Validating the CORE-10 as a mental health screening tool for prisoners

Gwen Lewis

Supervised by

Professor Reg Morris

May 2016

Thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Clinical Psychology at Cardiff University and the South Wales Doctoral Programme in Clinical Psychology

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 fulfilment of the requirements for the degree of D.Clin.Psy.

Signed ……(candidate) Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. 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 ……

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed …… (candidate) Date ……

Acknowledgments

I would firstly like to thank the participants who shared their personal stories and willingly gave up their time to take part in this research.

Secondly thank you to the staff at Cardiff and Parc prisons for enthusiastically supporting my idea, organising access and facilitating data collection. In particular thank you Sharron Jones and Karen Mills for your dedicated practical support, great organisation and patience

A huge thank you to my supervisor Professor Reg Morris who's expertise, dedication, enthusiasm and expeditious approach has been invaluable in guiding me through the thesis process.

A special thank you to Professor Mike McGuire who so kindly shared his wealth of knowledge, linked me with many a contact and provided me with invaluable feedback. Thank you also to Mike Hardy at the Welsh Government and Rob Heaton-Jones at NOMS for their on-going support of this research.

My utmost appreciation and thanks to Naomi Maturano, Jessica Corey and Nicola Lewis who skilfully and professionally supported with data collection.

Finally thank you to my Mother Jackie for her meticulous proof reading skills, and to my partner Kevin for his final checking and patience throughout.

Abstract

Background: Few mental health screening tools have been validated with prisoners and existing tools, do not assess severity of need in line with contemporary stepped care service models.

Aims: The current research aims to assess the CORE-10's psychometric reliability, validity and predictive accuracy as a screening tool for common (primary care) and severe (secondary care) mental health problems in prisoners.

Method: Cross –sectional study of 150 prisoners. All participants completed the CORE-10, Mini International Neuropsychiatric Interview version 6.0 (MINI) and the GHQ-12. Eighty-one participants repeated the CORE-10 and GHQ-12 two weeks later to assess re-test reliability. Clinical judgment data concerning referral for primary or secondary mental health services in prison were retrieved for each participant. Correlational, ROC and confirmatory factor analysis were utilised to assess the psychometric properties of the CORE-10 in comparison to the MINI, GHQ-12 and clinical judgment.

Results: Significant positive correlations were identified between the CORE-10 and all other measures of mental health. ROC analysis on the CORE-10 against the MINI 6.0 revealed significant areas under the curve for predicting both primary (AUC .85) and secondary care (AUC .76) level needs. At cut points of >6 for primary care and >10 for secondary care sensitivity was .88 and .83, with specificity of 64 and .61 respectively. Sensitivity and specificity of the CORE-10 was superior to current clinical judgment and the GHQ-12. Internal reliability (α .84-.89) was good and two-week re-test reliability (ICC=.83) moderate. Confirmatory factor analysis confirmed the CORE-10's original six-factor model to be a good fit.

Conclusions: The CORE-10 is an accurate screen for common and severe mental health problems in prisoners. The CORE-10 is a psychometrically robust tool for use with prisoners demonstrating convergent, discriminate and construct validity as well as good internal and retest reliability.

Table of Contents

List of tables

List of figures

1 Introduction

1.1 Focus of the thesis

Psychiatric morbidity within prison populations is considerably higher than in community populations (Fazel & Seewald, 2012; Singleton *et al*, 1998). Untreated mental health problems amongst prisoners have been linked to increased institutional violence, self-harm, suicide and reoffending (Martin *et al*., 2013). Contemporary policy has aimed to improve access to treatment for mental health problems in Wales for both common and severe mental health problems within a stepped care service model (Welsh Government, 2010; 2012a; 2012b; 2014). Secondary care mental health services typically treating severe mental illness (SMI) are well established within Welsh prisons. However, whilst Primary care services targeting common mental health problems have been embedded within the community, their implementation in Welsh prisons is in its infancy (Little, 2013). Accurately targeting such services to those in need is dependent on effective screening and identification of mental health problems. Reception health screening is integral to the identification of prisoners with mental health problems entering custody (Grubin *et al*., 2002; OHRN, 2010). Current screening processes emphasise identification of severe mental illness through assessment of historic risk factors (Grubin *et al*., 2002), but do not assess current distress which can be indicative of common mental health problems.

Clinical guidance and policy advocates improving screening through consistent unified use of structured assessment tools in screening procedures (NICE, 2011a; Welsh Government, 2014). No validated universal mental health screening tools are consistently utilised in Welsh prisons (Little, 2013). Furthermore, few mental health-screening tools have been validated with prisoners and the reliability of tools validated in community populations cannot be assumed since the higher base rate of mental health problems within prison populations changes the sensitivity, specificity and consequent predictive accuracy of such tools in this population (Weiner & Graham, 2003).

Psychometric tools that have been validated for use in prisons have not been assessed in terms of their utility for screening purposes and cut off points for distinguishing between common (primary) and severe (secondary care) [1] need in line with existing service models have not been established. For example, although the GHQ-12 has been validated for use with prisoners (Boothby e*t al*., 2010), its utility as a screening tool has not been assessed and valid cut off points for primary and secondary care levels of need have not been identified.

The CORE-10 is a brief self report screening and outcome measure of global distress designed for use in busy clinical settings (Barkham *et al*., 2010; Connell & Barkham, 2007). The CORE-10 consists of high and low intensity items and taps six factors: anxiety, depression, functioning, risk, trauma, physical symptoms and risk. The CORE-10 has been found to be feasible, reliable and valid in community settings (Barkham *et al*., 2013; Connell & Barkham, 2007), but has not been validated in a prison context.

