN by our group, its sensitivity was 96% and its specificity was 43% when applied to children <3 years old with ICI or cranial injury, admitted to *all* inpatient settings.(17) Taking the arbitrary probability cut-off of 50% PredAHT-2 had a lower sensitivity than PediBIRN, but a much higher specificity, and will categorize fewer AHT-negative cases as higher risk for AHT than PediBIRN.

Limitations

The study has a number of limitations. The majority of the predictive and outcome variables were collected prospectively; the predictive variable "apnoea", however, was extracted from medical records. Future validation studies should ensure that the six variables are collected prospectively, and should consider assessing their inter-rater reliability. Since case selection in our dataset was mostly based on ED identification, cases of possible AHT identified in a hospital ward or PICU would have been missed. One of the concerns in all studies of AHT is the possibility of creating a circular argument by defining the condition based on the features included in the tool. However, in this study AHT diagnosis was assigned by local multidisciplinary teams independent of the study. A cautious approach was taken, and if there was any doubt, a category of indeterminate was assigned. In addition the data for the study were extracted blinded to the case outcomes.

Conclusions

PredAHT-2 performed with reasonably high sensitivity and specificity when externally validated. The inclusion of probability estimates in incompletely investigated cases offers an opportunity for clinicians to consider the probability of AHT at different stages of the clinical assessment and whether or not further investigations are indicated.

Funding Source: The study was funded by grants from the National Health and Medical Research Council (project grant GNT1046727, Centre of Research Excellence for Paediatric Emergency Medicine GNT1058560), Canberra, Australia; the Murdoch Children's Research Institute, Melbourne, Australia; the Emergency Medicine Foundation (EMPJ-11162), Brisbane, Australia; Perpetual Philanthropic Services (2012/1140), Australia; Auckland Medical Research Foundation (No. 3112011) and the A + Trust (Auckland District Health Board), Auckland, New Zealand; WA Health Targeted Research Funds 2013, Perth, Australia; the Townsville Hospital and Health Service Private Practice Research and Education Trust Fund, Townsville, Australia; and supported by the Victorian Government's Infrastructure Support Program, Melbourne, Australia. FEB's time was part funded by a grant from the Royal Children's Hospital Foundation and the Melbourne Campus Clinician Scientist Fellowship, Melbourne, Australia and an NHMRC Practitioner Fellowship, Canberra, Australia. SRD's time was part funded by the Health Research Council of New Zealand (HRC13/556).

Financial Disclosure: The authors have no financial relationships relevant to this article to disclose.

Conflict of interest: Prof. Kemp and Dr. Cowley are part of the team that derived and validated the PredAHT tool. However, all data collection and analyses were undertaken independently of either of these authors.

Clinical Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12614000463673

Abbreviations: abusive head trauma (AHT); clinical prediction rule (CPR); confidence interval (CI); emergency department (ED); intracranial injury (ICI); paediatric intensive care unit (PICU); traumatic brain injury (TBI)

What is already known on this topic:

- PredAHT is one of only two published prediction rules determining the likelihood of abusive head trauma in children <3 years old with intracranial injury.
- PredAHT has not been validated in a broader setting outside Europe.

What this study adds:

- We externally validated the prediction tool in an Australian/New Zealand dataset.
- The tool performed with reasonably high sensitivity (74% (54%–89%)) and specificity (87% (73%–95%)) using a 50% probability cut-off.
- Performance was similar to that in the original validation study.

Contributors' Statement

Helena Pfeiffer: Ms. Pfeiffer contributed to the design of the study, conducted the review of medical records, carried out the initial analyses, drafted the initial manuscript and revised the article.

Laura E. Cowley: Dr. Cowley contributed to the design of the study, made substantial contributions to the interpretation and discussion of findings and the drafting of the manuscript, produced Figure 2, and critically revised the manuscript for important intellectual content.

Alison M. Kemp, John Cheek, Stuart R Dalziel, Meredith L Borland, Sharon O'Brien, Megan Bonish, Jocelyn Neutze, Ed Oakley, Louise Crowe, Mark D Lyttle and Silvia Bressan contributed to the design of the study, made substantial contributions to the interpretation and discussion of findings and critically revised the manuscript for important intellectual content.

Anne Smith: Dr. Smith contributed to the design of the study, supervised the categorization of cases and critically revised the manuscript for important intellectual content.

Stephen Hearps: Mr. Hearps contributed to the design of the study, carried out the initial analyses, drafted the tables and critically revised the manuscript for important intellectual content.

Franz E. Babl: Prof. Babl had the initial study idea, contributed to the design of the study, and critically revised the manuscript for important intellectual content. He takes responsibility for the paper as a whole.

REFERENCES

1. Duhaime AC, Christian CW, Rorke LR, Zimmerman RA. Nonaccidental head trauma in infants the “Shaken baby syndrome”. *N Engl J Med*. 1999;**338**:1822-9.
2. Klevens J, Leeb RT. Child maltreatment fatalities in children under 5: Findings from the National Violence Death Reporting System. *Child Abuse & Negl*. 2010;**34**:262-6.
3. Govind SK, Merritt NH. A 15 year cohort review of in-hospital pediatric trauma center mortality: A catalyst for injury prevention programming. *Am J Surg*. 2018;**216**:567-72.
4. Leventhal JM, Asnes AG, Pavlovic L, Moles RL. Diagnosing abusive head trauma: the challenges faced by clinicians. *Pediatr Radiol*. 2014;**44 Suppl 4**:S537-42.
5. Letson MM, Cooper JN, Deans KJ, *et al.* Prior opportunities to identify abuse in children with abusive head trauma. *Child Abuse & Negl*. 2016;**60**:36-45.
6. Jenny C, Hymel KP, Ritzen A, Reinert SE, Hay TC. Analysis of missed cases of abusive head trauma. *JAMA*. 1999;**281**:621-6.
7. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA*. 2000;**284**:79-84.
8. Cowley LE, Morris CB, Maguire SA, Farewell DM, Kemp AM. Validation of a Prediction Tool for Abusive Head Trauma. *Pediatrics*. 2015;**136**:290-8.
9. Maguire SA, Kemp AM, Lumb RC, Farewell DM. Estimating the probability of abusive head trauma: a pooled analysis. *Pediatrics*. 2011;**128**:e550-64.
10. Cowley LE, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO, Kemp AM. Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals. *Child Abuse & Negl*. 2018;**81**:192-205.
11. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. 2011;**45**:67.
12. Cowley LE, Farewell DM, Kemp AM. Potential impact of the validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A clinical vignette study. *Child Abuse & Negl*. 2018;**86**:184-96.
13. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;**61**:1085-94.
14. Steyerberg E. Clinical prediction models: a practical approach to development, validation and updating: Springer; 2009.
15. Babl FE, Borland ML, Phillips N, *et al.* Accuracy of PECARN, CATCH, and CHALICE head injury decision rules in children: a prospective cohort study. *Lancet*. 2017;**389**:2393-402.
16. Babl FE, Lyttle MD, Bressan S, *et al.* A prospective observational study to assess the diagnostic accuracy of clinical decision rules for children presenting to emergency departments after head injuries (protocol): the Australasian Paediatric Head Injury Rules Study (APHIRST). *BMC Pediatr*. 2014;**14**:148.
17. Pfeiffer H, Smith A, Kemp AM, *et al.* External Validation of the PediBIRN Clinical Prediction Rule for Abusive Head Trauma. *Pediatrics*. 2018;**141**.

18. Leeb RT, Paulozzi LJ, Melanson C, Simon TR, Arias I. Child Maltreatment Surveillance: Uniform Definitions for Public Health and Recommended Data Elements. *Centers for Disease Control & Prevention, Atlanta, Georgia*. 2008.
19. Dias MS, Boehmer S, Johnston-Walsh L, Levi BH. Defining 'reasonable medical certainty' in court: What does it mean to medical experts in child abuse cases? *Child Abuse & Negl*. 2015;**50**:218-27.
20. Levi BH, Brown G. Reasonable suspicion: a study of Pennsylvania pediatricians regarding child abuse. *Pediatrics*. 2005;**116**:e5-12.
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;**42**:377-81.
22. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Ann Intern Med*. 2015;**162**:735-6.
23. Chaiyachati BH, Asnes AG, Moles RL, Schaeffer P, Leventhal JM. Gray cases of child abuse: Investigating factors associated with uncertainty. *Child Abuse & Negl*. 2016;**51**:87-92.
24. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;**144**:201-9.
25. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;**338**:b606.
26. Hymel KP, Armijo-Garcia V, Foster R, *et al.* Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics*. 2014;**134**:e1537-44.
27. Hymel KP, Herman BE, Narang SK, *et al.* Potential Impact of a Validated Screening Tool for Pediatric Abusive Head Trauma. *J Pediatr*. 2015;**167**:1375-81 e1.
28. Hymel KP, Willson DF, Boos SC, *et al.* Derivation of a clinical prediction rule for pediatric abusive head trauma. *Pediatric Critical Care Medicine*. 2013;**14**:210-20.
29. Pfeiffer H, Crowe L, Kemp AM, *et al.* Clinical prediction rules for abusive head trauma: a systematic review. *Archives of Disease in Childhood*. 2018;**103**:776-83.
30. Kuppermann N, Holmes JF, Dayan PS, Hoyle JD, Jr., Atabaki SM, Holubkov R, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet*. 2009;**374:**1160-70.

Table 1 – Demographics and Epidemiology

	Total (n=87)		AHT positive (n=27)		AHT negative (n=45)		AHT indeterminate (n=15)	
Age (in years), n (%)								
< 1	61	(70)	21	(78)	28	(62)	12	(80)
1 – <2	17	(20)	3	(11)	12	(27)	2	(13)
2 – <3	9	(10)	3	(11)	5	(11)	1	(7)
Gender, n (%)								
Male	51	(59)	15	(56)	27	(60)	9	(60)
Female	36	(41)	12	(44)	18	(40)	6	(40)
PICU admission, n (%)	13	(15)	5	(19)	6	(13)	2	(13)
Neurosurgery, n (%)	26	(30)	9	(33)	10	(22)	7	(47)
Intubation, n (%)	13	(15)	4	(15)	8	(18)	1	(7)
ciTBI, n (%)	37	(43)	10	(37)	22	(49)	5	(33)
Mortality, n (%)	6	(7)	4	(15)	1	(2)	1	(7)
Length of stay (days)								
Median (IQR)	5	(3–10)	9.5	(7–18)	4	(3–6)	6	(4–14)

AHT abusive head trauma, PICU pediatric intensive care unit, ciTBI clinically important traumatic brain injury (using the Pediatric Emergency Care Applied Research Network definition[30]), IQR interquartile range

Table 2 – Presence of predictive variables

	Total (n=87)	AHT positive (n=27)			AHT negative (n=45)			AHT indeterminate (n=15)					
	N	**n**	**%**	**95% CI**	**n**	**%**	**95% CI**	**n**	**%**	**95% CI**	**OR for AHT[a]**	**95% CI**	***p***
Head or neck bruising													
Present	62	15	56	(35-75)	38	84	(71-94)	9	60	(32-84)	0.23	(0-1)	0.012
Absent	25	12	44	(25-65)	7	16	(6-29)	6	40	(16-68)			
Unknown	0	0	0	(0-13)	0	0	(0-8)	0	0	(0-22)			
Seizures													
Present	25	13	48	(29-68)	8	18	(8-32)	4	27	(8-55)	4.63	(2-14)	0.007
Absent	61	13	48	(29-68)	37	82	(68-92)	11	73	(45-92)			
Unknown	1	1	4	(0-19)	0	0	(0-08)	0	0	(0-22)			
Apnoea													
Present	11	6	22	(9-42)	2	4	(1-15)	3	20	(4-48)	6.14	(1-33)	0.046
Absent	76	21	78	(58-91)	43	96	(85-99)	12	80	(52-96)			
Unknown	0	0	0	(0-13)	0	0	(0-8)	0	0	(0-22)			
Rib fracture													
Present	8	6	22	(9-42)	0	0	(0-8)	2	13	(2-40)	14.81	(1-279)	0.024
Absent	58	21	78	(58-91)	24	53	(38-68)	13	87	(60-98)			
Unknown	21	0	0	(0-13)	21	47	(32-62)	0	0	(0-22)			
Long-bone fracture													
Present	12	9	33	(17-54)	1	2	(0-12)	2	13	(2-40)	8.50	(1-74)	0.034
Absent	48	18	67	(46-83)	17	38	(24-53)	13	87	(60-98)			
Unknown	27	0	0	(0-13)	27	60	(44-74)	0	0	(0-22)			
Retinal haemorrhage													
Present	21	16	59	(39-78)	0	0	(0-8)	5	33	(12-62)	58.83	(3-1.074)	<0.001
Absent	41	11	41	(22-61)	20	44	(30-60)	10	67	(38-88)			

Unknown	25	0	0	(0-13)	25	56	(40-70)	0	0	(0-22)

[a] Calculations exclude the indeterminate and unknown cases, substitution formula used for zero cell counts

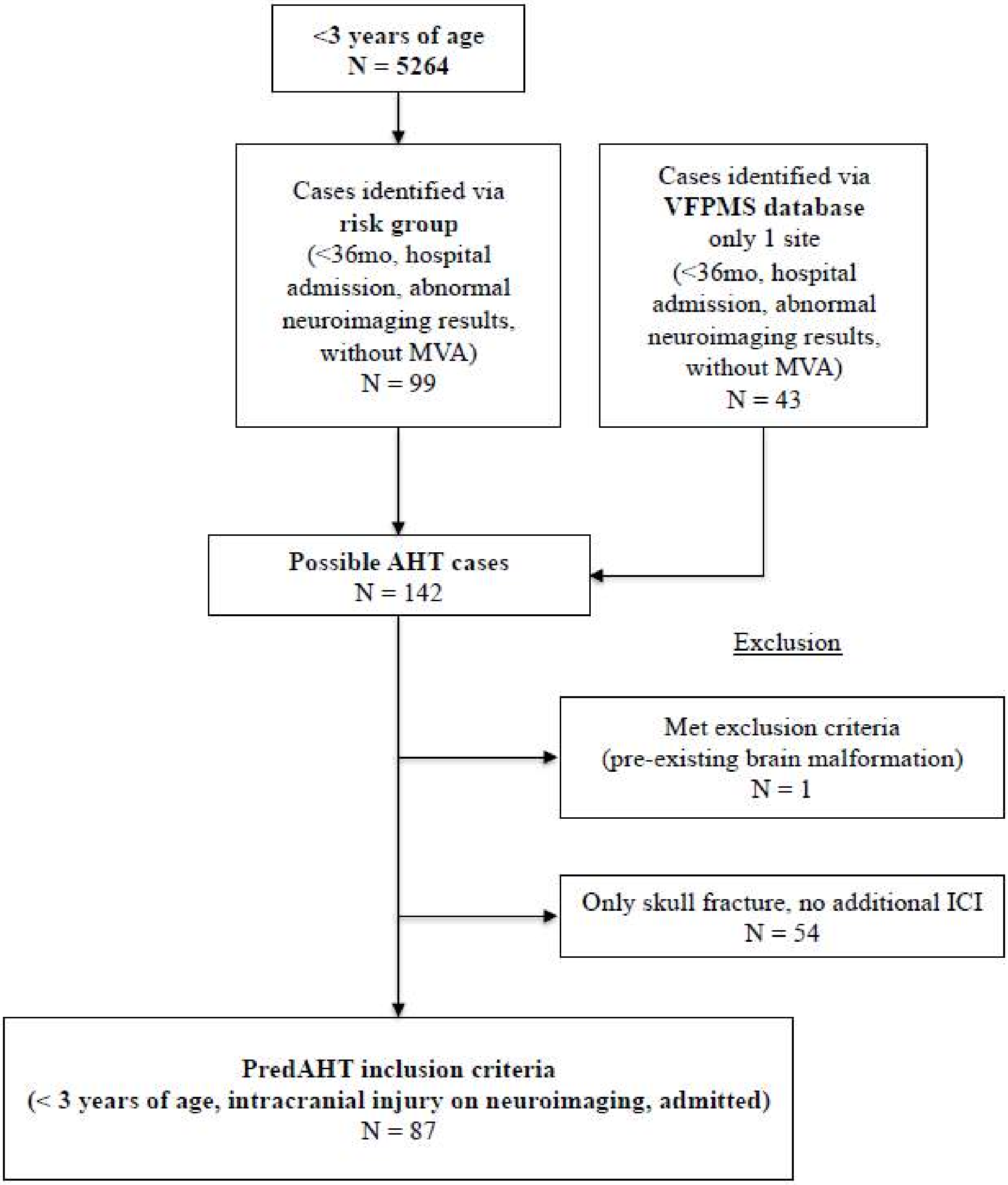
AHT=abusive head trauma, CI=confidence interval, p=Fisher's Exact

Table 3 – Performance of the PredAHT tool at three probability cut-offs

Applying PredAHT (indeterminate excluded)	**20% cut-off**		**50% cut-off**		**80% cut-off**	
	Outcome		Outcome		Outcome	
	AHT	nAHT	AHT	nAHT	AHT	nAHT
Higher risk of AHT	22	30	20	6	15	4
Lower risk of AHT	5	15	7	39	12	41
	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	81%	62%–94%	74%	54%–89%	56%	35%–75%
Specificity	33%	20%–49%	87%	73%–95%	91%	79%–98%
Positive predictive value	42%	29%–57%	77%	56%–91%	79%	54%–94%
Negative predictive value	75%	51%–91%	85%	71%–94%	77%	64%–88%
LR +	1.22	0.93–1.61	5.56	2.55–12.1	6.25	2.31–16.9
LR −	0.56	0.23–1.36	0.30	0.16–0.57	0.49	0.32–0.75

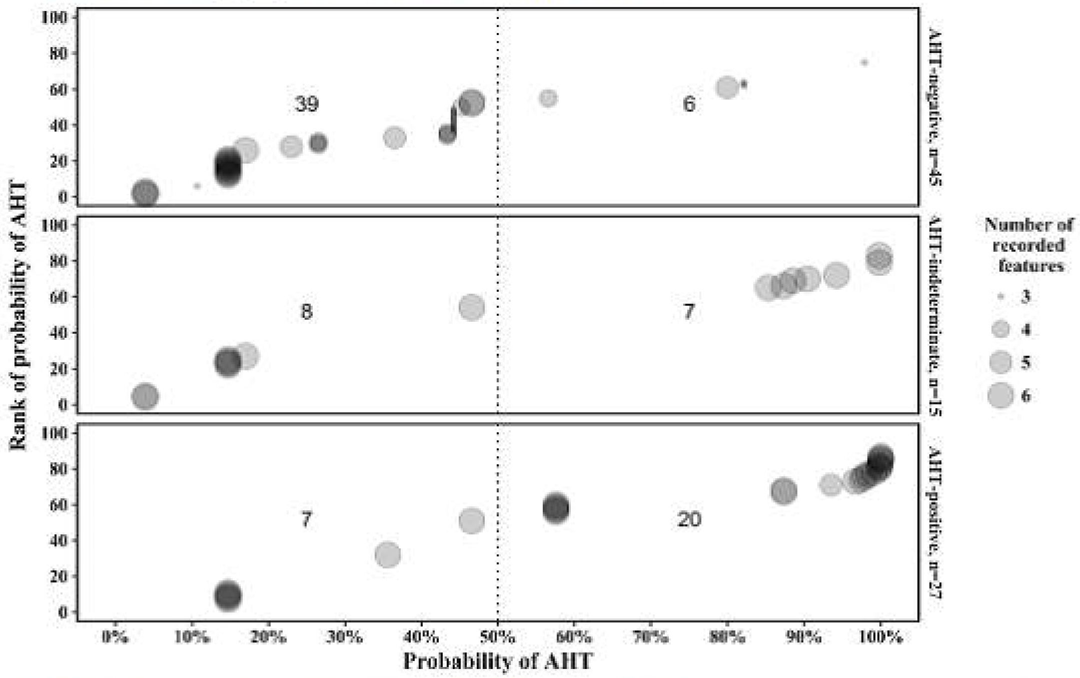
AHT abusive head trauma; nAHT no abusive head trauma, LR likelihood ratio

Figure 1 – Sources of possible abusive head trauma (AHT) cases



AHT abusive head trauma,, VFPMS Victorian Forensic Paediatric Medical Service, MVA motor vehicle accident, ICI intracranial injury

Figure 2 – Predicted probability of abusive head trauma (AHT) assigned by PredAHT for all 87 children with intracranial injury, by outcome and number of recorded features



The circles represent the calculated probability of AHT for each of the 87 children with intracranial injury included in the study.
The numbers on the figure correspond to the number of children categorized as higher or lower risk for AHT based on a 50% probability cut-off.
The size of the circles shows how many of the six features are recorded as present or absent. The smaller the circle, the more information is unknown and the less likely that a skeletal survey and/or an ophthalmology examination was undertaken.

Appendix 6. Published articles arising from the qualitative study presented in Chapter 5

1. **Cowley LE**, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO & Kemp AM (2018). Factors influencing child protection professionals' decision-making and multidisciplinary collaboration in suspected abusive head trauma cases: A qualitative study. *Child Abuse & Neglect,* 82: 178-191.
2. **Cowley LE**, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO & Kemp AM (2018). Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals. *Child Abuse & Neglect,* 81: 192-205.

Child Abuse & Neglect 82 (2018) 178–191

Contents lists available at ScienceDirect

Child Abuse & Neglect

journal homepage: www.elsevier.com/locate/chiabuneg

Factors influencing child protection professionals' decision-making and multidisciplinary collaboration in suspected abusive head trauma cases: A qualitative study

Laura E. Cowley*, Sabine Maguire, Daniel M. Farewell, Harriet D. Quinn-Scoggins, Matthew O. Flynn, Alison M. Kemp

Division of Population Medicine, School of Medicine, Cardiff University, Wales, United Kingdom

ARTICLE INFO

Keywords:
Child physical abuse
Abusive head trauma
Qualitative research
Child protection

ABSTRACT

Clinicians face unique challenges when assessing suspected child abuse cases. The majority of the literature exploring diagnostic decision-making in this field is anecdotal or survey-based and there is a lack of studies exploring decision-making around suspected abusive head trauma (AHT). We aimed to determine factors influencing decision-making and multidisciplinary collaboration in suspected AHT cases, amongst 56 child protection professionals. Semi-structured interviews were conducted with clinicians (25), child protection social workers (10), legal practitioners (9, including 4 judges), police officers (8), and pathologists (4), purposively sampled across southwest United Kingdom. Interviews were recorded, transcribed and imported into NVivo for thematic analysis (38% double-coded). We identified six themes influencing decision-making: 'professional', 'medical', 'circumstantial', 'family', 'psychological' and 'legal' factors. Participants diagnose AHT based on clinical features, the history, and the social history, after excluding potential differential diagnoses. Participants find these cases emotionally challenging but are aware of potential biases in their evaluations and strive to overcome these. Barriers to decision-making include lack of experience, uncertainty, the impact on the family, the pressure of making the correct diagnosis, and disagreements between professionals. Legal barriers include alternative theories of causation proposed in court. Facilitators include support from colleagues and knowledge of the evidence-base. Participants' experiences with multidisciplinary collaboration are generally positive, however child protection social workers and police officers are heavily reliant on clinicians to guide their decision-making, suggesting the need for training on the medical aspects of physical abuse for these professionals and multidisciplinary training that provides knowledge about the roles of each agency.

1. Introduction

Abusive head trauma (AHT) is the primary cause of fatal child abuse, and the majority of fatal head injuries in children aged less than two years are due to physical abuse (Gill et al., 2009). Morbidity for children who survive AHT is significant; a recent extended follow-up study of children who suffered severe AHT found that 40% presented with serious neurological impairment (Lind et al., 2016). AHT may go unrecognized in up to 30% of cases (Jenny, Hymel, Ritzen, Reinert, & Hay, 1999; Letson et al., 2016; Sheets et al.,

* Corresponding author.
E-mail addresses: laura_c24@hotmail.co.uk (L.E. Cowley), sabinemaguire@gmail.com (S. Maguire), FarewellD@cardiff.ac.uk (D.M. Farewell), Quinn-ScogginsHD@cardiff.ac.uk (H.D. Quinn-Scoggins), mattandflynn@gmail.com (M.O. Flynn), KempAM@cardiff.ac.uk (A.M. Kemp).

https://doi.org/10.1016/j.chiabu.2018.06.009
Received 5 March 2018; Received in revised form 7 June 2018; Accepted 11 June 2018
0145-2134/

2013) yet early detection of AHT can save lives; a seminal study indicated that 80% of deaths could have been prevented if AHT was recognized during a prior medical evaluation (Jenny et al., 1999).

Clinicians face unique diagnostic challenges in suspected child physical abuse cases (Leventhal, Asnes, Pavlovic, & Moles, 2014). In common with many areas of medicine, there is no gold-standard diagnostic test for AHT, and the history provided by the caregiver may be inaccurate or deliberately misleading. The stakes are high; if abuse is not identified, children may be re-injured, possibly fatally. Conversely, a wrongful diagnosis of abuse has profound emotional, societal and legal consequences for the families involved. Due to the complex nature of suspected abuse cases, clinicians must work with colleagues from other clinical sub-specialties (e.g. trauma surgeons, neuroradiologists and skeletal radiologists, ophthalmologists), child protection social workers (CPSWs), and professionals from law enforcement. These professionals must work together as a multidisciplinary team, to jointly determine the likelihood of AHT.

Despite this, studies have found that clinicians may lack the confidence to identify abuse (Flaherty et al., 2006), differ in their views of what constitutes a 'reasonable suspicion' or 'reasonable medical certainty' of abuse (Dias, Boehmer, Johnston-Walsh, & Levi, 2015; Levi & Brown, 2005), exhibit biases in their evaluations of AHT related to the family's socioeconomic status and race (Wood et al., 2010), and demonstrate inconsistencies in their investigations and diagnoses of suspected abuse (Anderst, Nielsen-Parker, Moffatt, Frazier, & Kennedy, 2016; Wood et al., 2012). In addition, the validity of AHT/shaken baby syndrome (SBS) as a medical diagnosis is constantly questioned, often falsely predicated on the premise that a "diagnostic triad" of subdural hemorrhages, retinal hemorrhages and encephalopathy defines AHT, and forms the basis of a clinical AHT diagnosis (Elinder et al., 2016; Lynøe et al., 2017; Rorke-Adams, 2011; Squier, 2011).

Much of the evidence regarding the barriers or facilitators to multidisciplinary working or the perceptions of professionals working in multidisciplinary teams in suspected abuse cases has been anecdotal, or has relied on case studies or surveys (e.g. Inkilä, Flinck, Luukkaala, Åstedt-Kurki, & Paavilainen, 2013; Sedlak et al., 2006). Furthermore, while surveys have been used to assess the factors affecting clinicians' decisions to *report* suspected abuse (e.g. Flaherty et al., 2006, 2008; Gunn, Hickson, & Cooper, 2005), these were all conducted in North America, and do not address decision-making processes in suspected AHT specifically. The primary aim of this study was to explore factors influencing decision-making in suspected AHT cases, amongst a variety of professionals involved. The secondary aim was to explore the working relationships between the different professional groups.

2. Methods

This was a qualitative semi-structured interview study. The study methods have been published previously (Cowley et al., 2018). The study received ethical approval from the Cardiff University School of Medicine Research Ethics Committee (Ref: 15/35). This study is reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines (Tong, Sainsbury, & Craig, 2007); a checklist is included in Appendix 1.

2.1. Participant recruitment

Purposive sampling and snowball sampling were used to recruit participants for this study. We targeted clinicians, CPSWs, legal practitioners, police officers and pathologists involved in suspected AHT cases across south west United Kingdom (UK). A list of potential participants was identified through personal contacts of the research team and organizational websites. Personal contacts and organizations were sent an information sheet to explain the study and were asked to suggest suitable participants for interview. A random selection of individuals from each professional group were then invited to take part. We recruited participants with different levels of child protection experience and seniority (Fig. 1). Individuals were contacted via email, with the exception of judges who were sent formal letters of invitation. In this study the term "clinician" refers to medical doctors and specialist nurses, who were sampled from three teaching hospitals and two district general hospitals across a range of specialties including pediatrics, radiology and neurosurgery. Most participating clinicians were consultants, with the exception of two associate specialists, two trainee doctors and one nurse. Judges had more child protection experience than barristers or solicitors, while forensic pathologists had more child protection experience than the pediatric pathologist. Senior CPSWs and police officers had more child protection experience than their junior counterparts.

2.2. Interview schedule development

The interview schedule was developed by two of the authors (LC and MF), discussed within the research team and revised accordingly (Appendix 2). Questions were derived from the existing research literature on the identification of AHT. The schedule was piloted with a police officer and a clinician, regarding the length, appropriateness, and content, and amended accordingly. The schedule comprised core open-ended questions, prompts and clarifying questions. Interviews explored participants' usual practice and decision-making in head-injury cases where AHT is suspected, and their experiences of multidisciplinary working. It was a guide rather than a definitive list, to allow exploration of additional topic areas that might be raised by participants. Early interview responses influenced questions asked in later interviews; the schedule was updated as data collection and analysis progressed and new topic areas were raised. We also explored the participants' attitudes towards the Predicting Abusive Head Trauma (PredAHT) clinical prediction rule; these results are reported elsewhere (Cowley et al., 2018).

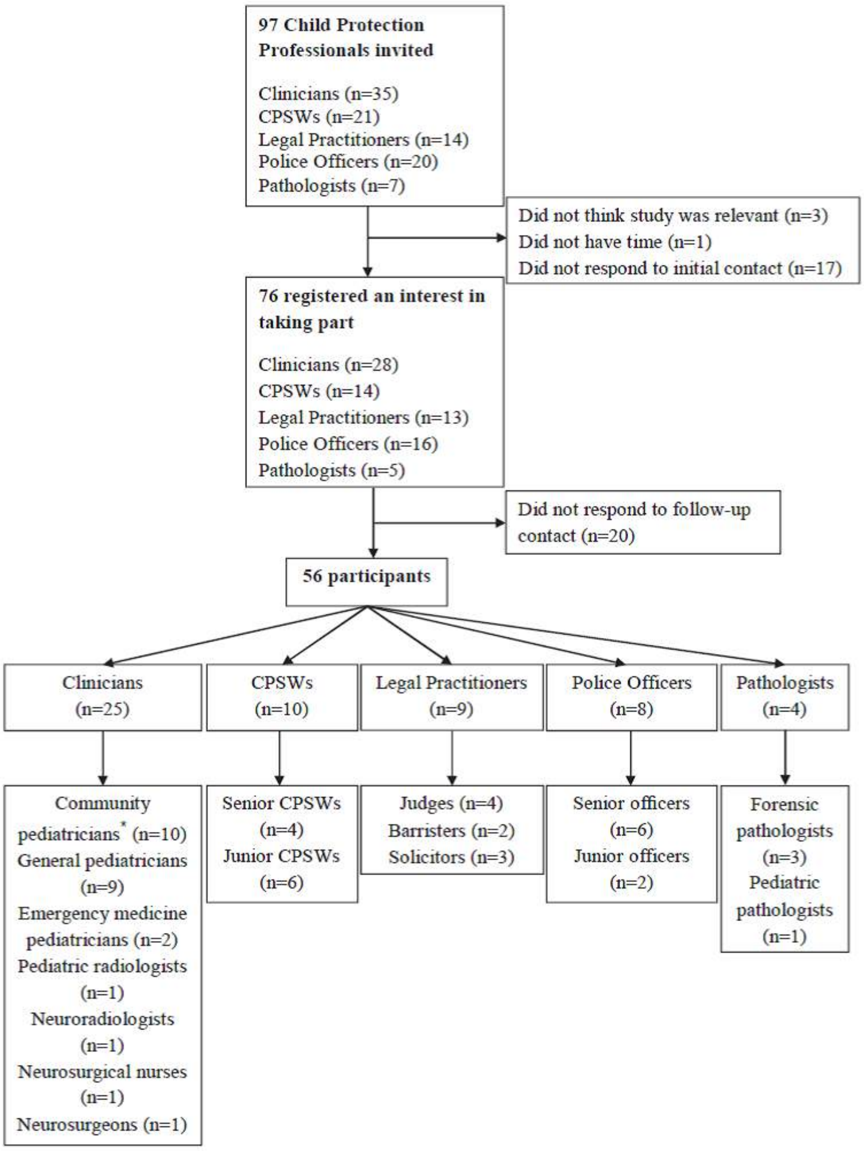


Fig. 1. Flowchart of child protection professionals participating in a qualitative study of decision-making in suspected abusive head trauma cases.
CPSWs = Child Protection Social Workers
*In the United Kingdom, community pediatricians are doctors with expertise in working with vulnerable groups of children and their care-givers, including those who are at risk of abuse or are being abused. They have additional training in safeguarding.

2.3. Data collection

Interviews were conducted by LC, a PhD student with training in qualitative research methods and qualitative interview techniques. No relationship was established between the interviewer and participants prior to the study. Informed consent was obtained, including permission for audio recording for verbatim transcription. When two participants declined to be audio recorded the interviewer made detailed notes of their responses. These were sent to the participants to check that they were a fair reflection of their views. The schedule was delivered to individuals or at two small group interviews (of three and five participants from the same professional group) where personal interaction between the participants was minimized, to maximize individual contributions.

Interviews lasted 45 min, and took place at the participants' workplace between June 2015 and September 2016. MF was also present to record relevant field notes such as participant non-verbal behavior and response to the interview, and critical reflections about the interview. No repeat interviews were conducted. In the interests of reflexivity, the interviewer considered how her own values and assumptions as a student researching decision-making in suspected AHT cases might influence the interviews or the interpretation of the findings. A reflective journal was kept in an attempt to minimize potential bias. In early interviews with clinicians, the interviewer was acutely aware of her status as a non-medical student with no medical training but nevertheless conducting PhD research in a medical topic, and how this may affect the power relationship between the researcher and the participant. Subsequently, to break down power imbalances, every effort was made to build a rapport with the participants and ensure that the interview was guided by them, while also staying on-topic. The researcher also made sure not to ask any leading questions or impose their own views on the participants.

2.4. Data analysis

Data analysis began shortly after the first interview using thematic analysis (Braun & Clarke, 2006). Data categories were arranged under overarching themes. A general inductive approach enabled the results to be guided by the aims and objectives of the research, and the raw data (Bryman & Burgess, 1994; Dey, 1993). The Framework Method was used to manage, summarize, display, and synthesize the data and to facilitate analysis (Gale, Heath, Cameron, Rashid, & Redwood, 2013). Analysis followed seven phases: transcription, familiarization, coding, developing an analytic framework, applying the analytic framework, charting data into framework matrices, and interpretation (Gale et al., 2013). Initial codes were generated independently by LC, MF and HQS. These were jointly grouped into clearly defined categories that were further arranged under themes. Discrepancies between coders were resolved by discussion and consensus. This process was undertaken in an attempt to minimize individual biases; 38% of the transcripts were independently double-coded. The joint analysis enabled the development of a preliminary analytic framework. Transcripts were imported into NVivo (QSR International Pty Ltd., 2014), to organize and manage the data and assist with data analysis. Quotes pertaining to each category were retrieved and 'charted' into thematic framework matrices. Interviews ceased when thematic saturation was achieved within each group of participants (clinicians, CPSWs, police officers, legal practitioners and pathologists), which was verified using the constant comparative method (Glaser & Strauss, 1967). The final phase of the analysis involved abstraction and interpretation of the data. Participants were not asked to provide feedback on the study findings. Analysis focused on identifying factors that were perceived to influence decision-making in cases of suspected AHT. Six major themes were identified: 'professional factors', 'medical factors', 'circumstantial factors', 'family factors', 'psychological factors', and 'legal factors'. Categories and their definitions are detailed in the final analytic framework (Table 1). The systematic synthesis of the data excerpts into thematic matrices enabled a review of the categories across cases, to identify barriers and facilitators to decision-making.

3. Results

Participant demographics and response rates are shown in Table 2 and Fig. 1. Data are presented using quotations, selected as examples of the themes and categories that emerged from the data. Within the quotations, square brackets represent text inserted for clarification. Word repetitions and irrelevant sections were removed and denoted by '...'.

3.1. Professional factors

Participants' **perceived role in the decision-making process** differed by professional group. All community and general pediatricians agreed that it is within their remit to come to a decision as to the likelihood that a child has suffered AHT, as part of a multidisciplinary team. Emergency medicine specialists, radiologists, the neurosurgeon and the pediatric pathologist would raise concerns with other colleagues, but not make the final diagnosis. Two forensic pathologists would provide a steer to other professionals, while one stated that it was not their job to make decisions about abuse. Barristers and solicitors, the neurosurgeon and one forensic pathologist emphasized that it is the **role of the court** and ultimately the judge to decide whether a child has suffered AHT.

"I suppose in every case you wonder whether that's happened, but it's not for us to ultimately make that decision, we just have to present the evidence and it's for the judge to make the decision at the end of the day." Legal Practitioner 2

The self-perceived role of CPSWs and judges in suspected AHT cases is to **protect the child from future harm**, rather than to determine whether AHT has occurred *per se*.

"What the [family] court has to decide is...has this child suffered significant harm? Or, does the evidence disclose, based on facts that you can find that there is a real possibility of significant harm in the future...so in terms of us deciding was this a non-accidental injury or an accidental injury, in some cases it won't make any difference to a decision that we have to make because you can have a very serious accident that will occur as the result of an inappropriate care-giver, or an unsafe care-giver, or a lack of supervision." Judge 3

CPSWs and police officers **are heavily reliant on medical professionals** to come to a decision as to whether AHT has occurred, and to guide their decision-making. This is due to their own lack of medical training and knowledge. The more experience these participants had investigating suspected AHT, the more knowledge they had.

"To support my decision-making I would rely heavily on what consultants are telling us, what the experts are telling us about those injuries and what the likely cause is, and what's acute, what's not, what's explained, what could potentially cause this. So yeah, major, major reliance on that clinical information, I can't emphasize that enough...I would rely quite heavily on that expert view, and the views of those medical professionals with child protection experience." Police Officer 7

Table 1
Analytic Framework.

Theme	Categories & Definitions
Professional factors	**Participants' perceived role in the decision-making process:** discussions regarding the participants' role in making a decision as to whether children in suspected abuse cases have suffered AHT; whether they feel it is within their remit to make such decisions and why; whether they form an opinion about the likelihood of AHT having taken place **Reliance on other professionals:** any comments relating to a reliance on others to identify AHT or direct participants' decision-making; any reasons why participants may rely on others such as medical professionals, e.g. due to a lack of medical knowledge, for information sharing or for a high quality clinical investigation; any difficulties associated with having to rely on others for information or guidance **Multidisciplinary collaboration:** any comments about the positive or negative aspects of working with other agencies e.g. discussions about the quality of the relationships between the professional groups; organizational barriers; delays; competing interests; disagreements between professionals **Resources:** any remarks regarding the availability of resources to support an investigation such as an adequate budget or staff with expertise in child protection work **Difficulty making the diagnosis:** any remarks about the ease or difficulty in making a diagnosis in suspected AHT cases; any reasons why a diagnosis of AHT may be easy or difficult to make **Confidence:** discussions regarding how confident the participants' feel working on AHT cases or making decisions regarding AHT; any reasons why participants' may feel confident or not i.e. the amount of experience or training they have had **Seeing 'the bigger picture':** any comments about having to piece together a 'jigsaw puzzle' of different types of evidence in order to understand the 'bigger picture'; any references to analyzing the different components of the investigation or considering a combination of different factors in order to reach a conclusion
Medical factors	**Clinical features:** any references to the clinical features that may influence decision-making such as bruising, fractures, burns or bites; any references to the 'triad' of injuries i.e. subdural hemorrhages, retinal hemorrhages and encephalopathy, any references to the medical literature or evidence-base around abuse-related injuries **Differential diagnoses:** any comments about differential diagnoses of AHT, or alternative explanations for injuries, e.g. accidental injury, or medical/genetic conditions such as bleeding disorders, osteogenesis imperfecta, glutaric aciduria etc. **Mechanisms of injury:** considerations of the manner or circumstance in which injuries may have occurred and how these considerations contribute to decision-making; any comments linking specific clinical features to possible injury mechanisms e.g. bruising associated with impact injuries, rib/chest injuries associated with compression forces **Severity of injuries:** comments regarding the severity of the injuries suffered by children as a factor affecting decision-making or the investigative process; perceptions of the seriousness of intracranial injuries in young children **Dealing with uncertainty:** any remarks about managing uncertainty in suspected AHT cases and how the degree of certainty impacts upon decision-making or the investigative/assessment process; discussions about so-called 'grey' cases where there is considerable uncertainty surrounding the diagnosis
Circumstantial factors	**Circumstances surrounding the incident:** discussions about the specific circumstances associated with the incident, including any witnesses to the event; details regarding the initial 999 call; examinations of the scene or surface where the incident purportedly occurred; comments about time to presentation at hospital including a delay in presentation; the behavior of the parents at the hospital or the scene and the parent-child interaction **History:** any discussions about the explanation for the child's injuries provided by the parents or carer, including whether the history given is consistent with the level of injury or the developmental stage of the child; or consistent across time and between caregivers
Family factors	**Social history:** any discussions regarding the social history of the family, including parental drug and alcohol use; parental mental health issues; domestic violence; previous involvement with social services; level of supervision of the child or previous history of neglect; socioeconomic status; and criminal history **Impact on the child/family:** any discussions regarding the impact that removing a child from the home or accusing a parent of AHT would have on the child and family **Working with the family:** anything relating to the challenges of working with the family during a suspected AHT case, and the need to act sensitively
Psychological factors	**Personal biases:** any remarks relating to disbelief or doubt that parents or carers are capable of inflicting injuries on their children; discussions of biases relating to the education level of the parents, socioeconomic or employment status, family structure or whether the family appears 'troubled' **Instinct:** any allusions to 'professional instincts' with regard to whether AHT has occurred, or instincts about a possible perpetrator, often referred to as a 'gut feeling' **Emotional factors:** any comments about the emotional or psychological impact of working on suspected AHT cases and how this may affect decision-making
Legal factors	**Identifying the perpetrator:** any comments about identifying a potential perpetrator in suspected AHT cases **Expert witnesses:** any discussions about working with or relying on expert witnesses; comments about disagreements between experts; remarks or interpretations about theories put forward by defense expert witnesses in an attempt to disprove cases

"We're not trained medically to know whether something's accidental or non-accidental. We can have an opinion on it, but it won't be based on research and training." CPSW 9

"I'm not a doctor, I'm not a medical expert, I'd want some clear guidance from the doctors about what they think, but having some experience now, no training, but some experience of dealing with these cases, I'm able to ask some relevant questions of the doctors." Police Officer 6

Judges rely on medical professionals to conduct a **timely and high quality clinical investigation** in suspected AHT cases, to facilitate the decision-making process in a court environment. In addition there is an expectation from legal practitioners and CPSWs that medical professionals are able to categorically determine the cause of the child's injuries by the clinical features alone.

"We have experience of saying to the medics to pin down to an absolute, 'this is the way it happened'…and they will always say…'I can't

Table 2
Demographics of child protection professionals participating in a qualitative study of decision-making in suspected abusive head trauma cases.

	Clinicians (N = 25)		CPSWs (N = 10)		Legal Practitioners (N = 9)		Police Officers (N = 8)		Pathologists (N = 4)	
	n	%	n	%	n	%	n	%	n	%
Gender										
Female	16	64	7	70	7	78	3	37.5	0	0
Male	9	36	3	30	2	22	5	62.5	4	100
Age group										
25–34	2	8	2	20	2	22	0	0	1	25
35–44	11	44	5	50	1	11	5	62.5	1	25
45–54	8	32	1	10	4	45	3	37.5	1	25
55–64	4	16	2	20	2	22	0	0	1	25
Ethnicity										
White British	19	76	10	100	8	89	8	100	4	100
White Other	4	16	0	0	1	11	0	0	0	0
Indian	2	8	0	0	0	0	0	0	0	0
Years in CP										
< 5	0	0	2	20	1	11	3	37.5	0	0
5–9	6	24	3	30	1	11	2	25	2	50
10–20	7	28	4	40	4	45	3	37.5	0	0
> 20	12	48	1	10	3	33	0	0	2	50
CP training										
Yes	25	100	10	100	3	33	7	87.5	4	100
No	0	0	0	0	6	66	1	12.5	0	0
Pediatric HI training										
Yes	18	72	1	10	3	33	4	50	3	75
No	7	28	9	90	6	66	4	50	1	25

CPSWs = child protection social workers, CP = child protection, HI = head injuries.

tell you that'." Legal Practitioner 1

However, clinicians and pathologists highlighted that other professional groups shouldn't be **relying solely on them** to come to a decision about suspected AHT.

"What I do with the police with these cases is actually say to them look, you shouldn't be relying on me. A case depends on lots of different bits of evidence." Pathologist 3

Many CPSWs, police officers and legal practitioners, including judges, said that decision-making is more difficult when medical professionals are unable to provide them with a **clear answer** as to whether AHT has occurred or not, or when they will not commit to a view either way.

"Often if there are clear injuries and the medics are actually saying that it is non-accidental then there is a clear process for us to follow. That makes it easier, it makes it a lot harder when health professionals are sitting on a fence." CPSW 3

However, CPSWs and police officers explained that the majority of the time medical professionals will **express their suspicions** to other agencies, which facilitates the investigative process, and if clinicians remain unsure, they would continue their investigations regardless.

"I mean, 9 times out of 10 it's fairly self-evident. I was able to have generally an open and honest discussion with the pediatrician, that pediatrician would say, 'In my opinion, this is what you've got. Either it's non-accidental or I'm concerned its non-accidental'. In which case they're both dealt with in pretty much the same way and investigated appropriately. It's fairly straight forward...But if they say, 'I'm unsure'. Then we still run with it anyway." Police Officer 3

"We usually get perhaps an initial medical report to say it's felt that these injuries are non-accidental...so initially you do get a concern that it is non-accidental." CPSW 6

Although clinicians do rely on other agencies to assist them in making decisions in suspected AHT cases, they seek **support and advice** from clinical colleagues to a greater extent.

"We can always speak to colleagues and we're never in it by ourselves...we're always in discussion with colleagues. I have never been in the situation where I've said, 'Right, I'm the only one making that call.' You're always discussing it with other people and so that certainly lessens the burden on you when you have to make those decisions. Even though you may be the person who is called to go to a strategy meeting or the case conference or actually to court you will have had those discussions with colleagues as well...you become confident because we reassure each other that's the diagnosis." Clinician 5

CPSWs and police officers overall have had positive experiences of **multidisciplinary collaboration**.

"We had, and we still do have, very good working relationships with health professionals, with pediatricians in particular." Police Officer 3

"Generally I find it's quite positive working with other agencies around safeguarding children." CPSW 2

"I think we've generally got a great relationship with the forensic pathology team." Police Officer 7

However, a handful of participants identified multidisciplinary working as a barrier to decision-making due to competing interests and disagreements between professionals both within and across agencies.

"Working with other agencies [is difficult] really, sometimes coming from competing backgrounds and also from here even you know, decision-making by managers is not always the same it can be varied." CPSW 3

CPSWs and police officers noted that **delays** can occur while the other agencies are carrying out their own assessments, which impacts on the overall investigation.

"We're guided a lot by medical staff; waiting for their statements to come through…It can take a long time. It can take months sometimes, you get an initial report but it's very much really not until towards the end where you really know what you've got." Police Officer 5

One influential factor for police officers was the amount of **resources** they are able to put into an investigation, reporting that more severe cases are better resourced.

"Do you say, 'This is definitely non-accidental'? In which case you're going to put a lot more resources in it. Or is it one of those really difficult ones to gauge and you know you're not going to have the budget to do absolutely everything." Police Officer 2

For some participants, coming to a decision about whether AHT has occurred is **extremely difficult**, while for others it is not.

"I think it is probably the most stressful and difficult set of decisions that one has to make in medicine. Partly due to the difficulty of coming to conclusions." Pathologist 3

"It's not that it's difficult, I think it just needs a lot of consideration and a lot of thought and weighing of the evidence rather than the actual decision being difficult. If the evidence is strong enough, I think the decision to be made is not that difficult." Clinician 6

Their **confidence** when investigating suspected AHT cases is strongly related to the amount of **experience** they have.

"I'm confident in dealing with the family, knowing my role, knowing the role of other professionals, but that might just be because I've been doing this for such a long time." Police Officer 5

"I'd say I was not very confident working on these cases, without a doubt, because I haven't worked on many physical abuse cases." CPSW 9

In terms of professional decision-making strategies, participants' discussed the importance of 'seeing the bigger picture' in suspected AHT cases, and **piecing together the evidence** from various different sources.

"It's a bit like a jigsaw puzzle to put together a number of different pieces of evidence to see if you can get any closer to the truth." Judge 3

3.2. Medical factors

Clinicians and pathologists **refer to the literature and evidence-base** on the different types of injury seen in abusive and non-abusive trauma when investigating suspected AHT cases, which gives them more confidence in their decisions.

"I've been through a lot a lot of the literature about it…so that has helped me in feeling maybe more confident about these cases." Clinician 7

"First of all I see whether there is any injury and decide what sort of injury it is, whether it's a blunt force injury or sharp force injury etcetera and then the distribution of the injuries on the body, and then relate the distribution that I find with what I know about the literature on different patterns of injury for assault or accident, falls." Pathologist 1

Participants mentioned a range of **clinical features** they understood to be concerning for abuse, including intracranial injuries, encephalopathy, retinal hemorrhages, fractures, apnea, seizures, spinal injury, and evidence of external injury such as bruising, bites or burns. They also recognized that **specific patterns of injury** are suspicious for abuse, including posterior rib fractures, metaphyseal fractures, patterned bruising and certain distributions of intracranial and retinal hemorrhages. Some also mentioned that fractures or intracranial bleeds of **different ages** are indicative of previous abuse or multiple incidents of abuse.

"If we are beginning to be concerned about abusive head trauma we would get an eye examination, so the presence of any retinal hemorrhages would be corroborative evidence, but particularly multi-layer widespread dot, blot and flare hemorrhages, other evidence of intraocular bleeding…I would be expecting or might see multiple focal thin layer subdural hemorrhages in different brain compartments." Clinician 10

Some participants, particularly police officers, have a high suspicion of AHT when the **"triad"** of subdural hemorrhages, encephalopathy and retinal hemorrhages is present. However, importantly, these were not the only features that these participants considered when coming to a decision about AHT.

"Once you get the triad of injuries and everything else, if you've got some attending injuries that appear to be evidence of abuse, in my view that would fuel the theory that there has been abusive head trauma. So for example if I had a child present in hospital with a head injury and they had bruising elsewhere on their body, that would make me more concerned about the head injury because of the evidence of abuse elsewhere." Police Officer 3

One CPSW alluded to the **"triad"** of injuries, demonstrating a **lack of training** on the clinical indicators of AHT, an outdated view of the features of the "triad" as diagnostic for AHT, and a lack of knowledge of the potential differential diagnoses of retinal hemorrhages.

"We haven't had proper training on this…I had training when I was studying my degree, but I had it drummed into me that if there's subdural hemorrhaging, retinal hemorrhage, it's abuse. Am I right in thinking that there can be no other organic cause for retinal hemorrhage?" CPSW 1

Participants emphasized the importance of ruling out organic medical conditions in children with suspected AHT, listing a variety of **differential diagnoses** they would consider, including blood clotting disorders, birth trauma, and glutaric aciduria among others.

"We would need the bloods, we would want to be screening for a significant coagulation disorder you know these kids often have a coagulation disorder after the event, so you need to confirm whether the coagulation disorder returns to normal after the child has been resuscitated. You would probably want to go back and re-examine the child looking for evidence of connective tissue disorders, you would want to review the family history, is there anybody in the family with a coagulation or connective tissue disorder." Clinician 10

One pathologist pointed out that **abuse can still occur** even when a child has an organic condition.

"I think that sometimes it is forgotten that even with natural pathology, it doesn't preclude there being something deliberate to go with it." Pathologist 2

Similarly, participants' often **link the injuries to the mechanism** purported by the care-givers, and deliberate over the plausibility of this.

"If I saw a head injury where there was not obviously impact, I would be looking for corroboration of the application of force somewhere, because as soon as you find a bruise or a fracture, or a graze or a split at whatever level in the body skin, soft tissues, the skeleton, you have what is undeniably the application of force and that helps you…the pattern of rib fractures indicates that there has been compression of the chest…the head injury may indicate impact or deceleration." Pathologist 3

"They could be toddling couldn't they, if they fell down the stairs from the top floor to the bottom they could have a brain injury couldn't they depending on the floor downstairs, they might have got a stone floor…but you wouldn't expect to have the other stuff there." CPSW 8

A clear factor influencing participants' decision-making is the **severity of the injuries** sustained. An intracranial bleed or rib fracture in a young child are viewed as serious injuries, and the more clinical features a child has, or the more impaired they are, the more likely participants' are to suspect AHT.

"My decision really would be based on the fact that I think you're talking about trauma here, a brain injury, if that's where we're looking at it, it's a really serious condition isn't it? So you've got to do everything you can to make sure that child is safe, and there's no risk that this could happen again." CPSW 9

Participants discussed dealing with the inherent **uncertainty in medicine**, and in the child protection arena in particular, stating that so-called 'gray' cases, where there is considerable uncertainty surrounding a diagnosis of AHT, are the most difficult.

"Medicine is rarely black and white, there are shades of gray in the middle of it, and often these are quite dark gray. You're pretty certain, there was a while where I felt slightly less sure, but most of them I felt reasonably confident, given if there's that constellation of injury, in the absence of an adequate explanation." Clinician 17

"Regularly in child protection we find ourselves in a situation where it's not 100% clear the person's guilty or the offence has happened, neither is it 100% clear that there is an innocent explanation, and left this gray area in between. Well there's still a risk, something possibly, or probably happened, but we can't prove it, can't rule it out, so where do we go with that…you know the gray area ones." Police Officer 6

It is hard for clinicians to convey to other agencies that a case **may not be clear-cut**.

"We are often trying to explain things to people who don't necessarily understand the uncertainty in medicine like police and social workers." Clinician 11

3.3. Circumstantial factors

Participants discussed the specific **circumstances surrounding the incident** in suspected AHT cases, including the explanation given for the child's injuries and details of their presentation to the hospital.

"What I'd be looking to do is looking at the accounts that have been given to the attending officers, the accounts given to the paramedics, what's been said on the 999 call [emergency number], what's been said when they first attend, usually they speak to the Accident & Emergency pediatrician, what they then say to the community pediatrician and my officer when they get there." Police Officer 1

The single most important factor that influences professionals' decision-making when AHT is suspected is the **history**, in particular whether the **mechanism of injury** is consistent with the type and severity of the injuries or the developmental stage of the child.

"Probably the single most important thing, is the detailed history. And the features of that, the description of what's happened, how possible and plausible that is, is it compatible with the injuries, is the child capable of what's described in terms of their developmental stage? So I think the detail of the history is really, really important." Clinician 12

However, participants find decision-making difficult when the history is consistent with the injuries, but the mechanism could nevertheless be **either abusive or accidental**.

"The difficult ones are where they come in and say 'I've fallen down the stairs with my baby' because you think if somebody has just lost it with a baby and smacked them against a wall, and is switched on and intelligent and actually quite manipulative, they probably would come up with a story of 'I've just dropped my baby'. So those ones are always a bit more difficult because you think it might be true, on the other hand it might not be." Clinician 13

When there is **no history of a traumatic event** whatsoever, participants have a very high index of suspicion for AHT.

"The lack of disclosure is a biggie, you know the child who presents with collapse and then you subsequently find that they have subdural hemorrhages or a fracture or broken ribs, that makes you very concerned that it's not the whole story being told to you." Clinician 24

Another influential factor is whether **the history is consistent** over time/between caregivers.

"I might be wanting to speak to a nurse, so 'You spoke to the parents when they first arrived, now you and the pediatrician have spoken to the parents here', is there any discrepancy between the two stories, or are they consistent, or is mum saying one thing and dad saying another thing?" Police Officer 6

A **delay in presentation** to hospital is an important risk factor for some participants.

"The other thing we always worry about is a delay. So we have had the odd few children that have presented a few days later because of a significant swelling and while that is possible that would raise a flag in your head." Clinician 16

Participants discussed the **behavior and appropriateness of the parents** and the interactions between the parent and the child throughout the investigation.

"We look at exploring the family dynamics, the response of the parents during the immediate child protection enquiry, the interview."

CPSW 1

3.4. Family factors

Participants discussed the importance of the families' **social history** when investigating suspected AHT cases. They talked about a wide range of issues including parental drug and alcohol use; parental mental health; domestic violence; previous involvement with social services; parent-child interactions; level of supervision of the child; neglect; socioeconomic status; and parental criminal history.

"I would be concerned if there was also then a family history of family violence, if I was getting background social history that there was known abuse in the past, or I guess if this baby had been more vulnerable for whatever reason, was maybe a pre-term or indeed if this was a mother who's quite young, not supported, new partner, and partner's not the biological father of this baby. They are things that I would actually…they'd be helping with the diagnosis. It wouldn't necessarily tip it but they would obviously add to my concern that my feeling is this is likely to be the case." Clinician 2

"Obviously if there's domestic violence, substance abuse, a history of neglect, that's obviously going to shoot up in terms of our assessment." CPSW 1

CPSWs and police officers place more emphasis on the **family setup** than the clinical factors.

"I would probably have a better understanding of the context in terms of the family scenario, levels of supervision and what it's actually like within the household." CPSW 5

However one police officer mentioned that he would give **less weight** to the social history of the family during his investigation.

"The social background is less important, because if we're investigating if there's been abuse or not, it's determined principally by the injuries, by an explanation, by the evidence, not by whether the parents are employed, whether the parents are smokers, or if the dad is an alcoholic, whether there has been domestic violence in the house, those are interesting background features, perhaps more likely to be prevalent in some cases than others, but it's not going to tell me abuse has happened or hasn't happened." Police Officer 6

Similarly, participants pointed out that a **lack of history with social services** or a lack of a criminal record does not rule out AHT.

"It's not always families that come revolving door, we have families that have not been known to us for years, or never been known and they've harmed a child." CPSW 4

A major factor influencing participants' decision-making in suspected AHT cases is the **impact on the family**. They discussed the impact of removing a child from the family home, and how intervening in a child's home life could be damaging for the child and family, particularly where a head injury is found to be non-abusive. The decision to remove a child from their parents is not taken lightly, as it may not be the best thing for the child.

"I know accidents happen with babies and children with the best will in the world and what you don't want to do is if a family is already traumatized by something that the child has experienced and they're doing the best for them, to add in the trauma of querying the abuse factor could just tear the family apart." CPSW 9

Clinicians, CPSWs and police officers find it difficult **working with the family** and having to treat parents as potential suspects or perpetrators when they are grieving or coping with a seriously unwell child. Participants talked about the need for sensitivity and the potential repercussions of falsely accusing a family of abuse.

"It's not so much the clinical diagnosis it's managing it and being the one who talks to the parents and is having to deal with their anxieties, their uncertainties and all their anger." Clinician 10

3.5. Psychological factors

Participants' decision-making in suspected AHT cases is influenced by their own **personal biases**, such as a disbelief that parents or care-givers from 'nice, middle-class families' are capable of inflicting injuries on their children.

"Well they shouldn't but if it looks like a really nice family that you couldn't imagine doing anything like that and that shouldn't influence you but it makes you think. People say 'Oh I've seen a case like this before' or they say 'No, no the family is too nice'. And other people will be saying 'But don't be fooled by it', all this goes on, I hear it all the time." Clinician 9

However, most participants **acknowledge these biases** and attempt to remain objective in their assessments.

"We always keep an open mind, we always continue to gather information and if there is new information, it will change our decisions." Clinician 10

CPSWs and judges find cases difficult when they only have medical evidence to rely on, and there are **no other risk factors** that they are able to identify within the family.

"There have been cases where we've removed children begrudgingly because of medical evidence and genuinely from the way the parents are with the child, their backgrounds you just don't think they did it. So that's very difficult ethically having to remove a child on the basis of a medical decision where there's nothing else to substantiate that." CPSW 1

However this CPSW also stressed that even in the **absence of other risk factors**, she would remain suspicious and continue with her investigations.

"I would be led very much by medical evidence and even if there was no other risk factors identified for that child, I would not be willing to take any risk on a case like that." CPSW 1

Clinicians, CPSWs, and particularly police officers are influenced by their "**gut instincts**" when conducting their investigations and assessments.

"I'd probably rely on my professional suspicion…my gut feeling…If I had an inkling something was not right then we would be doing more." Police Officer 3

Investigating suspected AHT is **emotionally demanding** and can be a barrier to remaining objective in these cases.

"There is emotion attached to them, so seeing children who are injured whether it is accidentally or deliberately, there's an emotional component to that. I find it difficult because I am intrinsically a relatively trusting and non-suspicious person and I've had to train myself to just take the emotion out of it, and deal with whatever facts are available." Clinician 12

3.6. Legal factors

Police officers and legal practitioners disclosed that **identifying the perpetrator** in suspected AHT cases is particularly difficult.

"The difficulty in my experience isn't identifying it, it's in establishing who's done it." Police Officer 3

Legal practitioners and especially judges, rely on **expert witnesses** to provide an interpretation of the clinical features, but noted that there are often disagreements and conflicting opinions between expert witnesses coming from different disciplines.

"There will sometimes be subtleties, particularly in the expert evidence that we get and you will have two extremely eminent experts sometimes from different disciplines, sometimes the neurosurgeon has a different view from the radiologist. I can remember doing a case in which they'd looked at the same scan and said I don't think we can agree what's there…So those are the difficulties that you have to encounter when you get a range of opinion on the interpretation of the medical evidence." Judge 3

Judges and police officers referred to the **various theories** that are put forward by the parents or the defense in an attempt to disprove cases of suspected AHT.

"Something needs to be looked at because this hasn't happened because the child has got gastro-esophageal reflux, which was one theory which used to be propagated at one stage because if a child had gastro-esophageal reflux it might stop breathing and that would lead to a rise in intracranial pressure which would then give rise to the bleed and we had that theory at one stage, not from the medics but that was one that was often propagated." Judge 3

4. Discussion

The findings from this study suggest that child protection professionals diagnose AHT based on knowledge of a wide range of clinical features described in the literature, features in the history, and risk factors within the family, after exclusion of potential differential diagnoses and discussion with colleagues from other specialties and disciplines. Barriers to identifying AHT included lack of experience, uncertainty, emotional factors, personal biases, the impact on the family and the fear of making an incorrect diagnosis, disagreements between professionals including expert witnesses, and alternative theories of causation proposed in court. Participants' experiences with multidisciplinary collaboration were reported as generally positive, however CPSWs and police officers reported being heavily reliant on clinicians to guide their decision-making, due to their own lack of medical training and knowledge. Facilitators to identifying AHT include support from colleagues, multidisciplinary working, knowledge of the literature and evidence-base, and "gut instinct". The strengths of this study lie in the wide range of professionals interviewed, the detail and depth of the data, and the robustness of the data analysis. Survey-based methods do not allow for such a detailed exploration of participants views (Fontana & Frey, 1994).

The results are consistent with barriers and facilitators influencing detection of physical abuse generally, and clinicians' decisions to report suspected abuse identified in previous studies. Flaherty et al. (2008) found that the decision to report suspected abuse was primarily influenced by the child's clinical and social history and physical examination findings, particularly if their injuries were inconsistent with the history or their developmental stage. Barriers to detecting (Regnaut, Jeu-Steenhouwer, Manaouil, & Gignon, 2015) and reporting (Gunn et al., 2005) abuse described previously include personal biases, the fear of being wrong and the subsequent impact on the family, uncertainty about the level of suspicion and the difficulty of establishing a diagnosis, while facilitators include support from colleagues and other agencies.

If a comprehensive evaluation reveals no other medical explanation for the child's injuries, clinicians must decide whether the injuries are accidental or abusive (Leventhal et al., 2014; Narang, 2011). In determining this, participants reported that one of the most important factors influencing their decision-making is whether the history of the mechanism of injury is consistent with the type and severity of the injuries seen, or the developmental stage of the child. This approach has been deemed both medically and legally valid (Narang, 2011), having been first described in a landmark article on the diagnosis of "battered child syndrome" (Kempe, Silverman, Steele, Droegemueller, & Silver, 1962).

Our findings clearly refute the claims of some recent literature that AHT is diagnosed based on the "triad" alone (Elinder et al., 2016; Lynøe et al., 2017) and echo the categorical statements made by experienced clinicians who do not diagnose AHT solely on the presence of the "triad" (Levin, 2017; Lucas et al., 2017; Ludvigsson, 2017; Narang & Greeley, 2017; Saunders et al., 2017). The misconception was the subject of a meeting convened by the Royal College of Pathologists in 2009 to consider the issues appertaining to the "triad" and the "unified hypothesis" in non-accidental head injury cases, following which legal guidance was issued from the UK Crown Prosecution Service (2011) on the prosecution approach to non-accidental head injury. This states that "the expert evidence finding of typical triad pathological features might **not be considered as diagnostic in itself** but simply as **strong evidence** that the injuries were non-accidental" (emphasis added). This view was reflected by clinicians and police officers in the current study. However, one senior CPSW described being taught at undergraduate level that the features of the "triad" are diagnostic for AHT. Although this may have been some time ago, this highlights how misconceptions become established, the differences between agencies and training gaps for social worker education in the clinical indicators and differential diagnoses of AHT, and suggests that their training should be regularly updated in line with the evolving evidence-base.

An important issue influencing child protection professionals' decision-making in suspected AHT cases is the proposal of scientifically unsupported alternative theories of causation for AHT in court. Judges and police officers alluded both to genuine diagnoses that lack scientific evidence to explain the injuries associated with AHT (e.g. Vitamin D deficiency) and unproven speculative hypotheses with no scientific evidence-base (e.g. that gastro-esophageal reflux causes intracranial pressure leading to intracranial hemorrhage). The use of these flawed theories has created controversy in the courtroom and the media regarding the diagnosis of AHT, and has serious consequences for the upholding of justice and the protection of children (Leventhal & Edwards, 2017). Several authors have suggested potential remedies for ensuring responsible expert medical testimony in AHT cases (Albert, Blanchard, & Knox, 2012; Holmgren, 2013; Leventhal & Edwards, 2017). Albert et al. (2012) recommended a comprehensive authoritative study of the strength of the medical evidence for AHT and the accuracy of AHT testimony, as well as tailored certification programs for medical professionals called upon to testify in court. Recently a consensus statement on AHT based on a thorough and comprehensive review of the literature and evidence-base was published, supported by nine pediatric and radiology international organizations, with the intention of helping jurors and judges to distinguish between "genuine evidence-based opinions of the relevant medical community from legal arguments or etiological speculations that are unwarranted by the clinical findings, medical evidence and evidence-based literature" (Choudhary et al., 2018). Leventhal and Edwards (2017) urge academic medical centers and professional societies to set standards for medical testimony in AHT cases, while Holmgren (2013) recommends peer review and quality control by responsible experts and disciplinary action against irresponsible and unethical experts.

Some participants stated that they are sometimes influenced by their "gut feeling" when investigating suspected AHT cases. "Gut feeling" has been defined as an intuition that something is wrong even in the absence of specific clinical indicators, or a sense of reassurance about a patient's condition and management in the absence of a definitive diagnosis (Stolper et al., 2011). The evidence-based medicine literature generally advises doctors against the use of intuitive reasoning, in order to avoid errors resulting from cognitive biases (Croskerry, 2003), and instead promotes the use of analytical models, clinical guidelines and decision tools (Sackett, Richardson, Rosenberg, & Haynes, 1997). A recent study demonstrated that child abuse pediatricians who had met the family and therefore had access to social intuition or "gut feelings" associated with a face-to-face encounter, were significantly less likely to perform adequate abuse evaluations for neuro-trauma and long-bone fracture compared to those who had not met the family (Keenan, Cook, Olson, Bardsley, & Campbell, 2017). However, studies have shown that intuition can outperform analytical reasoning in diagnostic decision-making (Dhaliwal, 2011), and that "gut feelings" may trigger the process of diagnostic reasoning, prompting clinicians to perform further investigations (Stolper et al., 2009). One qualitative study exploring the identification and management of child abuse found that Dutch healthcare professionals' intuitive "gut feelings" often formed the basis of a more objective investigation and triggered a systematic process of evidence gathering (Schols, de Ruiter, & Öry, 2013). Dhaliwal (2011) recommends that clinicians adhere to the principles of evidence-based medicine while also understanding when it is appropriate to "go with their gut".

The participants in this study did report that their decision-making in suspected AHT cases is influenced by their personal biases and emotions. However, that they are aware of these biases and their potential pitfalls is encouraging, as it provides opportunities for monitoring, reflection and deliberative efforts to minimize their negative effects (Laskey, 2014). Participants described the application of strategies recommended in the literature to avoid errors resulting from bias, including attempting to remain objective, consciously considering differential diagnoses, and collaborating with multidisciplinary colleagues (Laskey, 2014).

While many studies have evaluated the relationship between law enforcement and child protective services in suspected child abuse cases (e.g. Cross, Finkelhor, & Ormrod, 2005; Sedlak et al., 2006; Newman & Dannenfelser, 2005), comparatively few studies have assessed health professionals' perceptions of multidisciplinary working. Previous studies have described a hostile relationship between police officers and CPSWs, due to conflicting priorities and agendas, assumptions regarding the other's role, and time delays (Newman & Dannenfelser, 2005). Clinicians have also criticized social workers, describing them as unresponsive or inconsistent (Regnaut et al., 2015). In contrast, the majority of the participants in the current study described positive relationships and experiences with other agencies. Only a very small number of participants felt that multidisciplinary working is difficult due to competing interests and disagreements, while a handful of participants brought up time delays as significant barriers to the investigation. Overall, participants' views of multidisciplinary working indicated that police officers and CPSWs consider AHT to be a medical diagnosis, and are heavily reliant on clinicians decisions; many believe that medics can determine the cause of injures by clinical features alone, and it is difficult for other agencies when clinicians "sit on the fence". Conversely, clinicians find it difficult to convey medical uncertainty to other agencies. This finding echoes the results of a recent study exploring collaboration between pediatricians and CPSWs, which demonstrated that CPSWs rely on pediatricians' opinions regarding accidental and abusive bruising, but that pediatricians felt CPSWs harbored unrealistic expectations about the diagnostic value of a child protection medical examination to identify abusive bruising (Matthews, Kemp, & Maguire, 2017). Indeed, in some suspected AHT cases, the diagnosis may remain uncertain even after a thorough clinical and forensic investigation (Kelly, John, Vincent, & Reed, 2015; Leventhal et al., 2014). Clearly, joint training that provides knowledge about the individual roles and limitations of each agency would be valuable. In addition, clinicians should be prepared to provide a clear opinion about the likelihood of AHT to their non-medical colleagues and ensure that they have the necessary skills and experience required for the evaluation of children with suspected AHT (Christian & Committee on Child Abuse & Neglect, 2015). Interestingly, the current study found that pathologists defer to pediatricians to diagnose AHT; this may differ in the United States, where pathologists occupy a critical role in the medicolegal evaluation of AHT (Holmgren, 2013).

Participants discussed a range of social risk factors within the families, usually regarded as facilitators to reaching a decision about AHT. However, some felt that these factors impeded their decision-making, since a family without risk factors could be abusive, while a family with multiple risk factors may never harm their child. Previous research identified the presence of risk factors as a complicating factor in detecting child abuse for some clinicians (Schols et al., 2013), although a recent study found that children referred for abuse evaluations without certain risk factors were just as likely to be diagnosed with AHT as those with risk factors (Kelly et al., 2015).

Participants are more likely to suspect AHT and put greater resources into a case when the child's injuries are severe. However, it is well known that children can suffer repeated and escalated instances of abuse that eventually result in severe injury, and can sustain comparatively minor "sentinel" injuries such as isolated bruising or intra-oral injuries prior to a catastrophic injury (King, Kiesel, & Simon, 2006; Oral, Yagmur, Nashelsky, Turkmen, & Kirby, 2008; Petska, Sheets, & Knox, 2013; Sheets et al., 2013). Sheets et al. (2013) found that 30% of children diagnosed with AHT had previous sentinel injuries; where clinicians were aware of these injuries, either abuse was not suspected or was suspected but unsubstantiated, and their significance also went unrecognized by clinicians during the subsequent abuse evaluation.

4.1. Limitations

Most clinical participants were consultants based in teaching hospitals with considerable child protection experience; since participation was voluntary, these participants may have had a particular interest in the identification and investigation of AHT compared to other professionals who did not take part. The factors influencing decision-making and multidisciplinary working may be different for trainees or those with less experience. Similarly, results may have differed amongst other clinical subspecialties e.g. neurologists, intensivists, staff nurses or ophthalmologists, and only small numbers of specialists in radiology, neurosurgery and emergency medicine participated. In addition, the majority of the police officers had less than ten years child protection experience, while the majority of the legal practitioners had ten years or more child protection experience. However, qualitative research does not aim to make probabilistic generalizations to a population, but to arrive at logical, contextualized generalizations regarding the phenomenon under study (Polit & Beck, 2010). Further exploration of the factors influencing pathologists' decision-making in suspected AHT cases may be justified since only four pathologists participated and data saturation may not have been achieved with this group (Cowley et al., 2018). Our data represent the views and attitudes of professionals as recounted to the interviewer rather than observations of their practice, and participants may have felt obliged to give socially acceptable answers. Qualitative research inevitably relies on the researcher's interpretations, however we attempted to minimize subjective bias by using three trained qualitative researchers to double-code the data and resolve disagreements through discussion and consensus.

5. Conclusions

By directly seeking the views and practices of a wide range of child protection professionals investigating children with suspected AHT, we have contributed a deeper understanding of how these professionals make decisions and work together in these challenging cases. The findings contradict recent literature claiming that AHT is diagnosed based on the "triad" alone (Elinder et al., 2016; Lynøe et al., 2017). Rather, decision-making in AHT cases is complex and nuanced, and a diagnosis is arrived at only when all potential variables are carefully explored and considered, including clinical, historical, forensic and social features and potential differential diagnoses. The findings suggest that CPSWs and police officers may benefit from additional training in the medical aspects of physical abuse, and that joint training might provide a better understanding of the roles, expectations and limitations of each agency, thereby facilitating more effective collaboration.

Funding source

This work was supported by Health and Care Research Wales (grant number HS-14-24). The funders had no involvement in the study design, the collection, analysis, or interpretation of the data, the writing of the report or the decision to submit the article for publication.

Declaration of interest

The authors have no actual or potential conflicts of interest, or any financial, personal or other relationships relevant to this article to disclose.

Acknowledgments

We would like to thank Health and Care Research Wales for funding the study, and all of the professionals who took part.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.chiabu.2018.06.009.

References

Albert, D. M., Blanchard, J. W., & Knox, B. L. (2012). Ensuring appropriate expert testimony for cases involving the "shaken baby". *Journal of the American Medical Association, 308*(1), 39–40.
Anderst, J., Nielsen-Parker, M., Moffatt, M., Frazier, T., & Kennedy, C. (2016). Using simulation to identify sources of medical diagnostic error in child physical abuse. *Child Abuse & Neglect, 52*, 62–69.
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.
Bryman, A., & Burgess, R. (1994). *Analyzing qualitative data.* London: Routledge.
Choudhary, A. K., Servaes, S., Slovis, T. L., Palusci, V. J., Hedlund, G. L., Narang, S. K., et al. (2018). Consensus statement on abusive head trauma in infants and young children. *Pediatric Radiology*. http://dx.doi.org/10.1007/s00247-018-4149-1 Available from:.
Christian, C. W., & Committee on Child Abuse and Neglect (2015). The evaluation of suspected child physical abuse. *Pediatrics, 135*(5), e1337–e1354.
Cowley, L. E., Maguire, S., Farewell, D. M., Quinn-Scoggins, H. D., Flynn, M. O., & Kemp, A. M. (2018). Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals. *Child Abuse & Neglect, 81*, 192–205.
Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78*(8), 775–780.
Cross, T. P., Finkelhor, D., & Ormrod, R. (2005). Police involvement in child protective services investigations: Literature review and secondary data analysis. *Child Maltreatment, 10*(3), 224–244.
Crown Prosecution Service (2011). *Non accidental head injury cases (NAHI, formerly referred to as shaken baby syndrome [SBS]): Prosecution approach.* Available at:www.cps.gov.uk/legal/l_to_o/non_accidental_head_injury_cases/#a01.
Dey, I. (1993). *Qualitative data analysis.* London: Routledge.
Dhaliwal, G. (2011). Going with your gut. *Journal of General Internal Medicine, 26*(2), 107–109.
Dias, M. S., Boehmer, S., Johnston-Walsh, L., & Levi, B. H. (2015). Defining' reasonable medical certainty' in court: What does it mean to medical experts in child abuse cases? *Child Abuse & Neglect, 50*, 218–227.
Elinder, G., Eriksson, A., Hallberg, B., Lynöe, N., Rosén, M., & Sundgren, P. (2016). *Traumatic Shaking: The Role of the Triad in Medical Investigations of Suspected Traumatic Shaking. A Systematic Review.* Available atSwedish Agency for Health Technology Assessment and Assessment of Social Serviceshttp://www.sbu.se/en/publications/sbu-assesses/traumatic-shaking--the-role-of-the-triad-in-medical-investigations-of-suspected-traumatic-shaking/.
Flaherty, E. G., Sege, R. D., Griffith, J., Price, L. L., Wasserman, R., Slora, E., et al. (2008). From suspicion of physical child abuse to reporting: Primary care clinician decision-making. *Pediatrics, 122*(3), 611–619.
Flaherty, E. G., Sege, R., Price, L. L., Christoffel, K. K., Norton, D. P., & O'Connor, K. G. (2006). Pediatrician characteristics associated with child abuse identification and reporting: Results from a national survey of pediatricians. *Child Maltreatment, 11*(4), 361–369.
Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin, & Y. S. Lincoln (Eds.). *Handbook of qualitative research* (pp. 361–376). Thousand Oaks: Sage Publications.
Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13*, 117.
Gill, J. R., Goldfeder, L. B., Armbrustmacher, V., Coleman, A., Mena, H., & Hirsch, C. S. (2009). Fatal head injury in children younger than 2 years in New York City and an overview of the shaken baby syndrome. *Archives of Pathology & Laboratory Medicine, 133*, 619–627.
Glaser, B. G., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago, IL: Aldine Publishing Co.
Gunn, V. L., Hickson, G. B., & Cooper, W. O. (2005). Factors affecting pediatricians' reporting of suspected child maltreatment. *Ambulatory Pediatrics, 5*(2), 96–101.
Holmgren, B. K. (2013). Ethical issues in forensic testimony involving abusive head trauma. *Academic Forensic Pathology, 3*(3), 317–328.
Inkilä, J., Flinck, A., Luukkaala, T., Åstedt-Kurki, P., & Paavilainen, E. (2013). Interprofessional collaboration in the detection of and early intervention in child maltreatment: Employees' experiences. *Nursing Research and Practice,* 186414.
Jenny, C., Hymel, K. P., Ritzen, A., Reinert, S. E., & Hay, T. C. (1999). Analysis of missed cases of abusive head trauma. *Journal of the American Medical Association, 281*(7), 621–626.
Keenan, H. T., Cook, L. J., Olson, L. M., Bardsley, T., & Campbell, K. A. (2017). Social intuition and social information in physical child abuse evaluation and diagnosis. *Pediatrics, 140*(5), e20171188.
Kelly, P., John, S., Vincent, A. L., & Reed, P. (2015). Abusive head trauma and accidental head injury: A 20-year comparative study of referrals to a hospital child protection team. *Archives of Disease in Childhood, 100*(12), 1123–1130.
Kempe, C. H., Silverman, F. N., Steele, B. F., Droegemueller, W., & Silver, H. K. (1962). The battered-child syndrome. *Journal of the American Medical Association, 181*(1), 17–24.
King, W. K., Kiesel, E. L., & Simon, H. K. (2006). Child abuse fatalities: Are we missing opportunities for intervention? *Pediatric Emergency Care, 22*(4), 211–214.
Laskey, A. L. (2014). Cognitive errors: Thinking clearly when it could be child maltreatment. *Pediatric Clinics of North America, 61*(5), 997–1005.
Letson, M. M., Cooper, J. N., Deans, K. J., Scribano, P. V., Makoroff, K. L., Feldman, K. W., et al. (2016). Prior opportunities to identify abuse in children with abusive head trauma. *Child Abuse & Neglect, 60*, 36–45.
Leventhal, J. M., & Edwards, G. A. (2017). Flawed theories to explain child physical abuse – What are the medical-legal consequences? *Journal of the American Medical Association, 318*(14), 1317–1318.
Leventhal, J. M., Asnes, A. G., Pavlovic, L., & Moles, R. L. (2014). Diagnosing abusive head trauma: The challenges faced by clinicians. *Pediatric Radiology, 4*(Suppl. 4), S537–S542.
Levi, B. H., & Brown, G. (2005). Reasonable suspicion: A study of Pennsylvania pediatricians regarding child abuse. *Pediatrics, 116*(1), e5–e12.
Levin, A. V. (2017). The SBU report: A different view. *Acta Paediatrica, 106*(7), 1037–1039.
Lind, K., Toure, H., Brugel, D., Meyer, P., Laurent-Vannier, A., & Chevignard, M. (2016). Extended follow-up of neurological, cognitive, behavioral and academic outcomes after severe abusive head trauma. *Child Abuse & Neglect, 51*, 358–367.
Lucas, S., Bärtås, A., Edstedt Bonamy, A.-K., Törnudd, L., Wide, P., & Otterman, G. (2017). The way forward in addressing abusive head trauma in infants – Current perspectives from Sweden. *Acta Paediatrica, 106*(7), 1033.
Ludvigsson, J. F. (2017). Extensive shaken baby syndrome review provides a clear signal that more research is needed. *Acta Paediatrica, 106*(7), 1028.
Lynøe, N., Elinder, G., Hallberg, B., Rosén, M., Sundgren, P., & Eriksson, A. (2017). Insufficient evidence for 'shaken baby syndrome' – A systematic review. *Acta Paediatrica, 106*, 1021–1027.
Matthews, L., Kemp, A., & Maguire, S. (2017). Bruising in children: Exploring the attitudes, knowledge and training of child protection social workers and the interface with paediatricians regarding childhood bruising. *Child Abuse Review, 26*(6), 425–438.
Narang, S. A. (2011). Daubert analysis of abusive head trauma/shaken baby syndrome. *Houston Journal of Health, Law and Policy, 11*, 505–633.
Narang, S. K., & Greeley, C. S. (2017). Lynøe et al. - #theRestoftheStory. *Acta Paediatrica, 106*(7), 1047.
Newman, B. S., & Dannenfelser, P. L. (2005). Children's protective services and law enforcement: Fostering partnerships in investigations of child abuse. *Journal of Child Sexual Abuse, 14*(2), 97–111.
Oral, R., Yagmur, F., Nashelsky, M., Turkmen, M., & Kirby, P. (2008). Fatal abusive head trauma cases: Consequence of medical staff missing milder forms of physical abuse. *Pediatric Emergency Care, 24*(12), 816–821.
Petska, H. W., Sheets, L. K., & Knox, B. L. (2013). Facial bruising as a precursor to abusive head trauma. *Clinical Pediatrics, 52*(1), 86–88.
Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies, 47*(11), 1451–1458.
QSR International Pty Ltd (2014). *NVivo qualitative data analysis software; Version 10*.
Regnaut, O., Jeu-Steenhouwer, M., Manaouil, C., & Gignon, M. (2015). Risk factors for child abuse: Levels of knowledge and difficulties in family medicine. A mixed method study. *BMC Research Notes, 8*, 620.

Rorke-Adams, L. B. (2011). The triad of retinal haemorrhage, subdural haemorrhage and encephalopathy in an infant unassociated with evidence of physical injury is not the result of shaking, but is most likely to have been caused by a natural disease: No. *Journal of Primary Health Care, 3*(2), 161–163.
Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1997). *Evidence-based medicine: How to practice and teach EBM.* New York: Churchill Livingstone.
Saunders, D., Raissaki, M., Servaes, S., Adamsbaum, C., Choudhary, A. K., Moreno, J. A., et al. (2017). Throwing the baby out with the bath water – Response to the Swedish agency for health technology assessment and assessment of social services (SBU) report on traumatic shaking. *Pediatric Radiology, 47*(11), 1386–1389.
Schols, M. W. A., de Ruiter, C., & Öry, F. G. (2013). How do public child healthcare professionals and primary school teachers identify and handle child abuse cases? A qualitative study. *BMC Public Health, 13,* 807.
Sedlak, A. J., Schultz, D., Wells, S. J., Lyons, P., Doueck, H. J., & Gragg, F. (2006). Child protection and justice systems processing of serious child abuse and neglect cases. *Child Abuse & Neglect, 30*(6), 657–677.
Sheets, L. K., Leach, M. E., Koszewski, I. J., Lessmeier, A. M., Nugent, M., & Simpson, P. (2013). Sentinel injuries in infants evaluated for child physical abuse. *Pediatrics, 131*(4), 701–707.
Squier, W. (2011). The triad of retinal haemorrhage, subdural haemorrhage and encephalopathy in an infant unassociated with evidence of physical injury is not the result of shaking, but is most likely to have been caused by a natural disease: Yes. *Journal of Primary Health Care, 3*(2), 159–161.
Stolper, E., van Bokhoven, M., Houben, P., Van Royen, P., van de Wiel, M., van der Weijden, T., et al. (2009). The diagnostic role of gut feelings in general practice. A focus group study of the concept and its determinants. *BMC Family Practice, 10,* 17.
Stolper, E., Van de Wiel, M., Van Royen, P., Van Bokhoven, M., Van der Weijden, T., & Jan Dinant, G. (2011). Gut feelings as a third track in general practitioners' diagnostic reasoning. *Journal of General Internal Medicine, 26*(2), 197–203.
Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349–357.
Wood, J. N., Hall, M., Schilling, S., Keren, R., Mitra, N., & Rubin, D. M. (2010). Disparities in the evaluation and diagnosis of abuse among infants with traumatic brain injury. *Pediatrics, 126*(3), 408–414.
Wood, J. N., Feudtner, C., Medina, S. P., Luan, X., Localio, R., & Rubin, D. M. (2012). Variation in occult injury screening for children with suspected abuse in selected US children's hospitals. *Pediatrics, 130*, 853–860.

Child Abuse & Neglect 81 (2018) 192–205

Contents lists available at ScienceDirect

Child Abuse & Neglect

journal homepage: www.elsevier.com/locate/chiabuneg

Research article

Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals

Laura E. Cowley*, Sabine Maguire, Daniel M. Farewell, Harriet D. Quinn-Scoggins, Matthew O. Flynn, Alison M. Kemp

Division of Population Medicine, School of Medicine, Cardiff University, Wales, United Kingdom

ARTICLE INFO

Keywords:
Child physical abuse
Abusive head trauma
Clinical prediction tool
Qualitative research
Child protection

ABSTRACT

The validated Predicting Abusive Head Trauma (PredAHT) tool estimates the probability of abusive head trauma (AHT) based on combinations of six clinical features: head/neck bruising; apnea; seizures; rib/long-bone fractures; retinal hemorrhages. We aimed to determine the acceptability of PredAHT to child protection professionals. We conducted qualitative semi-structured interviews with 56 participants: clinicians (25), child protection social workers (10), legal practitioners (9, including 4 judges), police officers (8), and pathologists (4), purposively sampled across southwest United Kingdom. Interviews were recorded, transcribed and imported into NVivo for thematic analysis (38% double-coded). We explored participants' evaluations of PredAHT, their opinions about the optimal way to present the calculated probabilities, and their interpretation of probabilities in the context of suspected AHT. Clinicians, child protection social workers and police thought PredAHT would be beneficial as an objective adjunct to their professional judgment, to give them greater confidence in their decisions. Lawyers and pathologists appreciated its value for prompting multidisciplinary investigations, but were uncertain of its usefulness in court. Perceived disadvantages included: possible over-reliance and false reassurance from a low score. Interpretations regarding which percentages equate to 'low', 'medium' or 'high' likelihood of AHT varied; participants preferred a precise % probability over these general terms. Participants would use PredAHT with provisos: if they received multi-agency training to define accepted risk thresholds for consistent interpretation; with knowledge of its development; if it was accepted by colleagues. PredAHT may therefore increase professionals' confidence in their decision-making when investigating suspected AHT, but may be of less value in court.

1. Introduction

Abusive head trauma (AHT) has severe consequences for young children, and may be missed in the clinical setting (Jenny, Hymel, Ritzen, Reinert, & Hay, 1999; Letson et al., 2016). It is vital that AHT is accurately identified, to prevent further abuse from occurring and reduce the risk of falsely accusing innocent care providers. The evaluation and investigation of suspected AHT requires a multidisciplinary team approach whereby pediatricians collaborate with clinicians from other specialities, child protection social workers (CPSWs), law enforcement, and the justice system (Canadian Paediatric Society, 2007). Recent guidelines for the evaluation

* Corresponding author.
E-mail addresses: CowleyLE@cardiff.ac.uk (L.E. Cowley), sabinemaguire@gmail.com (S. Maguire), FarewellD@cardiff.ac.uk (D.M. Farewell), Quinn-ScogginsHD@cardiff.ac.uk (H.D. Quinn-Scoggins), mattandflynn@gmail.com (M.O. Flynn), KempAM@cardiff.ac.uk (A.M. Kemp).

https://doi.org/10.1016/j.chiabu.2018.04.022
Received 27 February 2018; Received in revised form 24 April 2018; Accepted 27 April 2018

Table 1
The six features included in the Predicting Abusive Head Trauma clinical prediction tool.

Feature	Description
Head or neck bruising	Any documented bruising to head or neck
Seizures	Any documented seizures from a single seizure to status epilepticus
Apnea	Any apnea documented in the initial history or during inpatient stay
Rib fracture	Any rib fracture documented after appropriate radiologic imaging
Long-bone fracture	Any long-bone fracture documented after appropriate radiologic imaging
Retinal hemorrhage	Any retinal hemorrhage documented after indirect ophthalmologic examination by a pediatric ophthalmologist

Published previously in Pediatrics (Cowley et al., 2015) and reproduced here with permission.

of suspected physical abuse recommend that medical records include a clear opinion about the likelihood of physical abuse and should elucidate specific levels of concern to aid police and CPSWs' investigations (Christian, Committee on Child Abuse and Neglect, & American Academy of Pediatrics, 2015), so that they can gauge the degree of certainty of AHT in each case (Canadian Paediatric Society, 2007).

Following a systematic review (Maguire et al., 2009), derivation (Maguire, Kemp, Lumb, & Farewell, 2011) and external validation studies (Cowley, Morris, Maguire, Farewell, & Kemp, 2015), we developed the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool (CPT). The derivation study used multivariable logistic regression and provided predicted probabilities of AHT for children (< 3 years old) with intracranial injury and each of 64 possible combinations of the presence or absence of six clinical features, detailed in Table 1 (Maguire, Kemp et al., 2011). PredAHT performed with a sensitivity of 72.3% and a specificity of 85.7% in the validation study, using a cut-off probability value of 50% (Cowley et al., 2015). However, in the clinical setting some investigations may not be undertaken to identify key features e.g. X Rays or ophthalmology for fractures or retinal hemorrhages respectively. Using data from the derivation study (Maguire, Kemp et al., 2011) we therefore estimated the probability of AHT when one or more features were unknown using multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011). PredAHT thus estimates the probability of AHT for all 729 permutations of the six clinical features depending on whether each is present, absent *or unknown*. A computerised version was built using Shiny, a web application framework for the R language and environment for statistical computing (Chang, Cheng, Allaire, Xie, & McPherson, 2015; R Core Team, 2015).

PredAHT can assist professionals when evaluating possible cases of AHT and contribute to decision-making at multiple stages of the assessment pathway, by demonstrating how combinations of clinical features combine to estimate different probabilities of AHT. A crucial aspect of CPT development is determining its acceptability to the population for whom it is intended (Brehaut et al., 2010; Reilly & Evans, 2006; Stiell & Wells, 1999). Even valid and reliable CPTs may not be accepted or used (Stiell & Wells, 1999). If a CPT proves to be acceptable in addition to demonstrating a positive impact on patient care, its long-term and widespread dissemination and implementation would be justified; if not, the CPT could undergo modification and further evaluation (Brehaut et al., 2010). There is a need to better understand whether PredAHT is acceptable to child protection professionals and whether or not it is likely to be used in practice. Therefore this study aimed to explore the acceptability of PredAHT with a variety of professionals involved in the child protection process.

2. Methods

This was a qualitative semi-structured interview study. The study received ethical approval from the Cardiff University School of Medicine Research Ethics Committee (Ref: 15/35). This study is reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines (Tong, Sainsbury, & Craig, 2007); a checklist is included in Appendix 1.

2.1. Participant recruitment

Purposive sampling and snowball sampling were used to recruit participants for this study. We targeted clinicians, CPSWs, legal practitioners, police officers and pathologists involved in suspected AHT cases across south west United Kingdom (UK). A list of potential participants was identified through personal contacts of the research team and organizational websites. Personal contacts and organizations were sent an information sheet to explain the study and were asked to suggest suitable participants for interview. A random selection of individuals from each professional group were then invited to take part. We recruited participants with different levels of child protection experience and seniority (Fig. 1). Individuals were contacted via email, with the exception of judges who were sent formal letters of invitation. In this study the term "clinician" refers to medical doctors and specialist nurses, who were sampled from three teaching hospitals and two district general hospitals across a range of specialties including pediatrics, radiology and neurosurgery. Most participating clinicians were experienced consultants or associate specialists, two were trainee doctors and one was a nurse. Senior CPSWs were team managers, who had a greater number of years of child protection experience than their junior counterparts. Senior police officers were inspectors or chief inspectors, and junior police officers were constables or sergeants. Judges had the greatest legal seniority with the responsibility of delivering the judgment on the evidence submitted by barristers or solicitors, and forensic pathologists had more experience of child protection investigations than the pediatric pathologist.