The primary aim of this thesis is to address this gap in the literature, policy and practice by determining the psychometric properties of the CORE-10 with a prisoner population. Specifically, its predictive validity for screening for both common and severe mental health problems in prisoners is assessed. This is achieved by comparing the CORE-10 against the GHQ-12 (a brief screen validated for use with prisoners), current practice in terms of referral to prison mental health services, and the MINI 6.0 gold standard diagnostic interview. Establishing the validity of the CORE-10 for screening purposes with prisoners offers the potential to improve screening processes in line with current service models, policy imperatives and clinical guidance (NICE, 2011a; Welsh Government, 2010; 2012b: 2014). In the long term improved screening may decrease the burden of untreated mental health problems in

[1] It is acknowledged that the terms 'common' and 'severe' mental health problems do not have clearly defined parameters and that there is substantial cross over between the two. However, for the purposes of the current thesis their use is consistent with service models in the UK and not subjected to theoretical scrutiny, as the emphasis of the thesis is on the utility of psychometric tools to support current service models, not on the theoretical validity of the terms and how they are applied.

prisoners, support streamlined referral pathways and consequently promote cost-savings.

1.2 Definitions of key terms[2]

Prison

In the United States (US), prisons typically detain prisoners who have been convicted and sentenced to a year or more in a correctional facility. In the UK, the term prison encompasses a range of establishments across the custodial estate including local remand, resettlement and high secure prisons as well as youth offending institutions. In the current thesis the term prison is used in line with the above definitions dependent on the country concerned.

Jail

In the US, jails typically detain individuals who are pre-trial or on remand or who have been sentenced to short periods of incarceration (Gangon, 2009). The term jail is not typically used in the United Kingdom (UK).

Common mental health problems

Definitions of common mental health problems vary. In the current thesis, the UK NICE (2011a) definition of common mental health problems or disorders, as including depression, generalised anxiety disorder (GAD), panic disorder, phobias, social anxiety disorder, obsessive-compulsive disorder (OCD) and post-traumatic stress disorder (PTSD) is adopted.

Severe mental illness (SMI)

There is no consistency of operational definitions for severe mental illness (Ruggeri *et al*., 2000). However, for the purposes of the current thesis SMI is defined as a diagnosis of major depressive disorder, bipolar i, bipolar ii, bipolar disorder not otherwise specified and mood disorder with psychosis or

[2] Throughout the thesis non client-centred language in terms of diagnostic labels and forensic latent variables are used in order to accurately reflect the literature drawn upon. It is however acknowledged that such language can be pejorative, stigmatising and unhelpful (Boyle & Johnstone, 2014), thus should read with a critical lens.

non-organic psychotic disorder in line with existing British research and policy (Grubin *et al*., 2002; OHRN, 2009; Welsh Government, 2014).

Screening

Screening in the current context refers to the brief assessment of mental distress, the results of which may indicate the need for further assessment. Broadly, screening is defined by the Medical Dictionary (2007) as the 'examination of a group to separate well persons from those who have an undiagnosed pathologic condition or who are at high risk, intended to determine suitability for a particular treatment modality'.

1.3 Literature review

The UK prison population on 18th March 2016 was 85,930, with the majority of those being adult males (Ministry of Justice, 2015a). On average a prison place costs £36,237 per prisoner per annum and reoffending rates remain high with 45% of adults released from prison being reconvicted within a year (Ministry of Justice, 2015b). Deaths in custody, self-harm and serious assaults in British prisons have increased in the last year (Ministry of Justice, 2015c), with identification and treatment of prisoners with mental health problems remaining a priority.

1.3.1 Psychiatric morbidity in the prison population

The prison population has steadily increased since 1993 (Ministry of Justice, 2014a), with arguments regarding the increase in proportion of those with mental illness detained in prisons emerging as far back as the 1970's following de-institutionalisation. Historically a number of theories have been proposed to explain the suggested increase in prevalence of mental health problems in the prison population. Theories concerning 'criminalisation of the mentally ill' (Abramson, 1972) suggest that following de-institutionalisation there was an increasing trend towards criminal justice responses to mentally disordered behaviour. Some have argued that a process of re-institutionalisation through imprisonment took place in the context of a paucity of community services (Priebe *et al*., 2005). Evidence to support the criminalisation theory is however equivocal (Fisher *et al*., 2006). Other

theories proposed to explain the increase in those with mental illness in custody include the 'psychiatricatization of criminality' (Davis, 1992) as well as prevailing punitive social attitudes towards criminality (Lamb & Weinberg, 1998) and the social construction of 'dangerousness' relating to mental illness (Link *et al*., 1999). Despite such theories, evidence regarding the increased prevalence of mental illness in prison populations is inconclusive (Lamb & Weinberg, 1998). In part, this is attributable to a lack of systematic high quality research over time, along with screening and service developments, which have improved the identification of mental illness in offenders. Contemporary literature tends to explain rates of mental illness in offender populations in light of complex interactions between social, legal, criminological and political factors (Fisher *et al*., 2006).