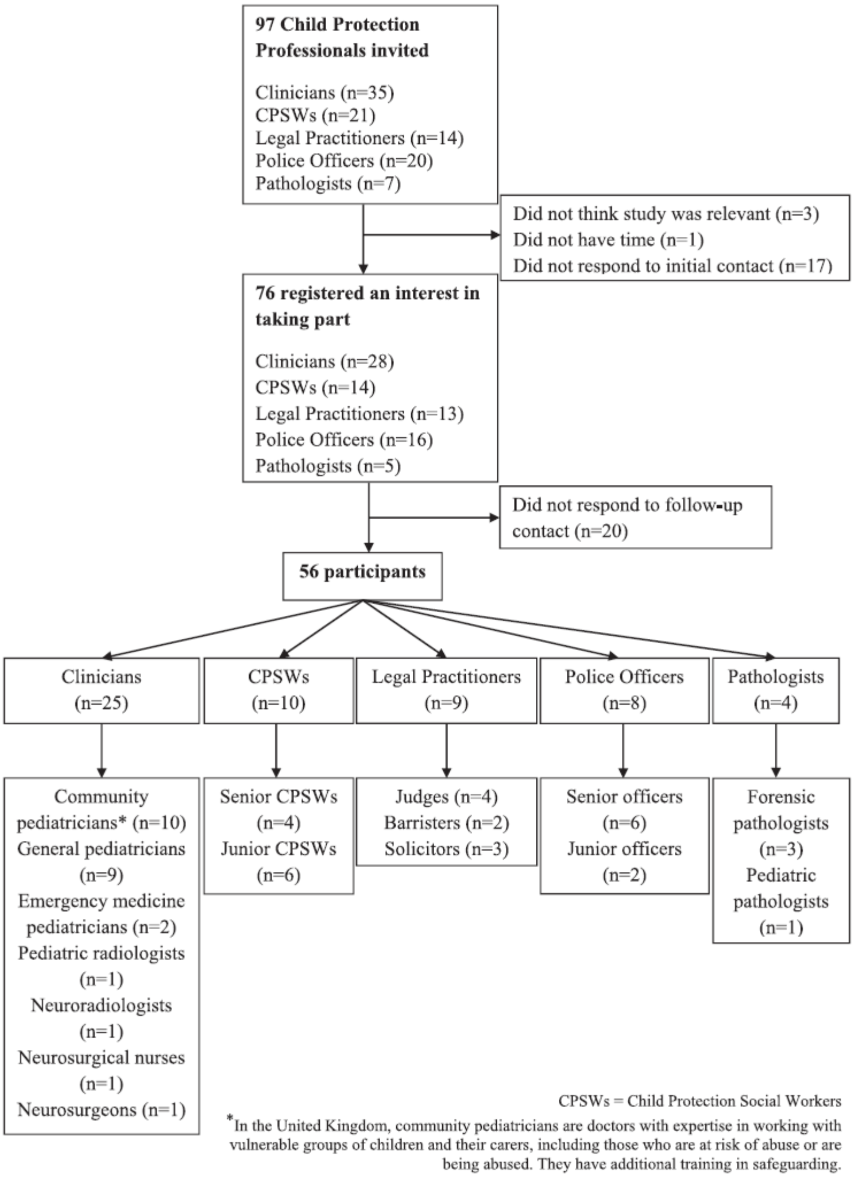


Fig. 1. Flowchart of child protection professionals participating in an evaluation of the acceptability of the Predicting Abusive Head Trauma clinical prediction tool.

2.2. Interview schedule development

The semi-structured interview schedule explored three main themes 1) participants' evaluations of PredAHT 2) participants' opinions about the optimal way to present the calculated probabilities 3) participants' interpretation of probabilities in the context of suspected AHT. The schedule was developed by LC and MF and revised following discussion with the research team (Appendix 2). Questions were derived following a review of the scientific literature on the identification of AHT, and the acceptability and evaluation of CPTs. The schedule was piloted with a police officer and a clinician, regarding the length, appropriateness, and content, and amended accordingly. The schedule included core open-ended questions, prompts and clarifying questions. It was a guide rather than a definitive list, to allow exploration of additional topic areas that might be raised by participants. Early responses influenced questions asked in later interviews; the schedule was updated as data collection and analysis progressed and new topic areas were raised.

2.3. Data collection

Interviews were conducted by LC, a PhD student with training in qualitative research methods and qualitative interview techniques. No relationship was established between the interviewer and participants prior to the study. Informed consent, and permission for audio recording for verbatim transcription were obtained before each interview. Participants received a demonstration of the computerised PredAHT and a brief description of its development and purpose. Clinicians saw a version of PredAHT that allows them to input their own "pre-test" (prior) probability of AHT, based on factors other than the six included features that are pertinent to each case e.g. purported history, clinical presentation or psychosocial features. The prior probability influences the post-test probability of AHT provided by PredAHT, and is otherwise set at 34%, the prevalence of AHT in the data used to derive the tool.

When two participants declined to be audio recorded the interviewer made detailed notes of their responses. These were sent to the participants to check that they were a fair reflection of their views. The schedule was delivered to individuals, or at two small group interviews (of three and five participants from the same professional group) where personal interaction between the participants was minimised, to maximise individual contributions from participants. Interviews lasted about 45 min and took place at the participants' workplace between June 2015 and September 2016. MF was present to record relevant field notes such as participant non-verbal behaviour and response to the interview, and critical reflections about the interview. No repeat interviews were conducted. In the interests of reflexivity, the interviewer considered how her own values and assumptions as a student involved in developing PredAHT might influence the interviews or the interpretation of the findings and a reflective journal was kept in an attempt to identify and minimize potential bias. It is acknowledged that the interviewer had a vested interest in PredAHT, hoping that it would be useful for the participants, and that they may have found it difficult to receive criticism or negative opinions about PredAHT. In early interviews with clinicians, the interviewer was acutely aware of her status as a psychology graduate, with no medical or legal training, but nevertheless conducting research in a medico-legal topic, and how this may affect the power relationship between the researcher and the participant. To break down power imbalances, every effort was made to build a rapport with the participants, and ensure that the interview was guided by them but stayed on-topic. The researcher made sure not to ask any leading questions, or impose their own views on the participants.

2.4. Data analysis

Data analysis began shortly after the first interview using thematic analysis (Braun & Clarke, 2006). A general inductive approach enabled the results to be guided by the aims and objectives of the research, and the raw data (Bryman & Burgess, 1994; Dey, 1993). The Framework Method was used to manage, summarise, display, and synthesise the data and to facilitate analysis (Gale, Heath, Cameron, Rashid, & Redwood, 2013). Analysis followed seven phases: transcription, familiarisation, coding, developing an analytic framework, applying the analytic framework, charting data into framework matrices, and interpretation (Gale et al., 2013). Initial codes were generated independently by LC, MF and HQS. These were jointly grouped into clearly defined categories that were further arranged under the three main overarching themes explored in the interviews. Discrepancies between coders were resolved by discussion and consensus. This process was undertaken in an attempt to minimize individual biases; 38% of the transcripts were independently double-coded. The joint analysis enabled the development of a preliminary analytic framework. Transcripts were imported into NVivo (QSR International Pty Ltd. Version 10, 2014), to organise and manage the data and assist with data analysis. Quotes pertaining to each category were retrieved and 'charted' into thematic framework matrices. Interviews ceased when thematic saturation was achieved within each group of participants (clinicians, CPSWs, police officers, legal practitioners and pathologists), which was verified using the constant comparative method (Glaser & Strauss, 1967). The final phase of the analysis involved abstraction and interpretation of the data. Participants were not asked to provide feedback on the study findings. Theme 1 is entitled 'participants' evaluations of PredAHT'. These were categorized as 'potential benefits of PredAHT', 'potential risks of PredAHT', 'provisos for the use of PredAHT', 'use of PredAHT in court', and 'clinicians' views about the practical use of PredAHT'. Theme 2 is entitled 'participants' opinions about how to present the calculated probabilities'. These were categorized as 'percentage probabilities versus broad risk categories', 'confidence intervals' and 'additional suggestions'. Theme 3 is entitled 'participants interpretation of probabilities in the context of suspected AHT'. These were categorized as 'threshold criteria', and 'comments about PredAHT scores'. The systematic synthesis of the data excerpts into thematic matrices enabled a final inspection of the categories across cases, to identify subcategories, i.e. the range of different elements being described under each category. All subcategories and their definitions are detailed in the analytic framework (Appendix 3).

3. Results

Participant demographics and response rates are shown in Table 2 and Fig. 1. Fifteen of twenty-five (60%) clinicians (six had previously used a CPT), 1/4 (25%) judges and 2/4 (50%) pathologists were familiar with and overall had a positive opinion of CPTs; none of the CPSWs, barristers, solicitors, or police officers were aware of them. Data are presented using quotations, selected as examples of the themes and categories that emerged from the data. Within the quotations, square brackets represent text inserted for clarification. Word repetitions and irrelevant sections were removed and denoted by '…'.

Table 2
Demographics of child protection professionals participating in an evaluation of the acceptability of the Predicting Abusive Head Trauma clinical prediction tool.

	Clinicians (N = 25)		CPSWs (N = 10)		Legal Practitioners (N = 9)		Police Officers (N = 8)		Pathologists (N = 4)	
	n	%	n	%	n	%	n	%	n	%
Gender										
Female	16	64	7	70	7	78	3	37.5	0	0
Male	9	36	3	30	2	22	5	62.5	4	100
Age group										
25–34	2	8	2	20	2	22	0	0	1	25
35–44	11	44	5	50	1	11	5	62.5	1	25
45–54	8	32	1	10	4	45	3	37.5	1	25
55–64	4	16	2	20	2	22	0	0	1	25
Ethnicity										
White British	19	76	10	100	8	89	8	100	4	100
White Other	4	16	0	0	1	11	0	0	0	0
Indian	2	8	0	0	0	0	0	0	0	0
Years in CP										
< 5	0	0	2	20	1	11	3	37.5	0	0
5–9	6	24	3	30	1	11	2	25	2	50
10–20	7	28	4	40	4	45	3	37.5	0	0
> 20	12	48	1	10	3	33	0	0	2	50
CP training										
Yes	25	100	10	100	3	33	7	87.5	4	100
No	0	0	0	0	6	66	1	12.5	0	0
Pediatric HI training										
Yes	18	72	1	10	3	33	4	50	3	75
No	7	28	9	90	6	66	4	50	1	25

CPSWs = child protection social workers, CP = child protection, HI = head injuries.

3.1. Theme 1: participants' evaluations of PredAHT

3.1.1. Potential benefits of PredAHT

PredAHT would be useful to support decision making as it is not influenced by personal feelings or opinions and could help reduce subjectivity in the assessment of risk.

> "I think they're a good idea because they can be completely evidence based, so it takes all your feelings out of it because it's…child protection, there's lots of emotions." Clinician 9

> "It would be helpful if a medical professional would have some confidence in saying it's an 85% chance because we would all understand what the chances were, because sometimes at strategy discussions you might get a pediatrician who will give an opinion, but as we all know, we take in messages in a different way. I might go back and record it in some way, the [child protection] social worker might go back and record it in a slightly different way." Police Officer 3

PredAHT would be helpful for heightening awareness of AHT, reinforcing or increasing concerns about possible AHT.

> "Where the number sits would help us to articulate that suspicion and perhaps work as a bit of a check. Probably in pushing it up, and highlighting to some people actually, you should be more suspicious because this is really unusual to get this combination." Police Officer 1

PredAHT could work both ways, encouraging participants to consider the possibility of non-abusive injury if PredAHT gave a low probability score for AHT.

> "It would also be helpful for us not to panic too much in the sense of there is the idea of accidental head injury as well. So it's also helpful for us to take a step backwards and not think right it's abusive trauma." CPSW 1

Clinicians, CPSWs and police officers said that PredAHT would provide them with reassurance or confidence that their concerns, suspicions or investigations were justified and that it would be useful to support their professional opinions.

> "I think as you used it more it would give you more confidence that actually, yes this is confirming that my level of suspicion is appropriate for the case…It would give you more confidence in making those decisions clinically." Clinician 15

The majority of clinicians would use PredAHT to back up their clinical opinion rather than to direct their decision-making.

> "I wouldn't use it just to sway my opinion, but if I had an opinion of whether it's abusive or not, and then, using this validated tool,

with the injuries found and the presence of head injury, it is likely, so that helps back up your opinion and hopefully then would add more weight to what you're saying." Clinician 21

However, some would find the score helpful if it did not agree with them.

"If there was a mismatch between my clinical opinion and the risk assessment tool that would cause me to stop and think and seriously consider whether I have gone down a bit of a blind alley with this and whether I need to stop and think again…It would be helpful just to reassure us that we are doing the right thing or maybe to cause us to stop and think actually perhaps we shouldn't walk away from this one." Clinician 10

PredAHT would be useful for explaining, justifying or rationalizing decision-making in suspected AHT cases.

"Family courts, criminal courts might want to know how have you arrived at this decision and if I was asked, well, these are the facts….and I've documented clearly why I've made a decision about something. Any tool I think that helps gives some…statistical interpretation for police, for social services, for the medical professionals, it's robust and trustworthy I suppose then, I only see that as a good thing because we're all accountable for the decisions we make." Police Officer 7

Clinicians and pathologists suggested that PredAHT may help to standardize or modify the clinical assessment of suspected AHT cases by prompting clinicians to perform investigations such as a skeletal survey or ophthalmoscopy in line with international standards, and to review the results of investigations already undertaken.

"Is this patient a patient that may have been abused and if so [the tool] triggers safeguarding procedures for siblings and it flags up this is a child who is going to need an ophthalmologist to look in her eyes, a skeletal survey, and a child protection pediatrician on call. If that triggers all of those things to take place that would be great." Pathologist 1

"We would just do all those investigations on anyone under 1, but it's in that 1–3 [age group], where you're just that bit more unsure, whereas should we be doing these things and it might actually guide us." Clinician 1

Pathologists and legal practitioners, including judges, could also appreciate the value of PredAHT for advocating further investigations, even if they would not find it useful themselves.

"I think if it can be used to ensure that front line clinicians are actually encouraged to undertake exploration of what they've got at a better level then we'd be saying yippee absolutely that's the best that you can do for us…If you can get clinicians to actually do what they should be doing when they should be doing it and triggering the protocols that need…because we see quite a lot of missed opportunities with initial investigation and you can't go back and do it again." Judge 2

Similarly, police officers and CPSWs said that PredAHT may help to justify further action within their respective agencies.

"If we've got a figure that says actually there's an 80% chance that there's abusive trauma, then that child isn't safe at all…we need to be taking pro-active action and that would I think be supportive." CPSW 3

PredAHT could contribute to 'the bigger picture', as part of a wider information gathering process. Many described PredAHT as a useful addition to the 'toolbox', or 'a piece of the jigsaw puzzle'.

"It would form one part of your prosecution case wouldn't it? It wouldn't be an enormous part but it could form a part of the evidence you'd built generally…I can't see for a minute that it wouldn't be useful." Police Officer 4

PredAHT could be used at multi-agency meetings or as part of information sharing to facilitate communication about the likelihood of AHT.

"It would be valuable for talking to the police, [child protection] social workers… just to say, 'Listen we've got this…' Because they will always say to you, 'Is there anything else it could be? Are we getting this wrong? Are we missing something medical?' I think when you're able to say with a degree of certainty, 'No, this is what it is because this is a validated tool. With this combination of injuries this is how confident we can be' I think it is going to be valuable for them as well…and you share it in the strategy meeting that would be very useful for me." Clinician 5

PredAHT would be useful for peer review or training. CPSWs in particular thought it was helpful to know that the six clinical features included in PredAHT are potential indicators of AHT.

"I think it will be very good for all child protection social workers dealing with these to know about these six things." CPSW 1

PredAHT may be most beneficial for so-called 'gray' cases, where there is considerable uncertainty surrounding the diagnosis, and most beneficial for those with the least experience in child protection.

"You get those ones where you think 'this is really not likely to be this' but you've got to go through the steps, and the ones where there clearly is likely to be a problem, so it's those gray ones in the middle where this might come in more useful than the clear cut ones." Clinician 1

"It might be helpful for someone who's never done pediatrics before and doesn't have the experience and the benefit of having done child protection work before and knowing these things…I think it is helpful for a very specific group of people." Clinician 20

3.1.2. Potential risks of PredAHT

Professionals may be over-reliant on PredAHT when making decisions in suspected AHT cases.

"If there was too much reliance placed on it at the beginning of an investigation and someone with little knowledge simply populated those fields present, absent, and came out with a low probability you know 14% or whatever and decided not to investigate, regardless of the presence of other factors not in your fields, then that would be foolish and dangerous, so the tool itself is not a disadvantage, it's how its uses could be." Police Officer 6

"I think there's a potential for people to make it the single most significant part of the decision making process so we'd end up sat in meetings and people would ignore most of what I said and say 'What does the tool say? Oh 67%, right that's the decision made'. That would be my worry that people would over-use it or overstate its importance." Clinician 12

A low score could instil a false sense of security and appropriate investigations might not be carried out in the face of a low probability score.

"If somebody is uncritically using this tool and they have got a child with an intracranial head injury, and head and neck bruising they are not worried then because the score says that it's less than 15% then that would be an incorrect use of this tool." Pathologist 1

However, reassuringly, all participants said that they would still carry out appropriate investigations if they received a low score from PredAHT, as there may be other features of the case that warrant further enquiry.

"If it was low like that, but there were previous allegations of abuse, dad had a violent background, that sort of stuff. That would sway me towards being quite concerned about this. So, I guess it's about the attendant circumstances around it. So if it was low, it would help me, but I would still look at the bubble around it and what's going on." Police Officer 3

Concern was expressed that PredAHT wouldn't be used as intended, alongside other known information about each case, and they agreed that it should never be used in isolation from other factors. Several clinicians and pathologists said that PredAHT was too reductionist and crude, comparing it to a box-ticking exercise.

"My main concern is people not taking into account the history or the other facts because they've got a big number on this score." Pathologist 2

The accuracy of PredAHT was questioned. The sensitivity and specificity that participants would be willing to accept was discussed, together with the implications of false positive and false negative predictions.

"The key thing is in how many cases is this wrong? And if it's wrong in any, then you've just got to ask yourself is this safe?" Pathologist 4

"Hopefully this will help us find all cases of abusive head trauma but there is a chance that we might label a non-abusive one as an abusive one as well, but I would probably weigh the benefit more than the risks, because if this is helping me to identify the really vulnerable children, I would rather use it...as long as I am protecting the vulnerable children, I would find it useful." Clinician 22

A few pathologists and clinicians, including a neurosurgeon, stated that they do not need a tool to make decisions about suspected AHT, and would therefore not use PredAHT.

"That's what we do in our brains, we put all the information together and spit out the probability based on our experience." Clinician 19

"It's something that is an irrelevance to me in that, one might take the view that this is an attempt to make my task less onerous by placing in my mind the conclusions of others, or their interpretation of the evidence, when it is my role to look at that evidence and the literature myself. This I fear might be regarded as a substitute for individual thought." Pathologist 4

Two clinicians were unsure about how much PredAHT would add to the investigative process, and could not say whether they would use it in practice or not.

"I have to say my initial thought looking at it is I'm not sure how much more it would add if you've done all the investigations already." Clinician 13

Some participants thought that there are important clinical and historical features missing from PredAHT, e.g. skull fractures, bruising patterns, spinal injury, or a history of trauma, and questioned why they were not included.

"You don't have a history of extraordinary trauma as an option. So the other thing is a non-declaration of the history that would be massive wouldn't it. Or no history of any injury...just woke up and the baby was like that. That would be a massive predictor I would imagine." Clinician 24

PredAHT cannot take into account specific details of the clinical features, including the severity of injury, injuries of different ages, and the precise locations and patterns of injuries, some of which may be highly specific for AHT.

"What about a healing fracture as opposed to a recent fracture and position of the fracture, and particularly rib fractures are they at the front of the chest of a child who has had resuscitation or are they posterior ribs, and it's this granularity that we are grilled

Table 3
Child protection professionals' provisos for the use of the Predicting Abusive Head Trauma clinical prediction tool.

Alongside professional judgment	*"I guess it's probably a combination of that along with a bit of professional judgement tied in… if you look at it in combination with other professional opinion, what else you know, what information you found out, then it could inform part of that pool of information."* **Police Officer 3** *"Yeah I think it's useful, it should not stop you from thinking I think you should still think outside the box and not 100% rely on it but I think as an additional tool to your clinical decision making, I think it is supportive and helpful."* **Clinician 7** *"This in addition to the rest of our assessment is really, really helpful."* **CPSW 1**
Definition of the six features	*"I would need a bit more information…you know I don't really know what retinal hemorrhages are."* **CPSW 6** *"Apnea, presumably you'd have a definition of how long that's for and stuff like that?"* **Pathologist 2**
If kept up to date	*"There's issues of keeping it up to date, you can't just do it once and then not revisit it, can you?"* **Clinician 3** *"To keep its credibility it would have to evolve with current thinking, so it's a continual process isn't it?"* **Police Officer 2**
Quality of the data on which PredAHT is based	*"I'd want to look at the original research and how the original cohort of patients were diagnosed with abusive head trauma and what's the robustness of that diagnosis in the first place, that the tool is then based on."* **Clinician 12** *"I would never use something like this without reviewing the publication and looking at the statistics and checking out that I was personally happy with the statistical analysis, because otherwise I'm just putting stuff into boxes."* **Pathologist 2**
Understanding how it works and how it is to be used	*"I would have to understand it and be able to explain it in court, so I'd need to come and have a little training session."* **Clinician 3** *"You would need to explain the unknown parts of it…as well you could do a small tutorial based on four or five cases if people want to get experience on how to use it."* **Clinician 14** *"It's important to understand what informs the figure, because otherwise it becomes a checklist…I think people need to have an understanding of what the tool is and how it is to be used."* **CPSW 5**
Agreeing accepted risk thresholds with colleagues	*"We can all have that figure and we can all explore then what that figure means to each independent agency, and what it means for that child and actually what safeguards need to be in place because of it."* **CPSW 3** *"I think within a team, there needs to be consistency as to what it's meaning at that point in time."* **Clinician 2** *"You want that consistency and agreement as to what the results could mean."* **Police Officer 1**
If accepted by colleagues	*"It would be only useful for us if it's accepted by the medical profession."* **Legal Practitioner 2** *"Whether it be the safeguarding board or the child death overview panel, you'd want something where the social workers and the pediatricians and the police all come together and agree that this is useful."* **Police Officer 1** *"That would give me the most confidence really if the medical professionals were on board with it."* **CPSW 6**

on and we have to take into account when we are giving our overall opinion but for a quick and dirty or 'Should I be contacting child protection services, should I be contacting our pediatrician on call for child protection?' This sort of thing is great." Pathologist 1

"Not all RHs are the same, you can have one in one eye and five in the other but if they're not in the layers that you'd expect them to be…you'd have to make some allowance for, not only the categories but subcategories of that…It needs to be more refined." Judge 2

Some pathologists and judges said that PredAHT may condition their decision-making or inadvertently introduce bias into the decision-making process.

"You would almost make it more difficult for the judge because the judge would then have to disentangle the expert opinion from either an apparent bias or an unconscious bias that might be established by the fact that the expert had looked at the clinical tool." Judge 3

3.1.3. Provisos for the use of PredAHT

Many participants would only use PredAHT with a proviso; e.g. alongside their professional judgment, with more information about the definition of the six features, if it was kept up to date, with knowledge of the quality of the data on which PredAHT is based, with an understanding of how it works and how it is to be used, after agreeing acceptable risk thresholds with multi-agency colleagues, and if it was accepted by their colleagues (Table 3).

3.1.4. Use of PredAHT in court

Most clinicians, CPSWs and police officers thought that PredAHT would be useful in court, because it is evidence-based and validated. CPSWs in particular felt that it would be useful in the family courts for future safeguarding of children, where the standard of proof is based upon the balance of probabilities.

"In the court arena I think it's going to be really very useful because it's not our hunch against the next doctor's hunch, you know? And I think people's general opinion that babies aren't injured by their carers and their parents…people don't want to hear that and they certainly don't want to believe it and acknowledge that this is happening, but if you've got a validated tool saying, 'Actually this is what has happened to this baby because of the other injuries that we've seen' then I think it's going to be very valuable indeed." Clinician 5

"It helps when going to court with the balance of probabilities if you can prove over 51%, that's the number I have in my head…

that's what we've got to convince evidence of a judge of." CPSW 2

However, pathologists and legal practitioners, including judges, expressed caution regarding its use in court, particularly in the criminal courts.

"What the criminal standard which is beyond reasonable doubt would make of that, because the decision in these circumstances would be that of a jury, again huge caution in thinking this is effectively steering a jury into saying it's 85%, it's beyond reasonable doubt therefore we've got no choice but to convict." Legal Practitioner 1

Some felt that PredAHT would be irrelevant because it cannot account for every detail of every case, and each case must be considered based on the entirety of the evidence.

"As lawyers we would probably want to treat it with extreme caution. Simply because this tool cannot cover every factor in every situation that we have to deal with." Legal Practitioner 1

Others remarked that PredAHT may not stand up under cross-examination, or that the defence will claim that their case falls into the reverse probability of non-AHT given.

"You would have to prove the tool in every case. You'll be cross examined about how it's been put together, how you've weighted the factors. There's always something that somebody can find if you're really trying to pull something apart. Then it goes out the window really evidentially." Judge 2

"We would be arguing well why isn't this one in the 15% of cases that suggests that it isn't non-accidental?" Legal Practitioner 1

In addition, some participants pointed out that PredAHT will not help to identify the perpetrator in suspected AHT cases.

"That doesn't help us with who caused it, it just says 'what's the probability of it being an abusive trauma' so there is that other element we have to consider." Legal Practitioner 1

Some clinicians and judges discussed historical child protection court cases that involved the use of statistical evidence, and the impact and implications of such cases on the subsequent acceptance of statistics in the courtroom.

"A slight worry any pediatrician will have, a study putting statistics up like that, is the way that [Roy] Meadow [UK pediatrician] was chopped down with statistics." Clinician 17

However, one judge thought that PredAHT would definitely be useful, in both the family and criminal courts.

"It will help to remind the courts and the experts that a certain combination of features does make abusive head trauma a more likely explanation...I think it would have the same role in the criminal courts. Even though the standard of proof is different, it would still be useful at the fact finding stage." Judge 4

Despite their reservations, the majority of the legal practitioners interviewed, including judges, agreed that they would probably take the PredAHT score into account if it was included as part of a medical court report.

"The way the courts see these matters from a child protection point of view is an analytical approach where you need the best evidence possible. If this is something that feeds into a medical report, by an expert who understands it, then I'm delighted to have that. Hopefully it can help to make the right decision for the family, because it is life changing." Judge 4

3.1.5. Clinicians' views about the practical use of PredAHT

All thought that PredAHT is simple to use, and not too time-consuming to complete.

"Time is always a disadvantage in getting people to fill these out sometimes, but this is relatively simple and straightforward so I don't imagine it being a huge issue." Clinician 14

The majority thought that PredAHT would be most useful for inpatients admitted to a ward or Pediatric Intensive Care Unit (PICU), and less useful in the Emergency Department (ED), where information about fractures and RHs is unlikely to be available. However, one emergency medicine pediatrician thought that PredAHT may have a role in the ED to prompt an initial referral to the safeguarding team.

"If they were unwell enough to go to PICU you may use that tool much less in the ED. If they were somebody that was going to a ward then you would probably use it more. I think it would depend on the patient and how sick they were." Clinician 16

There were different views regarding the stage of the assessment process that PredAHT would be most useful. Some would only use it once all relevant investigations were completed, to assist with report writing or reaching their final conclusions.

"In my opinion, there's not much point in using it if you have too many unknowns there…I personally would certainly like it for when I have to write my report." Clinician 7

However, others could see the value of PredAHT at multiple stages of the assessment process, and would use it more than once during a case to support their decision-making.

"I would probably use it as soon as I knew about the case, just to give me some idea, and then as more data is collected you could add it in and see how it changes your figure, and then you've got your last kind of figure then is what is going to be the most important one right at the end." Clinician 1

PredAHT could be completed by general and community pediatricians and intensivists, but most agreed that it should be completed by a consultant. One clinician thought that it should be a team exercise.

"Whether it would be something that would be used by the lead consultants in PICU or a general pediatrics consultant where they are thinking do we need to get the safeguarding team involved or not and then potentially I suppose used by a safeguarding consultant when it came to writing up." Clinician 8

Clinicians reflected on how PredAHT could be integrated into the clinical workflow and implemented in clinical practice. Although they acknowledged that each hospital has its own way of working, most thought that it would not be too difficult to incorporate the computerised version of PredAHT into existing hospital intranet systems. Some suggested including reminders or specific references to PredAHT on existing departmental or safeguarding paperwork.

"It could well go on to the intranet as an app…and maybe a reference to it as a little reminder on our safeguarding proforma." Clinician 4

Some clinicians could see the value of including their own prior probability score in the calculation, but the majority felt that this would introduce too much subjectivity into PredAHT and that they would need guidance on how to use this element of the tool.

"I think that's an important element to bring in because a lot of our decisions are often based around the history and does the history fit, is it consistent, stuff about whether they're presenting late, stuff about what the family background and social history is." Clinician 13

"I'd feel happier with the six features on its own, because I know that's very evidence based, isn't it, so that's fine. I do have an issue with the prior probability without some objectivity around it, because it's easy to think oh well, a child on the Child Protection Register and that ups my concerns and whereas a middle class family doesn't, and I think it is something you need to be really objective about." Clinician 3

3.2. Theme 2: participants' opinions about the optimal way to present the calculated probabilities

Precise percentage probabilities of abuse were preferred, rather than broad risk categories such as low, medium or high likelihood of abuse. While some suggested presenting both, they were unsure as to what percentages would equate to low, medium and high. Some participants felt that confidence intervals would be unnecessary, however others felt that they would be an important addition. Additional suggestions included background information/data about PredAHT, disclaimers, and visual aids (Table 4).

3.3. Theme 3: participants' interpretations of probabilities in the context of suspected AHT

3.3.1. Threshold criteria

Although all participants maintained that they have a very low threshold for suspicion of AHT in young children with intracranial injuries, their probability thresholds for suspecting abuse varied widely. Some participants would only feel confident to completely rule out AHT if the percentage probability was less than 1%, while others had higher thresholds.

"If it's something like in the thirties, gosh that's a really hard thing to factor in isn't it then, in terms of decision making, it still sounds quite high to me. You almost want it to be a 0.1% chance of it being an abusive head trauma to feel confident in your decision, because even at 30%, that's like one out of three families, that was abusive isn't it?" CPSW 9

"If that said to me there's a 1% chance then there's still a 1% chance. It's helping me it's not telling me there's no chance is it?" Police Officer 8

"I want to say sometimes I'm not happy about not taking any further action at times, but I'd have to be I don't know maybe 20%?" CPSW 2

Many participants simply could not put a figure on their threshold for abuse, stating that if there was any chance at all that it could be, then they would investigate further, and commenting that each case is dependent on the attendant circumstances around it. One CPSW indicated that often her risk judgments are very different to her colleagues', highlighting inconsistencies within the assessment process.

"I don't think I can put a figure on my threshold because it depends sometimes I look at something and think 'Why are we going out on this?' And then something else, 'We should've looked at that, why did we have all of this and we haven't done anything with it?'" CPS W 4

Participants also had different perceptions of what the expressions 'low', 'medium' or 'high' likelihood of abuse might mean in percentage terms.

Table 4
Child protection professionals' preferences for the presentation of the calculated probabilities from the Predicting Abusive Head Trauma clinical prediction tool.

Percentage probabilities versus broad risk categories	*"For a decision-making tool a percentage is spot on from my point of view, because I think everyone will understand it." **Police Officer 2*** *"A percentage just makes it a little bit more tangible, doesn't it, it's very real, because I find it difficult, I always have done, to quantify risk, so if it can be done for me, then yeah…" **Clinician 6*** *"I am happy with the percentage, I am confident with that because that is where we are at with safeguarding children, we are talking about the balance of probabilities and so we are talking anything over 50 per cent probability should lead to further action and further evaluation where you may still get information that shifts it the other way." **Clinician 10*** *"I'd personally prefer it if captured with likeliness and you might have very suspicious, strongly indicative or not likely." **Clinician 12*** *"I prefer it like that because low, medium, high can mean anything." **CPSW 4*** *"A percentage wouldn't be helpful in court, because it would actually lead to more uncertainty. There would always be an argument to have which would detract away from the purpose." **Judge 4*** *"I appreciate what it is, there could be a range but what would low, medium, high then be? I suppose that's the problem." **Police Officer 7***
Confidence intervals	*"I would like to know the variance, that's what we would like to know, that's what we would need to know…" **CPSW 3*** *"I do like the number as a percentage but I do like to know the confidence interval as well…" **Clinician 7*** *"I think it might be a bit too much information." **CPSW 6***
Additional suggestions	*"There should be a little caveat statement there saying that it can go both ways, the higher it is the more likely it is to be, but a low one doesn't exclude it." **Clinician 23*** *"I think you need some sort of disclaimer on it about this needs to be used as part of a full assessment." **Clinician 8*** *"You want to know where it's come from, what's the research basis behind it, what's the evidence behind it, how much can you trust it. Now that could be a short blurb and then links to the publications, the literature that supports this." **Clinician 2*** *"I would want to know the data behind it because it is obviously chunking and splitting the data in different ways, maybe if you have got all the individual data, you could list everyone who fell outside the non-accidental injury bracket?" **Clinician 24*** *"If you can say here's a big block of how many of these kids were deliberately injured compared to a little smidge of kids, you could almost support it perhaps with a quick graphic to go big block is battered kids, small block is unfortunate accident." **Police Officer 4***

"Less than 50 would be low. Maybe 50–70 medium and then over 70 high." CPSW 6

"I suppose low, medium and high can mean anything can't it, I'm guessing its maybe up to 30%, 60%, 90%." CPSW 8

3.3.2. *Comments about PredAHT scores*

After receiving a demonstration of PredAHT, participants offered their opinions on the probability scores that it gives for different combinations of features. They were told that intracranial injury with head/neck bruising alone gives a score of 14.7%. Some participants thought this score was low. Others interpreted the score to be an unacceptably high level of risk, while others still thought this figure could be interpreted in both ways.

"I think that's really low 14.7%." CPSW 4

"14.7%, what do I think of that? I still think it's bloody high." Police Officer 1

Participants were uncomfortable that PredAHT can give scores at or very close to 100% for certain combinations of features.

"99.6% and I've ticked rib fracture present, head and neck bruising present, apnoea present, seizure present. That I find hard you're saying essentially that's definitely abusive head trauma…I agree, I'd be very worried if I had that combination of features but I wouldn't say it's nearly a hundred." Pathologist 2

Others stated that PredAHT scores too low for certain combinations of features.

"So if you're fitting and have subdurals but you don't have RHs or any other markers, it's saying it's not non-accidental injury. I would be a lot more hawkish than that." Clinician 24

"I think that my findings were stronger than that score…the RH was very strong, in all layers…I might've hesitated if I'd seen 57%." Judge 2

4. Discussion

The study findings suggest that PredAHT would support the decision-making of clinicians, CPSWs, and police officers investigating suspected AHT, and provide them with greater confidence in expressing their opinion in the child protection and court setting. Benefits were perceived by junior and senior practitioners with different levels of child protection experience, and across all specialities with the exception of a neurosurgeon, although it was acknowledged that PredAHT may be most useful for those with the least child protection experience. Pathologists and legal practitioners, including judges, thought PredAHT to be useful as a screening tool for ruling in further clinical or multidisciplinary investigations, however with the exception of one judge, they expressed caution regarding its use in court. In practical terms, clinicians found PredAHT to be simple to complete and thought it would be straightforward to implement into existing hospital systems.

All professionals who come into contact with children and families have a duty to safeguard children and young people and should receive regular training to ensure that they are competent in their respective roles (Department for Education, 2015). Two UK government reports on social work interventions required in the child protection arena stated the need for the development of an evidence-based approach and learning culture, to inform good practice; one explicitly recommended the use of standardized tools to support decision-making and analysis of information about whether a child is suffering, or likely to suffer significant harm (Barlow, Fisher, & Jones, 2012). The other highlights the importance of improving the skills and knowledge of CPSWs (Department for Education, 2014). Our study highlighted gaps in the training and knowledge of professionals working in child protection; many were unaware that some of the clinical features included in PredAHT were indicators of AHT. This is consistent with a recent study that demonstrated gaps in knowledge and training about bruising amongst CPSWs (Matthews, Kemp, & Maguire, 2017). Improved communication within and between agencies is critical for identifying patterns and preventing further injury (Joughin, 2003) and has been recommended in the UK Safeguarding Children Research Initiative report (Davies & Ward, 2012). The current study confirmed that PredAHT would facilitate interagency communication about the likelihood of AHT.

Clinicians stated that PredAHT would give them more confidence in expressing their opinions about the likelihood of AHT in their court reports and in court settings. While pathologists and legal practitioners, including judges, appreciated the value of PredAHT for encouraging standardization of clinical investigations, and further clinical or social assessment, overall they expressed caution regarding its potential use in court. However, the majority would incorporate the probability score with all of the other evidence if it was provided as part of a medical report and PredAHT was accepted by the medical community. Previously, high profile cases involving the misinterpretation of statistics and probabilistic evidence in the courtroom have caused controversy (R v Cannings, 2004; R v Clark, 2000) and led to the development of a working group within the Royal Statistical Society in the UK, to improve the use of statistics in the administration of justice. They recommended a broad programme of education for judges, lawyers, and expert witnesses in probability theory and statistics (Aitken, Roberts, & Jackson, 2010). One judge and one pathologist were concerned that PredAHT would introduce cognitive bias into their decision-making, suggesting that they may not trust PredAHT to be a valid piece of evidence in their decision making.

The scientific literature confirms that there are specific patterns of intracranial injury, haemorrhagic retinopathy (Wright, 2017), and spinal injuries associated with AHT (Choudhary, Ishak, Zacharia, & Dias, 2014). Various psychosocial variables may also be influential (Pierce et al., 2017). Some participants wanted these additional features incorporated into PredAHT. However, a massive multi-center prospective study would be needed in order to add further variables. Meanwhile, the six clinical features in PredAHT are easily identifiable in the early phase of clinical assessment.

Participants had varying opinions about what percentage probabilities equate to the terms 'low', 'medium' or 'high' likelihood of abuse, and diverse probability thresholds for suspecting abuse, although participants stated that PredAHT helped them to quantify risk. This is consistent with previous studies that demonstrated that pediatricians struggle to define "reasonable suspicion" or "reasonable medical certainty" of abuse (Dias, Boehmer, Johnston-Walsh, & Levi, 2015; Levi & Brown, 2005). Thresholds in child protection social work have been the subject of much debate in recent years and are affected by a wide range of organisational factors, relationships with other professionals, and individual biases, heuristics and value systems (Platt & Turney, 2014). Similarly, participants postulated that PredAHT may be most useful for 'gray' cases, where there is significant uncertainty surrounding a diagnosis of AHT. In reality, the interpretation of the PredAHT score will depend upon individual perception of risk. PredAHT is designed to be an assistive tool rather than a decision rule, which typically recommends a direct course of action based on the results (Reilly & Evans, 2006); PredAHT provides no recommendations for professionals on what to do based on specific scores. Despite this, the majority of participants thought PredAHT would be useful for supporting their opinions and decision-making.

Previous research has identified barriers to the use of CPTs, some of which emerged from the interviews, such as scepticism of "cook-book" medicine, belief that clinical judgment is superior to the tool, distrust of the accuracy of the predictors and concern that the CPT does not address all relevant factors (Reilly & Evans, 2006). Reilly and Evans (2006) offer a number of strategies to overcome these barriers, including comparing clinical judgment with the CPT, and checking whether any excluded factors affect the CPTs predictions. Skull fracture was analysed within the original derivation study and did not discriminate between AHT and non-AHT (Maguire, Kemp et al., 2011). A study comparing PredAHT with clinical judgment is underway.

The findings have implications for the further development and implementation of PredAHT. Given that some participants wanted confidence intervals whilst others did not, it may be sensible to include an option to display these. Participants' identified a number of conditions under which they would use PredAHT, namely if it was accepted by their colleagues, used alongside their professional judgment, and if they understood how it works. These are consistent with a study exploring the acceptability of a tool to identify abusive or neglectful burns (Johnson, Hollén, Kemp, & Maguire, 2016). Any training on PredAHT would need to encompass these elements. Over-reliance on PredAHT, concern that it may be used improperly or failure to investigate appropriately if a low score is given, and the potential ramifications of "false positives" or "false negatives" emphasise the importance of providing clear guidance to practitioners about how PredAHT is intended to be used, namely as an assistive CPT, and *not* a diagnostic tool. Finally, PredAHT allows clinicians to factor in aspects of the history and other risk factors by incorporating their own prior probability of AHT into the calculation. Clinicians felt this element of PredAHT was subjective, and were unsure whether they would be comfortable estimating a prior probability of AHT in light of potential racial and socioeconomic bias (Wood et al., 2010). This suggests a lack of knowledge amongst clinicians of the evidence base regarding psychosocial risk factors for AHT. A recent survey found that less than half of health care professionals are adequately trained or prepared to identify risk factors associated with maltreatment (Foster, Olson-Dorff, Reiland, & Budzak-Garza, 2017). Training on how to estimate a prior probability of AHT should be incorporated into full implementation, as requested by clinicians.

4.1. *Strengths and limitations*

There are many derived CPTs for children; few are validated, while virtually none undergo impact analysis (Maguire, Kulik et al., 2011), and very rarely do investigators determine the acceptability of CPTs prior to their use. To our knowledge this study is one of only two studies (Johnson et al., 2016) exploring the acceptability of a CPT developed for use in child protection, and the first study to have done so with a wide range of professionals. Strengths of the current study include the rigorous data analysis methods employed, and the depth, detail, and richness of the data collected. Semi-structured interviews enabled the interviewer to build rapport with the participants (Fontana & Frey, 1994), and led to richer data than might be gained from a more structured approach, a survey or questionnaire. PredAHT is the only CPT that we are aware of that can estimate a predictive probability of AHT given different combinations of multiple clinical features, at various points along the assessment pathway. Other CPTs have been developed for use in the ED to assist clinicians deciding which high-risk children should undergo computed tomography (Berger et al., 2016), and for the PICU to help exclude AHT when negative (Hymel et al., 2014). However, whether clinicians or other practitioners would be prepared to use these CPTs in practice is unknown.

Despite the inclusion of a range of professional groups and clinical subspecialties, additional groups could have made valuable contributions, for example neurologists, intensivists, or ophthalmologists. Most of the clinicians were consultants working in teaching hospitals, and less experienced clinicians, and radiologists, neurosurgeons, and nurses, were under-represented. However it should be noted that probabilistic representativeness is not a goal of qualitative research (Popay, Rogers, & Williams, 1998). The fourth pathologist interview revealed some new insights that the other pathologists had not brought up, suggesting that data saturation may not have been reached within this professional group. This does not necessarily invalidate the findings for this group but rather means that further exploration of the topic may be warranted with these professionals (O'Reilly & Parker, 2012). Given the sensitive nature of the research, some respondents may have shown a social desirability bias and responded in a manner likely to be viewed as favourable by the researcher, however a number of participants exhibited an unfavourable view of PredAHT and were open about their opinions and intentions not to use it. Finally, qualitative research inevitably relies on the researcher's interpretations, however, subjective bias was minimized by using three trained qualitative researchers to double-code the data and resolve disagreements through discussion and consensus.

5. Conclusions

This evaluation has demonstrated that PredAHT is acceptable to child protection professionals across a range of disciplines assessing suspected AHT cases, and that they would be willing to use it as an adjunct to their decision-making. Although it may be most useful for those with the least child protection experience or knowledge, it is applicable to all professionals working in this area as it may help to reduce missed cases of AHT. These results confirm that the addition of a precise and objective evidence-based probability score that calculates the risk of AHT for child protection professionals is acceptable and potentially useful. This tool, when used in conjunction with a full clinical and social history, has the potential to standardize clinical assessment, and minimize subjectivity when weighing up the clinical features in cases of possible AHT. Feasibility work is underway to determine whether it is possible to evaluate the impact of PredAHT when it is applied in clinical practice. This will inform the planning and design of a formal impact analysis study and a long term implementation and dissemination plan to maximise uptake (Stiell & Wells, 1999).

Funding source

This work was supported by Health and Care Research Wales (grant number HS-14-24). The funders had no involvement in the study design, the collection, analysis, or interpretation of the data, the writing of the report or the decision to submit the article for publication.

Declaration of interest

The authors have no actual or potential conflicts of interest, or any financial, personal or other relationships relevant to this article to disclose.

Acknowledgments

We would like to thank Health and Care Research Wales for funding the study, and all of the professionals who took part.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.chiabu.2018.04.022.

References

Aitken, C. G. G., Roberts, P., & Jackson, G. (2010). *Fundamentals of probability and statistical evidence in criminal proceedings: Guidance for judges, lawyers, forensic scientists and expert witnesses*. London, UK: Royal Statistical Society.
Barlow, J., Fisher, J. D., & Jones, D. (2012). *Systematic review of models of analysing significant harm*. Oxford: Department for Education.
Berger, R. P., Fromkin, J., Herman, B., Pierce, M. C., Saladino, R. A., Flom, L., ... Kochanek, P. M. (2016). Validation of the Pittsburgh infant brain injury score for abusive head trauma. *Pediatrics, 138*(1), e20153756.
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.
Brehaut, J. C., Graham, I. D., Wood, T. J., Taljaard, M., Eagles, D., Lott, A., ... Stiell, I. G. (2010). Measuring acceptability of clinical decision rules: Validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. *Medical Decision Making, 30*(3), 398–408.
Bryman, A., & Burgess, R. (1994). *Analyzing qualitative data*. London: Routledge.
Canadian Paediatric Society (2007). *Multidisciplinary guidelines on the identification, investigation and management of suspected abusive head trauma*. Ottawa: Canadian Paediatric Society.
Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2015). *Shiny: Web application framework for R. R package version 0.11.1*. http://CRAN.R-project.org/package=shiny.
Choudhary, A. K., Ishak, R., Zacharia, T. T., & Dias, M. S. (2014). Imaging of spinal injury in abusive head trauma: A retrospective study. *Pediatric Radiology, 44*(9), 1130–1140.
Christian, C. W., Committee on Child Abuse and Neglect, & American Academy of Pediatrics (2015). The evaluation of suspected child physical abuse. *Pediatrics, 135*(5), e1337–e1354.
Cowley, L. E., Morris, C. B., Maguire, S. A., Farewell, D. M., & Kemp, A. M. (2015). Validation of a prediction tool for abusive head trauma. *Pediatrics, 136*(2), 290–298.
Davies, C., & Ward, H. (2012). *Safeguarding children across services: Messages from research*. London and Philadelphia: Jessica Kingsley.
Department for Education (2014). *Child protection, social work reform and intervention: Research priorities and questions*. London: Department for Education.
Department for Education (2015). *Working together to safeguard children: Statutory guidance on inter-agency working to safeguard and promote the welfare of children*. London: Department for Education.
Dey, I. (1993). *Qualitative data analysis*. London: Routledge.
Dias, M. S., Boehmer, S., Johnston-Walsh, L., & Levi, B. H. (2015). Defining 'reasonable medical certainty' in court: What does it mean to medical experts in child abuse cases? *Child Abuse & Neglect, 50*, 218–227.
Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin, & Y. S. Lincoln (Eds.). *Handbook of qualitative research* (pp. 361–376). Thousand Oaks: Sage Publications.
Foster, R. H., Olson-Dorff, D., Reiland, H. M., & Budzak-Garza, A. (2017). Commitment, confidence, and concerns: Assessing health care professionals' child maltreatment reporting attitudes. *Child Abuse & Neglect, 67*, 54–63.
Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13*(117), 1–8.
Glaser, B. G., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing Co.
Hymel, K. P., Armijo-Garcia, V., Foster, R., Frazier, T. N., Stoiko, M., Christie, L. M., ... Pediatric Brain Injury Research Network (PediBIRN) Investigators (2014). Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics, 134*(6), e1537–1544.
Jenny, C., Hymel, K. P., Ritzen, A., Reinert, S. E., & Hay, T. C. (1999). Analysis of missed cases of abusive head trauma. *Journal of the American Medical Association, 281*(7), 621–626.
Johnson, E. L., Hollén, L. I., Kemp, A. M., & Maguire, S. (2016). Exploring the acceptability of a clinical decision rule to identify paediatric burns due to child abuse or neglect. *Emergency Medicine Journal, 33*, 465–470.
Joughin, V. (2003). Working together for child protection in A&E. *Emergency Nurse, 11*(7), 30–38.
Letson, M. M., Cooper, J. N., Deans, K. J., Scribano, P. V., Makoroff, K. L., Feldman, K. W., ... Berger, R. P. (2016). Prior opportunities to identify abuse in children with abusive head trauma. *Child Abuse & Neglect, 60*, 36–45.
Levi, B. H., & Brown, G. (2005). Reasonable suspicion: A study of Pennsylvania pediatricians regarding child abuse. *Pediatrics, 116*(1), e5–e12.
Maguire, J. L., Kulik, D. M., Laupacis, A., Kuppermann, N., Uleryk, E. M., & Parkin, P. C. (2011). Clinical prediction rules for children: A systematic review. *Pediatrics, 128*(3), e666–e677.
Maguire, S. A., Kemp, A. M., Lumb, R. C., & Farewell, D. M. (2011). Estimating the probability of abusive head trauma: A pooled analysis. *Pediatrics, 128*(3), e550–e564.
Maguire, S., Pickerd, N., Farewell, D., Mann, M., Tempest, V., & Kemp, A. M. (2009). Which clinical features distinguish inflicted from non-inflicted brain injury? A systematic review. *Archives of Disease in Childhood, 94*(11), 860–867.
Matthews, L., Kemp, A., & Maguire, S. (2017). Bruising in children: Exploring the attitudes, knowledge and training of child protection social workers and the interface with paediatricians regarding childhood bruising. *Child Abuse Review, 26*(6), 425–438.
O'Reilly, M., & Parker, N. (2012). 'Unsatisfactory saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research, 13*(2), 190–197.
Pierce, M. C., Kaczor, K., Acker, D., Webb, T., Brenzel, A., Lorenz, D. J., ... Thompson, R. (2017). History, injury, and psychosocial risk factor commonalities among cases of fatal and near-fatal physical child abuse. *Child Abuse & Neglect, 69*, 263–277.
Platt, D., & Turney, D. (2014). Making threshold decisions in child protection: A conceptual analysis. *The British Journal of Social Work, 44*(6), 1472–1490.
Popay, J., Rogers, A., & Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research, 8*(3), 341–351.
QSR International Pty Ltd (2014). *NVivo qualitative data analysis software; Version 10*.
R Core Team (2015). *R: A language and environment for statistical computing (version 3.2.3)*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.
R v Clark, EWCA Crim 54, Oct. 2nd, 2000.
R v Cannings, EWCA Crim 01, Jan. 19th, 2004.
Reilly, B. M., & Evans, A. T. (2006). Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Annals of Internal Medicine, 144*(3), 201–209.
Stiell, I. G., & Wells, G. A. (1999). Methodologic standards for the development of clinical decision rules in emergency medicine. *Annals of Emergency Medicine, 33*(4), 437–447.
Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349–357.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.
Wood, J. N., Hall, M., Schilling, S., Keren, R., Mitra, N., & Rubin, D. M. (2010). Disparities in the evaluation and diagnosis of abuse among infants with traumatic brain injury. *Pediatrics, 126*(3), 408–414.
Wright, J. N. (2017). CNS injuries in abusive head trauma. *AJR American Journal of Roentgenology, 208*(5), 991–1001.