Despite debate regarding whether the prevalence of mental health problems has increased in Western prison populations in the last half a century, there is consensus that prevalence of mental illness is substantially higher in incarcerated compared to community populations (Birmingham *et al*., 1996; Fazel & Seewald, 2012; Linehan *et al*., 2005; Teplin, 1994). Several studies concerning prevalence of mental health problems in prisoners were published in the 1990's and early 2000's producing heterogeneous prevalence estimates (Fazel & Seewald, 2012). Early research in a US prison identified 33% of new receptions as having a current SMI or substance misuse disorder, increasing to 62% when considering lifetime prevalence (Teplin, 1994). Prevalence of disorders was higher in White prisoners and increased with age (Teplin, 1994). More recently, in a large sample of US jail detainees 14.5% were screened as having a current SMI (Steadman *et al*., 2009), while 25% self-reported a previously diagnosed mental health condition (Wilper *et al*., 2009). Worldwide meta-analyses have identified prevalence rates of 4% for psychosis, 10% for major depression and 65% for personality disorders in male prisoners (Fazel & Seewald, 2012). Rates of comorbid substance and mental disorder ranged between 20 and 43% (Fazel & Seewald, 2012). Such findings highlight the high prevalence of mental illness in incarcerated populations globally.

In the UK, Birmingham *et al*. (1996) found that in a sample of 569 prisoners on remand 26% had one or more current SMI and only a quarter of those were identified by exiting screening procedures. In a larger national survey the Office for National Statistics (Singleton *et al*.,1998) estimated that 55% of prisoners suffer from a neurotic disorder, 10% of prisoners exhibit psychotic symptoms and 65% have some form of personality disorder, with significantly higher rates among remanded compared to sentenced prisoners. Overall, 90% of prisoners were found to have one or more psychiatric disorder when substance misuse disorders were included (Singleton *et al*., 1998). In 2005, Burgha *et al*. reported that the prevalence of functional psychosis in the last year in British prisons was 10 times higher than the community prevalence rate. Moving away from diagnostically based estimates, Boothby e*t al*. (2010) using the GHQ-12 reported that 59% of adult male prisoners in a London local prison were clinically emotionally distressed. No research concerning the prevalence of mental health problems in British prisons seems to have been published in recent years, despite changes in the prison population. For example, the rapid increase in older prisoners (House of Commons, 2013) is likely to have had a significant effect: Fazel *et al*. (2001) reported that one in three older prisoners suffers from depression, with less than a fifth being treated with medication.

Heterogeneity in estimates of prevalence of mental health disorders amongst prisoners arises largely from the use of divergent methodological approaches across studies. There is a lack of consistency between studies in how mental disorder is defined (Fazel & Danesh, 2002), for example some focus on SMI (Teplin, 1994) while others encompass a broad range of disorders (Singleton *et al*., 1998). Some utilise structured diagnostic interviews as a criterion (Fazel & Seewald, 2012), whilst others rely on prisoners self-report of diagnosis (Wilper *et al*., 2009) or symptoms (Boothby *et al*., 2010). Temporal discrepancy also exists across studies in term of whether prevalence is estimated in terms of lifetime (Teplin, 1994) or current mental disorder (Birmingham *et al*., 1996). Definitions of current vary further, ranging from within the last year (Burgha *et al.,* 2005) to last few weeks (Boothby *et al*., 2010). In addition to methodological issues, cross cultural comparisons are

also compounded by differing constructions of mental illness, mental health service provision and legal frameworks across countries, all of which affect estimates of prevalence in incarcerated samples.

Methodological and conceptual issues preclude precise estimates of mental disorders in prisoner populations. Despite this, the literature consistently cites prevalence rates of mental disorders and psychological distress which substantially exceed those found in community populations. Research suggests that large numbers of prisoners suffer from mental health problems, which co-occur with a complex array of other substance misuse, learning and social needs (Bradley, 2009). Developing methods of effectively screening and subsequently treating prisoners with mental health problems is imperative for promoting prisoner wellbeing, improving institutional outcomes and decreasing costs.

1.3.2 Predictors of mental health problems and distress in prison

High prevalence rates of mental health problems in prisoners have provoked exploration of risk factors associated with the expression of mental health difficulties in custody. Adjustment to prison is thought to involve a complex interaction between individual and environmental factors (Clear & Sumter, 2002; Jiang & Winfree, 2006; Steiner & Wooldredge, 2008; Wright *et al.,* 1991). With regard to environmental factors access to more activities, low levels of 'environmental hassle' (Cooper & Berwick, 2001) and higher levels of perceived freedom and safety have been correlated with lower levels of distress in custody (Wright *et al*., 1991). Sentencing factors have also been cited as risk factors for mental health problems in prison, including being on remand or unsentenced (OHRN, 2010) and having a charge of homicide (Birmingham *et al*., 2000).

In terms of individual factors, a history of self-harm, previous psychiatric service contact and treatment (Birmingham *et al*., 2000; Grubin *et al*., 2002), substance misuse (DiCataldo *et al.,* 1995; Greenberg & Rosenheck, 2008; McNeil *et al*., 2005) and homelessness (Greenberg & Rosenheck, 2008; McNeil *et al*., 2005) are strongly associated with higher rates of mental health

problems in prison. Additionally, in a small-scale retrospective cross sectional study Hochstelter *et al*. (2004) reported that low self-control, race (being white) and previous trauma all directly predicted distress in prison. Mixed findings have also been reported from small scale studies with regard to the role of social support in predicting mental health problems in prison. Some studies have cited a lack of support outside of prison as a risk factor for distress (Cooper & Berwick, 2001), while others have reported having support such as a spouse increased distress in custody (Linquist, 2000). Overall, a number of environmental and individual factors interact in complex ways to predict individual adjustment to the prison environment. However, the interplay between risk factors is poorly understood and existing research is patchy with little exploration of current situational and personal risk factors. Research has however identified a number of consistent historic individual factors that predict mental health problems in custody which have been helpfully utilized in screening processes.

1.3.3 The effects of imprisonment on mental health

Exploration of risk factors concerning those presenting with mental health problems in prison has also led to debate regarding the effect of incarceration itself on mental health. Incarceration is typically considered a stressful experience which can increase distress amongst a population who typically have already experienced a constellation of hardships (Hochstetler *et al*., 2004). Negative effects of incarceration on mental health have been delineated (Gibbs, 1991; Haney, 2001; Linquist & Linquist, 1997), although there is great variation in prisoners' accounts of custody and its psychological effects (Hemmens & Marquart, 1999). Specifically, prison overcrowding has been associated with psychological distress (Evans, 2003; Werner & Keys, 1988). Nurse *et al's*. (2003) qualitative research in UK prisons revealed five factors which are detrimental to mental health in prison: isolation, drug misuse, poor relations with staff, bullying by other prisoners and staff and lack of family contact. Psychiatric symptoms may exacerbate such factors since they increase risk of segregation, victimization and disciplinary sanctions and decrease opportunities to access activities thus increasing risk of isolation (DiCataldo *et al*., 1995). Fellner (2006) and others (Andersen *et al*., 2003;

Haney, 2003) have also highlighted how isolation resulting from segregation can be particularly detrimental to prisoners with mental health problems, and can lead to a cycle of emergency treatment and re-segregation. Despite there being limited research directly concerning the impact of imprisonment on individuals with mental health problems, prisons are often deemed 'anti-therapeutic' environments (OHRN, 2010; Scott, 2004).

Research has however demonstrated that imprisonment does not have a universally detrimental impact on mental health (Botna & Gendreau, 1990), even amongst those with pre-existing mental health problems (Andersen *et al*., 2003; Blaauw *et al*., 2007; Taylor *at al*., 2010). In a prospective cohort study of 3097 prisoners received into British prisons Hassan *et al*. (2011) found that psychiatric symptoms did not deteriorate following two months of incarceration in both those with and without existing mental illnesses. Indeed, some studies have reported improvements in mental health following incarceration (Andersen *et al*., 2000; 2003; Taylor *at al*., 2010). Improvements in symptoms are typically attributed to increased structure, decreased exposure to substances, treatment of withdrawal symptoms, improved nutrition and removal of external stressors (Andersen *et al*., 2003; Blaauw *et al*., 2007; Hassan *et al*., 2011). These studies have however predominantly focused on the early stages of imprisonment, all with relatively short follow up periods of three months or less and limited controls for pre-existing characteristics. Overall, the effects of imprisonment on mental health are difficult to demonstrate and poorly understood since inmates enter prison with differing characteristics, many of which are often existing risk factors for poor mental health outcomes (Fazel & Lubbe, 2005; Hochstetler *et al*., 2004). Nevertheless, screening offers the opportunity to identify those in distress in order to target interventions, which may in turn ameliorate the potential of incarceration exacerbating symptoms.

1.3.4 Systemic impact of mental health difficulties in prison establishments

Failure to appropriately identify and support prisoners with mental health problems has not only a detrimental effect on individuals, but also a significant systemic impact on establishments with significant cost implications across

public services (McCrone *et al*., 2008). Undetected and untreated mental health problems amongst prisoners have been correlated to increased risk of self-harm and suicide, disciplinary infringements and violence as well as risk of victimisation and reoffending (Martin *et al*., 2013). Such systemic implications are discussed in detail in the following sections.

1.3.5 Self-harm

Self-harm, defined as any act of self-poisoning or self-injury carried out by an individual, irrespective of motivation (NICE, 2011b), has increased amongst male prisoners in recent years, with the highest numbers outside the female estate being seen in male category B local prisons (Hawtton *et al*., 2014; Ministry of Justice, 2014b). Hawtton *et al's*. (2014) case controlled longitudinal study across England and Wales' prison estate estimated that between five and six per cent of male prisoners self-harm, with repetition of self-harm common. This compares to just 0.6% of men in the community self-harming in the last year (Bebbington *et al*., 2010). For male offenders, being younger, white, being sentenced, serving a life sentence, being in a high-secure prison (Hawtton *et al*., 2014), having a diagnosis of personality disorder (Hillbrand *et al.,* 1994), hopelessness (Grey *et al*., 2003) and anger (Humber *et al*., 2013) have all been associated with increased incidence of self-harm. In terms of lethality, suicidal intent, suicidal ideation and depression in prisoners have been correlated with more lethal self-injurious behaviours, whereas psychopathy has been correlated with less lethal acts of self-harm (Lohner & Konrad, 2006). Studies concerning risk factors for self-injurious behaviour in custody are, however, limited, contradictory and confounded by significant divergences in operational definitions of self-injurious behaviour across studies (Lohner & Konrad, 2006). Furthermore, studies concerning self-harm typically rely on retrospective data which depends upon the accurate identification and recording of self-harm incidents by staff leaving significant room for under-detection (Hawton *et al*., 2014). Nevertheless, research has linked self-harm to a number of psychiatric disorders and a history of self-harm has been found to substantially increase the risk of suicide in custody (Fazel *et al*., 2008; Owens *et al*., 2002). As such mental health screening is

an important element supporting effective management and treatment of self-harm.