Appendix 7. Participant information sheet for legal practitioners, police officers, child protection social workers, and pathologists

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

NISCHR
Sefydliad Cenedlaethol ar gyfer Ymchwil Gofal Cymdeithasol ac Iechyd | National Institute for Social Care and Health Research

PARTICIPANT INFORMATION SHEET

1. **Study title**

Exploring the usefulness of a tool to estimate the probability of abusive head trauma in children less than three years of age

2. **Study team**

Researchers: Miss Laura Cowley (PhD student), Mr Matthew Flynn (visiting student). Supervisors: Professor Alison Kemp, Dr Sabine Maguire, Dr Daniel Farewell.

3. **Invitation**

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

4. **What is the purpose of the study?**

Abusive head trauma is one of the most serious types of child abuse and has severe consequences for young children. It is vital that it is identified accurately in order to prevent further harm occurring to children. It is also important that accidental injuries are not wrongly classified as abusive. An evidence-based tool has been developed to estimate the probability of abusive head trauma in head-injured children aged less than three years old. The aims of this study are to explore current practise and decision making in suspected child abuse cases, and to determine the potential usefulness of this tool for assisting with decision making in these cases.

5. **Why have I been chosen?**

You have been recruited for this study as you have been identified as a professional who may come into contact with a child with a head injury where abuse is suspected, and therefore you may potentially benefit from the use of this prediction tool.

Version: 1 **Date:** 22/04/2015

6. **Do I have to take part?**

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. You do not have to consent to be recorded but if you do you will be shown how to turn the recorder off in case you wish to do so at any point. Only the immediate study team (see above) will have access to the recording and transcript.

7. **What will happen to me if I take part / what do I have to do?**

You will be asked to answer questions regarding your decision making in cases involving children with a head injury where abuse is suspected. You will then be shown a tool that has been developed to predict the probability of abuse in head-injured children. The tool will be briefly explained to you, and you will then be asked to answer questions about your opinions on the usefulness of this tool. You will not be required to reveal any identifiable information about any individuals in the cases you may discuss. The interview will last approximately one hour and will be conducted in a location convenient for you.

8. **What about confidentiality?**

All information provided by participants will be treated confidentially and respectfully. Data will be held securely and access will be limited to the immediate study team (see above).

9. **Are there any risks?**

There are no risks involved in taking part in this study. You do not have to talk about any issues you do not want to discuss.

10. **Are there any benefits?**

This research study may not directly benefit you, but you will be shown a potential tool that you may find useful for your decision making if it was rolled out in the future. You will also be given a chance to reflect on your current practise. It will also give us a better understanding of the challenges faced by staff when involved in suspected abuse cases, and the usefulness of our prediction tool in practice.

11. **What will happen to the results of the research study?**

The results of the study will be written up as part of a PhD thesis, which is due to be submitted in September 2018. They will also be submitted for publication in an academic journal, likely to be published in 2016 or 2017 if accepted. The results may also be submitted to a conference as a scientific abstract or poster. Participants will be informed of when and where they can obtain a copy of the published results following publication, and if required and appropriate to do so, the researchers will forward on copies of the publication. Participants will not be identified in any report, conference talk, publication or thesis.

Version: 1 **Date:** 22/04/2015

12. Who is organising and funding the research?

This research is being organised by Miss Laura Cowley. Laura is a PhD student in the School of Medicine, Cardiff University. The PhD is being funded by the National Institute for Social Care and Health Research (NISCHR).

13. What if I am unhappy about any aspect of the study?

If you have any concerns or complaints about any aspect of the study please contact Professor Alison Kemp.

Email: KempAM@cardiff.ac.uk

14. Contact for further information

Should you require any further information or questions about the study, please contact Miss Laura Cowley.

Email: CowleyLE@cardiff.ac.uk

Version: 1 **Date:** 22/04/2015

Appendix 8. Participant information sheet for clinicians

PARTICIPANT INFORMATION SHEET

1. **Study title**

Exploring the usefulness of a tool to estimate the probability of abusive head trauma in children less than three years of age

2. **Study team**

Researchers: Miss Laura Cowley (PhD student)
Supervisors: Professor Alison Kemp, Dr Sabine Maguire, Dr Daniel Farewell.

3. **Invitation**

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

4. **What is the purpose of the study?**

Abusive head trauma is one of the most serious types of child abuse and has severe consequences for young children. It is vital that it is identified accurately in order to prevent further harm occurring to children. It is also important that accidental injuries are not wrongly classified as abusive. An evidence-based tool has been developed to estimate the probability of abusive head trauma in head-injured children aged less than three years old. The aims of this study are to explore current practise and decision making in suspected child abuse cases, and to determine the potential usefulness of this tool for assisting with decision making in these cases.

5. **Why have I been chosen?**

You have been recruited for this study as you have been identified as a professional who may come into contact with a child with a head injury where abuse is suspected, and therefore you may potentially benefit from the use of this prediction tool.

6. **Do I have to take part?**

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

Version: 1 **Date:** 21/08/2015

If you decide to take part you are still free to withdraw at any time and without giving a reason. You do not have to consent to be recorded but if you do you will be shown how to turn the recorder off in case you wish to do so at any point. Only the immediate study team (see above) will have access to the recording and transcript.

7. **What will happen to me if I take part / what do I have to do?**

You will be asked to answer questions regarding your decision making in cases involving children with a head injury where abuse is suspected. You will then be shown a tool that has been developed to predict the probability of abuse in head-injured children. The tool will be briefly explained to you, and you will then be asked to answer questions about your opinions on the usefulness of this tool. You will not be required to reveal any identifiable information about any individuals in the cases you may discuss. The interviews will be audio recorded with your permission. In addition, you will be given six hypothetical scenarios based on fictional data, and asked to provide your estimate of the likelihood of abuse and what child protection action you would take, if any. You will then be shown the score as calculated by the tool for each scenario, and asked whether this would change your estimates or potential actions. The whole study will last approximately forty-five minutes and will be conducted in a location convenient for you.

8. **What about confidentiality?**

All information provided by participants will be treated confidentially and respectfully. Data will be held securely and access will be limited to the immediate study team (see above).

9. **Are there any risks?**

There are no risks involved in taking part in this study. You do not have to talk about any issues you do not want to discuss.

10. **Are there any benefits?**

This research study may not directly benefit you, but you will be shown a potential tool that you may find useful for your decision making if it was rolled out in the future. You will also be given a chance to reflect on your current practise. It will also give us a better understanding of the challenges faced by staff when involved in suspected abuse cases, and the usefulness of our prediction tool in practice.

11. **What will happen to the results of the research study?**

The results of the study will be written up as part of a PhD thesis, which is due to be submitted in September 2018. They will also be submitted for publication in an academic journal, likely to be published in 2016 or 2017 if accepted. The results may also be submitted to a conference as a scientific abstract or poster. Participants will be informed of when and where they can obtain a copy of the published results following publication, and if required and appropriate to do so, the researchers will forward on copies of the publication. Participants will not be identified in any report, conference talk, publication or thesis.

12. **Who is organising and funding the research?**

This research is being organised by Miss Laura Cowley. Laura is a PhD student in the School of Medicine, Cardiff University. The PhD is being funded by the National Institute for Social Care and Health Research (NISCHR).

13. **What if I am unhappy about any aspect of the study?**

If you have any concerns or complaints about any aspect of the study please contact Professor Alison Kemp.

Email: KempAM@cardiff.ac.uk

14. **Contact for further information**

Should you require any further information or questions about the study, please contact Miss Laura Cowley.

Email: CowleyLE@cardiff.ac.uk

Appendix 9. Semi-structured Interview Schedule

Introduction

Hello my name is Laura and I will be interviewing you today. Thank you for being willing to take part in this project. Firstly, I would like to ask you for permission to audio record this interview. The main reasons for this are to ensure that the data collected is detailed and accurate and to facilitate data analysis. I would like to assure you that everything you say will remain completely confidential and only the immediate study team will have access to the audio recording and transcript.

I am going to be asking you some questions about the factors influencing your decision-making in suspected abusive head trauma cases. I will then show you a tool that the research team have developed to estimate the probability of abuse in head-injured children, and will ask you some questions about your thoughts on the usefulness of this tool. Do you have any questions before we proceed?

Participants' perceived role in the decision-making process

Have you ever been involved in a case concerning a child less than three years old with a head injury where abuse was suspected? *(explain what is meant by head injury – intracranial injury identified on neuroimaging)*

- Can you explain a little bit about what your role is in these cases?
- Is it your job to come to a decision as to whether the child has suffered abusive or accidental injury? **If no:** do you form an opinion about this regardless?

Factors influencing decision-making/multidisciplinary working in suspected AHT cases

What factors usually influence your decision-making in a child head injury case where abuse is suspected?

- Clinical factors? History given by caregiver? (*no history of trauma at all? Inconsistent history?*) Proposed mechanism of injury? (*consistent with developmental stage of child/severity of injuries?*) Family history? Child previously known to social services/previously attended hospital for injuries?
- Opinions of social services/police? Opinions of clinicians? Opinions of child abuse paediatricians? Advice from colleagues?
- Can you tell me more about why these particular factors influence your decision making?

- Can you tell me specifically how these factors influence your decision making?

How confident are you making a decision as to whether a head injury has been caused by abuse or an accident?

- Do you find these cases difficult?
- What are the challenges?

What information do you receive from other agencies when you are involved in a case?

- Results of clinical investigations? History given by caregiver? Proposed mechanism of injury? Family history? Child previously known to social services/previously attended hospital for injuries?
- How does this information help you with your decision-making?
- Can you describe your experiences with multidisciplinary working?

Explanation of the PredAHT clinical prediction tool

We have developed a clinical tool to estimate the probability of abusive head trauma in young children with head injuries, based on varying combinations of six clinical features. Each of the six features were included in the tool based on evidence from a systematic review of the literature, and were assigned a different weighting relative to their significance in a statistical model. The tool is intended for clinicians to complete, and we think it may be useful if they could communicate the results to other professionals who are involved in the child protection process, such as pathologists, police officers, child protection social workers, and legal professionals. It is intended for consideration alongside everything else that is known about each case, and should not be used as a diagnostic tool.

Participants' prior knowledge of clinical prediction tools

Are you familiar with clinical prediction/decision rules?

- **If yes:** What is your opinion of them in general?

Evaluations of PredAHT

Do you think PredAHT would be useful for your investigations/practise/decision making?

- Why/why not?
- Could you tell me specifically how it would be useful for you?
- Perhaps you haven't had much/any experience with suspected abusive head trauma cases before?
- Perhaps it would give you more confidence in your decisions?

Can you think of a recent case in which PredAHT would have been useful to you?

- In what way would it have been useful to you?

Can you think of any factors that would make you more or less likely to use PredAHT?

- Perhaps if you knew how it was developed?
- Perhaps if it was supported by your colleagues?
- Perhaps if you were able to use it alongside other evidence?

Do you think there would be any risks involved in using PredAHT?

- What risks?

Do you think PredAHT would assist you in your discussions with other professionals involved in a case?

- How?/Why not?
- Perhaps it would be useful in a strategy meeting?

What do you think about the score being presented in terms of a percentage probability?

- Is there any other way you would want the likelihood of abuse expressed? Why/Why not?
- Would you prefer it to be translated into a low, medium or high likelihood of abuse?
- Would you want to know the estimate of uncertainty around the score (confidence interval)?
- Do you have any other suggestions about how the results should be presented?

For clinicians only

When in the process of your investigations would PredAHT help you?

- At first presentation?
- Do you think it would be useful if you had missing data?
- Would it only be useful once all of the information about the features included in PredAHT was known?
- Would you use PredAHT to direct further examinations for example a skeletal survey?
- If further information became known regarding the features included in PredAHT, would you use it a second time to assess the change in the score?

Who do you think should complete PredAHT?

How do you think PredAHT could be integrated into existing hospital systems?

Interpretation of probabilities and risk categories associated with AHT

How great would the likelihood of abuse have to be in percentage terms for you to take further action?

What would a low, medium or high likelihood of abuse mean to you in percentage terms?

What does the phrase ‘on the balance of probabilities’ mean to you in percentage terms?

Closure

We seem to have covered a great deal of ground and you have been very patient. However do you think there is anything that we have missed out that might be relevant or important? Do you have any other comments about what we’ve discussed or about the research as a whole?

Appendix 10. Consent Form

CONSENT FORM

Title of Project:

Exploring the usefulness of a tool to estimate the probability of abusive head trauma in children less than three years of age

Please initial box

1. I confirm that I have read and understand the information sheet dated .... (version ....) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐

3. I agree to take part in the above study. ☐

4. I consent to the interview being audio-recorded   Yes ☐   No ☐

______________________	______________	__________________
Name of Participant	Date	Signature
______________________	______________	__________________
Researcher	Date	Signature

Version: 1   **Date:** 22/04/2015

Appendix 11. Demographics data collection form

Exploring the utility of a proposed clinical prediction tool to estimate the probability of abusive head trauma in children less than three years of age

Participant Number: *Date:* *Time:*

Location:

Demographics

Gender: Female ☐ Male ☐ **Age:**

18-24 years	☐
25-34 years	☐
35-44 years	☐
45-54 years	☐
55-64 years	☐
65 or older	☐

Full job title:……………………………………………….

Length of time involved in child protection: **……….**years ………. months

Ethnicity:

White: British ☐ Other ……………

Mixed: White & Black Caribbean ☐ White & Black African ☐ White & Asian ☐ Other …………

Black or Black British: Caribbean ☐ African ☐ Other ………………

Asian or Asian British: Indian ☐ Pakistani ☐ Bangladeshi ☐ Chinese ☐ Other…………….

Have you ever had any formal training in child protection?

Yes ☐ No ☐ Undergraduate ☐ Child protection level 1 ☐ Child protection level 2 ☐ Child protection level 3 ☐ Child protection level 4 ☐ Child protection level 5 ☐ Child protection level 6 ☐ In service training <half day ☐ Other ☐ Please specify________________________

Have you ever had any formal training in paediatric head injuries?

Appendix 12. Debriefing form

DEBRIEFING FORM

Exploring the usefulness of a tool to estimate the probability of abusive head trauma in children less than three years of age

Researcher: Laura Cowley

Thank you for participating in this study. The aims of this study are two-fold. The first aim is to explore current practise and decision making in child head-injury cases where abuse is suspected, with various professionals involved in the child protection process. Secondly, we aim to determine amongst these professionals the attitudes towards, acceptability and potential usefulness of a clinical tool to estimate the probability of abusive head trauma in children less than three years of age.

This study is important because identifying abusive head trauma is challenging. There are many professionals that may benefit from knowing the result of a clinical tool that can estimate the probability of abuse, including doctors, children's social care, law enforcement, pathologists, child protection nurses and the judiciary. Previous research has shown that in order for a clinical tool to be effective, its acceptability to those who use it must first be determined.

To our knowledge this is the first study to investigate qualitatively the acceptability of a clinical tool to identify abusive head trauma.

Further Reading

Cowley LE, Morris CB, Maguire SA, Farewell DM, Kemp AM. Validation of a prediction tool for abusive head trauma. *Pediatrics*. 2015;136(2)290-298.

Maguire SA, Kemp AM, Lumb RC, Farewell DM. Estimating the probability of abusive head trauma: a pooled analysis. *Pediatrics*. 2011;128(3):e550-e64.

Maguire S, Pickerd N, Farewell D, Mann M, Tempest V, Kemp AM. Which clinical features distinguish inflicted from non-inflicted brain injury? A systematic review. *Archives of Disease in Childhood*. 2009;94(11):860-7.

Contact Information

If you require any further information about the study please email Miss Laura Cowley at CowleyLE@cardiff.ac.uk

Version: 1 Date: 04/05/2015

Appendix 13. Cardiff University School of Medicine Research Ethics Committee Approval Letter

School of Medicine
Dean Professor John Bligh BSc MBChB MA(Lit) MMEd MD FRCGP HonFAcadMEd
Ysgol Meddygaeth
Deon Yr Athro John Bligh BSc MBChB MA(Lit) MMEd MD FRCGP HonFAcadMEd

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
School of Medicine
Heath Park
Cardiff CF14 4XN
meddean@cardiff.ac.uk
Prifysgol Caerdydd
Ysgol Meddygaeth
Parc Mynydd Bychan
Caerdydd CF14 4XN

Wednesday 27th May 2015

Laura Cowley and Matthew Flynn,
4th Floor, Neuadd Meirionnydd,
School of Medicine, Cardiff University,
Heath Park

Dear Laura and Matthew,

Re: Exploring the utility of a proposed clinical prediction tool to estimate the probability of abusive head trauma in children less than two years of age.

SMREC Reference Number: 15/35

This application was reviewed by the Committee on Wednesday 13th May 2015.

Ethical Opinion

On review, the Committee were happy to grant ethical approval provided that the under 18 age group was removed from the demographic information as it was felt that it was unlikely that any of your participants would be under 18.

Conditions of Approval

The Committee must be notified of any proposed amendments to the methodology and protocols outlined in your submission. Also, any serious or unexpected adverse reactions that may arise during the course of the study must be reported to the Committee.

Documents Considered

Document Type:	Version:	Date Considered:
Application Form	05/05/15	13/05/2015
Project Proposal	05/05/15	13/05/2015
Semi-structured Interview Schedule	V1 04/05/15	13/05/2015
Participant Information Sheet	V1 22/04/15	13/05/2015
Consent Form	V1 22/04/15	13/05/2015
Debriefing Form	V1 04/05/15	13/05/2015

With best wishes for the success of your study.

Yours sincerely,

Dr Jonathan Hewitt
Chair, School of Medicine Research Ethics Committee

cc: Professor Alison Kemp, Dr Sabine Maguire, Dr Daniel Farewell

Appendix 14. Cardiff University School of Medicine Research Ethics Committee Approval Letter following amendment request

School of Medicine
Dean Professor John Bligh BSc MBChB MA(Lit) MMEd MD FRCGP HonFAcadMEd
Ysgol Meddygaeth
Deon Yr Athro John Bligh BSc MBChB MA(Lit) MMEd MD FRCGP HonFAcadMEd

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
School of Medicine
Heath Park
Cardiff CF14 4XN
meddean@cardiff.ac.uk
Prifysgol Caerdydd
Ysgol Meddygaeth
Parc Mynydd Bychan
Caerdydd CF14 4XN

Friday 9th October 2015

Laura Cowley and Matthew Flynn,
4th Floor, Neuadd Meirionnydd,
School of Medicine, Cardiff University,
Heath Park

Dear Laura and Matthew,

Re: Exploring the utility of a proposed clinical prediction tool to estimate the probability of abusive head trauma in children less than two years of age.

SMREC Reference Number: 15/35

This application was reviewed by the Committee on Wednesday 13th May 2015. An amendment request was reviewed and approved on Friday 9th October 2015.

Ethical Opinion

On review, I can confirm that ethical approval has been granted for this study.

Conditions of Approval

The Committee must be notified of any proposed amendments to the methodology and protocols outlined in your submission. Also, any serious or unexpected adverse reactions that may arise during the course of the study must be reported to the Committee.

Documents Considered

Document Type:	Version:	Date Considered:
Application Form	05/05/15	13/05/2015
Project Proposal	05/05/15	13/05/2015
Semi-structured Interview Schedule	V1 04/05/15	13/05/2015
Participant Information Sheet	V1 22/04/15	13/05/2015
Consent Form	V1 22/04/15	13/05/2015
Debriefing Form	V1 04/05/15	13/05/2015
Email to Committee Secretary	21/08/2015	09/10/2015
Project Proposal	21/08/2015	09/10/2015
Example Hypothetical Scenario based on fictional data	21/08/2015	09/10/2015
Debriefing Form	V1 04/05/2015	09/10/2015
Participant Information Sheet	V1 21/08/2015	09/10/2015
Semi-structured Interview Schedule	V1 21/08/2015	09/10/2015

With best wishes for the success of your study.

Yours sincerely,

Dr Jonathan Hewitt
Chair, School of Medicine Research Ethics Committee

cc: Professor Alison Kemp, Dr Sabine Maguire, Dr Daniel Farewell

INVESTORS IN PEOPLE

Registered Charity, 1136855 *Elusen Gofrestredig*

Appendix 15. Excerpt from Reflexive Journal

Journal entry 01/07/2015

Group interview with five lawyers

At first the participants were giving simple yes or no answers, and it was looking like it might be difficult to coax much out of them. As the interview went on, they seemed to open up more, but I think because it was a group interview, they were each waiting for one of the others to answer first, and then they all tended to agree with one another rather than adding their own thoughts. There were two dominant participants, LP[B]1 and LP[B]2, and because they were initially doing most of the talking, perhaps the others felt that they didn't need to say as much.

The first half of my questions are related to how difficult it is to identify abusive head trauma, but some of the participants said that it was not their job to do this, which distracted me slightly. I would have liked to have found out more about how their opinions about each case impact upon their work and how they feel when they suspect that a child has been abused and yet they still have to represent the parents. I asked whether "there are any times when perhaps that judgment has crossed your mind and clouded your opinion of the case". LP[B]2 replied "in every case you wonder whether that has happened …" and LP[B]1 gave a detailed answer but I wish I had explored this further with the other participants.

When the participants were discussing the cases they had worked on, I wanted to know if the tool would have been useful in these cases, but they did not say much about this. When I first asked the question about whether the tool would have been useful in any recent cases they had dealt with, LP[S]1 brought up a case, and described the features that the child had. I asked if LP[S]1 could tell me more about the case but perhaps I should have asked again outright whether the tool would have been useful or not, because he continued to talk about the features and the other aspects of the case without reference to the tool. When I asked if the tool would have been useful considering they represent parents, it was LP[B]1 who answered but I would have liked to have had an answer from LP[S]1. LP[B]1 said that the tool would not have added anything over what they already knew and that it would not help them answer the question of who the perpetrator was.

When discussing the meaning of the phrase 'on the balance of probabilities,' LP[B]1 and LP[B]2 agreed that 50% was the percentage that they used to represent this phrase. I should have followed this up by asking what they would think of a case where the tool predicts a probability of, for example, 58%? LP[S]1 brought up the fact that the tool wouldn't take into account the severity of the injuries, and whether "damage was set out to be done".

Overall I felt that this was a successful interview. The participants appreciated the value of the tool but were unsure whether it would have had any impact on their decision-making in previous cases. Although every effort was made to elicit responses from all members of the group, there were two dominant participants. In future group interviews I will make a more concerted effort to ensure a balance of responses from the participants. In future interviews I will also try harder to ensure that the interview stays on topic, whilst still allowing participants to discuss issues that are important to them.

Appendix 16. Illustration of a snapshot of the thematic matrix for Theme 3: "Participants opinions on how to present the calculated probabilities", with categories as column headings and participants as rows

Theme - Participants' opinions on how to present the calculated probabilities		
	Category	
Participant	% vs. risk categories	Confidence intervals
C1	I quite like the fact that it is a figure because we never really have figures, if this is something that can give you a figure then that is definitely going to be really helpful to you, like I said we all use different words that all sort of mean the same thing, um, but when you've got a number and a percentage everyone knows exactly what that percentage means, so it's a more familiar term I suppose to everybody	I think ultimately it is that percentage
C2	I think stick with the percentages	Yeah, I think that would be a real additive, and I think again, that's the kind of world that people work in more. I mean obviously, they don't necessarily use confidence intervals, but I think there's always that ... it has to be a certain level of certainty as well, but what's the level of uncertainty that one can live with and what's one that you then feel that you're not confident in what it's throwing up for us
C3	the percentage is fine but then it's having an explanation for the meaning of that, isn't it, so say it was 34 percent, then that's something, if three children presented in a similar way, one of them it would be an abusive head trauma or something, it's putting it into, because you're using it in reports for non-medical professionals, or non-scientists, and sometimes we get confused about probability as well, so I don't, I think low, medium, that to me would fall back on the legal definitions of what's needed and so it's beyond reasonable doubt, whatever, and then family causes, (s/l evidence) as to probability, well, it has to be over 50 percent, doesn't it, so I think if you're going to do that, low, medium or high, I don't know, you need to link it to the legal evidence, the evidence required, yeah	We'd have to explain it, you see, we have to stand up in court basically and use it
C4	I would be happy with percentages	Clearly it would be helpful to have confidence intervals.Because again it gives an indication of a range, a rough idea and that's more helpful to particularly the more statistically minded of us
C5	I think the courts would love a number. I think people like numbers don't they? They say, "What's medium? What's a low probability?" That means, actually, that there is a slight chance that it is. I just think it's more dubious. I think it's less clear. Other people understand percentages and they go 50%, more than 50%. That's a bit more likely to be... I think if you were to say highs, mediums and lows, then I think that's confusing matters a bit	Yes. I think that – would that confuse things for people I don't know? But no I think that is statistically correct
C6	I quite like the percentage, personally, I think it just makes it a little bit more tangible, doesn't it, it's very real, because I find it difficult, I always have done, to quantify risk, so if it can be done for me, then yeah, I'd be, because the trouble with low, medium and high is it's, again, it'll mean different things to different people unless you've got very strict criteria of what low means and what medium means and what high means, whereas that becomes a lot, I just think it's a lot realler, if you know what I mean, that's not a very good word, but yeah	Probably, because it's very absolute, say 94.3, well, it's very nice to have that, it is a very stark number, so yeah, it's probably confidence intervals would be good, I'm not good with statistics at all, but yeah, I think a range would be helpful
C7	I do like the number as a percentage	I do like to know the confidence interval as well... if I say this has got an 85% probability, I do want to know that the confidence interval doesn't go down to 43%...I want to know that is quite tight, it would make me feel yeah more confident

Appendix 17. Analytic Framework

Theme	Category	Subcategories & Definitions
Factors influencing decision-making in suspected AHT cases	Professional factors	Participants' perceived role in the decision-making process: discussions regarding the participants' role in making a decision as to whether children in suspected abuse cases have suffered AHT; whether they feel it is within their remit to make such decisions and why; whether they form an opinion about the likelihood of AHT having taken place Reliance on other professionals: any comments relating to a reliance on others to identify AHT or direct participants' decision-making; any reasons why participants may rely on others such as medical professionals, e.g. due to a lack of medical knowledge, for information sharing or for a high quality clinical investigation; any difficulties associated with having to rely on others for information or guidance Multidisciplinary collaboration: any comments about the positive or negative aspects of working with other agencies e.g. discussions about the quality of the relationships between the professional groups; organisational barriers; delays; competing interests; disagreements between professionals Resources: any remarks regarding the availability of resources to support an investigation such as an adequate budget or staff with expertise in child protection work Difficulty making the diagnosis: any remarks about the ease or difficulty in making a diagnosis in suspected AHT cases; any reasons why a diagnosis of AHT may be easy or difficult to make Confidence: discussions regarding how confident the participants' feel working on AHT cases or making decisions regarding AHT; any reasons why participants' may feel confident or not i.e. the amount of experience or training they have had Seeing 'the bigger picture': any comments about having to piece together a 'jigsaw puzzle' of different types of evidence in order to understand the 'bigger picture'; any references to analysing the different components of the investigation or considering a combination of different factors in order to reach a conclusion
	Medical factors	Clinical features: any references to the clinical features that may influence decision-making such as bruising, fractures, burns or bites; any references to the 'triad' of injuries i.e. subdural haemorrhages, retinal haemorrhages and encephalopathy, any references to the medical literature or evidence-base around abuse-related injuries

		Differential diagnoses: any comments about differential diagnoses of AHT, or alternative explanations for injuries, e.g. accidental injury, or medical/genetic conditions such as bleeding disorders, osteogenesis imperfecta, glutaric aciduria etc. Mechanisms of injury: considerations of the manner or circumstance in which injuries may have occurred and how these considerations contribute to decision-making; any comments linking specific clinical features to possible injury mechanisms e.g. bruising associated with impact injuries, rib/chest injuries associated with compression forces Severity of injuries: comments regarding the severity of the injuries suffered by children as a factor affecting decision-making or the investigative process; perceptions of the seriousness of intracranial injuries in young children Dealing with uncertainty: any remarks about managing uncertainty in suspected AHT cases and how the degree of certainty impacts upon decision-making or the investigative/assessment process; discussions about so-called 'grey' cases where there is considerable uncertainty surrounding the diagnosis
	Circumstantial factors	Circumstances surrounding the incident: discussions about the specific circumstances associated with the incident, including any witnesses to the event; details regarding the initial 999 call; examinations of the scene or surface where the incident purportedly occurred; comments about time to presentation at hospital including a delay in presentation; the behaviour of the parents at the hospital or the scene and the parent-child interaction History: any discussions about the explanation for the child's injuries provided by the parents or carer, including whether the history given is consistent with the level of injury or the developmental stage of the child; or consistent across time and between caregivers
	Family factors	Social history: any discussions regarding the social history of the family, including parental drug and alcohol use; parental mental health issues; domestic violence; previous involvement with social services; level of supervision of the child or previous history of neglect; socioeconomic status; and criminal history Impact on the child/family: any discussions regarding the impact that removing a child from the home or accusing a parent of AHT would have on the child and family Working with the family: anything relating to the challenges of working with the family during a suspected AHT case, and the need to act sensitively
	Psychological factors	Personal biases: any remarks relating to disbelief or doubt that parents or carers are capable of inflicting injuries on their children; discussions of biases relating to the education level of the parents, socioeconomic or employment status, family structure or whether the family appears 'troubled'

		Instinct: any allusions to 'professional instincts' with regard to whether AHT has occurred, or instincts about a possible perpetrator, often referred to as a 'gut feeling' Emotional factors: any comments about the emotional or psychological impact of working on suspected AHT cases and how this may affect decision-making
	Legal factors	Identifying the perpetrator: any comments about identifying a potential perpetrator in suspected AHT cases Expert witnesses: any discussions about working with or relying on expert witnesses; comments about disagreements between experts; remarks or interpretations about theories put forward by defence expert witnesses in an attempt to disprove cases
Evaluations of PredAHT	**Potential benefits of PredAHT**	Objectivity: any perceptions of PredAHT as being free of personal biases or evidence-based, and the advantages of this for investigating suspected AHT cases Awareness: any comments regarding PredAHT as useful for heightening awareness of the possibility of AHT (or nAHT), or reinforcing, increasing (or decreasing) concerns or suspicions about possible AHT Reassurance: comments about how PredAHT could provide assurance or confidence that participants' concerns, suspicions or investigations (or lack thereof) are justified; accounts of how PredAHT may be useful to back up or support participants' professional opinions or judgment Rationalization of decisions: any comments regarding PredAHT as useful for helping participants to explain, justify or rationalize their decision-making in suspected AHT cases Standardization of clinical investigation: any comments regarding PredAHT as useful for prompting clinicians to perform a clinical work-up to look for fractures or retinal hemorrhages, or modifying the clinical investigation e.g. by double-checking results Justification for further action: any comments about PredAHT as useful for justifying further action, investigations or assessments, including clinical/social work investigations, requests for charging decisions, or additional resources Contributing to 'the bigger picture': discussions about PredAHT as an additional factor to be considered as part of the wider picture; comments about PredAHT being useful for piecing parts of the clinical information together Communication: discussions about PredAHT as useful for facilitating communication between professionals; comments about whether the participants would share the result of PredAHT with their colleagues; any references to how the scores might be discussed at multi-agency strategy meetings or as part of information sharing

		Training: discussions about the benefits of being aware of the six clinical features included in PredAHT as potential indicators for AHT, or of PredAHT being useful for peer review or training purposes Useful for ‘grey’ cases: discussions regarding how PredAHT may be beneficial when working on ‘grey’ cases, where there is considerable uncertainty surrounding the diagnosis Useful for the less experienced: any comments regarding the potential benefits of PredAHT for those who have had little experience working in the child protection arena
	Potential risks of PredAHT	Over-reliance: any concerns that professionals may place too much reliance on PredAHT to aid their decision-making in suspected AHT cases False reassurance from a low score: any remarks about a ‘low score’ e.g. 14% instilling a false sense of security; concerns that appropriate investigations would not be carried out if a low score was obtained May not be used as intended: any concerns that the tool would be improperly used; concerns that it may be used in isolation, without consideration of other factors relevant to AHT cases such as caregiver provided history or social history; concerns that the tool is too reductionist or crude Accuracy of PredAHT: any comments relating to the accuracy, sensitivity or specificity of PredAHT, discussions about false positives or false negatives and related implications Irrelevant: any comments about PredAHT being irrelevant or not particularly useful for participants’ decision-making, remarks that it may not add much to what is already known Features not included in PredAHT: any important features that the participants’ feel are missing and why e.g. skull fractures Age, number, location, pattern and severity of injuries: discussions regarding any information or details about suspected AHT cases that cannot be taken into account by PredAHT and the impact this may have on decision-making; comments relating to the inability of PredAHT to account for or distinguish between repeated or multiple injuries that may have been sustained over time, e.g. healing fractures; the precise location or pattern of the injuries e.g. posterior rib fracture; the seriousness of the injuries Introduces bias: remarks that PredAHT may condition decision-making or introduce apparent or unconscious bias
	Provisos for the use of PredAHT	If accepted by colleagues: comments regarding the acceptability of PredAHT to colleagues in the medical profession or colleagues in their own or other agencies as a stipulation for use Alongside professional judgement/other factors: any remarks regarding PredAHT being an additional piece of information to make use of in conjunction with other factors relating to the case as well as participants’ professional opinion

		If kept up to date: comments about the need for PredAHT to be regularly updated in light of current evidence Definition of the features: participants' understandings of the features included in PredAHT; remarks about the need for the features to be explicitly defined Understanding how PredAHT works: remarks about the desire to understand how PredAHT was developed, how it works and how it should be used; comments about training requirements Quality of the data: comments about the need to appraise the quality of the underlying data used to derive PredAHT Agreeing accepted risk thresholds: discussions about differing risk thresholds and the need for a consistent interpretation of the scores between colleagues and between agencies
	Practical use of PredAHT	Usability/simplicity: any comments about the ease of use of PredAHT, any potential barriers to completing it e.g. time/complexity Hospital settings it would be useful: comments about the settings in which PredAHT could be used e.g. district hospitals, the paediatric intensive care unit, the emergency department Stages of the assessment process: comments about the relative usefulness of PredAHT at the different stages of the investigative/assessment process; remarks about the best time point to use it Who should complete it: discussions about the best person to be responsible for completing PredAHT e.g. admitting consultant/safeguarding professional Integration into the clinical workflow: any remarks about how PredAHT might be implemented into existing hospital systems; comments about the most appropriate medium by which to use it e.g. computer/phone Prior probability: comments about estimating a prior probability of AHT and whether this would be difficult to do in practice; remarks about needing more information or guidance to complete this aspect of the tool; comments about the impact of incorporating a prior probability of AHT into the tool
	Use of PredAHT in court	Evidence-based: comments about PredAHT being useful in a court setting because it is based on evidence or has been validated Standards of proof: discussions about the value of PredAHT in court in relation to the standards of proof adhered to in different court settings; concern that a high score may equate to the term "beyond all reasonable doubt" and act as a deciding factor in a conviction Irrelevant: comments that PredAHT could not account for every factor in every case and so would be dismissed; comments that PredAHT can only take into account a limited number of clinical features

		Cross-examination: any remarks about counter-arguments that may arise as a result of using the tool in court e.g. criticisms of how the tool was developed, claims that the case falls into the reverse or 'flip' of the probability given by the tool; the need for those presenting the results to understand and explain how the tool was constructed Identifying the perpetrator: comments about the inability of PredAHT to identify a possible perpetrator in suspected AHT cases Historical child protection court cases involving statistical evidence: any discussions about previous court cases in child protection that have involved the use of quantitative tools or probability theory, and the impact of these cases on participants' thoughts about using PredAHT in court Medical court report: remarks about how the score could be useful as part of the wider medical report submitted to the courts
Participants' opinions about how to present the calculated probabilities	**Percentage probabilities vs. broad risk categories:**	Any comments regarding whether the results should be expressed as a precise numerical score, or using broad categories e.g. 'low, medium and high likelihood of abuse'; any explanations for participants' preferred choices
	Confidence intervals	Any discussions about participants' understanding of confidence intervals; any comments about whether or not they should be provided alongside the score and why
	Additional suggestions	Any other suggestions for expressing the results generated by PredAHT, including associated disclaimers, background information or visual aids
Participants' interpretation of probabilities in the context of suspected AHT	**Threshold criteria**	Any comments about participants' accepted probability thresholds for abuse and non-abuse; discussions about thresholds for continuing or stopping investigations/assessments; perceptions of the percentage equivalents and meanings of different risk categories such as 'likely' and 'unlikely'
	Comments on PredAHT scores	Comments about the percentages that PredAHT gives for different combinations of features; participants' opinions of the scores associated with any cases they are describing

Appendix 18. Samples of analysed transcripts from each of the professional groups

Themes, Categories, Subcategories

Legal practitioners

Group interview with Judges: LP[J]1, LP[J]2, LP[J]3

LP[J]3: Now I would agree with LP[J]2 there will sometimes be subtleties and particularly in the expert evidence that we get, and you will have two extremely eminent experts sometimes from different disciplines, sometimes the neurosurgeon has a different view from the radiologist. I can remember doing a case in which they had looked at the same scan and said I don't think we can agree what's there…

Interviewer: Right.

LP[J]3: …you know I don't think that's, that. I think that is that. So those are the difficulties that you have to encounter when you get a range of opinion on the interpretation of the medical evidence, and sometimes that can be further complicated by the experts saying "well in order to give you categorical decisions on this I would have wanted another scan in 28 days because that would then have helped me to be more precise about when this is likely to have happened", or "I could have tied it more neatly to a particular event".

Factors influencing decision-making
Legal factors
Expert witnesses

Interviewer: Yeah.

LP[J]3: What we tend to get quite often in non-accidental head injury cases is you will get the 999 call so in cases where something bad has happened to the child it takes a while from the event to the 999 call, to the ambulance coming, to them getting to the hospital. In other cases you have the panicked parent, you know something's happened to the child, the child has stopped breathing, the child has had a fall or whatever and you can hear you know that things are happening, I mean I once had a case in which I found that it wasn't an inflicted injury, where you could hear the mother saying "oh my god she's being sick", so there were things going on, on the 999 tape that

Factors influencing decision-making
Circumstantial factors
Circumstances surrounding the incident

allowed you to better evaluate the situation, because for us the difficulty is the only people who are there is the person who, if it's an inflicted injury, caused the injury. They don't usually do it when somebody else is looking because the classic is the degree of force that is used is a degree of force that an objective bystander would say "if you carry on doing that this child is going to be injured". So they don't do it when somebody else is present, they may be in the home but they are coming on to the event after it's actually happened.

Factors influencing decision-making
Circumstantial factors
Circumstances surrounding the incident

Interviewer: Yeah.

LP[J]3: So essentially you have to evaluate 'do I believe the account of the person who has said this is how it happened'? Usually that's inconsistent with what the experts are saying could be the progression of the child's condition, and it's a bit like a jigsaw puzzle, you're having to put together a number of different pieces of evidence to see if you can get any closer to the truth. Very often we have to give judgements in which we say "I can't say what happened to the child I cannot make any finding as to the precise facts of the circumstances in which the injury was occasioned, but this was a degree of force that this child ought not to have been subjected to, it was occasioned by 'X' and I therefore regard it as an injury that was deliberately caused in the sense that there was a direct link between the act of an adult and the condition that the child was left with", but sometimes that's as far as you can go.

Factors influencing decision-making
Circumstantial factors
History

Factors influencing decision-making
Professional factors
Seeing 'the bigger picture'

Factors influencing decision-making
Professional factors
Participants' perceived role in the decision-making process

LP[J]2: I agree yeah I agree I mean you know what you're looking for really especially in terms of the 999 calls is appropriate distress and an explanation that fits with what is in the differential diagnosis as to causation I mean that's what you're looking for isn't it? Sometimes you get "nothing happened" literally I mean...nothing happened the child just went down and started to falter, didn't breathe, started to fit...

Factors influencing decision-making
Circumstantial factors
Circumstances surrounding the incident
History
Medical factors
Differential diagnoses

Police Officer: PO1

PO1: …it's just so complicated with child injuries. You know there can be natural explanations for it and of course nobody ever wants to think that parents do this to children.

Factors influencing decision-making
Medical factors
Differential diagnoses
Psychological factors
Personal biases

Interviewer: Yeah.

PO1: So you know, you need something objective telling you, well okay, you might not want to think that these parents do this, but experience tells us they do when these are presented. It's more likely than not that it was inflicted, and I think, you know, perhaps for me, I've seen a few cases so I'm more cynical, but other Detective Inspectors who in the middle of the night get this, you know it would perhaps help them to think along the lines of "actually, I'm suspicious, because chances are this isn't natural causes or accidental".

Evaluations of PredAHT
Potential benefits of PredAHT
Objectivity
Useful for the less experienced

Interviewer: Okay. So if, for example, a clinician came to you and said I've got this case, it's 85% abuse, would you want to know which of the features have gone into deciding that score?

PO1: Kind of, but I wouldn't, um…well I would but I would be getting the full medical report anyway. It might be a post-mortem report. If it was a post-mortem, obviously you get the initial report when you… if it was anything like this you would be doing a forensic post-mortem. So you would be there at the post-mortem getting it first hand from the pathologist then you would have the initial report, which would list all the detail and some interpretation to it. So I wouldn't really need to know too much detail about that.

Interviewer: Yeah.

PO1: And again if it was um, if it was a paediatrician calling me from the hospital about a child who hadn't died, I wouldn't specifically want to know these because I wouldn't be needing to know if there was head or neck bruising. I would want to know about every mark on that child and what it was. Um, the skeletal survey, you know I'd want to know detail of what injuries there were, where they were, so yeah you would want an absolute detailed breakdown of everything medically about the child. Not just whether there was a rib fracture.

Evaluations of PredAHT
Potential risks of PredAHT
Age, number, location, pattern & severity of injuries
Features not included in PredAHT

Interviewer: Okay, yeah. So can you think of a recent case or not necessarily recent, just any case where this might have been useful for you? It may be not so useful in the one with the genetic …maybe in perhaps a non-fatal case?

PO1: Um, I think it might have been useful in that case. Just from our point of view, we were suspicious from as soon as we found the um, the subdural haemorrhage, but um, yeah… I mean the numbers are quite surprising.

Interviewer: Yeah 99.8%.

PO1: Yeah and I'm not even sure that the experienced paediatricians and pathologists would have put it quite as that unlikely to see all these things presented. Um, I'm not sure it would have changed much of what we had done…but it might have made us look at things slightly different. But yeah, it would have been useful just to have that.

Factors influencing decision-making
Medical factors
Clinical features

Participants' interpretations of probabilities in the context of suspected AHT
Comments on PredAHT scores

Evaluations of the tool
Potential risks of PredAHT
Accuracy of PredAHT

Interviewer: Yeah.

Child protection social worker: SW1

Interviewer: Was it your job then to make a decision regarding whether abuse had occurred or not in this case?

SW1: No we were very much led by the paediatricians so whilst we were responsible for safeguarding the child in terms of it then going into the family courts, because whenever we get a case of suspected abuse, a head injury it's very, very likely that we would be going into the court arena and then you know, the judge will then decide whether there needs to be expert witnesses and the paediatricians report will always kind of dictate our decision making, whereas we normally would remove the child and the siblings on a kind of basis of there's no other medical cause for that type of injury, so it has to be, that's our assumption then we would look at safeguarding the siblings as well, even if they haven't got any injuries. So even though we do make decisions about safeguarding, we're led very much by the paediatricians and what we're hoping for is for a paediatrician to step up and be willing to say yes this is non-accidental injury. Sometimes we don't get that, but with this type of trauma normally it's quite clear cut that there's no other explanation.

Factors influencing decision-making
Professional factors
Reliance on other professionals
Medical factors
Differential diagnoses

Factors influencing decision-making
Professional factors
Participants' perceived role in the decision-making process

Interviewer: Okay, so I know you said it's not really your job to make those decisions but do you form an opinion yourself in these cases as to whether it is abuse or not? That's what I'm trying to get at.

SW1: Yes, yes I suppose we look at exploring the family dynamics, the response of the parents during the immediate child protection enquiry, the interview and we look at the dynamics and obviously, you do get a gut feeling sometimes about a perpetrator. Unfortunately in all of my cases that I've had of that type of situation, or with bruising or fractures it's all with unfortunately the father, or the mum's partner and it's sometimes quite obvious in the initial interview that they present in a certain way, and often they won't admit, obviously, and then it comes out much later on, but yeah you

Factors influencing decision-making
Family factors
Social history
Circumstantial factors
Circumstances surrounding the incident
Psychological factors
Instinct

know we go on our professional judgement as well and we look at the whole context, so in that sense we are able to form an opinion that we think it was highly likely that this was abuse because of so and so, it might be the family have got a chronology, a history of involvement with us that we would consider as well. There may be a history of neglect, or domestic violence, we consider all of that as well.

Factors influencing decision-making
Professional factors
Seeing 'the bigger picture'
Family factors
Social history

Interviewer: How difficult do you find it working on these cases? What are the challenges?

SW1: Very, very difficult, I've had a similar case where it was a middle class family, and a child with a head injury and I had a gut feeling that you sometimes feel that these parents didn't do this, or this hasn't happened, and unfortunately because there's no other medical explanation at this time as far as I'm aware for this type of symptom in a baby, we have to remove the child, and sometimes that child won't go back because there's no other evidence, and sometimes you do get this feeling, that actually what if it was something else and actually these parents didn't do it and there's been cases where we've removed children, we're doing it begrudgingly because of medical evidence and you genuinely from the way the parents are with the child, their backgrounds you just don't think. So that's very difficult ethically having to remove a child on the basis of a medical decision where there's nothing else to substantiate that, and that does happen but at the end of the day if there is no other medical evidence, you know you can't take that, that risk. So it is difficult sometimes.

Factors influencing decision-making
Psychological factors
Instinct
Personal biases
Medical factors
Differential diagnoses
Family factors
Impact on the family
Professional factors
Reliance on other professionals
Difficulty making the diagnosis

Pathologist: PA1

Interviewer: Okay, so in these cases then was it your job to make a decision regarding whether abuse had occurred or not?

PA1: Um, only in the sense that the people investigating the deaths are wanting a steer on whether that deceased child was abused or not. So the pathological examination is looking for evidence of injury for starters and once you have found evidence of injury you have to decide what the distribution of injury says in terms of how the injuries might have been sustained.

Factors influencing decision-making
Professional factors
Participants' perceived role in the decision-making process
Medical factors
Mechanisms of injury

Interviewer: So thinking about some of the cases that you've been involved in, can you tell me a little bit about the factors that contributed to your decision making?

PA1: Uh as to whether the injuries that I've found are likely to have been from an assault or something else?

Interviewer: Yes.

PA1: So I um, first of all see whether there is any injury and decide what sort of injury it is, whether it's a blunt force injury or sharp force injury etcetera and then the distribution of the injuries on the body and uh then relate the distribution that I find with what I know about the literature on different patterns of injury for assault or accident, falls. So lots of different potential explanations for an injury or a set of injuries and then my job is really to put all of the pathological findings together in the context of what is known about the case and other peoples' evidence like radiologists, haematologists etcetera, and then try and put it all together and give the investigators a steer on what I think is going on.

Factors influencing decision-making
Professional factors
Seeing 'the bigger picture'
Medical factors
Mechanisms of injury
Clinical features

Interviewer: Yeah okay, so are you able to tell me a little bit about the specific patterns of injuries that you look for when you're trying to make these decisions?

PA1: Uh any injuries which are not confined to bony prominences of the body um, could potentially be from an assault. Specific areas of the body like the ears are difficult to injure um, in a fall for example unless you fall on a piece of furniture with projecting edges um, so injuries on ears in children and particularly in infants are concerning.

Factors influencing decision-making
Medical factors
Clinical features

Interviewer: Yeah.

PA1: Any injuries that have got a specific pattern in them um, because that allows me to make sure that I document those injuries in such a way that if a surface of interest is recovered later on then a comparison can be made later on photographs of each of those surfaces, um, injuries to certain parts of the body which are less commonly injured accidentally, genitalia, anus um, certain injuries on limbs so on the ulna borders of the arms might be from um, defence type injuries. Injuries to fingers, finger nails which have been torn um, but essentially it's not a specific, not a single individual injury that says to me this is assault above any other explanation it's where are the injuries, what sort of injuries are they, bruises, abrasions are they burns do they look like they've been caused by something with a shape to it or a pattern um, and it's an integration of all of those types of injuries where they are and how many are there so it's not one particular thing.

Factors influencing decision-making
Medical factors
Clinical features
Mechanism of injury
Professional factors
Seeing 'the bigger picture'

Interviewer: How difficult do you find it to make a decision regarding you know whether abuse has occurred or not in these cases?

PA1: It's very difficult, and can be impossible based on the pathology alone.

Factors influencing decision-making
Professional factors
Difficulty making the diagnosis

Clinician: C5

C5: I think it would be useful. I think we are able to make those decisions because the reports from the other consultants who are part of this investigative process will also give their opinions. So, you know, you have the ophthalmologist, if there's several retinal haemorrhages...They all say this is indicative of non-accidental injury. The neurologist wants to say, you know, this is indicative, highly suspicious of non-accidental injury. So you've got all these other clinicians who are saying the same thing as you and so you're pretty confident that everybody's saying the same and this can't be anything else based on the story that you've been given. We have never given any numbers for that confidence. When...It would be really useful if when we're talking to parents and explaining, "this is why and how we have come to this decision. This is not a hunch. This is a validated tool and with this combination of features then this is what it is."

Factors influencing decision-making
Professional factors
Reliance on other professionals

Evaluations of PredAHT
Potential benefits of PredAHT
Objectivity

Interviewer: Yeah.

C5: Secondly, in the court arena I think it's going to be really very useful because it's not our hunch against the next doctor's hunch, you know? And I think peoples' general opinion that babies aren't injured by their carers and their parents...You know people don't want to hear that and they certainly don't want to believe it and acknowledge that this is happening, but if you've got a validated tool saying, "actually this is what has happened to this baby because of the other injuries that we've seen" then I think it's going to be a very valuable indeed.

Evaluations of PredAHT
Use of PredAHT in court
Evidence-based
Potential benefits of PredAHT
Objectivity

Factors influencing decision-making
Psychological factors
Personal biases

Interviewer: Okay. That's great. So you're saying that it would be valuable to talk to the parents and...

C5: It would be valuable for talking to the police, social workers...Just to say, "listen we've got this..." Because they will always say to you, "is there anything else it could be? Are we getting this wrong? Are we missing something medical?" I think when you're able

Evaluations of PredAHT
Potential benefits of PredAHT
Communication

to say with a degree of certainty, “no, this is what it is because this is a validated tool. With this combination of injuries this is how confident we can be” then I think it is going to be valuable for them as well.

Evaluations of PredAHT
Potential benefits of PredAHT
Reassurance

Interviewer: Okay, but it would be less valuable for you among your clinical colleagues, is that what you're saying?

C5: Well we’re not using it and I think [pause]. No, I don’t think it’s going to be less valuable. I think we’re pretty happy with what we are saying now because in these situations they’re so awful that everybody’s generally saying the same thing. We rarely have a child with this pattern of investigations and results when somebody’s saying this is definitely not abuse. We don’t have anybody who is dissenting. I can’t think of that happening. Everybody seems to be saying the same thing whenever these cases come about. That’s not to say that this isn’t going to be valuable but I think we’ve probably...Unless you...We can prove that some of these are slipping through the net because we haven’t got a tool. I think we are dealing with them as abusive head injuries anyway.

Evaluations of PredAHT
Potential risks of PredAHT
Irrelevant

Interviewer: Yeah.

C5: But I think having a percentage figure and a number makes us happier that this is what we’re saying.

Interviewer: Okay.

C5: So I don’t think it’s less valuable then but we’re not used to using it. I think we could use it and I think we would find that the confidence that gives us very valuable.

Evaluations of PredAHT
Potential benefits of PredAHT
Reassurance

Appendix 19. Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist

Developed from:

Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

Item number	Guide questions/description	Reported in
Domain 1: Research team and reflexivity		
Personal Characteristics		
Interviewer/facilitator	Which author/s conducted the interview or focus group?	Laura Cowley Methods – Data Collection
Credentials	What were the researcher's credentials? E.g. PhD, MD	MSc Neuropsychology BSc (Hons) Psychology
Occupation	What was their occupation at the time of study?	PhD student Methods – Data Collection
Gender	Was the researcher male or female?	Female
Experience and training	What experience or training did the researcher have?	The researcher received substantial experience with qualitative research methods in her undergraduate and postgraduate degrees, and undertook a number of qualitative research projects as part of these. This experience was supplemented with the following recent training courses: "Interviewing in Social Science Research" (2015), "Qualitative Analysis Software" (2015), "Qualitative Data Analysis" (2016) and "Interpreting and writing up your Qualitative findings" (2016) Methods – Data Collection
Relationship with participants		

Relationship established	Was a relationship established prior to study commencement?	No Methods – Data Collection
Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	Participants were informed that the research study was being conducted as part of the researcher's PhD project via the Information Sheet
Interviewer characteristics	What characteristics were reported about the interviewer/facilitator? e.g. bias, assumptions, reasons and interests in the research topic	The interviewer is a PhD student researching abusive head trauma and considered how her assumptions may influence the interviews and findings Methods – Data Collection
Domain 2: study design		
Theoretical framework		
Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	Thematic analysis based on a general inductive approach Methods – Data Analysis
Participant selection		
Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	Purposive and snowball sampling to identify professionals involved in suspected AHT cases Methods – Participant recruitment
Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	Email, or letters to judges Methods – Participant recruitment
Sample size	How many participants were in the study?	56 Table 1 and Figure 1
Non-participation	How many people refused to participate or dropped out? Reasons?	97 invited 76 registered 56 took part Figure 1
Setting		
Setting of data collection	Where was the data collected? E.g. home, clinic, workplace	Participants' workplace Methods – Data Collection
Presence of non-participants	Was anyone else present besides the participants and researchers?	Yes MF to record field notes Methods – Data Collection
Description of sample	What are the important characteristics of the	Table 1

	sample? e.g. demographic data, date	
Data collection		
Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	The schedule included open-ended questions, prompts and clarifying questions and was piloted with two people Methods – Interview Schedule Development
Repeat interviews	Were repeat interviews carried out? If yes, how many?	No Methods – Data Collection
Audio/visual recording	Did the research use audio or visual recording to collect the data?	Audio recording Methods – Data Collection
Field notes	Were field notes made during and/or after the interview or focus group?	Yes Methods – Data Collection
Duration	What was the duration of the interview or focus group?	45 minutes Methods – Data Collection
Data saturation	Was data saturation discussed?	Yes data saturation was verified using the constant comparative method Methods – Data Analysis
Transcripts returned	Were transcripts returned to participants for comment and/or correction?	Only for two people who declined to be audio recorded Methods – Data Collection
Domain 3: analysis and findings		
Data analysis		
Number of data coders	How many data coders coded the data?	Three Methods – Data Analysis
Description of the coding tree	Did authors provide a description of the coding tree?	The analytic framework is provided in Appendix 3
Derivation of themes	Were themes identified in advance or derived from the data?	Derived inductively from the data Methods – Data Analysis
Software	What software, if applicable, was used to manage the data?	NVivo 10 Methods – Data Analysis
Participant checking	Did participants provide feedback on the findings?	No Methods – Data Analysis
Reporting		
Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each	Quotations were presented and each participant was identified according to their

	quotation identified? e.g. participant number	professional group and participant number Results
Data and findings consistent	Was there consistency between the data presented and the findings?	The use of the constant comparative method ensured that quotations under each theme and category were reviewed for consistency and coherence Results
Clarity of major themes	Were major themes clearly presented in the findings?	All themes and categories identified during data analysis were presented in the results Results
Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Yes, discrepant cases and minor themes are discussed throughout the results Results

Appendix 20. Additional participant quotations

Theme	Category	Sub-category	Finding	Quotation(s)
Factors influencing decision-making in suspected AHT cases	Professional factors	Participants' perceived role in the decision-making process	The self-perceived role of CPSWs and judges in suspected AHT cases is to protect the child from future harm, rather than to determine whether AHT has occurred per se.	"If the injuries were caused, and if they weren't caused, are there other factors which would mean that this child is still at risk of harm? Sometimes I'll have a lot of concern of other factors…even if it was found that the injuries weren't caused, or the judge decided that they weren't caused. Sometimes there's other factors that might mean that the children can't go home." **CPSW 6**
Factors influencing decision-making in suspected AHT cases	Professional factors	Reliance on other professionals	CPSWs and police officers are heavily reliant on medical professionals to come to a decision as to whether AHT has occurred, and to guide their decision-making.	"Even though we do make decisions about safeguarding, we're led very much by the paediatricians and what we're hoping for is for a paediatrician to step up and be willing to say yes this is non-accidental injury." **CPSW 1**
Factors influencing decision-making in suspected AHT cases	Professional factors	Reliance on other professionals	Clinicians and pathologists highlighted that other professional groups shouldn't be relying solely on them to come to a decision about suspected AHT.	"It is the information provided by other people like the safeguarding nurse and the health visitor and the primary care professionals, and the school and police and social services [too]." **Clinician 12**
Factors influencing decision-making	Professional factors	Reliance on other professionals	Many CPSWs, police officers and legal practitioners, including judges, said that decision-making is more difficult when medical professionals are unable	"In all honesty, the only difficulty comes when a clinician sits on the fence." **Police Officer 6**

in suspected AHT cases			to provide them with a clear answer as to whether AHT has occurred or not, or when they will not commit to a view either way.	"One of the difficulties for the forensic process is even the most eminent expert witnesses who come to give evidence will say 'Well in medicine we can never say never'." **Judge 3**
Factors influencing decision-making in suspected AHT cases	Professional factors	Reliance on other professionals	Although clinicians do rely on other agencies to assist them in making decisions in suspected AHT cases, they seek support and advice from clinical colleagues to a greater extent.	"I'm heavily reliant on colleagues within Neurology and Neuroradiology as well as other agencies as well." **Clinician 4**
Factors influencing decision-making in suspected AHT cases	Professional factors	Multidisciplinary collaboration	A handful of participants identified multidisciplinary working as a barrier to decision-making due to competing interests and disagreements between professionals both within and across agencies.	"We can be relatively concerned and social services can go 'I'm not really concerned and I've been working with the family and we know them', which they may well be right about." **Clinician 13**
Factors influencing decision-making in suspected AHT cases	Professional factors	Multidisciplinary collaboration	CPSWs and police officers noted that delays can occur while the other agencies are carrying out their own assessments, which impacts on the overall investigation.	"There have been times when we've been delayed, because we're waiting for the police and it's meant real delay for this family." **CPSW 9**
Factors influencing decision-making in suspected AHT cases	Professional factors	Confidence	Participants' confidence when investigating suspected AHT cases is strongly related to the amount of experience they have.	"I think I am getting to the point in my career where I have had a lot of experience of doing them and I am probably as confident as I am going to get." **Clinician 12**
Factors influencing decision-making	Professional factors	Seeing 'the bigger picture'	In terms of professional decision-making strategies, participants' discussed the importance of 'seeing the bigger picture' in suspected AHT cases,	"I guess it is the whole picture and the intricate details of the whole picture, that often comes in stages. Usually then it is piecing all of that information together to

in suspected AHT cases			and piecing together evidence from various different sources.	see what kind of picture you've got." **Clinician 12** "We tend to equate it to pieces of the puzzle. We bring the pieces of the puzzle to the table, put it together, and when you can see them together you understand the picture." **Police Officer 3**
Factors influencing decision-making in suspected AHT cases	Medical factors	Clinical features	Some participants, particularly police officers, have a high suspicion of AHT when the "triad" of subdural haemorrhages, encephalopathy and retinal haemorrhages is present. However, importantly, these were not the only features that these participants considered when coming to a decision about AHT.	"The thing is, if they've got the triad or whatever, those features and they've come with an intracranial injury and then there's another fracture, and there's no clear alternative medical explanation, they're actually very easy in the sense of the medical diagnosis is clear…child abuse can be easy to diagnose when there's other features, particularly like unexplained fractures and bruising." **Clinician 3**
Factors influencing decision-making in suspected AHT cases	Medical factors	Severity of injuries	A clear factor influencing participants' decision-making is the severity of the injuries sustained. The more clinical features a child has, or the more impaired they are, the more likely participants' are to suspect AHT.	"Physical findings…if they have immediate evidence of severe brain dysfunction, so they are encephalopathic with seizures or they are needing life support, they are severely impaired and needing resuscitation." **Clinician 10**
Factors influencing decision-making in suspected AHT cases	Circumstantial factors	History	The single most important factor that influences participants' decision-making when AHT is suspected is the history, in particular whether the mechanism of injury is consistent with the type and severity of the injuries or the developmental stage of the child.	"If the specific explanation for the injuries doesn't explain the findings then that also makes me more concerned." **Pathologist 1**

Factors influencing decision-making in suspected AHT cases	Circumstantial factors	History	Participants find decision-making difficult when the history is consistent with the injuries, but the mechanism could nevertheless be either abusive or accidental.	"If you've got a parent saying that they've fallen down the stairs then you would know that the injuries would be consistent with that, it doesn't mean to say it happened that way." **Police Officer 2**
Factors influencing decision-making in suspected AHT cases	Circumstantial factors	History	An influential factor is whether the history is consistent over time and/or between caregivers.	"It tends to be the story from the family and whether that story is consistent, so sometimes the story can change over time and it will become more elaborate." **Clinician 15**
Factors influencing decision-making in suspected AHT cases	Family factors	Social history	Participants discussed the importance of the families' social history when investigating suspected AHT cases. They talked about a wide range of issues including parental drug and alcohol use; parental mental health; domestic violence; previous involvement with social services; parent-child interactions; level of supervision of the child; neglect; socioeconomic status; and parental criminal history.	"You have to look at the child in context of the family, so you have to decide if there are any if you like what we would call dynamic risks factors so any parental behaviour which fluctuates for example, drugs, alcohol, mental health issues, domestic violence. Any social services background in relation to that child, or older children." **Judge 3** "We have of course put a lot of weight on the social factors as well concerning the family and what we can learn about them, the background intelligence information." **Police Officer 8**
Factors influencing decision-making in suspected AHT cases	Family factors	Social history	CPSWs and police officers place more emphasis on factors within the family than clinical factors.	"I think those of us who have dealt with several of these and who have sort of made a career of child protection have a better than average understanding of brain injury, but we don't begin to get the complexity of how they might present or how they might

				have arisen. And so whether we like it or not we're really influenced by what else we know. We are." **Police Officer 4**
Factors influencing decision-making in suspected AHT cases	Family factors	Impact on the family	A major factor influencing participants' decision-making in suspected AHT cases is the impact on the family. They discussed the impact of removing a child from the family home, and how intervening in a child's home life could be damaging for the child and family, particularly where a head injury is found to be non-abusive. The decision to remove a child from their parents is not taken lightly, as it may not be the best thing for the child.	"It's a hell of a thing you know to point your finger at the parents of the child and then look at removing the child from the family and you know that's not always necessarily the best thing for the family at all, but this is an overriding need to safeguard." **Police Officer 3**
Factors influencing decision-making in suspected AHT cases	Family factors	Working with the family	Clinicians, CPSWs and police officers find it difficult working with the family and having to treat parents as potential suspects or perpetrators when they are grieving or coping with a seriously unwell child. Participants talked about the need for sensitivity and the potential repercussions of falsely accusing a family of abuse.	"You have to be very careful because it could be something that is not abuse so you've got to be very careful about how you deal with the family as well…you just have to deal with it very sensitively and very carefully." **Police Officer 5**
Factors influencing decision-making in suspected AHT cases	Psychological factors	Personal biases	Participants' decision-making in suspected AHT cases is influenced by their own personal biases, such as a disbelief that parents or carers from 'nice, middle-class families' are capable of inflicting injuries on their children.	"The PRUDiC [Procedural Response to Unexpected Deaths in Childhood] process came to the conclusion there's no concerns basically because the parents were both social class one and professionals, one of whom was a doctor, and basically the attitude seemed to be well this doesn't fit

				the normal demographic for child abuse." **Pathologist 2** "A lot of clinicians wouldn't believe it 'Oh they are such nice parents, they can't have done it, it must have been an accident'." **CPSW 8**
Factors influencing decision-making in suspected AHT cases	Psychological factors	Personal biases	Most participants acknowledge their personal biases and attempt to remain objective in their assessments.	"I think the minor injuries I don't feel confident about because you often end up basing it on other stuff as well; are they already known to social services, or are there previous concerns. But that's not to say those kind of families can't have an accident like anybody else and that somebody who's never met social services hasn't done something in a fit of anger." **Clinician 13**
Factors influencing decision-making in suspected AHT cases	Psychological factors	Personal biases	CPSWs and judges find cases difficult when they only have medical evidence to rely on, and there are no other risk factors that they are able to identify within the family.	"If you've got unimpeachable parents with no history nothing to suggest that they might be predisposed to injuring a child you would then scrutinise the clinical evidence you've got with a fine toothcomb and if the clinical evidence was equivocal so the clinical evidence didn't provide you the answer and there was potentially a benign explanation for it, then on the balance of probabilities it might be possible that it was an inflicted injury, but on the balance of probabilities, it wasn't." **Judge 3**
Factors influencing	Psychological factors	Gut instincts	Clinicians, CPSWs, and particularly police officers are influenced by their	"Obviously, you do get a gut feeling sometimes about a perpetrator." **CPSW 1**

decision-making in suspected AHT cases			"gut instincts" when conducting their investigations and assessments.	"Sometimes what we get is like a case who just presents unresponsive, so where you wouldn't have any of the other features…So at that point, it would be my gut feeling that will help me, I think." **Clinician 22**
Factors influencing decision-making in suspected AHT cases	Legal factors	Expert witnesses	Legal practitioners and especially judges, rely on expert witnesses to provide an interpretation of the clinical features, but noted that there are often disagreements and conflicting opinions between expert witnesses coming from different disciplines.	"There are interesting arguments between the paediatric neurologists and the radiologists. I would always rely on expert evidence, but you do get eminent doctors disagreeing." **Judge 4** "The ophthalmologists were in conflict, the paediatrician was in conflict, the radiologists were, we had a number of conflicting opinions." **Legal Practitioner 1**
Factors influencing decision-making in suspected AHT cases	Legal factors	Expert witnesses	Judges and police officers referred to the various theories that are put forward by the parents or the defence in an attempt to disprove cases of suspected AHT.	"I think the challenge is when we're trying to charge people. For every expert you get that will argue this could only have been done as a result of a really significant, deliberate trauma inflicted on that child, you'll have some equally qualified well-paid clever bod who will say, 'No there's actually all these reasons for it'." **Police Officer 4** "In the weeks running up to trial, we were suddenly served a 62 page defence document which just threw in all types of different scenarios that could have happened…So, there were things sort of thrown to us from the side of various

				hypotheses of how the injuries could have been caused." **Police Officer 2**
Participants' evaluations of PredAHT	Potential benefits of PredAHT	Reassurance	Clinicians, CPSWs and police officers said that PredAHT would provide them with reassurance or confidence that their concerns, suspicions or investigations were justified and that it would be useful to support their professional opinions.	"It's another piece of research, another piece of evidence to back up your concerns." **CPSW 6** "It would definitely give me a higher degree of confidence." **Police Officer 8**
Participants' evaluations of PredAHT	Potential benefits of PredAHT	Standardization of clinical investigation	Clinicians and pathologists suggested that PredAHT may help to standardise or modify the clinical assessment of suspected AHT cases by prompting clinicians to perform investigations such as a skeletal survey or ophthalmoscopy in line with international standards, and to review the results of investigations already undertaken.	"It would be very useful for the paediatricians to tell me the score, as it would cause me to review the CT scans again if the probability came out low." **Clinician 25**
Participants' evaluations of PredAHT	Potential benefits of PredAHT	Standardization of clinical investigation	Pathologists and legal practitioners, including judges, could also appreciate the value of PredAHT for advocating further investigations, even if they would not find it useful themselves.	"I can see a lot of value in terms of screening and education, and the decision-making that happens in living children for child protection services." **Pathologist 3** "In terms of ruling in further investigations I think it's enormously helpful." **Judge 3** "I can see its particular value from the clinicians' point of view about deciding what investigations need to be done." **Legal Practitioner 1**

Participants' evaluations of PredAHT	Potential benefits of PredAHT	Contributing to 'the bigger picture'	PredAHT could contribute to "the bigger picture", as part of a wider information gathering process. Many described PredAHT as a useful addition to the "toolbox", or "a piece of the jigsaw puzzle".	"It is a part of the jigsaw piece rather than the final answer." **Clinician 8**
Participants' evaluations of PredAHT	Potential benefits of PredAHT	Communication	PredAHT could be used at multi-agency meetings or as part of information sharing to facilitate communication about the likelihood of AHT.	"I could see us discussing that with colleagues, discussing that at multi professional strategy meetings definitely." **Police Officer 5**
Participants' evaluations of PredAHT	Potential benefits of PredAHT	Training	PredAHT would be useful for peer review or training. CPSWs in particular thought it was helpful to know that the six clinical features included in PredAHT are potential indicators of AHT.	"I'm just thinking of colleagues actually because the role I'm in now is running peer reviews and I think for some of the newer consultants coming through, and particularly the ones who are maybe not as familiar with our work here; I think this would be quite helpful because they're not always linking [the six clinical features] together." **Clinician 2**
Participants' evaluations of PredAHT	Potential risks of PredAHT	Over-reliance	Professionals may be over-reliant on PredAHT when making decisions in suspected AHT cases.	"Straight off the top of my head, the disadvantage is over reliance isn't it of saying 'Well we've looked at this tool, therefore, that's the way it's gone'. I'd just be worried that you would just become a bit too compliant with it and just complacent and just sort of say 'Well we just always do it this way'." **CPSW 9**
Participants' evaluations of PredAHT	Potential risks of PredAHT	False reassurance	A low score could instil false reassurance, and appropriate investigations might not be carried out in the face of a low probability score.	"What, a 14.7% chance of it being caused then yeah…it's not telling me it's not, it's not saying minus 14.7% so I wouldn't want people less industrious, proactive, or

				passionate about what they do to hide behind it. It's not just the police that could do this. I wouldn't want any partners either to say well that's only come up 14.7%...we had one last week that was 97% we don't need to worry about this one...that would be my only fear but like you say it's not to be used on its own is it? You know there's a wide range of factors." **Police Officer 8**
Participants' evaluations of PredAHT	Potential risks of PredAHT	False reassurance	Reassuringly, all participants said that they would still carry out appropriate investigations if they received a low score from PredAHT, as there may be other features of the case that warrant further enquiry.	"It wouldn't be a case that we would go 'Well it's only 14%, therefore, it's definitely accidental, therefore, that child can go home with the family and we're not going to be worried', we'd still be saying 'Okay well it looks unlikely that this was caused deliberately but we'll still probably be involved with that family while we're doing our own assessment'." **CPSW 9**
Participants' evaluations of PredAHT	Potential risks of PredAHT	May not be used as intended	Concern was expressed that PredAHT wouldn't be used as intended, alongside other known information about each case, and they agreed that it should never be used in isolation from other factors. Several clinicians, pathologists and CPSWs said that PredAHT was too reductionist and crude, comparing it to a box-ticking exercise.	"My one drawback is you don't want to go too much into the tick box thing of 'Well we've done this test, we've done that test, and that's fine the child can go home, or the child has definitely been harmed by the parents', so it has to be used in the right way." **CPSW 9** "I think my main concern is it is quite crude in that the diagnosis of abusive head trauma is not a tick box exercise and the trouble is,

				this reduces it to a tick box exercise." **Clinician 19**
Participants' evaluations of PredAHT	Clinicians' views about the practical use of PredAHT	Usability/simplicity	All clinicians thought that PredAHT is simple to use, and not too time-consuming to complete.	"I think it looks very user-friendly." **Clinician 6**
Participants' evaluations of PredAHT	Clinicians' views about the practical use of PredAHT	Who should complete it	Participants discussed who should complete the tool. PredAHT could be completed by general and community paediatricians and intensivists, but most agreed that it should be completed by a consultant. One clinician thought that it should be a team exercise.	"I worry about making one person in charge of anything, and I think that if we are working in a collaborative manner with the whole team perhaps the whole team should have access to it…I think if you leave it up to one person then the whole point of making a decision as a team is less relevant then." **Clinician 9**
Participants' evaluations of PredAHT	Use of PredAHT in court	Irrelevant	Pathologists and legal practitioners, including judges, felt that PredAHT would be irrelevant because it cannot account for every detail of every case, and each case must be considered based on the entirety of the evidence.	"A percentage won't help because it will be attacked on the basis of you do not have within that percentage all the data that we have in this case." **Pathologist 3**

Appendix 21. Published article arising from the vignette study presented in Chapter 6

Cowley LE, Farewell DM, Kemp AM (2018). Potential impact of the validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A clinical vignette study. *Child Abuse & Neglect*, 86: 184-196.

Child Abuse & Neglect 86 (2018) 184–196

Contents lists available at ScienceDirect

Child Abuse & Neglect

journal homepage: www.elsevier.com/locate/chiabuneg

Research article

Potential impact of the validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A clinical vignette study

Laura E. Cowley*, Daniel M. Farewell, Alison M. Kemp

Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, Wales, United Kingdom

ARTICLE INFO

Keywords:
Abusive head trauma
Child physical abuse
Clinical prediction tool
Child protection

ABSTRACT

Background: The validated Predicting Abusive Head Trauma (PredAHT) tool estimates the probability of abusive head trauma (AHT) in children < 3 years old with intracranial injury.
Objective: To explore the impact of PredAHT on clinicians' AHT probability estimates and child protection (CP) actions, and assess inter-rater agreement between their estimates and between their CP actions, before and after PredAHT.
Participants and Setting: Twenty-nine clinicians from different specialties, at teaching and community hospitals.
Methods: Clinicians estimated the probability of AHT and indicated their CP actions in six clinical vignettes. One vignette described a child with AHT, another described a child with non-AHT, and four represented "gray" cases, where the diagnosis was uncertain. Clinicians calculated the PredAHT score, and reported whether this altered their estimate/actions. The 'think-aloud' method was used to capture the reasoning behind their responses. Analysis included linear modelling, linear mixed-effects modelling, chi-square tests, Fisher's exact tests, intraclass correlation, Gwet's AC_1 coefficient and thematic analysis.
Results: Overall, PredAHT significantly influenced clinicians' probability estimates in all vignettes ($p < 0.001$), although the impact on individual clinicians varied. However, the influence of PredAHT on clinicians' CP actions was limited; after using PredAHT, 9/29 clinicians changed their CP actions in only 11/174 instances. Clinicians' AHT probability estimates and CP actions varied somewhat both before and after PredAHT. Qualitative data suggested that PredAHT may increase clinicians' confidence in their decisions when considered alongside other associated clinical, historical and social factors.
Conclusions: PredAHT significantly influenced clinicians' AHT probability estimates, but had minimal impact on their CP actions.

1. Introduction

It is the responsibility of all clinicians to act upon suspicions of abusive head trauma (AHT), to investigate cases fully, and where necessary to refer cases to children's services. Clinicians from a range of pediatric specialties must piece together all available information and make a decision, based upon the balance of probabilities, about the likelihood of AHT (Colbourne, 2015). Ultimately, distinguishing between AHT and non-abusive head trauma (nAHT) relies on a forensic assessment of the clinical and investigation findings in the context of the history given, and a thorough consideration of the differential diagnoses, and requires a multidisciplinary team approach.

* Corresponding author at: Division of Population Medicine, School of Medicine, Neuadd Meirionnydd, Heath Park, Cardiff University, CF14 4YS, United Kingdom.
E-mail addresses: CowleyLE@cardiff.ac.uk (L.E. Cowley), FarewellD@cardiff.ac.uk (D.M. Farewell), KempAM@cardiff.ac.uk (A.M. Kemp).

https://doi.org/10.1016/j.chiabu.2018.09.017
Received 28 July 2018; Received in revised form 14 September 2018; Accepted 20 September 2018
Available online 09 October 2018
0145-2134/

Table 1
The six features included in the Predicting Abusive Head Trauma clinical prediction tool.

Feature	Description
Head or neck bruising	Any documented bruising to head or neck
Seizures	Any documented seizures from a single seizure to status epilepticus
Apnea	Any apnea documented in the initial history or during inpatient stay
Rib fracture	Any rib fracture documented after appropriate radiologic imaging
Long-bone fracture	Any long-bone fracture documented after appropriate radiologic imaging
Retinal hemorrhage	Any retinal hemorrhage documented after indirect ophthalmologic examination by a pediatric ophthalmologist

Published previously in Pediatrics (Cowley et al., 2015) and reproduced with permission.

In the United Kingdom (UK), Child Abuse Pediatrics is not a clinical subspecialty as in the United States (US). Most cases of suspected AHT are referred to a community pediatrician for expert child protection (CP) advice. These are doctors who have specialist training in CP and safeguarding. Numerous studies have demonstrated variability in clinicians' confidence and experience in identifying child abuse; their perceived likelihood of abuse; the investigations and evaluation strategies used; and diagnostic decisions made (e.g. Anderst, Nielsen-Parker, Moffatt, Frazier, & Kennedy, 2016; Flaherty et al., 2006; Laskey, Sheridan, & Hymel, 2007; Lindberg, Lindsell, & Shapiro, 2008; Wood et al., 2010).

The Predicting Abusive Head Trauma (PredAHT) clinical prediction tool (CPT) was developed to assist clinicians in deciding which children < 3 years old with intracranial injury (ICI), require additional specialist clinical, multidisciplinary and multiagency investigations for possible AHT (Cowley, Morris, Maguire, Farewell, & Kemp, 2015; Maguire, Kemp, Lumb, & Farewell, 2011). It is intended for use by any clinician involved in the evaluation of children where AHT may be considered within the differential diagnosis, alongside their clinical judgment and in combination with all other information about each case. The derivation study gave predicted probabilities of AHT for 64 possible combinations of the presence or absence of six clinical features, detailed in Table 1 (Maguire et al., 2011). In an external validation study the sensitivity of PredAHT was 72.3% and the specificity was 85.7% using a 50% probability cut-off (Cowley et al., 2015).

More recently we developed a computerized version of PredAHT (Cowley et al., 2018). In summary, we used our derivation dataset (Maguire et al., 2011) to estimate the probability of AHT when one or more features were unknown using multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011). We previously used this technique in our validation study as it was found to be the best available approach in comparison to alternative imputation methods (Cowley et al., 2015). We then calculated likelihood ratios for each combination of features, to allow clinicians to incorporate their own prior probability of AHT based on factors unaccounted for by PredAHT e.g. purported history, clinical presentation or psychosocial features. The "baseline" prior probability is 34%, which is simply the prevalence of AHT in the data used to derive the tool. PredAHT thus provides predicted probabilities and likelihood ratios for all 729 potential combinations of the six clinical features, depending on whether each is present, absent *or unknown*. PredAHT could therefore contribute to decision-making at multiple points along the assessment pathway, according to the extent of information available about each of the six features.

There are three main stages to the development of CPTs; derivation, validation, and impact analysis to determine their impact on clinician behavior and patient care (McGinn et al., 2000). In addition, it is recommended that preparatory work is undertaken prior to a formal experimental impact analysis study, to assess the acceptability of the tool and the feasibility of conducting such a study in clinical practice (Wallace et al., 2011). A recent qualitative study concluded that PredAHT was acceptable to a range of CP professionals and could potentially increase professionals' confidence in their decision-making (Cowley et al., 2018). The current study explores the potential impact of PredAHT on clinicians' judgments and decision-making in simulated clinical scenarios.

Experimental vignette methodologies are ideal for analyzing medical decisions that necessitate judgment around sensitive topics and under conditions of uncertainty (Aguinis & Bradley, 2014; Evans et al., 2015). Fictitious yet plausible vignettes, designed through systematic manipulation and control of variables, allow researchers to measure multiple predictors of clinician behavior, maximizing internal and external validity (Aguinis & Bradley, 2014; Evans et al., 2015), and to assess the quality of clinical practice in complex medical situations (Peabody et al., 2004; Rousseau, Rozenberg, & Ravaud, 2015). Using six clinical vignettes, this study aimed to explore the impact of PredAHT on clinicians' probability estimates of AHT, and their proposed CP actions, assessing the rationale behind their responses, and the degree of agreement between clinicians' judgments both before, and after, using PredAHT.

2. Methods

This was a vignette-based cross-sectional survey of clinicians involved in the assessment of young children with ICI, where AHT is amongst the differential diagnosis. The concurrent 'think-aloud' method was used to capture participants' rationale for their responses to the vignette questions. This method instructs participants to articulate their thoughts and feelings as they perform a task, and is based on the assumption that an individual's cognitive processes are directly accessible as verbal data (Ericsson & Simon, 1999). It is often used to study clinicians' diagnostic reasoning and clinical decision-making alongside vignettes (e.g. Skånér, Backlund, Montgomery, Bring, & Strender, 2005; Thackray & Roberts, 2017). The study therefore adopted a convergent mixed methods approach, using qualitative methods to gain a comprehensive understanding of the quantitative results (Creswell, 2013). The study received ethical approval from the Cardiff University School of Medicine Research Ethics Committee (Ref: 15/35).

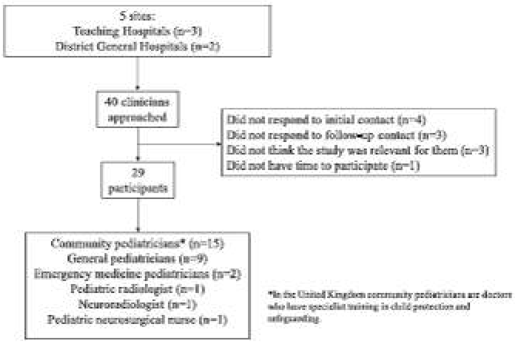


Fig. 1. Flowchart of clinicians participating in a vignette-based study investigating the potential impact of the Predicting Abusive Head Trauma clinical prediction tool.

2.1. *Participant recruitment*

Participants were recruited via email using purposive and snowball sampling. We targeted clinicians across south west United Kingdom (UK) with experience evaluating young children with ICI where AHT is a possible cause. A list of potential participants was identified through personal contacts of the research team who were sent an information sheet to explain the study and asked to suggest clinicians who were eligible to take part. A random selection of 40 clinicians from this list with different levels of CP experience and seniority were then invited to participate (Fig. 1). In this study the term "clinician" refers to medical doctors and specialist nurses, who were sampled from three teaching hospitals and two district general (community) hospitals across a range of specialties including pediatrics, radiology and neurosurgery.

2.2. *Vignette design*

Six clinical vignettes were designed according to methodological recommendations and best practices described in the literature and reported in Supplementary material in Appendix 1 (Aguinis & Bradley, 2014; Evans et al., 2015). All described children < 3 years old with ICI evident on neuroimaging. Table 2 lists the key features of each vignette; Supplementary Table 1 includes the full vignettes as presented to clinicians. Each vignette was comprised of two sections. Section one included the child's age, gender, any history of trauma or social history, and the characteristics of the ICI. Section two included the clinical information required to complete PredAHT, namely; whether the six clinical features were present, absent or unknown.

Two vignettes were based on real cases. One described a child with confirmed AHT ("V1:AHT"), the other a child with confirmed nAHT ("V2:nAHT"). Demographic details were altered to protect the identity of the children. We hypothesized that PredAHT would have the greatest impact on decision-making in "gray" cases, where there is uncertainty surrounding the diagnosis (Chaiyachati, Asnes, Moles, Schaeffer, & Leventhal, 2016). The remaining four vignettes were designed to represent such cases. We created two gray cases ("V3:AHT*" and "V4:nAHT*") by altering elements of the history and social history from "V1:AHT" and "V2:nAHT" but keeping the clinical features the same. Similar approaches have been taken in previous studies (Anderst et al., 2016; Laskey et al., 2007). In "V3:AHT*" the child is older than in "V1:AHT", and it is developmentally plausible that a short fall occurred. The incident was unwitnessed, and the clinical features and severity of the injuries appear discordant with the mechanism of injury (Jenny, 2014; Maguire et al., 2013; Sturm, Knecht, Landau, & Menke, 2009). In "V4:nAHT*", there are inconsistencies within the history, a delay in presentation, plus social concerns within the family that may increase suspicion of AHT in comparison to "V2:nAHT". Two further gray cases ("V5:ICI-only" and "V6:missing") were developed around one of the most challenging clinical scenarios whereby a baby has ICI with no additional clinical features suggestive of abuse. "V6:missing" is almost identical to "V5:ICI-only", but neither skeletal radiology nor ophthalmology examination were undertaken. This vignette was created to explore the effects of missing data and the imputation feature of PredAHT.

2.3. *Data collection*

Written informed consent was obtained from all participants. The researcher explained how PredAHT was developed and validated, and described its various features and intended purpose, to each participant. Participants completed the six vignettes in a random sequence, to account for possible order effects. The data collection procedure is outlined in Fig. 2 and took approximately 45 min. Participants first estimated their own prior probability of AHT for each vignette based on the information given in section 1. They then estimated their Time 1 probability of AHT and Time 1 proposed CP action for each vignette, based on further information given in section 2. The PredAHT score was then calculated for each vignette using the clinicians' prior probabilities, and the clinical details in section 2. Finally, participants estimated their Time 2 probability of AHT and Time 2 proposed CP action for each vignette, after seeing the PredAHT score. CP actions were aligned with three categories of concern (Table 3), as per National Institute for

Table 2
Key features of each of the six clinical vignettes.

Vignette	Information given in Section 1		Information given in Section 2						PredAHT Results	
	Presentation, History and Social History	CT Scan Results	B	A	S	RF	LBF	RH	PredAHT Probability[a]	PredAHT Likelihood ratio
1:AHT	3 months old Lethargy, vomiting No history of trauma	HII affecting both cerebral hemispheres, brainstem and thalami Hyperdense SDH at the vertex	No	Yes	Yes	?	?	Yes	98.4%	118.79
2:nAHT	23 months old No delay in presentation Fall from a chair at a height of 1.5 metres onto a tiled floor Consistent history between parents and over time	Frontal lobe hyperdense SDH Linear, undisplaced skull fracture of left frontal parietal bone	Yes	No	No	No	No	No	14.2%	0.32
3:AHT*	14 months old Lethargy, vomiting No delay in presentation Unwitnessed short fall onto wooden floor Consistent history over time	HII affecting both cerebral hemispheres, brainstem and thalami Hyperdense SDH at the vertex	No	Yes	Yes	?	?	Yes	98.4%	118.79
4:nAHT*	23 months old Six hour delay in presentation to the hospital Initially no history of trauma Possible fall from a chair at a height of 1.5 metres onto a tiled floor Domestic violence concerns Previous children's services involvement	Frontal lobe hyperdense SDH Linear, undisplaced skull fracture of left frontal parietal bone	Yes	No	No	No	No	No	14.2%	0.32
5:ICI-only	3 months old Lethargy, vomiting Rolled off the sofa onto the floor	Multiple small bilateral SDHs	No	No	No	No	No	No	3.7%	0.08
6:missing	3 months old Lethargy, vomiting Rolled off the sofa onto the floor	Multiple small bilateral SDHs	No	No	No	?	?	?	10.4%	0.22

B, head/neck bruising; A, apnea; S, seizures; RF, rib fractures; LBF, long-bone fractures; RH, retinal hemorrhages, PredAHT, Predicting Abusive Head Trauma tool; HII, hypoxic ischemic injury; SDH, subdural hemorrhage.
[a] This was calculated using the "baseline" prior probability of 34%, the prevalence of abusive head trauma in the data used to derive the tool.

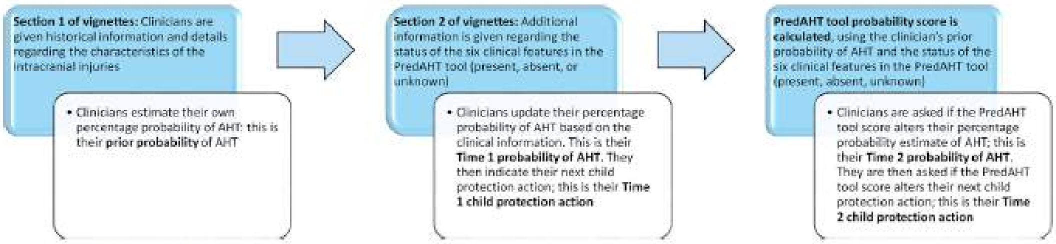


Fig. 2. Flowchart of data collection procedure.
AHT = abusive head trauma, PredAHT = Predicting Abusive Head Trauma tool.

Health and Care Excellence child maltreatment guidelines (National Collaborating Centre for Women's & Children's Health (UK), 2009). Text boxes were included for comments and participants were asked to verbalize their thought processes when deciding on their estimated probabilities of AHT and their proposed CP actions. If participants paused for longer than a few seconds, the researcher reminded them to keep thinking aloud. Otherwise, all interaction between the participants and researcher was minimized so as not to interrupt the participants' flow of thoughts. This enabled participants' verbalizations to be transcribed by the researcher in real time.

Table 3
Possible child protection actions and associated categories of concern in line with National Institute for Health & Care Excellence (NICE) child maltreatment guidelines.

Indicated child protection action		Category
No further child protection action		No concern (abuse excluded)
Investigate further:	Discuss with line manager	Concern (abuse considered)
	Discuss with child protection colleague	
	Gain collateral information from other agencies and health disciplines (e.g. health visitor)	
	Order further investigations (please specify)	
Refer to children's services		Suspicion (abuse suspected)

2.4. Quantitative analysis

Statistical analyses were performed using R software version 3.2.3 (R Core Team, 2015); $p < 0.05$ was considered statistically significant. We used linear modelling and linear mixed effects modelling to analyze the impact of PredAHT on clinicians' probability estimates of AHT, and chi-square tests and Fisher's exact tests to analyze the impact of PredAHT on clinicians' proposed CP actions. We used the intraclass correlation coefficient (ICC) to assess inter-rater reliability between clinicians' probability estimates of AHT, and Gwet's AC_1 coefficient and jackknifing to assess inter-rater reliability between clinicians' proposed CP actions. Further detail is provided in Supplementary material in Appendix 2.

2.5. Qualitative analysis

Participants' verbal data and free text comments were classified into themes using thematic analysis (Braun & Clarke, 2006). Thematic analysis has been used to analyze 'think-aloud' data in a number of studies (e.g. Thackray & Roberts, 2017). Analysis entailed grouping codes into categories, and further arranging categories under overarching themes. This involved six phases including 1) familiarization with the data 2) generating initial codes 3) searching for themes 4) reviewing themes 5) defining and naming themes and 6) writing up the results (Braun & Clarke, 2006). To enhance the trustworthiness and rigor of the thematic analysis, a purposeful approach was adopted (Nowell, Norris, White, & Moules, 2017). The first author developed an analytic framework that was amended as new data were collected; all categories and their definitions are detailed in the framework (Supplementary material in Appendix 3). Findings were discussed at research team meetings; disagreements regarding data interpretation were resolved by consensus. In the interests of reflexivity, the researcher considered how her own values and assumptions as a student involved in developing PredAHT might influence the interpretation of the findings.

3. Quantitative results

3.1. Response rates and participant demographics

All vignettes were completed by 29 clinicians in a fully-crossed design between April–September 2016. Twenty-four of the clinicians also took part in a qualitative study on the acceptability of PredAHT (Cowley et al., 2018). Response rates are shown in Fig. 1. Participant demographics are reported in Table 4.

3.2. Descriptive statistics

There were no missing data and no obvious order effects. Table 5 shows clinicians' mean probability estimates of AHT for each vignette.

3.3. Impact of PredAHT on clinicians' probability estimates of AHT

The PredAHT score significantly influenced clinicians' AHT probability estimates in all vignettes ($p < .001$). Fig. 3 shows the estimated linear model slope coefficients $\hat{\beta}$ and 95% confidence intervals for each vignette. Higher slope coefficients $\hat{\beta}$ indicate a greater impact of PredAHT on clinicians' AHT probability estimates. PredAHT had the greatest impact on clinicians' probability estimates of AHT in "V3:AHT*" and the least impact in "V5:ICI-only". Mixed modelling revealed a significant impact of PredAHT on clinicians' probability estimates of AHT overall across vignettes ($\hat{\beta} = 0.35$, SE = 0.07, $p < .001$, 95% CI 0.21–0.50). PredAHT appeared most influential for those based at teaching hospitals, for those other than general or community pediatricians, for younger clinicians, for clinicians with the least CP experience and no formal training in pediatric head injuries, and for trainee doctors, however these differences were not statistically significant. Variation in the slope coefficients $\hat{\beta}$ was greater between clinicians than between vignettes (Supplementary Fig. 1). This means that the impact of PredAHT was reasonably consistent across vignettes, but varied between individual clinicians.

Table 4
Demographics and characteristics of clinicians participating in the vignette study.

Demographics / Characteristics	Community Paediatricians (N = 15)		General Paediatricians (N = 9)		Other Specialty (N = 5)	
	n	%	n	%	n	%
Gender						
Female	15	100	2	22.2	4	80
Male	0	0	7	77.8	1	20
Age group						
25–34	0	0	1	11.1	1	20
35–44	5	33.3	4	44.4	3	60
45–54	6	40	3	33.3	0	0
55–64	4	26.7	1	11.1	1	20
Ethnicity						
White British	12	80	6	66.7	4	80
White Other	2	13.3	1	11.1	1	20
Indian	1	6.7	2	22.2	0	0
Years in CP						
5–9	3	20	2	22.2	2	40
10–20	4	26.7	3	33.3	1	20
> 20	8	53.3	4	44.4	2	40
CP training						
Yes	15	100	9	100	5	100
No	0	0	0	0	0	0
Paediatric HI training						
Yes	11	73.3	4	44.4	5	100
No	4	26.7	5	55.6	0	0
Hospital Type						
Teaching	11	73.3	5	55.6	5	100
District general	4	26.7	4	44.4	0	0
Seniority						
Consultant	8	53.3	9	100	3	60
Associate specialist	5	33.3	0	0	0	0
Trainee doctor	2	13.3	0	0	1	20
Senior staff nurse	0	0	0	0	1	20

CP = child protection, HI = head injuries.