1.3.6 Suicide

There has been a substantial increase in self-inflicted deaths in custody in England and Wales in recent years, with 75 self-inflicted deaths in custody in 2013 (Ministry of Justice, 2014b). Rates of suicide have been shown to be five times higher within male prison populations compared to their community comparators (Fazel *et al*., 2005). Likelihood of death by suicide in men recently released from prison is also eight times higher compared to their general population counterparts (Pratt *et al*., 2006).

A number of risk factors for suicide and suicidal behaviour in prison have been identified. Prison based variables in terms of previous spells in prison, being in custody for less than 30 days, being in their current prison for less than 30 days, serving a life sentence, being unsentenced, single cell accommodation and being bullied in prison have been associated with near lethal or lethal suicide attempts (Fazel *et al*., 2008; Fruehwald *et al.*, 2004; Rivlin *et al*., 2010; Rivlin *et al*., 2013). Psychiatric disorders have been consistently related to suicide (Cavanagh *et al*., 2003). In prisoners, psychiatric disorders - particularly psychosis, neurosis and drug dependence - are highly related to suicidal thoughts and to both lifetime suicide attempts and attempts in the last year (Jenkins *et al*., 2005; Rivlin *et al*., 2010; Shaw *et al*., 2004). Suicidal thoughts and attempts in prisoners, as in the community (Meltzer *et al*., 2002) are also correlated to social factors including lack of social support, poor education and previous adversity such as being in care (Jenkins *et al*., 2005; Rivlin *et al*., 2010; 2013). Risk of successful suicide is associated with self-harm (especially amongst older prisoners and those who repeatedly seriously self-harm) and previous suicidal behaviour in prisoners (Fruehwald *et al.*, 2004; Hawtton *et al*., 2014). Clearly, mental health problems and self-harm, along with other prison based and social risk factors, contribute to male prisoners' increased risk of suicide.

Despite research consistently identifying a number of risk factors for suicide in prisoners, studies relating to suicide are confounded by a number of methodological limitations. For obvious ethical reasons and logistical constraints due to the low overall incidence of suicide, many studies concerning suicide are retrospective in design (Fazel *et al*., 2008; Fruehwald *et al*., 2004). As such, data is often incomplete or not available particularly surrounding dynamic clinical variables (Fazel *et al*., 2008). Research on suicide in prisons is typically dependent on prison reporting systems, where some variables like self-harm are likely to be underreported (Hawton *et al*., 2014). Some studies have overcome this by focusing on near lethal attempts (Rivlin *et al*., 2010; 2013), although such studies still depend on participants' retrospective recall to a time of acute stress. With regard to studies concerning the relationship between psychiatric disorder and suicide, variation exists in definitions and measurement of psychiatric disorder. Furthermore, due to the low incidence of suicide and lack of available data, studies are often unable to explore variation within risk factors - for example, whether specific diagnoses are linked to increased risk of suicide. Despite such limitations research has consistently found significant relationships between mental ill health, suicide and self-harm in prisoners. Consequently, quickly identifying and supporting prisoners with mental health problems is vital to reduce the risk of self-harm and suicide in prions.

1.3.7 Violence and disciplinary infringements

Prisoners with untreated mental health problems may not only constitute a risk to themselves, but also to others and the prison environment. Historically, research has produced mixed results regarding the association between mental health problems and violence (Elbogen & Johnson, 2009; Silver *et al*., 2008; Steadman *et al*., 1998). Research specifically with prisoners has also failed to consistently find significant linear associations between mental illness and institutional violence (Adams, 1983; Torch & Adams, 1986; DiCataldo *et al*., 1995; Ditton, 1999; McCorkle, 1995). For example DiCataldo *et al*. (1995) US study of 514 maximum-security prisoner found that prisoners who screened positively for a SMI on the Referral Decision Scale (RDS) were

significantly more likely to be found with a weapon in their possession and those with schizophrenia had significantly more recorded incidents in their first 90 days of incarceration. However, there was no difference in rates of incidents during later time periods. Additionally, McReynolds & Wasserman (2008) found that incarcerated youths with psychiatric disorders were less likely to have disciplinary infractions than their counterparts without psychiatric disorders. Mixed results are propagated in part by methodological issues surrounding inconsistent definitions and measurement of mental illness, reliance on self-report, the American centric samples and poor controlling for confounding variables (e.g. DiCataldo *et al*., 1995). Furthermore, mental illness and violence share many of the same risk factors (Aneshenesel, 1992; Felson *et al*., 2012), which makes it difficult to distinguish causal from spurious finding and potentially contributes to mixed findings.

Today there is a growing body of research indicating that untreated mental health problems with co-morbid substance misuse increase risk of violence in community settings (Swartz *et al*., 1998). Replicating patterns identified in community samples, prisoners with a dual diagnosis are more at risk of being both perpetrators and victims of assaults (Wood & Buttaro, 2013). Research has also begun to uncover a more complex relationship between mental health and violence in prisoners. Specifically paranoid thinking, psychosis and depression have been identified as predictors of aggressive and non-aggressive offending in custody, whereas anxiety disorders are not (Felson *et al*., 2012). Furthermore, Walters (2011) found that 'criminal thinking' significantly mediated the relationship between major mental illness and physical aggression in US prisoners, although mental illness was not formally assessed.