Table 5
Means, standard deviations, and minimum and maximum values of clinicians' probability estimates of AHT for each of the six vignettes.

	Summary statistic	V1: AHT	V2: nAHT	V3: AHT*	V4: nAHT*	V5: ICI-only	V6: missing
Prior probability	Mean	80.28	32.45	72.34	64.34	78.28	77.93
	SD	(14.54)	(20.00)	(17.16)	(16.94)	(14.90)	(13.20)
	Min–Max	40–98	5–80	30–90	30–95	40–100	50–100
Time 1 probability	Mean	91.31	33.97	89.38	61.41	61.28	78.34
	SD	(9.38)	(21.97)	(12.02)	(19.82)	(24.61)	(13.25)
	Min–Max	60–100	5–90	50–100	30–99	10–100	50–95
Time 2 probability	Mean	95.06	26.72	95.61	54.36	54.55	72.00
	SD	(6.71)	(21.43)	(5.92)	(20.92)	(27.30)	(20.95)
	Min–Max	75–100	0–90	75–100	20–99	10–100	18–100

3.4. Impact of PredAHT on clinicians' proposed CP actions

The majority of clinicians would have referred to children's social care at both Time 1 and Time 2 in all cases except "V2:nAHT", where most clinicians elected to investigate further (Fig. 4). However, 9/29 (31%) clinicians changed their proposed CP action in 11/174 (6%) instances after using PredAHT (Supplementary Fig. 2). Chi-square and Fisher's exact tests revealed no significant associations between a change in action and any demographic variables (age, specialty, hospital type, years of CP experience, pediatric head injury training, seniority). In four instances where their probability of AHT increased after using PredAHT, clinicians escalated

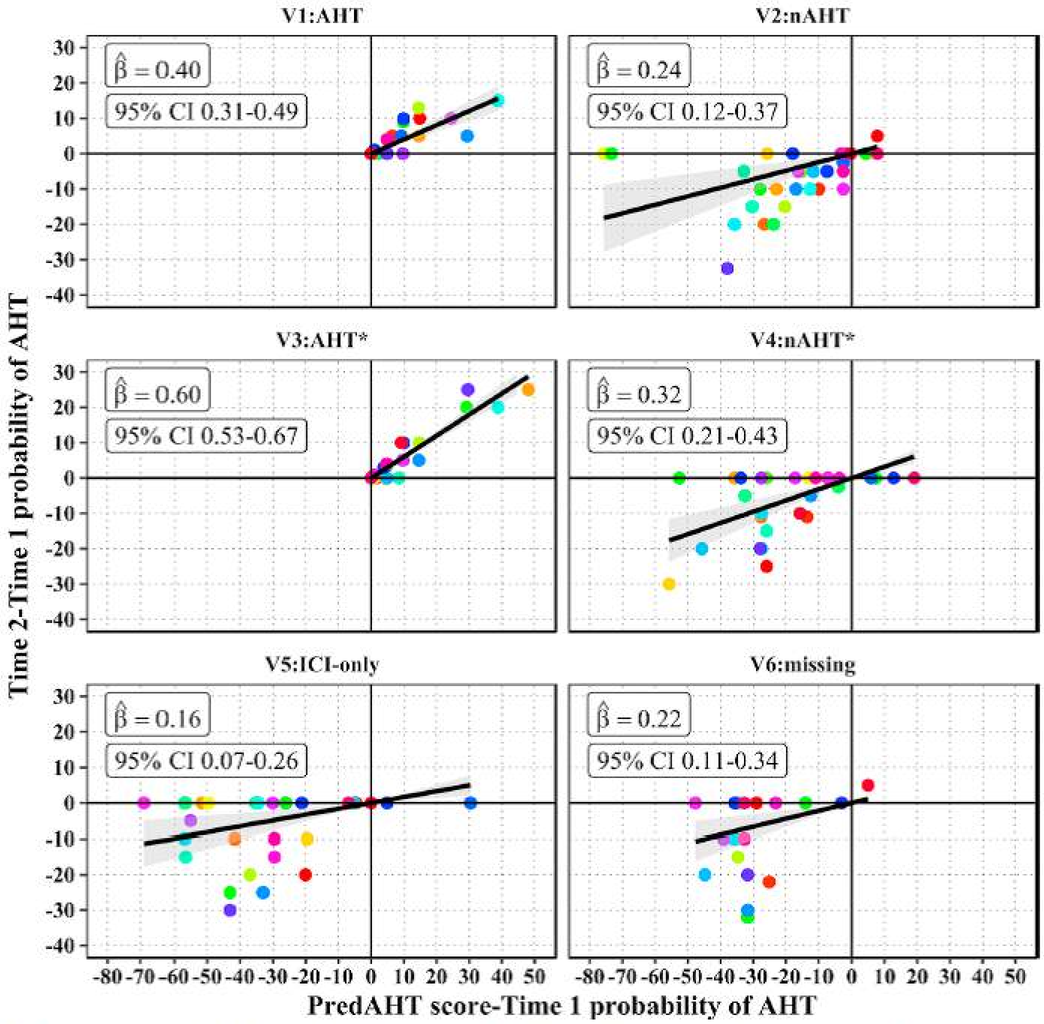


Fig. 3. The impact of the Predicting Abusive Head Trauma tool on clinicians' probability estimates of AHT for each of the six vignettes. Colored dots represent different clinicians. Higher coefficients $\hat{\beta}$ indicate a greater impact of PredAHT on clinicians' probability estimates of AHT. Points at 0 on the x-axis indicate no difference between the clinicians' Time 1 probability estimate and the PredAHT score. Points at 0 on the y-axis indicate no change in clinicians' probability estimates of AHT from Time 1 to Time 2. Points greater than 0 on the y-axis indicate an increase in clinicians' probability estimates of AHT from Time 1 to Time 2. Points less than 0 on the y-axis indicate a decrease in clinicians' probability estimates of AHT from Time 1 to Time 2.

their proposed CP actions. In four instances where their probability of AHT decreased after using PredAHT, clinicians downgraded their proposed CP actions. Three clinicians changed their proposed CP actions despite not altering their own probability estimate of AHT after using PredAHT.

3.5. Inter-rater reliability of clinicians' probability estimates of AHT

Inter-rater agreement of clinicians' prior and Time 1 probabilities was "fair" according to published guidelines (Cicchetti, 1994); prior ICC 0.55 (95% CI 0.31 – 0.88); Time 1 ICC 0.59 (95% CI 0.35 – 0.90). Agreement at Time 2 increased to "good"; ICC 0.66 (95% CI 0.42 – 0.92). However, this difference did not reach statistical significance.

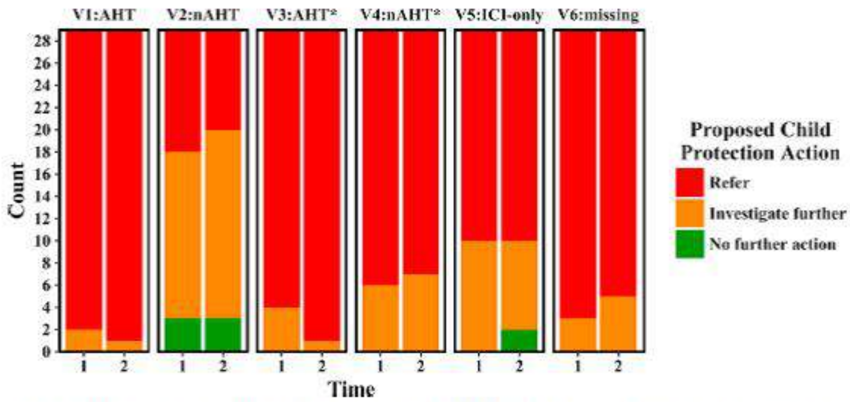


Fig. 4. Clinicians' proposed Time 1 and Time 2 child protection actions for each of the six clinical vignettes.

3.6. Inter-rater reliability of clinicians' proposed CP actions

Inter-rater agreement of their CP actions was "fair" under Gwet's model (Gwet, 2012); Time 1 AC_1 coefficient 0.59 (95% CI 0.33–0.85); Time 2 AC_1 coefficient 0.61 (95% CI 0.32–0.90).

3.7. Impact of clinicians' prior probabilities on the PredAHT score

Supplementary Fig. 3 compares the PredAHT score, given a baseline prior probability of 0.34, with the scores obtained when clinicians' prior probabilities were incorporated. Scores incorporating clinicians' prior probabilities were similar to what would be obtained given the baseline prior for "V1:AHT", "V2:nAHT" and "V3:AHT*". However, PredAHT scores with clinicians' priors were higher than PredAHT scores using the baseline prior for "V4:nAHT*", "V5:ICI-only" and "V6:missing".

4. Qualitative results

Four overarching themes were identified: clinicians' rationale for their responses, evaluations of PredAHT, interpretations of probabilities, and comments on the vignettes. Data are presented using quotations, selected as examples of the themes that were generated from the data.

4.1. Rationale for responses

Clinicians' comments confirmed that they found the "gray" cases difficult to classify: *"I find these really difficult, the 3-month-old rolling off the sofa." Clinician 13, "V5:ICI-only" & "V6:missing".*

For most clinicians, the presence of a concerning social history increased suspicion of AHT in "V4:nAHT*" as compared to "V2:nAHT". This explains why participants' estimated probabilities of AHT were higher on average for "V4:nAHT*" than for "V2:nAHT" even though the clinical features were the same: *"It shows what informs you in these cases, because for me the social services involvement and domestic violence are important." Clinician 20, "V4:nAHT*".* However, some clinicians placed more weight on the lack of additional clinical features concerning for AHT: *"The lack of clinical features is more important to me than the history here." Clinician 25, "V4:nAHT*".*

Almost all of the clinicians were highly suspicious of AHT in "V3:AHT*" due to the concerning clinical features, although the history was potentially less concerning than in "V1:AHT" (no history of trauma). Clinicians stated that the history did not match the severity of the injuries sustained: *"I am not happy with the history as 14-month old children fall a lot and don't get subdural hemorrhages." Clinician 8, "V3:AHT*".*

Clinicians gave reasons as to why they disagreed with the tool and confirmed why PredAHT had the lowest impact on their probability estimates for "V5:ICI-only" and "V6:missing": *"This is where the tool takes away some of the subtlety in the history, this is where I would say I don't care what it says." Clinician 12, "V5:ICI-only".* Although clinicians were informed about the imputation strategy built into PredAHT to account for missing investigations, they were reluctant to change their probability estimates from Time 1 to Time 2 in "V6:missing" because they didn't have the full clinical picture, and stated that PredAHT might act as a prompt for ordering further investigations: *"That's the reason for doing the whole package isn't it because if these things are absent it brings you right down again." Clinician 10, "V6:missing".*

Clinicians' reasons for their estimated probabilities of AHT and proposed CP actions included knowledge of the clinical features indicative of AHT and non-AHT. Clinical knowledge sometimes increased clinicians' suspicions of AHT: *"Retinal hemorrhages would increase my suspicion." Clinician 16, "V1:AHT".* Sometimes it decreased their suspicions: *"I would not be too concerned as the chair is very*

Table 6
Participants' reported that PredAHT increased their confidence in their decision-making in the vignette cases.

Clinician ID and specialty	Vignette ID	Quote
Clinician 16 Community pediatrician	V2:nAHT	*"I would still need more information about the cheek bruising but the low score (24%) would reassure me."*
Clinician 9 Community pediatrician	V3:AHT*	*"The history just doesn't fit with the level of trauma…the score helps to remind you that you are right to be concerned and helps you not to be too sensitive about the family."*
Clinician 10 Community pediatrician	V6:missing	*"The 7% would make me much more confident that this is an accident."*
Clinician 27 General pediatrician	V1:AHT	*"I think mostly where it helps is reassuring you."*
Clinician 16 Community pediatrician	V3:AHT*	*"My estimate is very close to PredAHT, so I wouldn't change my actions but my agreement with PredAHT would give me more confidence in expressing my opinion to multiagency colleagues."*
Clinician 17 Other specialty	V2:nAHT	*"Bruising to the cheeks made me worried but the tool would then reassure me to pull it back down."*
Clinician 25 General pediatrician	V3:AHT*	*"It would be helpful at the end to validate my opinion that probably it is abuse."*

high, it is a linear undisplaced skull fracture and that type of floor is quite a hard floor." *Clinician 16, "V2:nAHT"*. Other times it contributed to their uncertainty about a case: *"Left parietal skull fracture, the most common skull fracture in both abused and non-abused children." Clinician 21, "V2:nAHT"*. Lack of clinical knowledge also contributed to uncertainty in estimating the probability of AHT: *"See this is going into detail about the eye findings some of which I don't know the significance of." Clinician 15, "V1:AHT"*.

Clinicians considered the age and developmental stage of the child when estimating the probability of AHT: *"A three-month old can't roll so the history is immediately suspicious." Clinician 9, "V5:ICI-only" & "V6:missing"*.

An important factor influencing some clinician's probability estimates of AHT was a consistent history. However this was less important when there were other concerning features present: *"Even though the story is consistent the history is still dodgy and the neuroimaging features are suspicious." Clinician 25, "V3:AHT*"*.

When completing the vignettes, clinicians deliberated over the purported mechanism of injury and whether this was consistent with the features observed: *"I would be worried that there's no bruising because that means there's no impact." Clinician 21, "V3:AHT*"*.

4.2. Evaluations of PredAHT

Participants talked about the potential benefits of PredAHT while completing the vignettes. Overall, 27/29 clinicians would find PredAHT useful in their practice: *"This would undoubtedly be extremely useful." Clinician 6*. However two clinicians were unsure: *"I think this would be more useful for older children but I'm not sure it actually adds much." Clinician 15*. Clinicians would find PredAHT useful as they do not usually think in terms of probability when assessing risk: *"I never give percentages, even in court I would say that we don't talk in those terms, and that's why I think the tool is going to be helpful." Clinician 5*.

Many clinicians felt reassured by PredAHT, and reported that it gave them more confidence in their decisions, even if they did not change their CP actions based on the score (Table 6).

Participants also discussed the potential risks of PredAHT. Some thought that variables relating to the history should be included in the tool: *"There's no factor for the lack of history is there which is key isn't it?" Clinician 3*. Others felt that the tool cannot account for the subtleties that are often present in individual cases, or that since it cannot account for all possible indicators for abuse, a low score may provide false reassurance: *"The cheek bruising is really worrying, it shows that PredAHT can't take into account nuances with just yes and no answers." Clinician 9, "V2:nAHT"*. Some participants also discussed at length the need to understand how PredAHT works, and the importance of critically appraising the quality of the data that it is based on: *"We would need to know where the figures in the tool came from, and to make sure they are correct." Clinician 22*.

4.3. Interpretation of probabilities

Participants talked about their probability thresholds for investigation and referral as justification for their proposed CP actions. One clinician would refer all cases she considered to have a 50% risk or greater of AHT to social services, but would investigate cases she thought carried a lower probability of AHT: *"All that matters for referral is whether it's over 50% or not." Clinician 3*.

Many clinicians were interested in exploring the estimated post-test probabilities that PredAHT provided based on different prior probabilities. Some were shocked by the impact the prior probability had on the PredAHT score: *"I'm shocked by how much my prior probabilities have affected the scores. This makes me think I might be too hawkish about abuse." Clinician 26*. However other clinicians justified their high estimated prior probabilities due to the neuroimaging features in the vignettes: *"I can only say a 90% prior probability for all of these vignettes because if there is a subdural hemorrhage, to me that's a really high probability." Clinician 5*. Some questioned how they might estimate their prior probability in practice and mentioned that in reality some of the clinical features included in PredAHT may be incorporated in their prior probability estimates: *"That's interesting then to see how my gut feeling is coming in. Really I'm estimating the prior probability without knowing all the information. What are we taking into account with our prior probability in practice and what is our evidence for that?" Clinician 10*.

4.4. Comments on vignettes

Comments on the details of the vignettes themselves revealed important information about clinicians' behavior when assessing suspected AHT. Some questioned why certain investigations were or were not performed e.g. why a skeletal survey was not performed in "V1:AHT" and "V3:AHT*": *"You would still need to do a skeletal survey even if the probability is already high." Clinician 14, "V1:AHT"*. Other asked why a skeletal survey and ophthalmology exam were ordered in "V2:nAHT": *"I'm not sure I would have done any of these tests in this case!" Clinician 15, "V2:nAHT"*.

Some clinicians talked about additional investigations they would perform: *"I don't know why you keep missing the bloods out!" Clinician 16*. Similarly, many participants reported needing more detail on the history in order to make more informed probability estimates or CP decisions: *"The problem is you would want so much more information. I would assess if they could roll in the department." Clinician 19, "V6:missing"*. In addition some participants wanted more detail on the clinical features in order to assess whether the mechanism was plausible: *"What side is the cheek bruising and is the bruising to the scalp the same side as the head injury?" Clinician 19, "V2:nAHT"*. The majority of clinicians were concerned about the cheek bruising in "V2:nAHT" and "V4:nAHT*", and wanted more information about the pattern and mechanism of the bruising. This explains why some participants' estimated probabilities of AHT were high for "V2:nAHT", despite the fact that this vignette represented a confirmed case of non-AHT: *"It would depend on the pattern of bruising to the cheeks." Clinician 4, "V2:nAHT"*

Finally, while considering the probability of AHT, clinicians discussed a variety of possible differential diagnoses, not detailed in the vignettes, that they would rule out in practice: *"He wouldn't have hit his head that badly just falling on a floor, unless he has got some bleeding disorder or something." Clinician 10, "V3:AHT*"*

5. Discussion

In this vignette study, statistical modelling demonstrated that PredAHT significantly influenced clinicians' AHT probability estimates in all vignettes. Interestingly however, clinicians' proposed CP actions were only influenced by PredAHT in a minority of cases, and PredAHT did not significantly improve the overall agreement between clinicians' AHT probability estimates or their proposed CP actions. Despite this, the 'think-aloud' data showed that 27/29 clinicians would find PredAHT useful in their practice, and that it provided them with greater confidence in their decisions in the vignette cases, confirming the findings of the recent qualitative study on the acceptability of PredAHT (Cowley et al., 2018). However, it was evident that clinicians were influenced by a variety of social, historical and clinical factors in each case, emphasizing the need to consider the PredAHT probabilities in the context of these associated factors.

PredAHT had the greatest impact in "V3:AHT*" and "V1:AHT". This suggests that PredAHT may act to increase clinicians' suspicions when there are several clinical features indicative of AHT and that it may help clinicians to remain objective during their assessment of a young child with ICI. PredAHT had the least impact in "V5:ICI-only", where the history and presentation was concerning, but due to the absence of any additional clinical features, the PredAHT score was low (3.7% at baseline). Reassuringly, this suggests that clinicians were not simply following PredAHT, but were considering factors that it cannot account for. Similarly, a number of clinicians reported disregarding the low PredAHT score (14.2% at baseline) for "V2:nAHT", due to concerns about the cheek bruising, which is a recognized indicator of physical abuse (Kemp, Maguire, Nuttall, Collins, & Dunstan, 2014). Even those who felt reassured by the score would have requested further information about this feature.

Despite being aware of the imputation strategy built into PredAHT to account for missing data, the tool had minimal impact on clinicians' probability estimates in "V6:missing". This highlights the importance of obtaining an ophthalmology exam and skeletal survey whenever AHT is suspected, in line with international recommendations (The Royal College of Ophthalmologists & the Royal College of Paediatrics & Child Health, 2013; The Royal College of Radiologists & the Royal College of Paediatrics & Child Health, 2008). Qualitative analysis suggested that PredAHT may help to standardize investigations in suspected AHT by highlighting the clinical significance of fractures and retinal hemorrhages, and the influence these features, if known, would have on the PredAHT score.

Although PredAHT significantly influenced clinicians' probability estimates of AHT, there were only 11/174 instances where clinicians changed their proposed CP action after seeing the score. For analysis purposes, we collapsed the categories of CP action, however some clinicians who elected to investigate further would have conducted additional investigations after seeing the PredAHT score. With the exception of "V2:nAHT", clinicians mean probability estimates of AHT exceeded 50% at all time points, and many clinicians did not change their actions as they had already elected to investigate/refer to children's services at Time 1.

It was evident that probabilities are interpreted differently by different people, and clinicians have different thresholds on which they act. There is little professional agreement as to what equates to a "reasonable suspicion" of abuse, varying in one study from a probability of 10%–35%, 40%–50% or 60%–70% and for a smaller group to > 75% (Levi & Brown, 2005). In another study, 51% of participants defined the term "reasonable medical certainty" in the context of child abuse as ≥90% probability, 30% defined it as ≤50% probability and 2% used a definition of ≤25% probability (Dias, Boehmer, Johnston-Walsh, & Levi, 2015). Furthermore, Flaherty et al. (2008) found that clinicians only reported 73% of the children that they thought were likely or very likely abused to children's services, and only 24% of children that they thought were possibly abused. Other studies have found that improving clinicians' judgments of disease probability does not necessarily change or improve their treatment decisions and may have an unpredictable effect on clinicians' behavior; one possibility is that clinicians' CP actions in this study were not based on probabilistic thresholds (Poses, Cebul, & Wigton, 1995). This is consistent with the observation that some clinicians changed their CP actions after seeing the PredAHT score, but not their probability estimate of AHT. Alternatively, this finding could suggest that PredAHT may help to reduce the uncertainty around clinicians' point estimates of the probability of AHT, and give them more confidence in their decisions; this was confirmed by the qualitative data, where many clinicians stated that they felt reassured by the tool even if they did not change their proposed CP action.

This study found that clinicians' AHT probability estimates for each vignette varied somewhat. This finding is consistent with other vignette studies evaluating the likelihood of abuse amongst clinicians. One such study asked US pediatricians to rate 16 cases of pediatric traumatic brain injury on a seven point scale ranging from definitive unintentional injury to definitive inflicted injury, and found they were unable to agree on the cause of the injuries in half of the scenarios (Laskey et al., 2007). Lindberg et al. (2008) found extensive variability between experienced CP pediatricians when estimating the likelihood of abuse in video vignettes of cases referred to a hospital child abuse team, using three rating scales and a percentage probability.

The PredAHT score with clinicians' priors was higher than the baseline score for "V4:nAHT*", "V5:ICI-only" and "V6:missing". Allowing clinicians to incorporate their prior probabilities of AHT enables them to take into account factors that PredAHT does not. Although higher prior probabilities may lead to higher PredAHT scores in some cases, this should prompt further investigation and may help to circumvent the possibility of false reassurance provided by a low score. However, it is important that clinicians' prior probabilities are evidence-based, to minimize the possibility of false accusations of abuse. Some clinicians were alarmed by the impact their prior probability of AHT had on the PredAHT score and questioned how they would estimate a prior probability in practice. Taken together, these results reinforce findings from a qualitative study on the acceptability of PredAHT (Cowley et al., 2018), that any training on PredAHT would need to incorporate guidance on estimating a prior probability of AHT.

The actual impact of PredAHT on clinicians' probability estimates of AHT and subsequent CP actions is likely to differ in clinical practice (Reilly & Evans, 2006). It is not yet known whether clinicians will use PredAHT, whether they will use it accurately, or what actions they may take in practice based on specific probability scores. Importantly, PredAHT was designed as an assistive tool, and the qualitative analysis highlighted that whilst clinicians felt that it would be useful in practice to support their decision-making, they also confirmed the importance and value of further essential information. PredAHT is *not* a diagnostic tool, and unlike a directive tool, PredAHT does not recommend a direct course of action based on the results. A directive, validated, highly sensitive 4-variable screening tool for AHT has been developed for use in the pediatric intensive care unit (PICU), to minimize missed cases of AHT and exclude AHT when negative (Hymel et al., 2014). When one or more of four clinical or neuroradiological variables are present in an acutely head-injured infant or young child, a thorough abuse evaluation is recommended. A recent potential impact study of this tool suggested that it may improve the identification of AHT in the intensive care setting (Hymel et al., 2015). However, in order to determine whether PredAHT or the 4-variable PICU tool can change clinician behavior for the better, and to determine their impact on relevant outcomes, formal impact analysis studies are required for both.

5.1. Strengths and limitations

A strength of this study is the use of mixed methods; asking clinicians to articulate the reasoning behind their responses to the vignettes allowed for a meaningful interpretation of the quantitative data. Another strength is that the experimental control afforded by vignette studies permits assessment of the vignette factors' causal effect on the dependent variable. This enhances internal validity compared to traditional surveys (Aguinis & Bradley, 2014; Evans et al., 2015; Steiner, Atzmüller, & Su, 2016). The 'think-aloud' data provided additional evidence of internal validity, because clinicians confirmed that their probability estimates differed as a result of the factors manipulated in the vignettes. Since vignettes differ from real life situations, vignette studies are often criticised due to potential limitations in external validity (Steiner et al., 2016). Clinicians' responses may have been decontextualized from the types of responses they may have made in highly pressured or difficult real life situations, where decision-making does not just depend on a rational analysis of the features of a case. We used only six vignettes, yet there are many more scenarios in which PredAHT could be applied. Most participants were consultants, and half were community pediatricians with considerable CP experience; results may have been different amongst trainee doctors or other specialties involved in the assessment of suspected AHT. To maximize external validity, participants were randomly sampled from a larger pool of potential participants, which extends external validity at least to the target population of clinicians involved in suspected AHT cases (Steiner et al., 2016). The selectivity of vignettes and how this is interpreted by participants can generate valuable data in itself; in this study, clinicians' comments about clinical investigations, elements of the history, or differential diagnoses not detailed in the vignettes revealed insights about the factors influencing their judgments and decision-making in suspected AHT cases.

One possible limitation is that the order of the information presented in the vignettes may not have reflected clinical reality. For example, in practice it is likely that clinicians would have the information regarding apnea, seizures, and head/neck bruising prior to a child undergoing neuroimaging to look for possible ICI, and they may not gather information regarding the social history until later on in the assessment process. The information was presented as such because clinicians' estimated prior probability of AHT should not be based on the clinical features included in PredAHT but on the other features of a case that PredAHT cannot account for. The qualitative results revealed that in reality, it may be difficult for clinicians to estimate a prior probability of AHT excluding the clinical features in PredAHT once the presence or absence of these are already known.

While the impact of PredAHT differed by clinician demographic variables, these findings were not statistically significant; a larger study would be required to further examine these observed trends. Measures were taken to reduce potential subjectivity of qualitative data analysis and bias by involving the research team in data analysis and encouraging researcher reflexivity. Although participants were randomly sampled from a larger list of possible participants, such a sample is not as representative of the population as a probability random sample (Palinkas et al., 2015). Therefore there may be some degree of underestimation of the population variance and overstatement of statistical significance. We do not interpret our p values literally but treat them as a guide for further exploration.

6. Conclusion

This study has demonstrated that PredAHT had a significant impact on clinicians' AHT probability estimates, showing that clinicians are willing to alter their own probability estimate of AHT when exposed to a validated CPT. However, clinicians' proposed CP actions were only influenced by the tool in a minority of cases. Additional research is required to assess the actual impact of PredAHT in clinical practice.

Funding source

This work was supported by Health and Care Research Wales (grant number HS-14-24). The funders had no involvement in the study design, the collection, analysis, or interpretation of the data, the writing of the report or the decision to submit the article for publication.

Declaration of interest

The authors have no actual or potential conflicts of interest, or any financial, personal or other relationships relevant to this article to disclose.

Acknowledgements

We would like to thank Dr Kilem Li Gwet for his advice and direction regarding the statistical analysis of the data, Dr Sabine Maguire for editorial advice, Health and Care Research Wales for funding the study (grant number HS-14-24), and all of the clinicians who took part.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.chiabu.2018.09.017.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*(4), 351–371.

Anderst, J., Nielsen-Parker, M., Moffatt, M., Frazier, T., & Kennedy, C. (2016). Using simulation to identify sources of medical diagnostic error in child physical abuse. *Child Abuse & Neglect, 52*, 62–69.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Chaiyachati, B. H., Asnes, A. G., Moles, R. L., Schaeffer, P., & Leventhal, J. M. (2016). Gray cases of child abuse: Investigating factors associated with uncertainty. *Child Abuse & Neglect, 51*, 87–92.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.

Colbourne, M. (2015). Abusive head trauma: Evolution of a diagnosis. *BCMJ, 57*, 331–335.

Cowley, L. E., Morris, C. B., Maguire, S. A., Farewell, D. M., & Kemp, A. M. (2015). Validation of a prediction tool for abusive head trauma. *Pediatrics, 136*(2), 290–298.

Cowley, L. E., Maguire, S., Farewell, D. M., Quinn-Scoggins, H. D., Flynn, M. O., & Kemp, A. M. (2018). Acceptability of the Predicting Abusive Head Trauma (PredAHT) clinical prediction tool: A qualitative study with child protection professionals. *Child Abuse & Neglect, 81*, 192–205.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage Publications, Inc.

Dias, M. S., Boehmer, S., Johnston-Walsh, L., & Levi, B. H. (2015). Defining' reasonable medical certainty' in court: What does it mean to medical experts in child abuse cases? *Child Abuse & Neglect, 50*, 218–227.

Ericsson, K. A., & Simon, H. A. (1999). *Protocol analysis: Verbal reports as data*Cambridge, MA: MIT Press.

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., ... Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical & Health Psychology, 15*(2), 160–170.

Flaherty, E. G., Sege, R. D., Griffith, J., Price, L. L., Wasserman, R., Slora, E., ... Binns, H. J. (2008). From suspicion of physical child abuse to reporting: Primary care clinician decision-making. *Pediatrics, 122*(3), 611–619.

Flaherty, E. G., Sege, R., Price, L. L., Christoffel, K. K., Norton, D. P., & O'Connor, K. G. (2006). Pediatrician characteristics associated with child abuse identification and reporting: Results from a national survey of pediatricians. *Child Maltreatment, 11*(4), 361–369.

Gwet, K. L. (2012). Benchmarking the agreement coefficient. In K. L. Gwet (Ed.). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters* (pp. 121–147). Gaithersburg, MD: Advanced Analytics, LLC.

Hymel, K. P., Armijo-Garcia, V., Foster, R., Frazier, T. N., Stoiko, M., Christie, L. M., ... Wang, M. (2014). Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics, 134*(6), e1537–e1544.

Hymel, K. P., Herman, B. E., Narang, S. K., Graf, J. M., Frazier, T. N., Stoiko, M., ... Wang, M. (2015). Potential impact of a validated screening tool for pediatric abusive head trauma. *Journal of Pediatrics, 167*(6), 1375–1381 e1.

Jenny, C. (2014). Alternate theories of causation in abusive head trauma: What the science tells us. *Pediatric Radiology, 44*(Suppl. 4), S543–S547.

Kemp, A. M., Maguire, S. A., Nuttall, D., Collins, P., & Dunstan, F. (2014). Bruising in children who are assessed for suspected physical abuse. *Archives of Disease in Childhood, 99*(2), 108–113.

Laskey, A. L., Sheridan, M. J., & Hymel, K. P. (2007). Physicians' initial forensic impressions of hypothetical cases of pediatric traumatic brain injury. *Child Abuse & Neglect, 31*(4), 329–342.

Levi, B. H., & Brown, G. (2005). Reasonable suspicion: A study of Pennsylvania pediatricians regarding child abuse. *Pediatrics, 116*(1), e5–e12.

Lindberg, D. M., Lindsell, C. J., & Shapiro, R. A. (2008). Variability in expert assessment of child physical abuse likelihood. *Pediatrics, 121*(4), e945–e953.

Maguire, S. A., Kemp, A. M., Lumb, R. C., & Farewell, D. M. (2011). Estimating the probability of abusive head trauma: A pooled analysis. *Pediatrics, 128*(3), e550–e564.

Maguire, S. A., Lumb, R. C., Kemp, A. M., Moynihan, S., Bunting, H. J., Watts, P. O., ... Adams, G. G. (2013). A systematic review of the differential diagnosis of retinal haemorrhages in children with clinical features associated with child abuse. *Child Abuse Review, 22*(1), 29–43.

McGinn, T. G., Guyatt, G. H., Wyer, P. C., Naylor, C. D., Stiell, I. G., & Richardson, W. S. (2000). Users' guides to the medical literature: XXII: How to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA, 284*(1), 79–84.

National Collaborating Centre for Women's and Children's Health (UK) (2009). *When to suspect child maltreatment. NICE clinical guidelines, No. 89*. London: RCOG Press July.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*, 1–13.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration & Policy in Mental Health, 42*(5), 533–544.

Peabody, J. W., Luck, J., Glassman, P., Jain, S., Hansen, J., Spell, M., ... Lee, M. (2004). Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Annals of Internal Medicine, 141*(10), 771–780.

Poses, R. M., Cebul, R. D., & Wigton, R. S. (1995). You can lead a horse to water–improving physicians' knowledge of probabilities may not affect their decisions. *Medical Decision Making, 15*(1), 65–75.

R Core Team (2015). *R: A language and environment for statistical computing (Version 3.2.3)*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Reilly, B. M., & Evans, A. T. (2006). Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Annals of Internal Medicine, 144*(3), 201–209.

Rousseau, A., Rozenberg, P., & Ravaud, P. (2015). Assessing complex emergency management with clinical case-vignettes: A validation study. *PloS One, 10*(9), e0138663.

Skånér, Y., Backlund, L., Montgomery, H., Bring, J., & Strender, L. E. (2005). General practitioners' reasoning when considering the diagnosis heart failure: A think-aloud study. *BMC Family Practice, 6*, 4.

Steiner, P. M., Atzmüller, C., & Su, D. (2016). Designing valid and reliable vignette experiments for survey research: A case study on the fair gender income gap. *Journal of Methods and Measurement in the Social Sciences, 7*(2), 52–94.

Sturm, V., Knecht, P. B., Landau, K., & Menke, M. N. (2009). Rare retinal haemorrhages in translational accidental head trauma in children. *Eye, 23*(7), 1535–1541.

Thackray, D., & Roberts, L. (2017). Exploring the clinical decision-making used by experienced cardiorespiratory physiotherapists: A mixed method qualitative design of simulation, video recording and think aloud techniques. *Nurse Education Today, 49*, 96–105.

The Royal College of Ophthalmologists & the Royal College of Paediatrics & Child Health (2013). *Abusive head trauma and the eye in infancy*. London: RCO & RCPCH.

The Royal College of Radiologists & the Royal College of Paediatrics & Child Health (2008). *Standards for radiological investigations of suspected non-accidental injury*. London: RCR & RCPCH March.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.

Wallace, E., Smith, S. M., Perera-Salazar, R., Vaucher, P., McCowan, C., Collins, G., ... Fahey, T. (2011). Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). *BMC Medical Informatics & Decision Making, 11*, 62.

Wood, J. N., Hall, M., Schilling, S., Keren, R., Mitra, N., & Rubin, D. M. (2010). Disparities in the evaluation and diagnosis of abuse among infants with traumatic brain injury. *Pediatrics, 126*(3), 408–414.

Appendix 22. Methodological recommendations and best practices for designing vignette studies

Recommendations for vignette content

Developed from:

Evans, S.C., Roberts, M.C., Keeley, J.W., Blossom, J.B., Amaro, C.M., Garcia, A.M,…Reed, G.M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical & Health Psychology, 15*(2), 160–70.

Recommendation number	Vignettes should	Reported:
1	Derive from the literature and/or clinical experience	Vignette design
2	Be clear, well-written and carefully edited	Vignette design; vignettes were reviewed by supervisory team and edited accordingly, and piloted before use. See vignettes
3	Not be longer than necessary (typically between 50 and 500 words)	See vignettes
4	Follow a narrative, story-like progression	See vignettes Initial information presented in section 1 followed by additional clinical details in section 2
5	Follow a similar structure and style for all vignettes in the study	See vignettes All vignettes followed a similar style and structure
6	Use present tense (past tense only for history and background information)	See vignettes All written in present tense
7	Avoid placing the participant "in the vignette" (e.g. as first or third-person character)	See vignettes Participants were not "placed in the vignette" but were asked to answer survey questions as they would in clinical practice
8	Balance gender and age across vignettes	See vignettes,
9	Be as neutral as possible with respect to cultural and socio-economic factors, unless these are included among the experimental variables	See vignettes Cultural and socio-economic factors were not included as variables nor mentioned

10	Resemble real people, not a personification of a list of symptoms or behaviours	See vignettes
11	Be relatable, relevant, and plausible to participants	Vignette design. The vignettes were piloted and were felt to be clear and to reflect plausible cases. This was further confirmed by the ‘think-aloud’ technique
12	Avoid “red herrings”, misleading details, and bizarre content	See vignettes There was no misleading or bizarre content, vignettes were designed to represent plausible cases
13	Highlight the key variables of interest, facilitating experimental effects	Changes to key variables were indicated in italics
14	Facilitate participant engagement and thinking by including vague or ambiguous elements	Four vignettes were designed as “grey” cases to introduce uncertainty into the decision and stimulate reasoning. This was confirmed by the ‘think-aloud’ technique
15	Cover all pertinent variables (or omit selected variables for specific purposes)	It was not possible to cover all pertinent variables, the omission of certain information led to useful insights in itself, as confirmed by the qualitative analysis of the ‘think-aloud’ data

Best practice recommendations for designing and implementing experimental vignette methodology studies

Developed from:

Aguinis, H., & Bradley, K.J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*(4), 351–371.

Item number	Guide questions/description	Reported:
Planning an EVM study		
Decision Point 1	Deciding whether EVM is a suitable approach	Introduction
Decision Point 2	Choosing the type of EVM	Paper people study
Decision Point 3	Choosing the type of research design	Within-person fully-crossed design where all clinicians completed all vignettes
Decision Point 4	Choosing the level of immersion	Written vignette only
Decision Point 5	Specifying the number and levels of the manipulated factors	Three factors each with two levels (Concerning history yes/no, concerning social history yes/no, missing data yes/no)
Decision Point 6	Choosing the number of vignettes	Six
Implementing an EVM study		
Decision Point 7	Specifying the sample and number of participants	Clinicians from a variety of specialities involved in suspected AHT cases. 40 were approached to take part
Decision Point 8	Choosing the setting and timing for administration	At the participants workplace in a single session
Decision Point 9	Choosing the best method for analysing the data	Linear models and linear mixed effects models
Reporting results of an EVM study		
Decision Point 10	Choosing how transparent to be in the final presentation of results and methodology	See methods and results. Full vignettes provided plus detailed description of their derivation, the analysis and the results

Appendix 23. Consent form used in the vignette study

CONSENT FORM

Title of Project:

Exploring the utility of a proposed clinical prediction tool to estimate the probability of abusive head trauma in children less than three years of age

Name of Researcher:

Please initial box

1. I confirm that I have read and understand the information sheet dated .... (version ....) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐

3. I agree to take part in the above study. ☐

4. I consent to direct, anonymised, quotes being used as part of a thesis, publication in an academic journal, or conference abstract, presentation or poster. ☐

5. I consent to being audio-recorded Yes ☐ No ☐

Name of Participant	Date	Signature
Researcher	Date	Signature

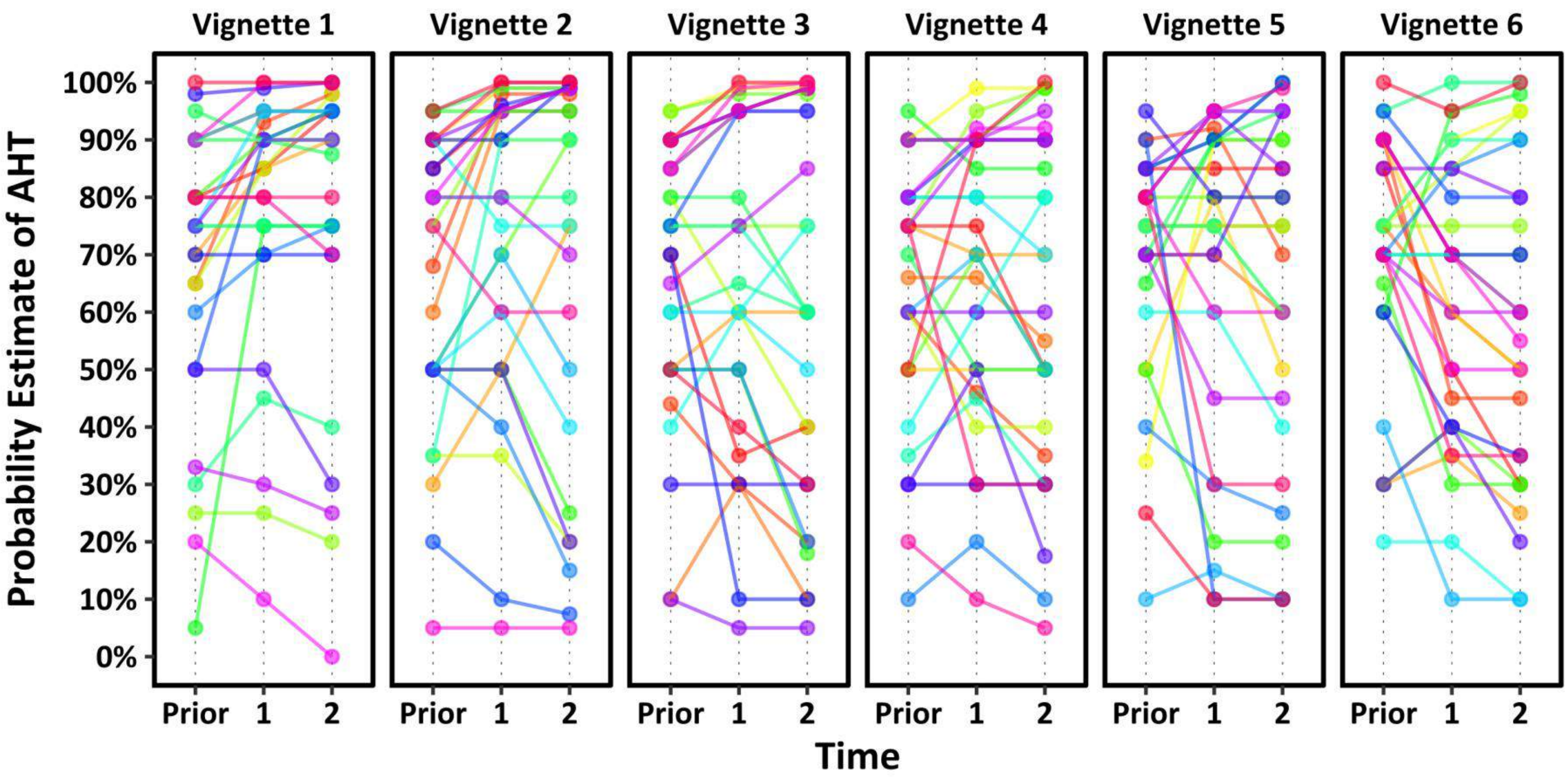
Version: 3 **Date:** 19/01/2016

Appendix 24. Analytic framework

Theme	Category	Definitions
Rationale for responses	**Gray cases difficult**	Any comments about the ease or difficulty of estimating the probability of AHT or deciding on proposed child protection actions for the "gray" cases; any reasons why the "gray" cases were difficult to classify
	Impact of social history	Any comments about the impact the social history had on participants estimated probabilities or proposed child protection actions; any comparisons between "V2:nAHT" and "V4:nAHT*"
	History doesn't match level of trauma	Any discussions about the impact the history had on participants estimated probabilities or proposed child protection actions in "V1:AHT" and "V3:AHT*"; any comparisons between "V1:AHT" and "V3:AHT*"
	Agreement/disagreement with tool	Any reasons why participants disagreed with the PredAHT score and therefore did not change their probability estimates or proposed child protection actions at Time 2. Any reasons why participants agreed with the PredAHT score and therefore did change their probability estimates or proposed child protection actions at Time 2
	Knowledge of clinical features	Any comments about the impact participants' knowledge of the clinical features indicative of AHT and nAHT had on their probability estimates or proposed child protection actions
	Developmental stage	Any considerations about the child's age and developmental stage when completing the vignettes
	Consistent history	Any discussions about the impact a consistent or inconsistent history had on participants probability estimates or proposed child protection actions
	Mechanism of injury	Any considerations about the proposed mechanism of injury and whether this was consistent with the clinical features and level of trauma observed
Evaluations of PredAHT	**Potential benefits**	Any discussions about whether PredAHT would be useful for participants in their clinical practice and why; comments about how PredAHT might help participants to quantify risk; comments about how PredAHT could reassure participants that their suspicions (or lack thereof) are justified and provide them with confidence in their opinions
	Potential risks	Any discussions about the potential risks or downsides of using PredAHT including comments about important features missing from PredAHT; comments about potential false reassurance from a low score; comments about how PredAHT cannot take into account all potential indicators of abuse or

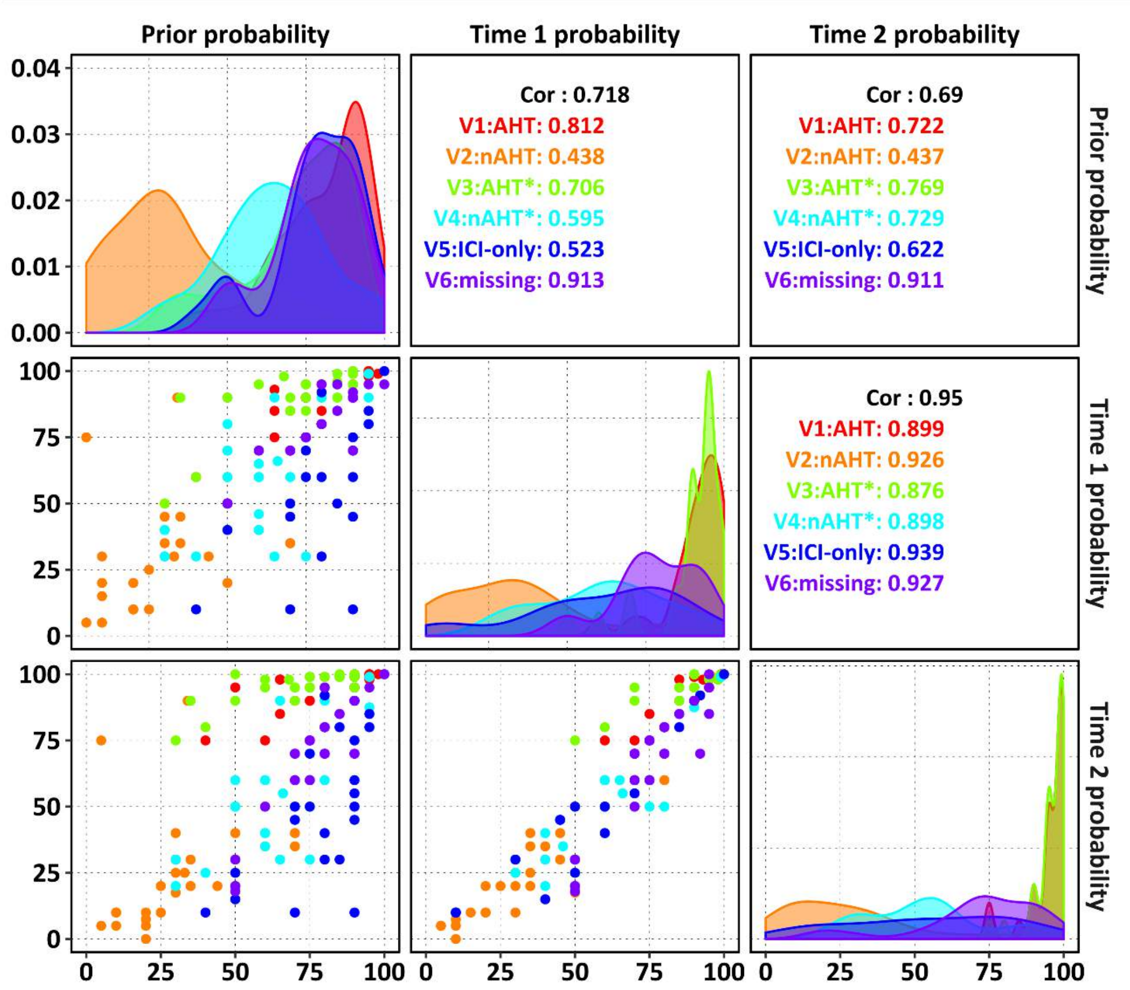
		nuances in individual cases; comments about the need to understand and explain how PredAHT works, and appraise the quality of the underlying data and the accuracy of the scores
Interpretations of probabilities	**Threshold probability**	Any comments about participants' accepted probability thresholds for investigation and referral of suspected AHT
	Impact of the prior probability	Any discussions about the impact participants estimated prior probabilities had on the post-test probability provided by PredAHT; any reasons participants gave for their prior probabilities; discussions about how participants would estimate a prior probability in practice and the information they would use to do this
Comments on details of the vignettes	**Investigations**	Any comments about why certain investigations were or were not performed; comments about additional investigations participants would order that are not detailed in the vignettes
	Detail of the history/clinical features	Any discussions about needing additional detail about the history or clinical findings in order to estimate the probability of AHT, including the age and pattern of clinical findings or more detail on the proposed mechanism of injury
	Differential diagnoses	Any comments about the differential diagnoses, not detailed in the vignettes, that participants would rule out in practice

Appendix 25. Visual inspection of the raw data based on the order the vignettes were completed by participants

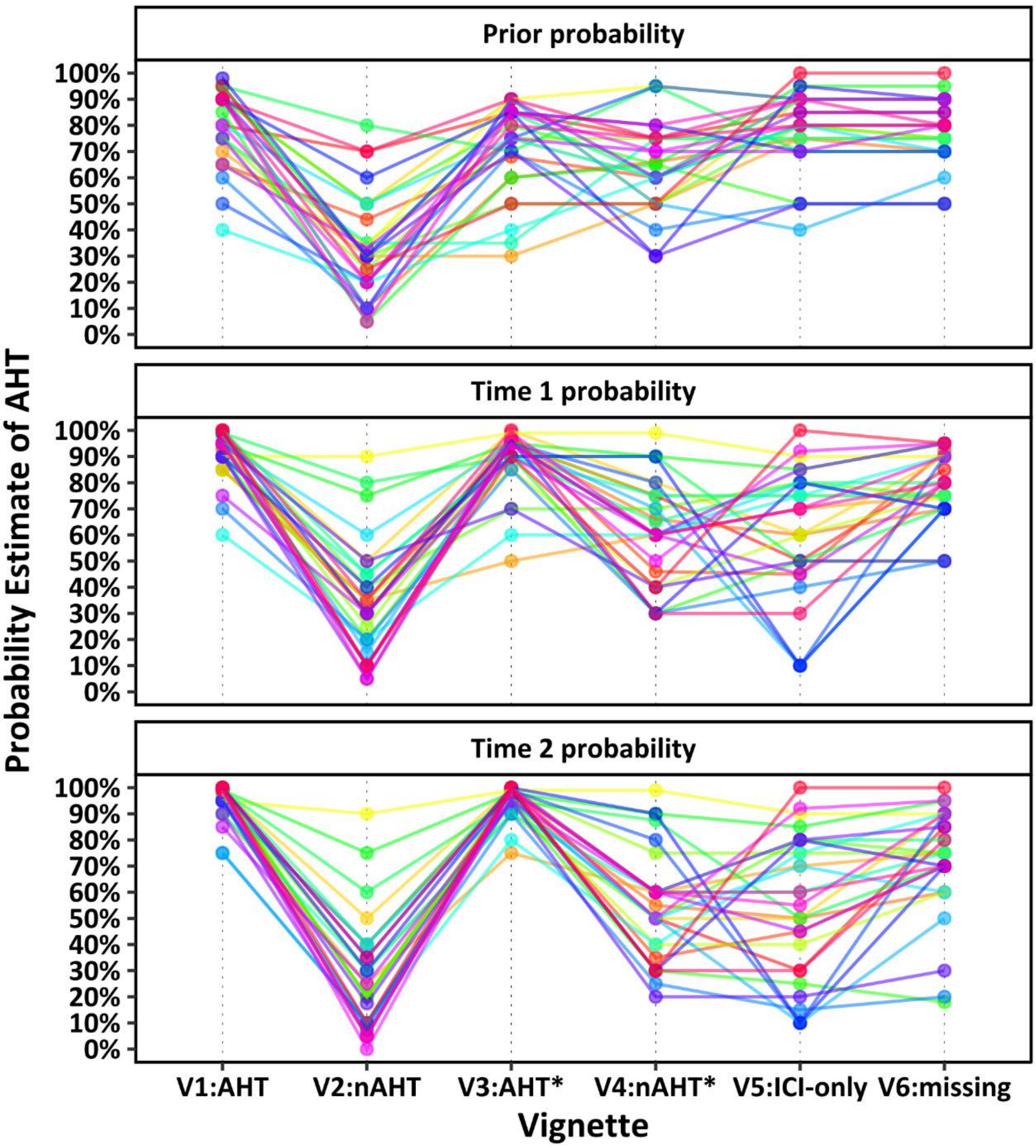


Each coloured line represents a different clinician. Vignette numbers represent the order in which the vignettes were completed by the participants.