This more recent, if relatively small, body of evidence suggests that some mental health problems, particularly if undetected and comorbid with substance misuse (Fellner, 2006), can increase the risk of violence. However, the relationship between mental illness and violent and non-violent disciplinary infractions in custody is complex with mediating systemic and individual factors at play. Some instances of violence and disciplinary

problems may be symptomatic of an underlying mental health problem (Adams, 1986; Grey *et al*., 2003). Prison disciplinary officers are often poorly trained to distinguish between inmates 'acting out' and the behavioural manifestations of mental illness, which are frequently inappropriately dealt with via disciplinary mechanisms (Fellner, 2006). Poor training and a reluctance to consider mental health considerations in disciplinary hearings may contribute to high levels of disciplinary actions amongst mentally ill prisoners (Fellner, 2006). Screening can thus offer an opportunity for early identification and intervention of those with mental health problems to reduce their risk of subsequent violence (for some groups) and disciplinary infringements, which place a safety and resource burden on prisons.

1.3.8 Victimisation

Prisoners with mental health problems may be at increased risk of displaying violent behaviour and disciplinary infringements, but are also more likely be victimized in custody - like their counterparts in the community (Goodman *et al*., 2001; Maniglio, 2009). Unusual behaviours amongst prisoners with mental illness may irritate other prisoners or make them appear easy targets for exploitation increasing their risk of victimisation and bullying (Blaauw *et al*., 2001; Olgoff *et al*., 1994). Wolff *et al.* (2007) found that six month self-reported prevalence rates of sexual assault by other inmates or staff were substantially higher among mentally ill as compared to non-mentally ill prisoners. Furthermore, prisoners with dual diagnosis and a history of childhood physical or sexual assault have been found to be at increased risk of physical and sexual abuse in prison (Wolff & Shin, 2009). In turn, victimisation in itself has also been found to predict PTSD and depressive symptoms in prisoners (Hochstetler *et al*., 2004). Such studies often however fail to control for confounding variables such as perpetration of violence, which can in turn increase risk of victimization (Wood & Buttaro, 2013). Accurate identification of prisoners with mental health problems may aid appropriate location of vulnerable individuals thus potentially reducing the risk of victimisation (Nicholls *et al*., 2005).

1.3.9 Reoffending

Untreated mental health problems may not only present difficulties in custody but may also impede successful rehabilitation. Substantial research has focused on demographic and criminogenic predictors of reoffending, however there is scant research concerning psychiatric determinates of reoffending (Grann *et al*., 2008).

Research to date has reported mixed findings with regard to reoffending amongst ex-prisoners with mental health problems. Some studies have reported higher rates of reconviction amongst those with a SMI compared to controls without a SMI (Silver *et al*., 1989), while others have found no significant difference in reconviction rates (Feder, 1991; Lovell *et al*., 2002). More recently, in a longitudinal study of nearly 80,000 US prisoners, Baillargeon *et al*. (2009) found that prisoners with serious mental illness were significantly more likely to have multiple incarcerations over the six year follow up period. As with many studies regarding reconviction, conflicting findings, variations in follow up periods and definitions of reoffending (for example self-reported reoffending and recorded reconviction) (Ministry of Justice, 2015b) make drawing conclusions from this small body of studies difficult. Nevertheless, there is consensus that untreated mental illness can significantly impede ex-prisoners community reintegration, which may contribute to reoffending (Baillargeon *et al*., 2010; Birmingham, 1999). Identification of prisoners with mental health needs in prisons allows for continuity of care arrangements to be put in place to support prisoners with the transition between custody and community and potentially ameliorate their risk of reoffending (Birmingham, 1999).

1.3.10 Mental health services in prisons

Historically, health care provision within prisons has been heavily criticised with prisoners' complex mental health needs and risk issues concerning self – harm and suicide thought to be inadequately met (Gunn *et al*., 1991; HMIP, 1996; Hughes, 2000; Sim, 1994). Recent years have, however, seen a policy imperative toward tacking health inequality, including a drive toward achieving equivalence of mental and physical health care in custody, with prisoners

being entitled to the same range and quality of services as available in the community (DoH & HMPS, 2001). In line with this, responsibility for planning and commissioning health services in prisons was transferred from the Prison Service to the NHS in 2006.

Prior to this, prisoners with mental health needs were treated by poorly trained health care officers, prison primary care services, general practitioners and health care units in prisons (OHRN, 2010; Reed, 2003). At the turn of the century, work had already begun to improve specialist mental health care for prisoners, with emphasis on a whole prison approach to meeting the needs of those with mental health problems (DoH & HMPS, 2001). *Changing the Outlook* saw the introduction of mental health in-reach teams (MHIRT) in prisons offering specialist wing-based services to meet the needs of and divert prisoners with severe and enduring mental illness (DoH & HMPS, 2001). MHIRTs aimed to mirror the Community Mental Health Team (CMHT) model, despite equivocal outcomes for the CMHT model in community settings (Burns, 2001; Simmonds *et al.*, 2001; Tyrer *et al*., 2003). Although little evaluation of prison MHIRTs has been conducted (OHRN, 2010), research that exists suggests they had a beneficial impact in terms of improving communication, waiting times, treatment and decreasing stigma and self-harm (Armitage *et al*., 2003; Brooker *et al*., 2005). However, MHIRTs were found to be treating a broad range of mental health problems including common mental health problems, with little face-to-face intervention and dealing with a high incidence of self-harm, with many prisoners with SMI still not being identified or accepted onto in-reach caseloads (Brooker *et al.,* 2005; Meiklejohn *et al*., 2004; OHRN, 2009; Steel *et al*., 2007). As such, significant numbers of prisoners with mental health needs continue to remain unidentified and untreated presenting both ethical and risk issues.

Ameliorating some responsibility for suicide prevention and self-harm from MHIRTs, in 2007 Assessment, Care in Custody and Teamwork (ACCT) procedures were introduced in English and Welsh prisons. ACCT is a multi-disciplinary approach to suicide prevention involving prison officers, other agencies and health care taking joint responsibility for caring for those at risk

to themselves. At the same time, in the community 2007 saw the large-scale announcement of the Improving Access to Psychological Therapies (IAPT) initiative for treating depression and anxiety in England (Department of Health, 2008). In 2009 the first IAPT *Offenders Positive Practice Guidance* was published (and updated in 2013) outlining the potential benefits and challenges of delivering IAPT services to offenders, including those in English prisons (NHS England, 2009; 2013). The guidance reflected recommendations from across reports, which whilst highlighting the considerable progress made in secondary care prison mental health, stipulated the need for the development of robust models of primary mental health service in prisons in order to meet the needs of those with common mental health problems (Appleby, 2010; Bradley, 2009; HMIP, 2007; OHRN, 2009; OHRN, 2010).

In Wales, the Mental Health Measure 2010 similarly legislated to provide expanded primary care services for common mental health problems including psychological therapies for those in the community and custody (Welsh Government, 2012a; Welsh Government 2012b). The later mental health needs assessment (Little, 2013) across the Welsh prison estate, however, identified patchy implementation with a lack of primary care mental health provision and interventions. Drawing on Little's (2013) findings, the *Policy Implementation Guidance (PIG): Mental Health Services for Prisoners* (Welsh Government, 2014) now specifies as minimum requirements for primary care: comprehensive assessment, short term interventions, provision of information and support to prisoners and other workers as well as co-ordination of onward referral where appropriate. Although largely consistent with the previous MHIRT remit, the PIG also outlines minimum requirements for secondary care in terms of: specialist assessment, care-coordination, provision of evidence based interventions for those with severe and enduring mental-health problems and responsibility for urgent assessment and transfer of prisoners who may require admission under the Mental Health Act. The PIG thus aims to address the diverse needs of those with low-level common mental health problems, as well as those with more severe and chronic presentations.

At both of the prisons sampled, mental health care pathways reflect a stepped care model including primary care mental health provision and MHIRT secondary provision aimed at treating a range of needs. Primary mental health teams typically treat lower level needs, have multiple points of access in line with national guidance (NICE, 2011a) and act as gatekeepers to the MHIRTs. Gauging the severity of mental health need is imperative to offering evidence-based interventions through the appropriate step of the pathway. Structured tools support an assessment of severity of need, with clinical guidance recommending specific tools for differing presentations (NICE, 2011a). Research has suggested that structured tools offer a more accurate assessment of severity of disorders than clinical judgment which has been found to be unreliable (Kendrick *et al*., 2005). Use of systematic screening tools in primary care mental health contexts has also been linked to improved clinical outcomes, for example reduced depressive symptoms (Bower *et al*., 2006). However, differing screening tools can produce divergent severity ratings and consequent referral rates as well as failing to account for broader bio-psychosocial factors which may impact severity of symptoms. As such applying rigid protocols based on screening scores alone is not recommended (Cameron *et al*., 2008). However, screening to support clinical differentiation between common and severe mental health problems has the potential to support existing care pathways structured around the Mental Health measure.

1.3.11 Mental health screening in prisons

Underlying the concept of screening is the assumption that early detection of a condition will improve clinical outcome. Martin *et al*. (2016) highlights that screening has the potential to be beneficial where there are high prevalence rates and low prior detection rates. However, when prevalence rates are low screening may be ineffective, if not harmful (Martin *et al*., 2016). The high prevalence rate of mental health problems in prisoners, poor prior detection and the multiple negative outcomes of undetected mental health problems in prisoners previously outlined highlights the conceptual relevance of screening with this population. Wilson & Junger's (1968) classical screening criteria cite that screening is appropriate when: the condition is an important health

problem, there is an accepted treatment, facilities for diagnosis and treatment are available, there are recognisable early symptoms, there is suitable testing that is acceptable to the population, the development of the condition is adequately understood, there is a policy on who to treat as patients, the cost of case finding is economically balanced and case-finding is an on-going process. Prevalence data, the significant impact of mental health problems in prisons, their associated cost burden and the availability of prison based mental health services suggest that prisoner mental health conforms to Wilson and Junger's (1968) criteria for appropriate screening.

Despite mental health screening for prisoners rationally fitting with most screening principles, there are also a number of conceptual issues in relation to mental health screening that need to be considered. Firstly, in screening there is an assumption that there are early or hidden symptoms that can be detected. With physical health screening such as for cancer there are clear symptomatic biomarkers that can be identified; however, with regard to mental health there is no established 'latent' phase and diagnosis is dependent on the presence of subjective symptoms. Secondly, screening requires a suitable test. Whilst physical health tests typically assess an independent objective marker of disease (e.g. blood sugar levels in diabetes), mental health diagnoses are dependent on the testing of the presence of the subjective symptoms that form arbitrary diagnostic categories. As such, although psychometric tests are well established in the assessment of mental health, their subjectivity brings into question their suitability. Finally, there is little evidence concerning whether mental health screening is economically and clinically effective or not (Richardson *et al*., 2015). Despite these conceptual issues, practically screening affords the opportunity to sift through large numbers of individuals with minimal resources to identify a subset of individuals who may have a mental health problem and offer further assessment and treatment. However, screening tools have a short shelf life (Benedek *et al*., 2009) and historically prison mental health screening tools have not been frequently updated.