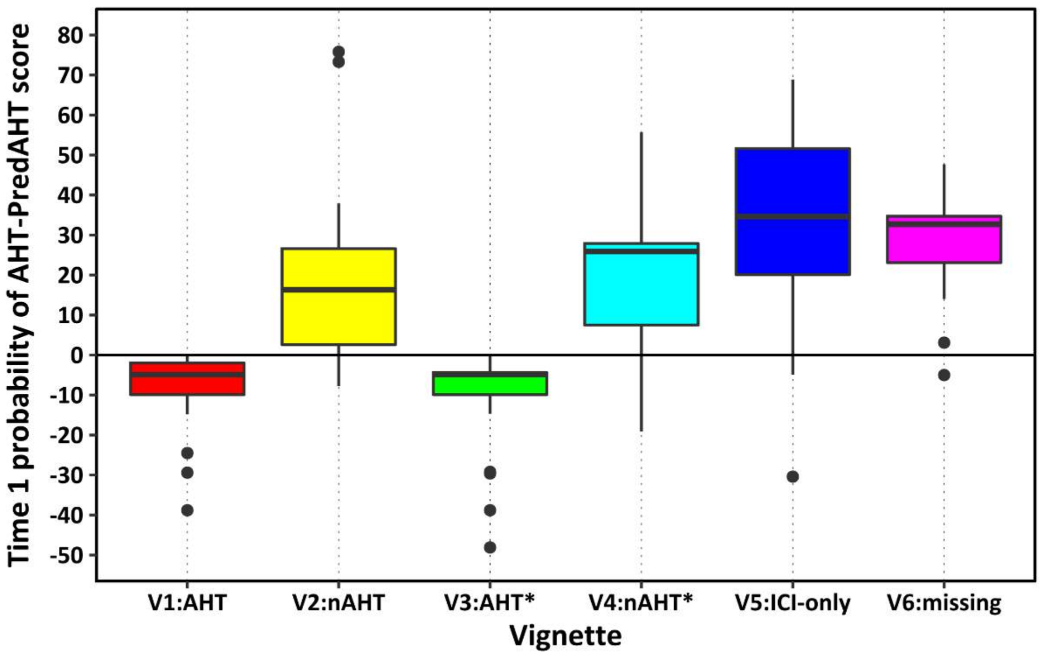
Appendix 26. Scatterplot matrix



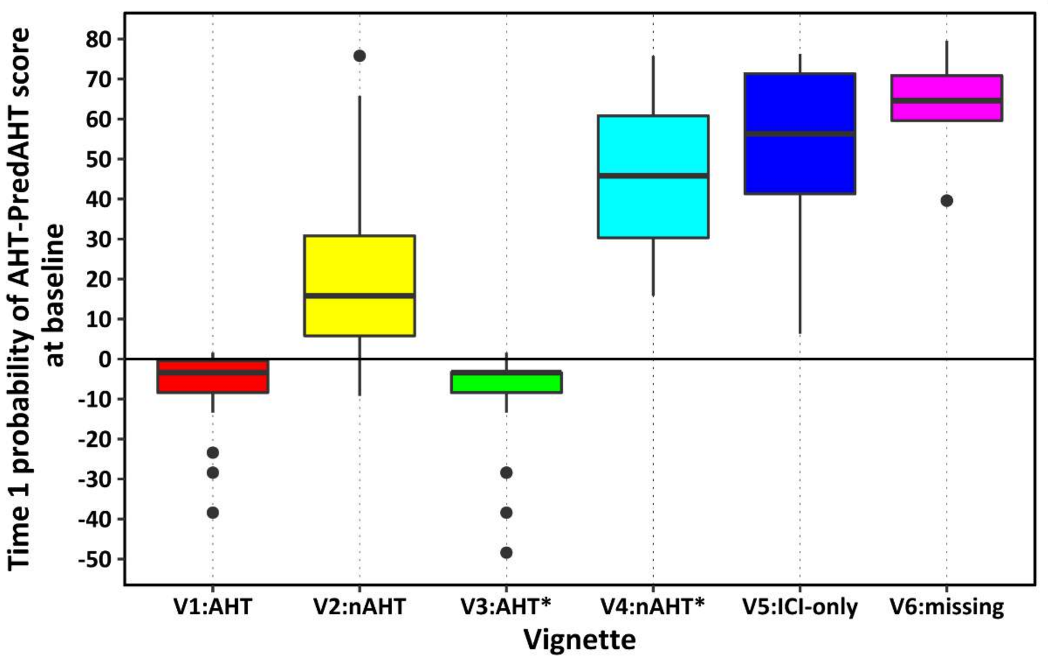
Appendix 27. Parallel coordinates plot



Appendix 28. Boxplot comparing clinicians' Time 1 probabilities of AHT with the PredAHT score (incorporating clinicians' prior probabilities)

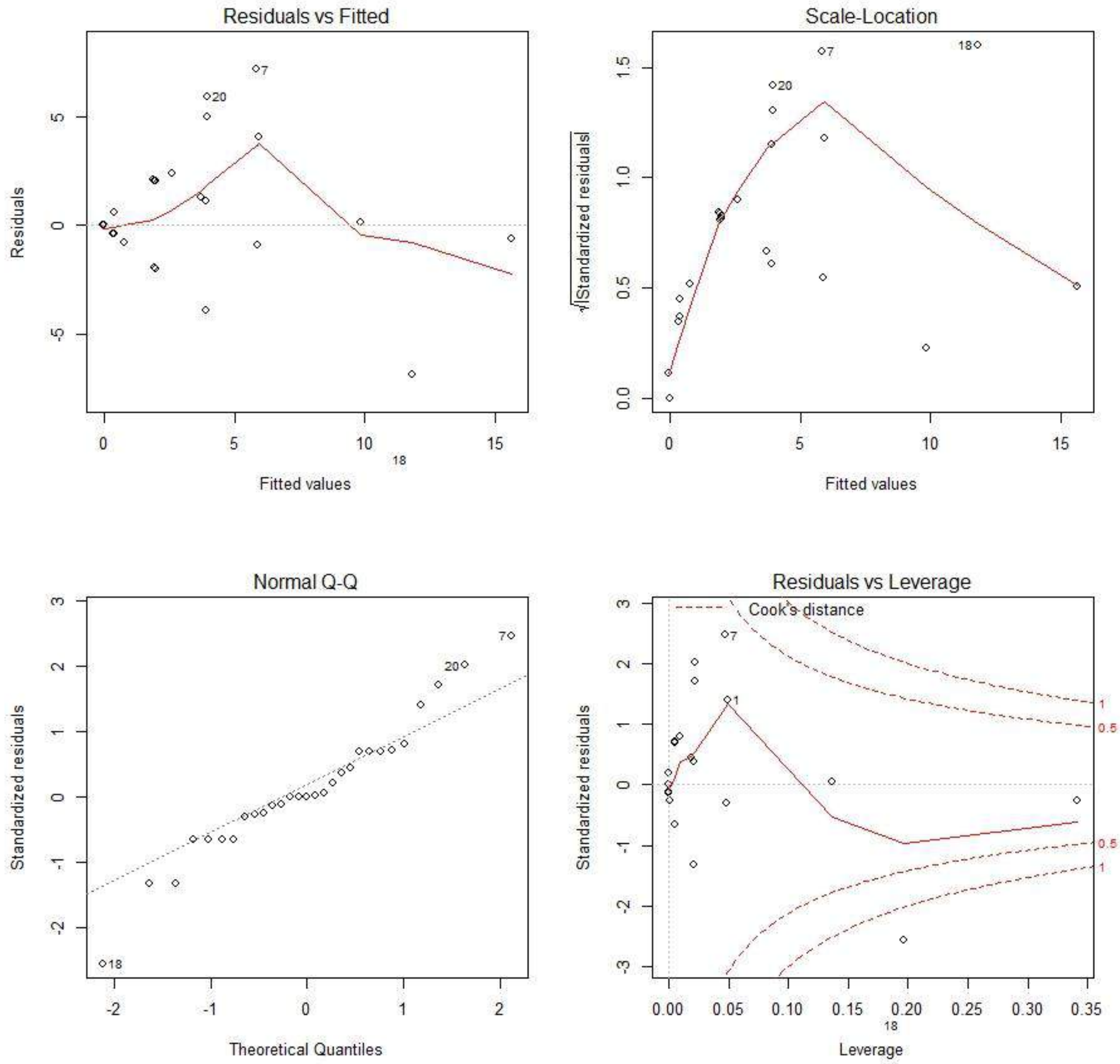


Appendix 29. Boxplot comparing clinicians' Time 1 probabilities of AHT with the PredAHT score using the baseline prior

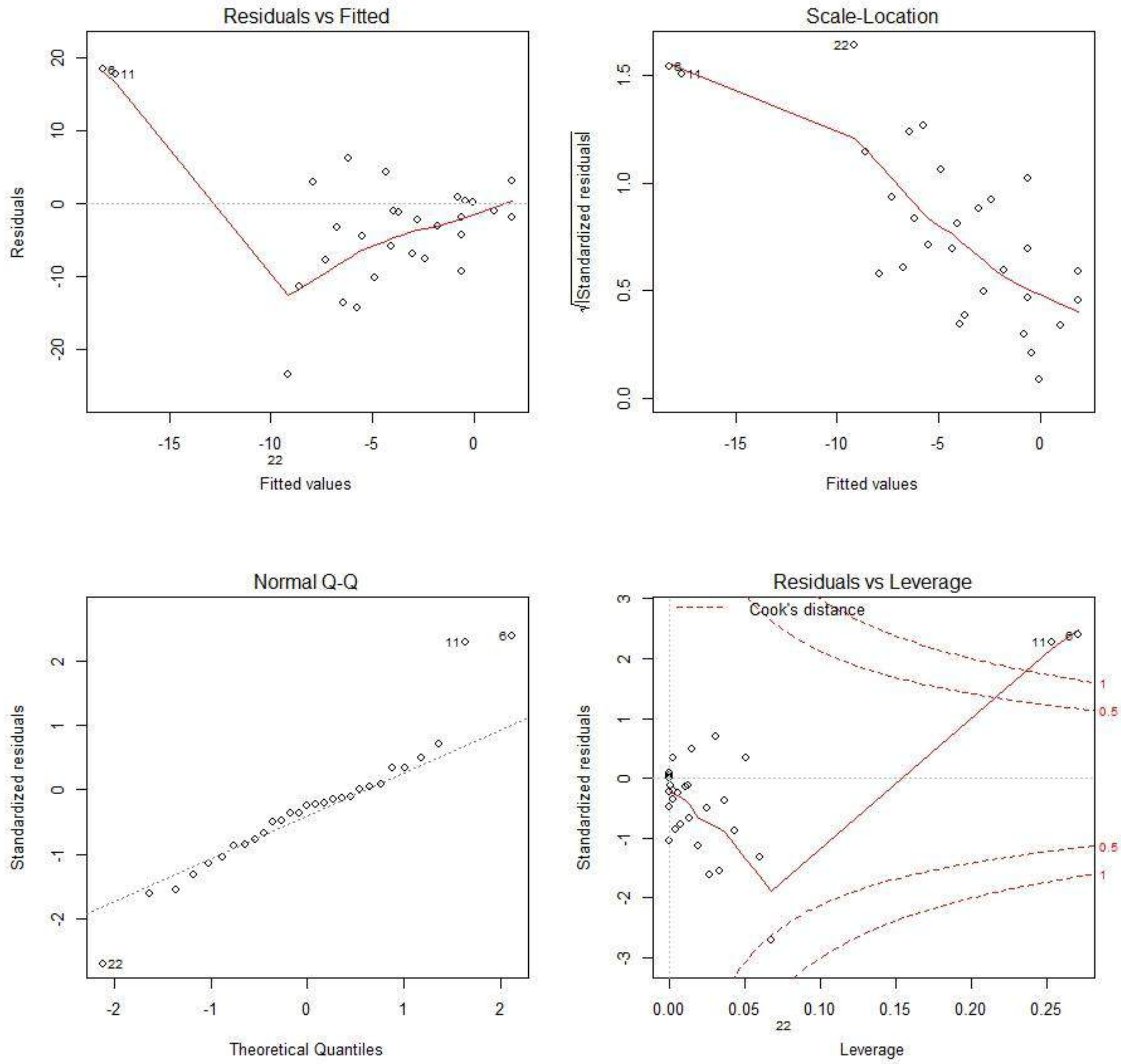


Appendix 30. Model diagnostic plots for the six linear models

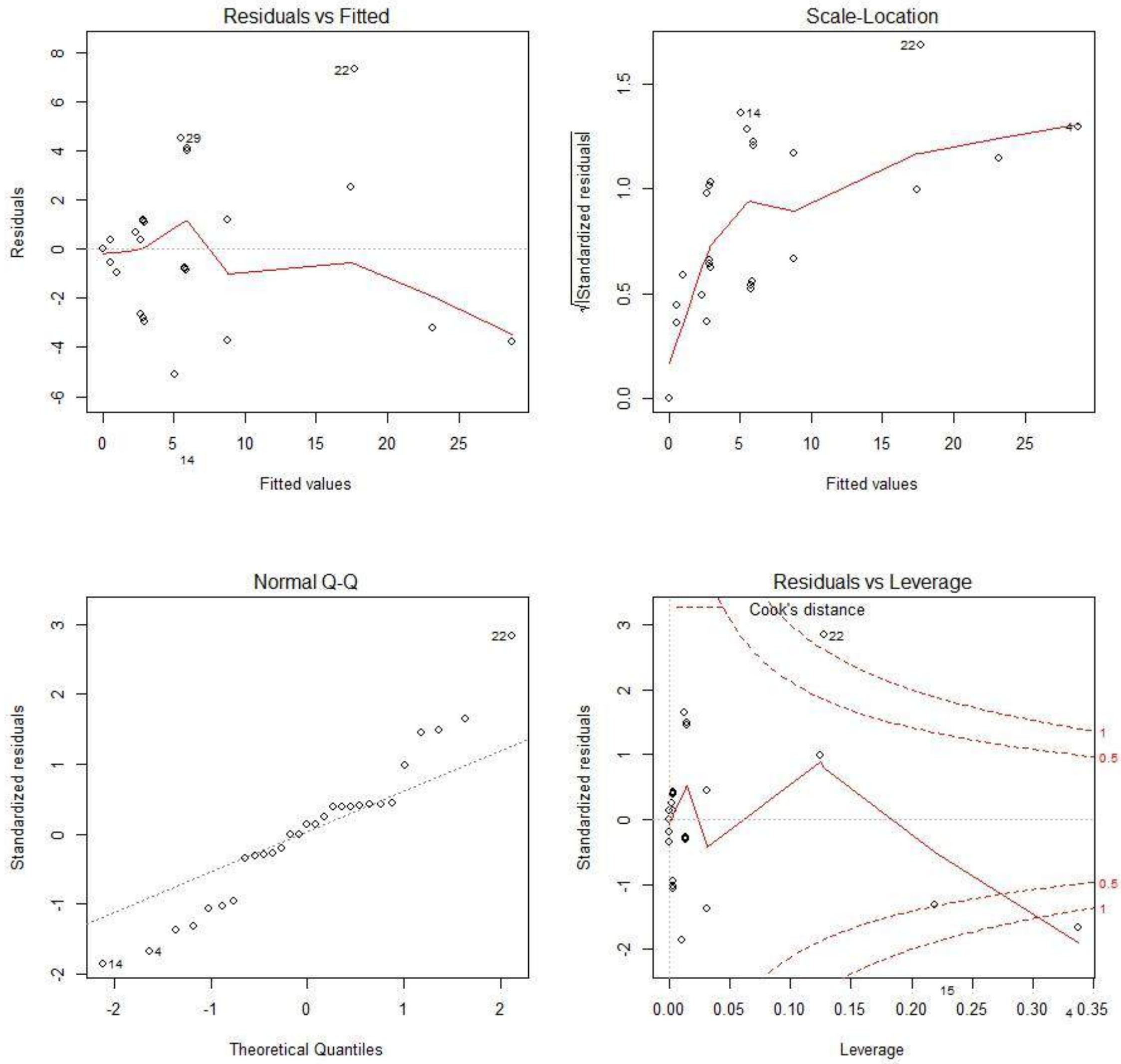
Diagnostic plots for V1:AHT



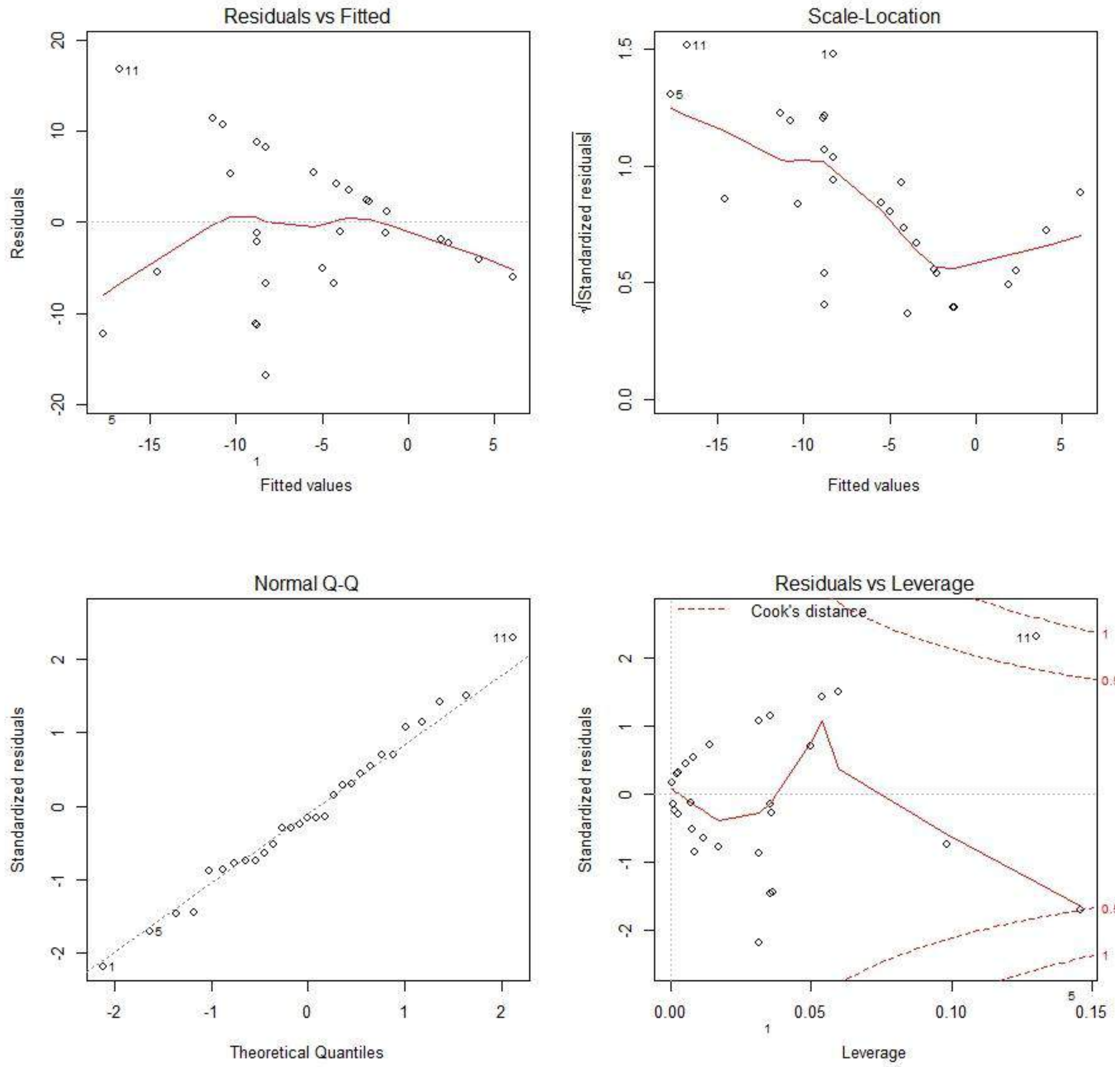
Diagnostic plots for V2:nAHT



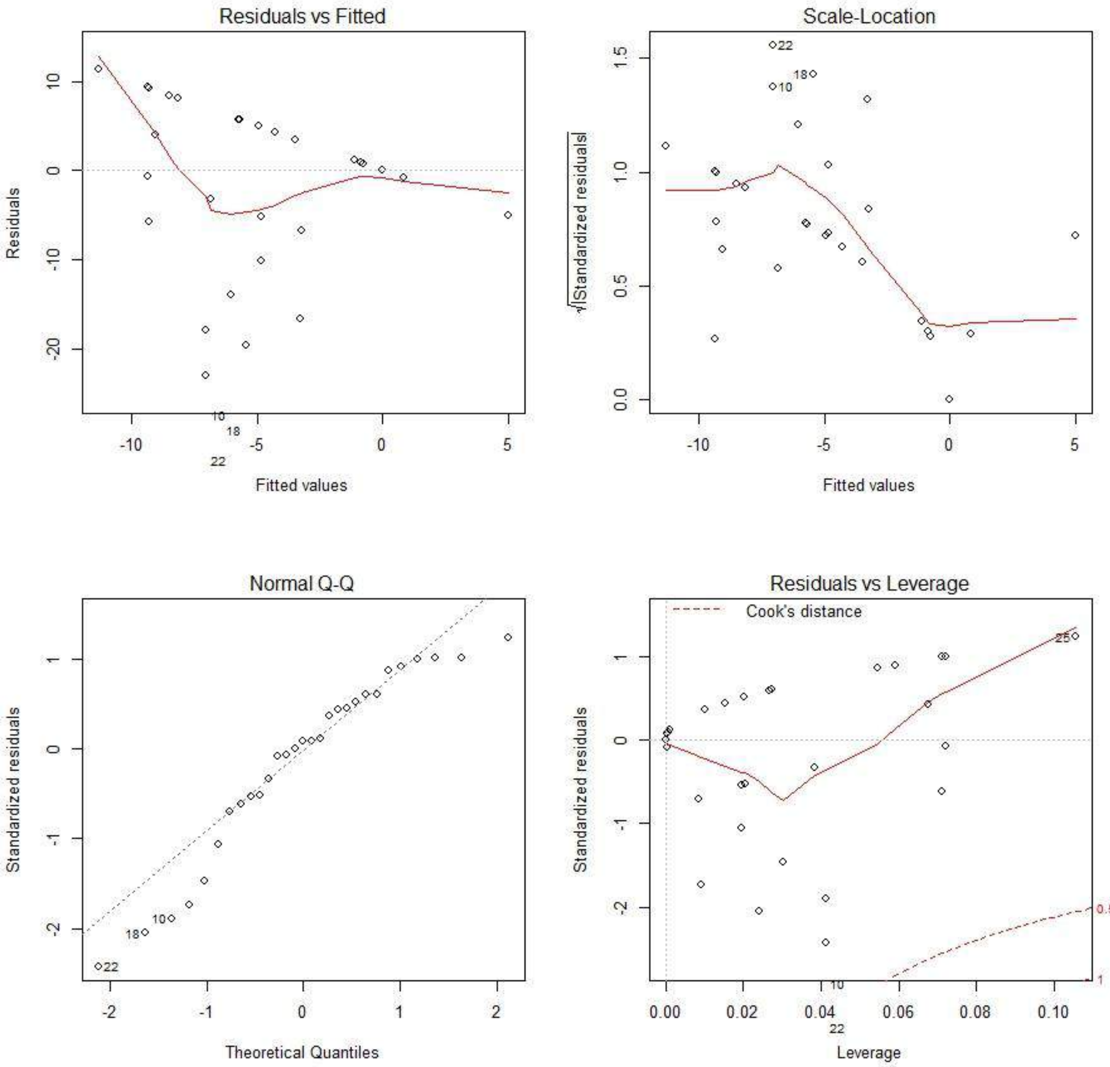
Diagnostic plots for V3:AHT*



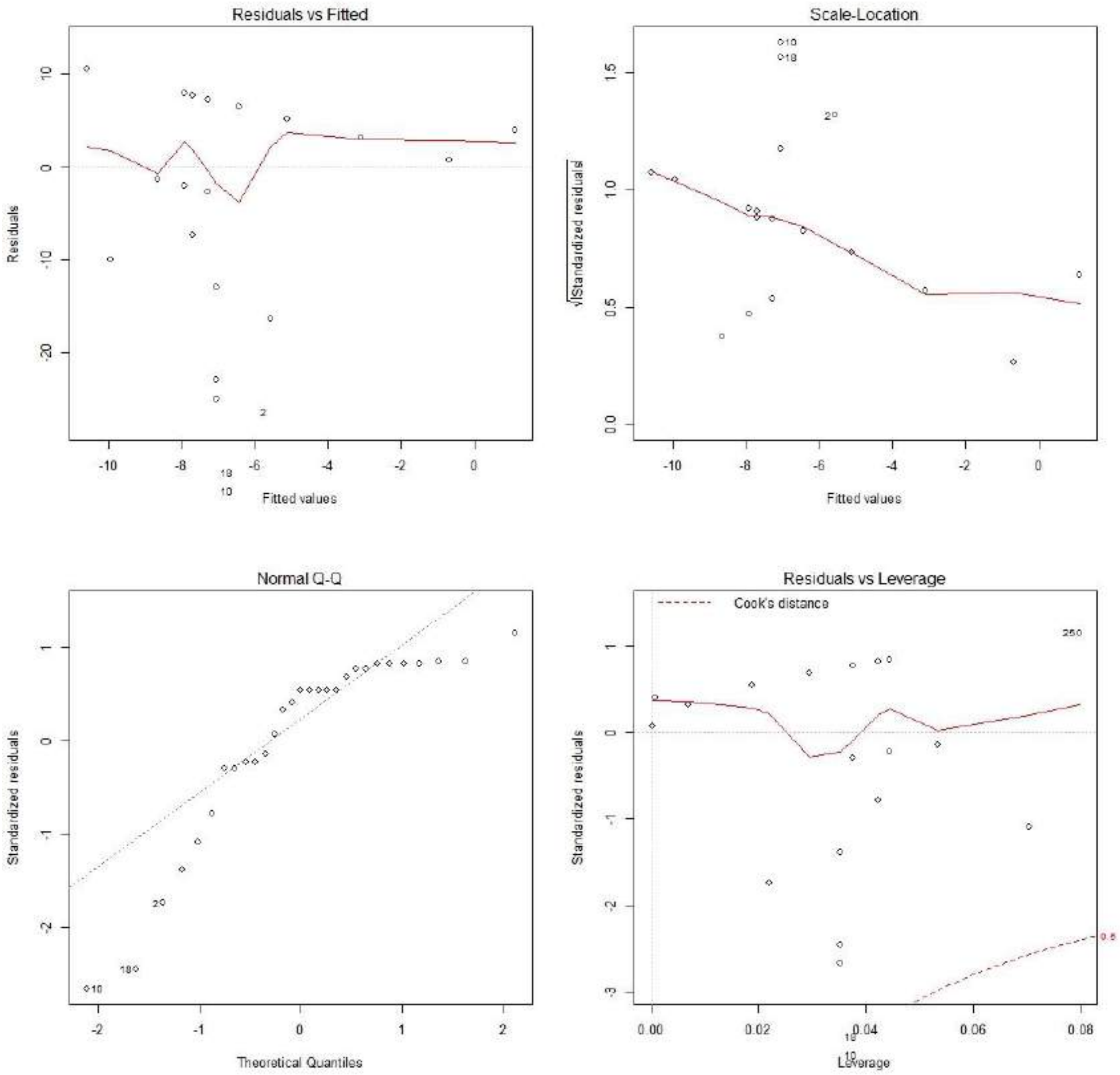
Diagnostic plots for V4:nAHT*



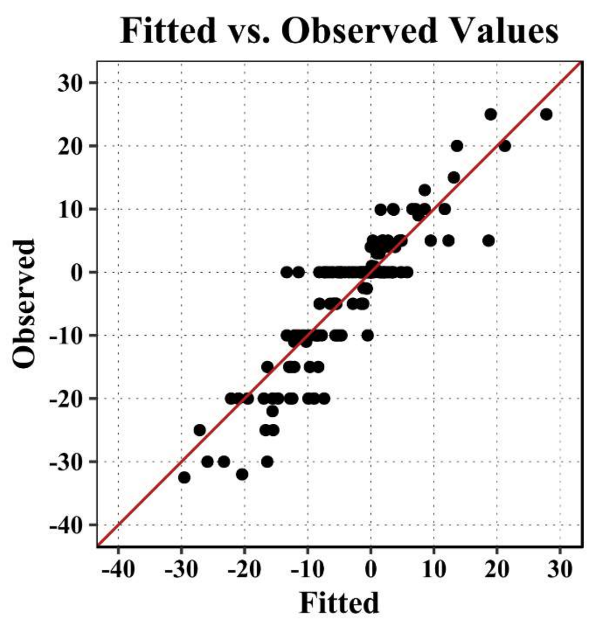
Diagnostic plots for V5:ICI-only

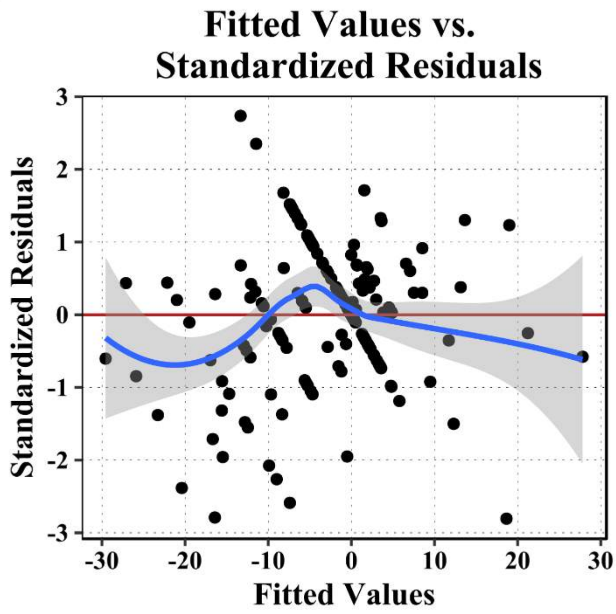


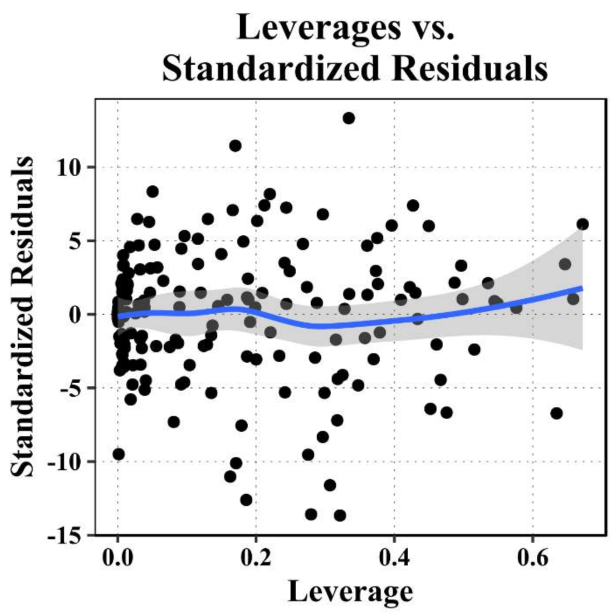
Diagnostic plots for V6:missing

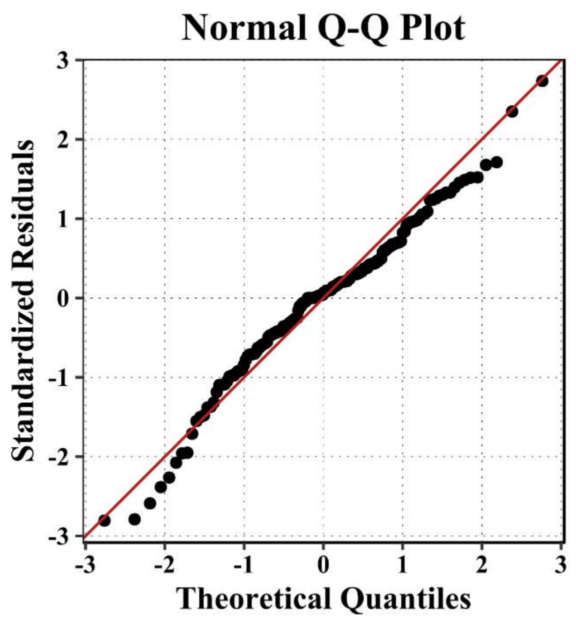


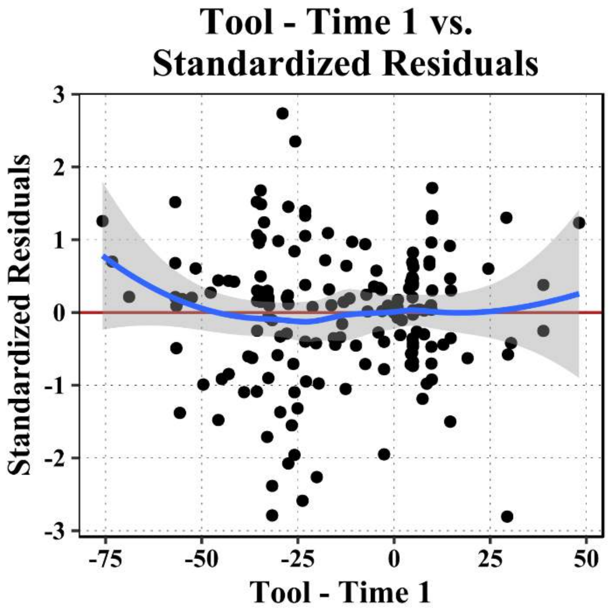
Appendix 31. Diagnostic plots for the linear mixed effects model

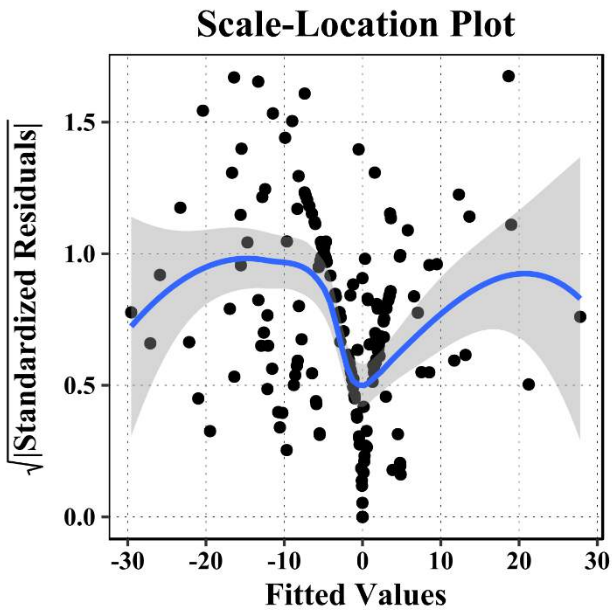


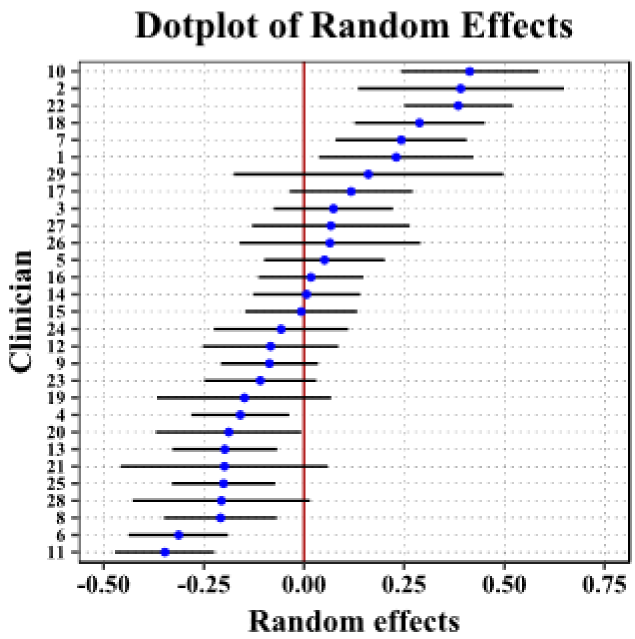


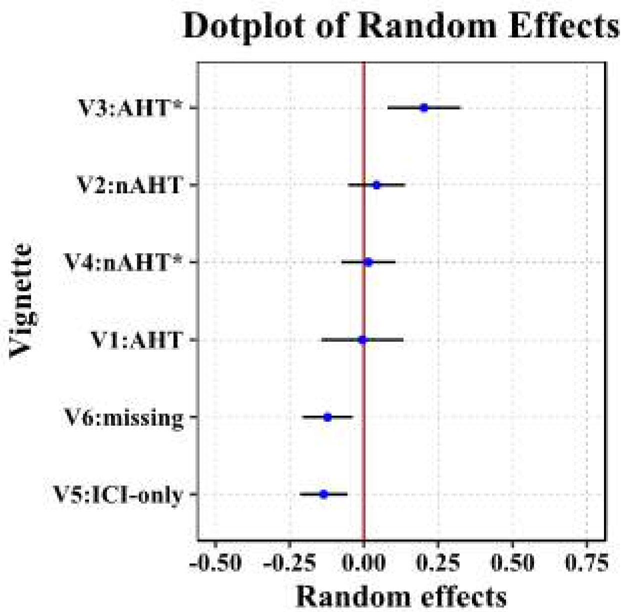

Leverages vs.
Standardized Residuals
Standardized Residuals
10
5
0
-5
-10
-15
0.0
0.2
0.4
0.6
Leverage


Normal Q-Q Plot
Standardized Residuals
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Theoretical Quantiles


Tool - Time 1 vs.
Standardized Residuals
Standardized Residuals
Tool - Time 1
3
2
1
0
-1
-2
-3
-75
-50
-25
0
25
50


Scale-Location Plot
√|Standardized Residuals|
Fitted Values
1.5
1.0
0.5
0.0
-30
-20
-10
0
10
20
30


Dotplot of Random Effects
Clinician
10
2
22
18
7
1
29
17
3
27
26
5
16
14
15
24
12
9
23
19
4
20
13
21
25
28
8
6
11
-0.50
-0.25
0.00
0.25
0.50
0.75
Random effects


Dotplot of Random Effects
Vignette
V3:AHT*
V2:nAHT
V4:nAHT*
V1:AHT
V6:missing
V5:ICI-only
-0.50
-0.25
0.00
0.25
0.50
0.75
Random effects

Appendix 32. Clinicians' proposed child protection actions at Time 1 and Time 2 by vignette

Time 1		V1: AHT N=29 (% total)	V2: nAHT N=29 (% total)	V3: AHT* N=29 (% total)	V4: nAHT* N=29 (% total)	V5: ICI-only N=29 (% total)	V6: missing N=29 (% total)
No concern	**No further CP action**	0 (0%)	3 (10%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Concern (abuse considered)	**Discuss with line manager**	5 (17%)	3 (10%)	4 (13%)	5 (17%)	3 (10%)	5 (17%)
	Discuss with CP colleague	19 (65%)	12 (41%)	19 (65%)	17 (58%)	18 (62%)	17 (58%)
	Gain collateral information from other agencies/disciplines	19 (65%)	21 (72%)	20 (68%)	21 (72%)	21 (72%)	21 (72%)
	Order further investigations	26 (89%)	9 (31%)	24 (82%)	13 (44%)	17 (58%)	24 (82%)
Suspicion (abuse suspected)	**Refer to social services**	27 (93%)	11 (37%)	25 (86%)	23 (79%)	19 (65%)	26 (89%)
	Skeletal Survey	24 (82%)	3 (10%)[a]	20 (68%)	1 (3%)[a]	4 (13%)[a]	20 (68%)
	Bone Scan	7 (24%)	3 (10%)	5 (17%)	2 (6%)	5 (17%)	7 (24%)

Further Investigations Specified	**MRI**	7 (24%)	1 (3%)	7 (24%)	2 (6%)	5 (17%)	7 (24%)
	Ophthalmology	N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	20 (68%)
	Haematology	16 (55%)	7 (24%)	14 (48%)	11 (37%)	12 (41%)	13 (44%)
	Full CP work-up	1 (3%)	1 (3%)	2 (6%)	1 (3%)	2 (6%)	3 (10%)
	Other[c]	4 (13%)	3 (10%)	2 (6%)	3 (10%)	4 (13%)	1 (3%)
Time 2		**V1: AHT N=29 (% total)**	**V2: nAHT N=29 (% total)**	**V3: AHT* N=29 (% total)**	**V4: nAHT* N=29 (% total)**	**V5: ICI-only N=29 (% total)**	**V6: missing N=29 (% total)**
No concern	**No further CP action**	0 (0%)	3 (10%)	0 (0%)	0 (0%)	2 (6%)	0 (%)
Concern (abuse considered)	**Discuss with line manager**	5 (17%)	3 (10%)	4 (13%)	5 (17%)	3 (10%)	5 (17%)
	Discuss with CP colleague	19 (65%)	12 (41%)	19 (65%)	15 (51%)	18 (62%)	19 (65%)
	Gain collateral information from other agencies/disciplines	19 (65%)	21 (72%)	20 (68%)	21 (72%)	20 (68%)	20 (68%)
	Order further investigations	26 (89%)	9 (31%)	24 (82%)	13 (44%)	16 (55%)	23 (79%)

Suspicion (abuse suspected)	**Refer to social services**	28 (96%)	9 (31%)	28 (96%)	22 (75%)	19 (65%)	24 (82%)
Further Investigations Specified	**Skeletal Survey**	24 (82%)	3 (10%)[a]	20 (68%)	1 (3%)[a]	4 (13%)[a]	20 (68%)
	Bone Scan	7 (24%)	3 (10%)	5 (17%)	2 (6%)	5 (17%)	7 (24%)
	MRI	7 (24%)	1 (3%)	7 (24%)	2 (6%)	5 (17%)	7 (24%)
	Ophthalmology	N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	20 (68%)
	Haematology	16 (55%)	7 (24%)	14 (48%)	11 (37%)	12 (41%)	13 (44%)
	Full CP work-up	1 (3%)	0 (0%)	2 (6%)	1 (3%)	1 (3%)	2 (6%)
	Other[c]	4 (13%)	4 (13%)	2 (6%)	3 (10%)	4 (13%)	1 (3%)

N.B column totals do not add to 29 as clinicians were able to choose multiple options. CP = child protection, [a] Refers to a repeat skeletal survey, [b] Children in these vignettes had already undergone an ophthalmoscopy exam, [c] Other investigations include a metabolic screen, discussions with neurologists/radiologists, medical photography and reviewing the history

Appendix 33. Patient data collection form

INVESTIGATIONS						
		Y/N	**NUMBER OF SCANS**	**DATES OF SCANS**	**NUMBER OF DAYS AFTER ADMISSION**	**RESULTS**
CT:	Head					
	Spine					
	Thorax					
	Abdomen					
	Other (please state)					
MRI:	Head					*Include detail of MRI e.g. the pulse sequence, T1 or T2-weighted, diffusion-weighted etc.:*
	Spine					

<table>
<tr><td colspan="2"></td><td colspan="4">Other (please state)</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2" rowspan="2">CXR:</td><td colspan="4">PA</td><td></td><td></td><td></td><td></td><td rowspan="2">Include details of location imaging performed i.e. ED, PICU, radiology dept:</td></tr>
<tr><td colspan="4">LATERAL OBLIQUE</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">FOLLOW UP CXR:</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">ADDITIONAL XR:</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">SKELETAL SURVEY:</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">FOLLOW UP SKELETAL SURVEY</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">RADIONUCLIDE BONE SCAN</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>INDIRECT
(Y/N)</td><td colspan="2">DILATED
PUPILS
(Y/N)</td><td>PERFORMED
BY
(SpR/CON)</td><td>RETCAM
(Y/N)</td><td>STANDARDIZED
PROFORMA?
(Y/N)</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td></td><td colspan="2"></td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><th colspan="8">HISTORY</th></tr>
<tr><td rowspan="4">EXPLANATION GIVEN (Y/N)

EXPLANATION CONSISTENT BETWEEN CAREGIVERS? (Y/N/Unknown)

EXPLANATION CONSISTENT OVER TIME? (Y/N/Unknown)

INJURIES CONSISTENT WITH MECHANISM? (Y/N/Unknown)</td><td rowspan="4">DETAIL OF EXPLANATION:</td><td>WITNESSED INJURY? (Y/N/Unknown)</td><td>WITNESSED BY WHOM?</td><td>FALL</td><td>ACTIVITY PRIOR TO FALL</td><td>POSITION PRIOR TO FALL (STANDING/SITTING/STANDING ON OBJECT/IN ARMS)</td><td>SURFACE OF IMPACT</td></tr>
<tr><td rowspan="2"></td><td rowspan="2"></td><td>HEIGHT (<0.6M/0.6-1M/1-1.5M/1.5-2M/>2M)</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>STATE HEIGHT:………………</td></tr>
<tr><td>COMMENT ON PRECEDING EVENTS:</td><td>ADMITTED ABUSE? (Y/N)</td><td colspan="4">IMMEDIATE SYMPTOMS AFTER INJURY (Y/N):</td></tr>
</table>

MOTOR VEHICLE COLLISION	PASSENGER (FRONT/BACK)	PEDESTRIAN (STRUCK WHILE STANDING)	PEDESTRIAN (STRUCK WHILE CYCLING)	COMMENTS
CRUSH INJURY	OBJECT THAT LANDED ON CHILD	WITNESSED	BY WHOM?	COMMENTS
CHILD'S DEVELOPMENTAL STAGE AT INJURY	(NOT INDEPENDENTLY MOBILE/SITTING/ROLLING/CRUISING/CRAWLING/WALKING)			

NEUROIMAGING FINDINGS								
		INTERHEMISPHERIC	MULTIPLE	OVER CONVEXITY	SINGLE	POSTERIOR FOSSA	COMMENT	
SDH	Y							
	N							

APPEARANCE ON CT *(give dates of CT if > 1)*	HYPERDENSE			HYPODENSE	MIXED DENSITY
SAH	Y		COMMENT (NUMBER/SITE/NATURE)		
	N				
EDH	Y		COMMENT (NUMBER/SITE/NATURE)		
	N				
CEREBRAL OEDEMA	Y		COMMENT (NUMBER/SITE/NATURE)		
	N				
FOCAL PARENCHYMAL INJURY	Y		COMMENT (NUMBER/SITE/NATURE)		
	N				
HYPOXIC ISCHAEMIC INJURY	Y		COMMENT (NUMBER/SITE/NATURE)		

	N			
SKULL FRACTURE	Y		COMMENT (NUMBER/SITE/NATURE)	
	N			

OPTHAMOLOGICAL FINDINGS (LEFT EYE)								
		UNILATERAL	DISTRIBUTION	LAYERS OF RETINA INVOLVED	NUMBER OF RH	SIZE	ADDITIONAL FEATURES	COMMENTS
RETINAL FINDINGS	Y							
	N							

OPTHAMOLOGICAL FINDINGS (RIGHT EYE)								
		UNILATERAL	DISTRIBUTION	LAYERS OF RETINA INVOLVED	NUMBER OF RH	SIZE	ADDITIONAL FEATURES	COMMENTS
RETINAL FINDINGS	Y							
	N							

ADDITIONAL FEATURES				
RIB FRACTURES	Y	NUMBER	LOCATION (ANT/LAT/POS)	COMMENT (VISIBLE ON FIRST/SUBSEQUENT IMAGES/AT PM)
	N			
	NOT KNOWN			
ADDITIONAL FRACTURES	Y	COMMENT (BONE/TYPE/AGE OF FRACTURE)		
	N			
	NOT KNOWN			
APNOEA	Y	DOCUMENTED BY (HCP/CARER/PARENT)		
	N			

SEIZURES	Y	DOCUMENTED BY (HCP/CARER/PARENT)	PRE ADMISSION (Y/N)	ON ADMISSION (Y/N)	POST ADMISSION (Y/N)	TYPE (FOCAL/GLOBAL/STATUS EPILEPTICUS)	COMMENTS
	N						

HEAD AND NECK BRUISING	Y	NUMBER	LOCATION	ADDITIONAL BRUISING TO REST OF BODY	Y	COMMENT (LOCATION/NATURE)
			MARK ON ATTACHED DIAGRAM (i)		N	
	N					
PETECHIAE	Y					
	N					

Appendix 34. Participant Information Sheet

PARTICIPANT INFORMATION SHEET

1. Study title

Feasibility of an impact analysis of a clinical prediction tool to estimate the probability of abusive head trauma in children less than three years of age

2. Study team

Researchers: Miss Laura Cowley (PhD student)
Supervisors: Professor Alison Kemp, Dr Sabine Maguire, Dr Daniel Farewell.

3. Invitation

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

4. What is the purpose of the study?

Abusive head trauma is one of the most serious types of child abuse and has severe consequences for young children. It is vital that it is identified accurately in order to prevent further harm occurring to children. It is also important that accidental injuries are not wrongly classified as abusive. An evidence-based tool has been developed to estimate the probability of abusive head trauma in head-injured children aged less than three years old. The aim of this study is to evaluate the feasibility of conducting an impact analysis of the tool in clinical practice.

5. Why have I been chosen?

You have been recruited for this study as you have been identified as a professional who may come into contact with a child with a head injury where abuse is suspected, and therefore you may potentially benefit from the use of this prediction tool.

6. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. You do not have to consent to be recorded but if you do you will be shown how to turn the recorder off in case you wish to do so at any point. Only the immediate study team (see above) will have access to the recording and transcript.

7. What will happen to me if I take part / what do I have to do?

You will be asked to give your estimate of the probability of abuse in children hospitalised to your site with an intracranial injury identified on neuroimaging. You will also be asked what your next child protection action will be regarding each case (if any). You will then be given an iPad on which to complete the clinical prediction tool; for each child, data regarding the presence or absence of six clinical features (head or neck bruising, rib fractures, long bone fractures, seizures, apnoea and retinal haemorrhages) will be entered. The tool will generate a percentage probability of abuse, based on the combinations of features entered, and you will be asked whether this would change your own estimate or your next child protection action, and the reasons for your answers. At the end of the study period you will be contacted and asked to complete a short survey and to partake in an interview about the usefulness of the tool.

8. What about confidentiality?

All information provided by participants will be treated confidentially and respectfully. Data will be held securely and access will be limited to the immediate study team (see above).

9. Are there any risks?

We are not aware of any major risks that would preclude the study. The tool is not intended to be directive or to make specific recommendations, but is intended to be assistive and to act as an adjunct to clinical decision making. The tool is not intended to replace clinical judgement. The information provided by calculating the tool's score is based on clinical features that clinician's will have considered within their decision making, albeit without a numerical estimate of probability. If the investigator has any concerns about a case, these will be brought to the attention of the academic supervisor, who will discuss these with the treating clinician.

10. Are there any benefits?

You will have the opportunity to use a rigorously developed and validated tool that you may find useful for your decision making. The study will give us an understanding of the feasibility of conducting a large scale impact analysis of the tool, and of the acceptability of the tool to clinicians.

11. What will happen to the results of the research study?

The results of the study will be written up as part of a PhD thesis and submitted for publication in an academic journal. The results may also be submitted to a conference as a scientific abstract or poster. Participants will be informed of when and where they can obtain a copy of the published results following publication, and if required and appropriate to do so, the researchers will forward on copies of the publication. Participants will not be identified in any report, conference talk, publication or thesis.

12. Who is organising and funding the research?

This research is being organised by Miss Laura Cowley. Laura is a PhD student in the School of Medicine, Cardiff University. The PhD is being funded by the National Institute for Social Care and Health Research (NISCHR).

13. What if I am unhappy about any aspect of the study?

If you have any concerns or complaints about any aspect of the study please contact Professor Alison Kemp.

Email: KempAM@cardiff.ac.uk

14. Contact for further information

Should you require any further information or questions about the study, please contact Miss Laura Cowley.

Email: CowleyLE@cardiff.ac.uk

Appendix 35. Consent form used in the feasibility study

CONSENT FORM

Title of Project:

Feasibility of an impact analysis of a clinical prediction tool to estimate the probability of abusive head trauma in children less than two years of age

Name of Researcher:

Please initial box

1. I confirm that I have read and understand the information sheet dated .... (version ....) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐

3. I agree to take part in the above study. ☐

4. I consent to direct, anonymised, quotes being used as part of a thesis, publication in an academic journal, or conference abstract, presentation or poster. ☐

5. I consent to being audio-recorded Yes ☐ No ☐

____________________	____________	________________
Name of Participant	Date	Signature
____________________	____________	________________
Researcher	Date	Signature

Version: 1 **Date:** 21/10/2015

Appendix 36. Clinician data collection form

Centre:
Case ID number: *(allocated by research team)*
Clinician ID number: *(allocated by research team)*
Date: _____/_____/_____
Time: _____:_____ *(24 hrs)*

Data Collection Form

Study title: Feasibility of an impact analysis of a clinical prediction tool to estimate the probability of abusive head trauma in children less than three years of age

Clinician Demographics

Gender: Female ☐ Male ☐

Age: 18-24 years ☐ 25-34 years ☐ 35-44 years ☐ 45-54 years ☐

55-64 years ☐ 65 or older ☐

Ethnicity:

White: British ☐ Other ……………………………………………

Mixed: White & Black Caribbean ☐ White & Black African ☐

White & Asian ☐ Other ……………………………………………

Black or Black British: Caribbean ☐ African ☐ Other ……………………………………………

Asian or Asian British: Indian ☐ Pakistani ☐ Bangladeshi ☐ Chinese ☐

Other ………………………………………

Designation ***(e.g. doctor, nurse)*** ………………………………………………

Speciality:…………………………………………………………

Grade: ……………**Years since graduation:** ……………

Clinical Features in the child ***(to be completed by researcher)***

Any bruising to the head and/or neck:	Present ☐	Unknown ☐	Absent ☐
Any seizure (single, multiple or status epilepticus):	Present ☐	Unknown ☐	Absent ☐
Any apnoea in the history and/or during inpatient stay:	Present ☐	Unknown ☐	Absent ☐
Any long bone fracture seen on imaging:	Present ☐	Unknown ☐	Absent ☐
Any rib fracture seen on imaging:	Present ☐	Unknown ☐	Absent ☐
Any retinal haemorrhage seen on ophthalmology exam:	Present ☐	Unknown ☐	Absent ☐

Likelihood of abuse

What is your estimate of the probability of abuse, in percentage terms?............................%

What will your next child protection action be?

No concern (abuse excluded):

No further child protection action ☐

Concern (abuse considered):

Discuss with line manager ☐

Discuss with child protection colleague ☐

Gain collateral information from other agencies and health disciplines ☐

Order further investigations ☐ ***give details:***

……………………………………………………………………………

Suspicion (abuse suspected):

Refer to social services ☐

Other ☐ ***(please describe):***………………………………………………………………

Comments:

Researcher to record prior probability _____%

Having seen the score provided by the tool, what is your estimate of the probability of abuse now, in percentage terms?.....................................%

Having seen the score provided by the tool, what will your next child protection action be?

No concern (abuse excluded):

No further child protection action ☐

Concern (abuse considered):

Discuss with line manager ☐

Discuss with child protection colleague ☐

Gain collateral information from other agencies and health disciplines ☐

Order further investigations ☐ ***give details:***

....

Suspicion (abuse suspected):

Refer to social services ☐

Appendix 37. Cardiff University study sponsorship

Research and Innovation Services
Director Geraint W Jones
Gwasanaethau Ymchwil ac Arloesi
Cyfarwyddwr Geraint W Jones

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
7th Floor
30 - 36 Newport Road
Cardiff CF24 0DE
Wales UK
Tel *Ffôn* +44(0)29 2087 5834
Fax *Ffacs* +44(0)29 2087 4189
Prifysgol Caerdydd
Llawr 7
30 - 36 Heol Casnewydd
Caerdydd CF24 0DE
Cymru Y Deyrnas Unedig

28th October 2015

Professor Alison Kemp
Institute of Primary Care and Public Health
Cardiff University School of Medicine
5th Floor, Neuadd Meirionnydd
Heath Park
Cardiff
CF14 4YS

Dear Professor Kemp,

Feasibility of an impact analysis of a clinical prediction tool to estimate the probability of abusive head trauma in children less than two years of age

I understand that you are acting as Academic Supervisor for the above PhD project to be conducted by Laura Cowley.

I confirm that Cardiff University agrees in principle to act as Sponsor for the above project, as required by the Research Governance Framework for Health and Social Care.

Scientific (Peer) Review
I can also confirm that Scientific (Peer) Review has been obtained from Health and Care Research Wales (the funder).

Insurance
The necessary insurance provisions will be in place prior to the project commencement. Cardiff University is insured with UMAL. Copies of the insurance certificate are attached to this letter.

Approvals

On completion of your IRAS form (for NHS REC and NHS R&D approvals), you will be required to obtain signature from the Sponsor ('Declaration by the Sponsor Representative').

Please then submit the project to the following organisations for approvals:

- The Confidentiality Advisory Group (CAG);
- An NHS Research Ethics Committee;
- Health & Care Research Wales Permissions Coordinating Unit (formerly known as NISCHR PCU) -to arrange host organisation R&D approval for Welsh NHS sites).

Once Research and Innovation Services has received evidence of the above approvals, the University is considered to have accepted Sponsorship and your project may commence.

Roles and Responsibilities
As Chief Investigator you have signed a Declaration with the Sponsor to confirm that you will adhere to the standard responsibilities as set out by the Research Governance Framework for Health and Social Care. In accordance with the University's Research Governance Framework, the Chief Investigator is also responsible for ensuring that each research team member is qualified and experienced to fulfill their delegated roles including ensuring adequate supervision, support and training.

If your study is adopted onto Health & Care Research Wales Clinical Research Portfolio you are required to upload recruitment data onto the portfolio database.

Contracts

- No research-specific tasks delegated to NHS Host Organisation (staff acting as participants);
- Roles and responsibilities are adequately detailed in the research protocol – no contract required;

- The student researcher will be in possession of an NHS Research Passport, prior to accessing NHS data.

May I take this opportunity to remind you that, as Chief Investigator, you are required to:

- ensure you are familiar with your responsibilities under the Research Governance Framework for Health and Social Care;
- undertake the study in accordance with Cardiff University's Research Governance Framework and the principles of Good Clinical Practice;
- ensure the Research complies with the Data Protection Act 1998;
- inform Research and Innovation Services of any amendments to the protocol or study design, including changes to start /end dates and ensure any such amendments are submitted to, and approved by, the relevant bodies (e.g. RECs and/or R&D offices);
- co-operate with any audit inspection of the project files or any requests from Research & Innovation Services for further information.

You should quote the following unique reference number in any correspondence relating to sponsorship for the above project:

SPON 1471-15

This reference number should be quoted on all documentation associated with this project.

Yours sincerely

Dr K J Pittard Davies
Head of Research Governance and Contracts
Direct line: +44 (0) 29208 79274
Email: resgov@cardiff.ac.uk

Cc: Ms. Laura Cowley.

Appendix 38. Ethical approval letter

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

Gwasanaeth Moeseg Ymchwil
Research Ethics Service

Ariennir gan Lywodraeth Cymru
Funded by Welsh Government

Wales REC 3
Health and Care Research Support Centre
Castlebridge 4
15-19 Cowbridge Road East
Cardiff CF11 9AB

Telephone : 029 2078 5735
E-mail : corinne.scott@wales.nhs.uk
Website : www.hra.nhs.uk

21 January 2016

Professor Alison Kemp
Cochrane Institute of Primary Care and PH,5th Floor, Neuadd Meirionnydd,
Heath Park
Cardiff CF14 4YS

Dear Professor Kemp

Study title:	**Feasibility of an impact analysis of a clinical prediction tool to estimate the probability of abusive head trauma in children less than three years of age**
REC reference:	**16/WA/0003**
Protocol number:	**SPON 1471-15**
IRAS project ID:	**167217**

Thank you for responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to make a request to postpone publication, please contact the REC Manager, Dr. Corinne Scott, corinne.scott@wales.nhs.uk.

Confirmation of ethical opinion
On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion
The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).
Guidance on applying for NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the

R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials
All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database within 6 weeks of recruitment of the first participant (for medical device studies, within the timeline determined by the current registration and publication trees).

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to contest the need for registration they should contact Catherine Blewett (catherineblewett@nhs.net), the HRA does not, however, expect exceptions to be made. Guidance on where to register is provided within IRAS.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites
NHS sites
The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Approved documents
The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Covering letter on headed paper [Covering letter]	1	24 November 2015
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Insurance & Indemnity]	1	10 November 2015
Interview schedules or topic guides for participants [Interview schedule with version no]	1	17 November 2015
IRAS Checklist XML [Checklist_24112015]		24 November 2015
IRAS Checklist XML [Checklist_26112015]		26 November 2015
IRAS Checklist XML [Checklist_19012016]		19 January 2016
Letter from funder [Letter from funder]	1	06 August 2014
Letter from sponsor [Letter from Sponsor]	1	28 October 2015
Letter from statistician [Institute statistician Peer Review comments]	1	03 March 2014
Other [Dr. Daniel Farewell CV]	1	10 November 2015
Other [Letter from funder RE project amendment]	1	12 December 2014
Other [Letter regarding meeting attendance]	1	19 January 2016
Participant consent form [Consent Form Version 3 tracked changes]	3	19 January 2016
Participant consent form [Consent Form Version 3 clean copy]	3	19 January 2016
Participant information sheet (PIS) [Information Sheet Version 3 tracked changes]	3	19 January 2016
Participant information sheet (PIS) [Information Sheet Version 3 clean copy]	3	19 January 2016

REC Application Form [REC_Form_24112015]		24 November 2015
Referee's report or other scientific critique report [Institute Peer Review comments]	1	03 March 2014
Research protocol or project proposal [Research Protocol Version 2]	2	20 November 2015
Summary CV for Chief Investigator (CI) [Prof. Alison Kemp CV]	1	10 November 2015
Summary CV for student [Laura Cowley CV]	1	24 November 2015
Summary CV for supervisor (student research) [Dr Sabine Maguire CV]	1	10 November 2015
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Summary of protocol]	1	03 November 2015
Validated questionnaire [Survey]	1	16 November 2015

Statement of compliance
The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review
Reporting requirements
The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback
The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training
We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

16/WA/0003	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project.

Yours sincerely

Dr. Corinne Scott
Senior Ethics Service Manager, Health and Care Research Wales
pp Dr. Pete Wall, Chair

Enclosures: "After ethical review – guidance for researchers"

Copy to: Helen Falconer, Cardiff University
Christopher Fegan, Cardiff and Vale University Health Board
Confidentiality Advise Team

Appendix 39. Ethical approval letter for amendment

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

Gwasanaeth Moeseg Ymchwil
Research Ethics Service

Ariennir gan Lywodraeth Cymru
Funded by Welsh Government

Wales REC 3
Sixth Floor, Churchill House
17 Churchill Way
Cardiff CF10 2TW

Telephone : 029 2037 6829
Fax : 029 2037 6824
E-mail : corinne.scott@wales.nhs.uk
Website : www.hra.nhs.uk
Fax:

29 July 2016

Professor Alison Kemp
Professor of Child Health
Cardiff University
Cardiff Child Protection Systematic Reviews,
Cochrane Institute of Primary Care and PH,
5th Floor, Neuadd Meirionnydd,
Heath Park,
Cardiff CF14 4YS

Dear Professor Kemp

Study title:	**Feasibility of an impact analysis of a clinical prediction tool to estimate the probability of abusive head trauma in children less than three years of age**
REC reference:	**16/WA/0003**
Protocol number:	**SPON 1471-15**
Amendment number:	**1**
Amendment date:	**12 July 2016**
IRAS project ID:	**167217**

The above amendment was reviewed [at the meeting of the Sub-Committee held on 29 July 2016.

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents
The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-CTIMP)	1	12 July 2016
Research protocol or project proposal	3	12 July 2016

Membership of the Committee
The members of the Committee who took part in the review are listed on the attached sheet.

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance
The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

16/WA/0003:	Please quote this number on all correspondence

Yours sincerely

Dr. Corinne Scott
Senior Ethics Service Manager
Health and Care Research Wales

pp **Mrs. Monika Hare**
Vice Chair

Email: corinne.scott@wales.nhs.uk

Enclosures: List of names and professions of members who took part in the review

Copy to: Christopher Fegan, Cardiff and Vale University Health Board
Mrs Helen Falconer
Confidentiality Advise Team

Wales REC 3

Attendance at Sub-Committee of the REC meeting on 29 July 2016

Committee Members:

Name	Profession	Present	Notes
Mrs Monika Hare	Vice Chair / Lay member	Yes	Chaired meeting
Dr Richard Walker	Alternate Vice Chair / Lay Plus member	Yes	
Dr Pete Wall	Chair / Clinical Physiologist	Yes	

Also in attendance:

Name	Position (or reason for attending)
Dr Corinne Scott	Senior Ethics Service Manager

Appendix 40. Confidentiality Advisory Group Approval Letter

NHS
Health Research Authority
Confidentiality Advisory Group

Skipton House
80 London Road
London
SE1 6LH

Tel: 020 797 22557
Email: HRA.CAG@nhs.net

21 March 2016

Professor Alison Kemp
Professor of Child Health
5th floor, Neuadd Meirionnydd
Heath Park
Cardiff
CF14 4YS

Dear Professor Kemp

Application title:	**Feasibility of an impact analysis of a clinical prediction tool to estimate the probability of abusive head trauma in children less than three years of age**
CAG reference:	**16/CAG/0022**
IRAS project ID:	**167217**
REC reference:	**16/WA/0003**

Thank you for your research application, submitted for approval under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 to process patient identifiable information without consent. Approved applications enable the data controller to provide specified information to the applicant for the purposes of the relevant activity, without being in breach of the common law duty of confidentiality, although other relevant legislative provisions will still be applicable.

The role of the Confidentiality Advisory Group (CAG) is to review applications submitted under these Regulations and to provide advice to the Health Research Authority on whether an application should be approved, and if so, any relevant conditions. This application was considered at the CAG meeting held on 25 February 2016.

Health Research Authority

The Health Research Authority, having considered the advice from the Confidentiality Advisory Group as set out below, has determined the following:

1. The application is approved, subject to compliance with the standard and specific conditions of approval.

This letter should be read in conjunction with the outcome letter dated 10 March 2016.

Context

Purpose of application

Purpose of application

This application from Cardiff University set out the purpose of assessing the Feasibility of using an impact analysis tool to predict abuse. Abusive Head Trauma is the leading cause of death and disability in young children who have suffered abuse; however it can be difficult to tell if a head injury is the result of abuse. The researchers have created a tool to facilitate clinicians in this judgement. They will collect:

1. The probability assessment provided by the tool;
2. The probability assessment of the clinician pre and post use of the tool [with consent of the clinician];
3. The child protection decisions of the clinician pre and post use of the tool [with consent of the clinician, s251 is not sought for this]; and
4. Whether the child was deemed abused or not.

Support for activities 2 and 3, above, was not requested.

The researchers will identify cases by attending child protection and radiology peer review meetings. The notes will then be consulted to discover the six clinical indicators the tool uses as the basis for its prediction (seizure, apnoea, long-bone fracture, rib fracture. Head/neck bruising, and retinal haemorrhages). These cases will be given a study identifier and the link between study identifier and NHS number will be retained separately by the CI to allow data-linkage for follow-up.

Clinicians will then be approached to take part in the study. If they agree they will be asked to make an assessment pre and post use of the tool, and what their child protection decisions are pre and post use of the tool.

The research student will follow up each case to determine abuse/non-abuse. This will be determined by the decision of the multidisciplinary assessment conducted by the clinicians, social workers, police, and other relevant agencies at a strategy meeting, case conference, or child death case review meeting. These outcomes will be determined from the children's case notes or child protection peer review meeting. Abuse will also be confirmed if witnessed or admitted to. Where abuse is not suspected (and therefore no multidisciplinary assessment meeting is held), the clinician will be followed up after six months to determine if they have any child protection concerns – unless there is an independently witnessed accidental injury or confirmed underlying organic disease. Follow up data will be obtained from health visitor records.

In addition, they will collect quantitative process data (e.g. the number of eligible patients) and qualitative process data (e.g. researcher observations, structured field notes arising from informal conversations and stakeholder organisations.)

A recommendation for class 1 and 6 support was requested to cover the activity specified in the application.

Confidential patient information requested

Access was requested to access case note data (NHS number, name, and date of birth) relating to head trauma in fifty 0—3 year olds presenting with head injury; from Bristol Royal Hospital for Children and University Hospital Wales.

Confidentiality Advisory Group advice conclusion

The CAG agreed that the minimum criteria under the Regulations appeared to have been met and that there was a public interest in projects of this nature being conducted, and therefore advised recommending support to the Health Research Authority, subject to compliance with the specific and standard conditions of support as set out below.

Clarifications required

1. Clarification is required as to what data will be recorded on the i-pad and what security will be in place for this information. What organisation is responsible for the security and management of data on the i-pad? **Received, letter dated 15 March 2016**
2. Clarification should be provided as to how and by whom data would be extracted from the Health Visitor Records. **Received, letter dated 15 March 2016**

Specific conditions of support

1. An update on patient and public involvement to be included in the first annual review submitted to CAG. **Confirmed, letter dated 15 March 2016**
2. CAG receipt of a favourable opinion from a Research Ethics Committee. **Received, letter dated 21 January 2016**
3. Confirmation from the IGT Team at the Health and Social Care Information Centre of suitable security arrangements via Information Governance Toolkit (IGT) submission.. **Confirmation received by e-mail dated 24 November 2015, version 12, 97%**

As the above conditions and clarifications have been accepted and/or met, this letter provides confirmation of final approval. I will arrange for the register of approved applications on the HRA website to be updated with this information.

Annual review

Please note that your approval is subject to submission of an annual review report to show how you have met the conditions or report plans, and action towards meeting them. It is also your responsibility to submit this report on the anniversary of your final approval and to report any changes such as to the purpose or design of the proposed activity, or to security and confidentiality arrangements. An annual review should be provided no later than 21 March 2017 and preferably 4 weeks before this date. If at any stage you no longer require support under the Regulations as you will cease processing confidential patient information without consent you should inform the Confidentiality Advice Team of this in writing as soon as possible.

Reviewed documents

The documents reviewed at the meeting were:

Document	*Version*	*Date*
CAG application from (signed/authorised)		12 January 2016
Confidentiality policy	1.1	29 June 2015
Covering letter on headed paper		13 January 2016
Data Protection Registration [Z6549747]		18 March 2002
Other [Data flow diagram]		
Other [Peer review]		03 March 2014
Other [Letter confirming employment]		19 January 2016

Other [System Level Security Policy]	1	12 November 2013
Other [Information Security Policy]	1	14 October 2013
Other [IG Toolkit confirmation]		24 November 2015
Other [IG Toolkit]	12	
Other [Checklist]		
Patient Information Materials [Consent form]	3	19 January 2016
Patient Information Materials [Information Sheet]	4	04 February 2016
REC favourable opinion letter and all correspondence [Provisional Opinion]		15 January 2016
REC favourable opinion letter and all correspondence [Further Information Favourable Opinion]		21 January 2016
REC favourable opinion letter and all correspondence [Response to Provisional Opinion]		19 January 2016
Research protocol or project proposal	2	20 November 2015
Write recommendation from Caldicott Guardian (or equivalent) of applicant's organisation		19 January 2016

Membership of the Committee

The members of the Confidentiality Advisory Group who were present at the consideration of this item or submitted written comments are listed below.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R & D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Yours sincerely

On behalf of the HRA.

Christopher Ward
Senior Confidentiality Advisor
Email: HRA.CAG@nhs.net

Enclosures: *List of members who considered application*
Standard conditions of approval

Copy to: corinne.scott@wales.nhs.uk

Confidentiality Advisory Group meeting 25 February 2016

Group Members:

Name	*Present*
Professor Jennifer Kurinczuk	Yes
Dr Tony Calland	Yes
Ms Hannah Chambers	Yes
Dr Patrick Coyle	Yes
Dr Lorna Fraser	Yes
Dr Rachel Knowles	Yes
Mr David Smallacombe	No
Dr Mark Taylor	Yes
Mr Marc Taylor	Yes

Standard conditions of approval

The approval provided by the Health Research Authority is subject to the following standard conditions.

The applicant will ensure that:

1. The specified patient identifiable information is only used for the purpose(s) set out in the application.

2. Confidentiality is preserved and there are no disclosures of information in aggregate or patient level form that may inferentially identify a person, nor will any attempt be made to identify individuals, households or organisations in the data.

3. Requirements of the Statistics and Registration Services Act 2007 are adhered to regarding publication when relevant.

4. All staff with access to patient identifiable information have contractual obligations of confidentiality, enforceable through disciplinary procedures.

5. All staff with access to patient identifiable information have received appropriate ongoing training to ensure they are aware of their responsibilities.

6. Activities are consistent with the Data Protection Act 1998.

7. Audit of data processing by a designated agent is facilitated and supported.

8. The wishes of patients who have withheld or withdrawn their consent are respected.

9. The Confidentiality Advice Team is notified of any significant changes (purpose, data flows, data items, security arrangements) prior to the change occurring.

10. An annual report is provided no later than 12 months from the date of your final confirmation letter.

11. Any breaches of confidentiality / security around this particular flow of data should be reported to CAG within 10 working days, along with remedial actions taken / to be taken.

Appendix 41. R&D approval letter

Bwrdd Iechyd Prifysgol Caerdydd a'r Fro
Cardiff and Vale University Health Board

Ysbyty Athrofaol Cymru
University Hospital of Wales

Heath Park,
Cardiff, CF14 4XW
Phone 029 2074 7747
Fax 029 2074 3838
Minicom 029 2074 3632

Parc Y Mynydd Bychan,
Caerdydd, CF14 4XW
Ffôn 029 2074 7747
Ffacs 029 20743838
Minicom 029 2074 3632

Tel: 029 20746986
Fax: 029 20745311
CAV_Research.Development@wales.nhs.uk

From: Professor C Fegan
R&D Director
R&D Office, 2nd Floor TB2
University Hospital of Wales
Cardiff
CF14 4XW

29 March 2016

Professor Alison Kemp
Clinical Professor of Child Health
5th Floor Neuadd Meirionnydd
Heath Park
Cardiff
CF14 4YS

Dear Professor Kemp

Cardiff and Vale UHB Ref and Study Title: 15/RPM/6359 : Feasibility Of Impact Analysis Of A Clinical Prediction Tool To Estimate The Probability Of Abusive Head Trauma In Children Less Than Three Years Of Age

IRAS Project ID: 167217

Clinical Research Portfolio Ref: 20762

The above project was forwarded to Cardiff and Vale University Health Board R&D Office by the Health and Care Research Wales Permissions Coordinating Unit. A Governance Review has now been completed on the project.

Documents approved for use in this study are:

Document	Version	Date
R&D Form	5.1.0	21/12/2015
SSI Form	5.2.0	06/01/2016
Protocol	2	20/11/2015
Summary, Synopsis Diagram	1	03/11/2015
Interview Schedule topic guides	1	17/11/2015
Validated Questionnaire: Survey	1	16/11/2015
Participant Consent Form	3	19/01/2015
Participant Information Sheet	4	04/02/2016

I am pleased to inform you that the UHB has no objection to your proposal. You have informed us that Cardiff University is willing to act as Sponsor under the Research Governance Framework for Health and Social Care.

Please accept this letter as confirmation of permission for the project to begin within this UHB.

In order to comply with Health and Care Research Wales reporting requirements, you must inform the R&D Office of the date which this site opens to recruitment and the date that the first patient is recruited at this site. Please email this information to CAV research.development@wales.nhs.uk

As your study is adopted onto the UKCRN Clinical Research Portfolio, it is a condition of this NHS research permission that you are required to either upload recruitment data onto the portfolio database, or forward recruitment data to the Chief Investigator to be uploaded.

Please ensure that you email portfolio@wales.nhs.uk with notification when the study has opened to recruitment so that the study can be added to the Health and Care Research Wales Clinical Research Portfolio.

During recruitment accrual data will need to be submitted on a monthly basis to the UKCRN database. Failure to do so may result in the withdrawal of R&D approval. Uploading recruitment data will enable Health and Care Research Wales to monitor research activity within NHS organisations, leading to NHS R&D allocations which are activity driven. Systems have been set up to streamline and make this process as automated as possible. Details on how to upload accrual data are available at http://www.crn.nihr.ac.uk/can-help/funders-academics/nihrcrn-portfolio/portfolio-user-guides/. Please contact portfolio@Wales.nhs.uk if help is required.

May I take this opportunity to wish you success with the project and remind you that as Chief / Principal Investigator you are required to:

- Inform the R&D Office if this project has not opened within 12 months of the date of this letter. Failure to do so may invalidate R&D approval.
- Inform the Health and Care Research Wales Permissions Coordinating Unit and the UHB R&D Office if any external or additional funding is awarded for this project in the future
- Ensure that all study amendments are submitted to the Health and Care Research Wales Permissions Coordinating Unit by the Chief Investigator
- Ensure the Health and Care Research Wales Permissions Coordinating Unit is notified of the study's closure
- Ensure that the study is conducted in accordance with all relevant policies, procedures and legislation
- Provide information on the project to the UHB R&D Office as requested from time to time, to include participant recruitment figures

Yours sincerely,

Professor Christopher Fegan
R&D Director / Chair of the Cardiff and Vale Research Review Service (CaRRS)

CC **R&D Lead Dr Philip Connor, Children's Services**
Sponsor: Helen Falconer, Cardiff University
Student: Miss Laura Crowley, Cardiff University
Academic Supervisor: Dr Daniel Farewell, University Hospital of Wales
Academic Supervisor: Dr Sabine Maguire, University Hospital of Wales
Finance: Anthony Williams, University Hospital of Wales
Clinical Board Assistant Heads of Finance: Catherine David, Children & Women, University Hospital Wales

Appendix 42. Letter of Access

GIG CYMRU NHS WALES | Bwrdd Iechyd Prifysgol Caerdydd a'r Fro
Cardiff and Vale University Health Board

Ysbyty Athrofaol Cymru
University Hospital of Wales
Heath Park,
Cardiff, CF14 4XW
Phone 029 2074 7747
Fax 029 2074 3838
Minicom 029 2074 3632

Parc Y Mynydd Bychan,
Caerdydd, CF14 4XW
Ffôn 029 2074 7747
Ffacs 029 20743838
Minicom 029 2074 3632

Research and Development Office

Email:
CAV_research.development@wales.nhs.uk

24th March 2016

Ms Laura Cowley
7 Sachville Court
Sachville Avenue
Heath
Cardiff

Dear Ms Cowley,

Letter of access for research issued by Cardiff and Vale University Health Board

Title of Agreed Research Project: Feasibility of an impact analysis of a tool to predict abusive head trauma.

R&D Reference: 15/RPM/6359

Agreed Duties to be Undertaken: Attendance at radiology / child protection peer review meetings, extracting data from case notes, testing prediction tool with clinicians, surveying and interviewing clinicians.

This letter confirms your right of access to conduct research through Cardiff and Vale University Health Board for the purpose and on the terms and conditions set out below. This right of access commences on **29th March 2016** and ends on **30 September 2017** unless terminated earlier in accordance with the clauses below.

You have a right of access to conduct such research as confirmed in writing in the letter of permission for research from this NHS organisation. Please note that you cannot start the research until the Principal Investigator for the research project has received a letter from us giving permission to conduct the project.

The information supplied about your role in research at Cardiff and Vale University Health Board has been reviewed and you do not require an honorary research contract with this NHS organisation. We are satisfied that such pre-engagement checks as we consider necessary have been carried out.

You are considered to be a legal visitor to Cardiff and Vale University Health Board premises. You are not entitled to any form of payment or access to other benefits provided by this NHS organisation to employees and this letter does not give rise to any other relationship between you and this NHS organisation, in particular that of an employee.

While undertaking research through Cardiff and Vale University Health Board, you will remain accountable to your employer **Cardiff University** but you are required to follow the reasonable

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

instructions of **Prof Alison Kemp** in this NHS organisation or those given on her/his behalf in relation to the terms of this right of access.

Where any third party claim is made, whether or not legal proceedings are issued, arising out of or in connection with your right of access, you are required to co-operate fully with any investigation by this NHS organisation in connection with any such claim and to give all such assistance as may reasonably be required regarding the conduct of any legal proceedings.

You must act in accordance with Cardiff and Vale University Health Board policies and procedures, which are available to you upon request, and the Research Governance Framework.

You are required to co-operate with Cardiff and Vale University Health Board in discharging its duties under the Health and Safety at Work etc Act 1974 and other health and safety legislation and to take reasonable care for the health and safety of yourself and others while on Cardiff and Vale University Health Board premises. You must observe the same standards of care and propriety in dealing with patients, staff, visitors, equipment and premises as is expected of any other contract holder and you must act appropriately, responsibly and professionally at all times.

If you have a physical or mental health condition or disability which may affect your research role and which might require special adjustments to your role, if you have not already done so, you must notify your employer and Cardiff and Vale University Health Board Research and Development Office prior to commencing your research role at the Health Board.

You are required to ensure that all information regarding patients or staff remains secure and *strictly confidential* at all times. You must ensure that you understand and comply with the requirements of the NHS Confidentiality Code of Practice (http://www.dh.gov.uk/assetRoot/04/06/92/54/04069254.pdf) and the Data Protection Act 1998. Furthermore you should be aware that under the Act, unauthorised disclosure of information is an offence and such disclosures may lead to prosecution.

You should ensure that, where you are issued with an identity or security card, a bleep number, email or library account, keys or protective clothing, these are returned upon termination of this arrangement. Please also ensure that while on the premises you wear your ID badge at all times, or are able to prove your identity if challenged. Please note that this NHS organisation accepts no responsibility for damage to or loss of personal property.

We may terminate your right to attend at any time either by giving seven days' written notice to you or immediately without any notice if you are in breach of any of the terms or conditions described in this letter or if you commit any act that we reasonably consider to amount to serious misconduct or to be disruptive and/or prejudicial to the interests and/or business of this NHS organisation or if you are convicted of any criminal offence. You must not undertake regulated activity if you are barred from such work. If you are barred from working with adults or children this letter of access is immediately terminated. Your employer will immediately withdraw you from undertaking this or any other regulated activity and you MUST stop undertaking any regulated activity immediately.

Your substantive employer is responsible for your conduct during this research project and may in the circumstances described above instigate disciplinary action against you.

Cardiff and Vale University Health Board will not indemnify you against any liability incurred as a result of any breach of confidentiality or breach of the Data Protection Act 1998. Any breach of the Data Protection Act 1998 may result in legal action against you and/or your substantive employer.

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

If your current role or involvement in research changes, or any of the information provided in your Research Passport changes, you must inform your employer through their normal procedures. You must also inform your nominated manager in this NHS organisation.

Yours sincerely

L. Hathaway

Mrs Lee Hathaway

Registration and Permissions Improvement Manager
On behalf of Cardiff and Vale University Health Board

CC **HR department of the substantive employer**
Manager at Cardiff and Vale UHB

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

Appendix 43. Approval from University Hospitals Bristol NHS Foundation Trust

University Hospitals Bristol NHS
NHS Foundation Trust

Dr Giles Haythornthwaite
Bristol Royal Hospital for Children
Paul O'Gorman Building
Upper Maudlin Street
Bristol
BS2 8BJ

Research and Innovation
University Hospitals Bristol NHS Foundation Trust
Education & Research Centre Level 3
Upper Maudlin Street
Bristol BS2 8AE

Tel: 0117 342 0233
Fax: 0117 342 0239

research@uhbristol.nhs.uk
http://www.uhbristol.nhs.uk/research-innovation

31/03/2016

NHS Permission for Research has been granted for the study detailed below at University Hospitals Bristol NHS Foundation Trust (UH Bristol). Permission is subject to any conditions and is effective from 31/03/2016 until 30/09/2017.

Dear Dr Haythornthwaite

RE: Feasibility of an impact analysis of a tool to predict abuse - R&D Number CH/2015/5028

NHS permission for the above research has been granted on the basis of the application submitted and a favourable opinion from an authorised REC.

Permission is granted on the understanding that the study is conducted in accordance with the Research Governance Framework, Good Clinical Practice, and NHS Trust policies and procedures. As Principal Investigator it is your responsibility to ensure you and your team read, understand and abide by the relevant research related policy and procedures as applicable to your study; these can be found at http://www.uhbristol.nhs.uk/research-innovation/our-research/strategy,-policy-sops/

Please note, any training in UH Bristol Trust research SOPs must be documented for all team members as applicable to their role, prior to study start. Any queries or inconsistencies with trial or departmental SOPs should be discussed with the R&I office before study start.

It is also a condition of NHS Permission at this site that local recruitment data is uploaded to the EDGE system and the study record is kept up-to-date. Please contact the Research Management Office if you are unsure how to do this.

The following conditions must be met prior to recruitment commencing:

• A site file is set-up and delegation log established.

UH Bristol is required to monitor research to ensure compliance with the Research Governance Framework and other legal and regulatory requirements. For further details about monitoring arrangements please contact the Research Management Office. The Research Management Office will monitor recruitment on an on-going basis and can provide support and advice if you are experiencing problems in meeting your targets within the agreed time frame.

The Research Management Office should be notified of any urgent safety measure taken in order to protect research participants against any immediate hazard to their health or safety. This should be within the same time frame as notification to the REC and any other regulatory bodies and should include the reasons why the measures were taken and any plan for further action.

NHS indemnity is provided for the period of permission given above. Requests for changes to the period of permission (e.g. an extension of the study) must be made to the Research Management Office before permission ceases with an explanation as to why the change is being sought.

All amendments (including changes to the local research team) need to be submitted in accordance with regulatory and national requirements which can be found on IRAS. Please note if we are sponsoring this study separate notification of an amendment already authorised by us as sponsor for submission to the regulatory bodies is not required, the sponsor authorisation will cover R&D acknowledgement of the amendment at this trust. The Research Management Office also needs to be notified if there are any changes to the study status.

We wish you every success with this study.

Yours sincerely,

Diana Benton
Head of Research and Innovation/Deputy Director of Research

Copy to:

POINT OF CONTACT : Laura Cowley

SPONSOR: Cardiff University (Miss Helen Falconer)

Appendix 44. Letter of Access to University Hospitals Bristol

University Hospitals Bristol NHS
NHS Foundation Trust
Jobs@UHBristol
Resourcing
The Courtyard
Old Building, BRI
Lower Maudlin Street
Bristol
BS2 8HW
Tel: 0117 342 5000 – opt 1
HonoraryRequests@uhbristol.nhs.uk

PRIVATE AND CONFIDENTIAL
Laura Cowley
7 Sachvile Court
Sachville Avenue
Cardiff

31 March 2016

Dear Laura,

Letter of access for UH Bristol R&D Number: CH/2015/5028
Study Title: Feasibility of an impact analysis of a tool to predict abuse.

This letter confirms your right of access to conduct research through UH Bristol or the purpose and on the terms and conditions set out below. This right of access commences on 31/03/2016 and ends on 30/09/2017 unless terminated earlier in accordance with the clauses below.

You have a right of access to conduct such research as confirmed in writing in the letter of permission for research from this NHS organisation UH Bristol. Please note that you cannot start the research until the Principal Investigator for the research project has received a letter from us giving permission to conduct the project.

The information supplied about your role in research at has been reviewed and you do not require an honorary research contract with this NHS organisation. We are satisfied that such pre-engagement checks as we consider necessary have been carried out.

You are considered to be a legal visitor to UH Bristol premises. You are not entitled to any form of payment or access to other benefits provided by this NHS organisation to employees and this letter does not give rise to any other relationship between you and this NHS organisation, in particular that of an employee.

While undertaking research through UH Bristol you will remain accountable to your employer the University of Bristol but you are required to follow the reasonable instructions of Dr Giles Haythornthwaite in this NHS organisation or those given on her/his behalf in relation to the terms of this right of access.

Where any third party claim is made, whether or not legal proceedings are issued, arising out of or in connection with your right of access, you are required to co-operate fully with any investigation by this NHS organisation in connection with any such claim and to give all such assistance as may reasonably be required regarding the conduct of any legal proceedings.

You must act in accordance with UH Bristol policies and procedures, which are available to you upon request, and the Research Governance Framework.

You are required to co-operate with UH Bristol in discharging its duties under the Health and Safety at Work etc Act 1974 and other health and safety legislation and to take reasonable care for the health and safety of yourself and others while on UH Bristol premises. You must observe the same standards of care and propriety in dealing with patients, staff, visitors, equipment and premises as is expected of any other contract holder and you must act appropriately, responsibly and professionally at all times.

You are required to ensure that all information regarding patients or staff remains secure and *strictly confidential* at all times. You must ensure that you understand and comply with the requirements of the NHS Confidentiality Code of Practice (http://www.dh.gov.uk/assetRoot/04/06/92/54/04069254.pdf) and the Data Protection Act 1998. Furthermore you should be aware that under the Act, unauthorised disclosure of information is an offence and such disclosures may lead to prosecution.

You should ensure that, where you are issued with an identity or security card, a bleep number, email or library account, keys or protective clothing, these are returned upon termination of this arrangement. Please also ensure that while on the premises you wear your ID badge at all times, or are able to prove your identity if challenged. Please note that this NHS organisation accepts no responsibility for damage to or loss of personal property.

We may terminate your right to attend at any time either by giving seven days' written notice to you or immediately without any notice if you are in breach of any of the terms or conditions described in this letter or if you commit any act that we reasonably consider to amount to serious misconduct or to be disruptive and/or prejudicial to the interests and/or business of this NHS organisation or if you are convicted of any criminal offence. Your substantive employer is responsible for your conduct during this research project and may in the circumstances described above instigate disciplinary action against you.

UH Bristol will not indemnify you against any liability incurred as a result of any breach of confidentiality or breach of the Data Protection Act 1998. Any breach of the Data Protection Act 1998 may result in legal action against you and/or your substantive employer.

If your current role or involvement in research changes, or any of the information provided in your Research Passport changes, you must inform your employer through their normal procedures. You must also inform your nominated manager in this NHS organisation.

Yours sincerely

Amelia Armstrong
Recruitment Coordinator