1.3.12 The development of mental health screening in prisons

In the UK routine health screening of new prisoners has been carried out since the passing of the Prisons Act 1886. Traditionally this involved a medical officer and subsequently a doctor reviewing each prisoner received into custody with minimal standardisation in approach (Birmingham *et al.,* 1996; 1997). The utility and effectiveness of this practice was, however, brought into question, with research revealing that traditional reception screening processes resulted in substantial duplication, whilst still failing to identify three quarters of prisoners with severe mental health disorders (Birmingham *et al*., 1997; 2000). As such, screening processes were subsequently reviewed and a standardised approach implemented.

1.3.13 Current prison health screening procedures in England and Wales

Reception-screening procedures in English and Welsh prisons were reviewed and an updated process was piloted and rolled out across the estate at the turn of the century (Grubin *et al*., 1999; 2002; HM Prison Service, 2006; Ministry of Justice, 2011). The new approach involves an initial standardised triage screening for immediate physical health, mental health, substance misuse and risk needs within 24 hours of reception, followed by a general health assessment within a week of reception by an appropriately trained professional (Grubin *et al*., 2001). In practice however, there is substantial variation across establishments in the implementation of standardised processes (Lewis & Meek, 2013; OHRN, 2008). An evaluation of the reception screening process revealed that staff generally perceived the current standardised tool to be least effective for identifying risk of suicide, self-harm and mental health problems (OHRN, 2008). Specifically, reliance on historic indicators of mental health problems rather than a focus on here and now presentation was raised as a concern (OHRN, 2008). Such concerns are not surprising given the disparity between historic emphasis in the current screening process and the current policy emphasis on treating presenting problems with least invasive most effective intervention first (NICE, 2011a).

In Wales it has been noted that, while initial assessments are relatively embedded, more comprehensive secondary assessment of mental health

requires development (Little, 2013). Furthermore, the National Service Model for primary care mental health now stipulates the need for eligible practitioners to undertake mental health assessments[3] (Welsh Government, 2011). Consequently, existing mental health screening procedures and tools in prisons have failed to keep abreast of service developments. Despite this, the reception health screening process has been established as integral to the identification of prisoners with mental health problems entering custody (Birmingham *et al*., 1997; Grubin *et al*., 2001; OHRN, 2010; Watson *et al*., 2004). Furthermore, current policy advocates improving screening processes, including consistent unified use of structured assessment tools in screening procedures (Welsh Government, 2014).

1.3.14 Benefits of mental health screening in prisons

Prisoners are typically received into custody with a complex array of health needs (Bradley, 2009). Despite high levels of need, offenders frequently do not access health services in the community (Harty *et al*., 2003; DoH, 2002), yet make extensive use of health services in prison (Feron *et al*., 2005; Marshall *et al*., 2001). Marshall *et al*. (2001) reported that male British prisoners consulted the doctor three times more frequently and other health care professionals seventy seven times more frequently than their community comparators. As such health screening in prison offers an invaluable opportunity to identify and subsequently treat offenders unmet physical and mental health needs.

From an institutional perspective, screening at point of reception offers the first opportunity to detect those with mental health problems who may be at elevated risk of self-harm, suicide and indiscipline if left unidentified and untreated (Felson *et al*., 2012; Grey *et al*., 2003; Hawtton *et al*., 2014). Furthermore, screening allows for identification of those with less severe mental health problems, but who may have significant difficulty adjusting to prison and thus be at elevated risk of self-harm and suicide early in custody

[3] Eligible practitioners are defined as: a qualified/registered social worker, a level 1 or 2 RMN or learning disability nurse, a registered occupational therapist, a registered practitioner psychologist or a registered medical practitioner (Welsh Government, 2011).

(Dooley *et al*, 1990; Paton & Borrill, 2004; Slade & Edelman, 2014; Topp, 1979).

Screening is also vital to ensure continuity of care between community and custody. Screening allows for the identification of individuals who may already be receiving mental health services and treatment in the community and require on-going input in custody to maintain their wellbeing (OHRN, 2008). In the most severe cases, screening offers the opportunity to identify those who require diversion out of the criminal justice system into mental health facilities. Once in residence on prison landings prisoners with mental health problems are less likely to come to the attention of services and those who are the most unwell may be the least likely to seek out help (Meiklejohn *et al*., 2004). Indeed, whether symptoms were identified at reception has been identified as a significant predictor of subsequent access to mental health treatment in custody (Teplin, 1990).

One way in which attempts to improve screening can be made is through the use of structured tools (Watson *et al*., 2004). Use of structured tools increases sensitivity of screening processes by linking questions asked in practice to predicators of mental health problems determined by research. Structured tools also decrease discrepancy in assessment approach and accuracy of identification between and within professionals of differing levels of competence. Greehalgh *et al*. (2005) has proposed that as well as improving detection, use of patient reported outcome measures can impact on patient treatment by eliciting information which can lead to the development of shared goals, increase patient adherence and clinicians appropriate targeting and monitoring of interventions. As such the on-going use of structured screening tools has the potential to improve both detection and treatment outcomes for patients.

1.3.15 Problems with screening in prisons

Despite the potential benefits of screening for mental health problems on entry to prisons, screening cannot be assumed to be unanimously beneficial. There is no evidence to demonstrate under what conditions screening

improves outcomes compared to current practice (Martin *et al*., 2013). In the worst case, screening could be harmful in that it may take up considerable resource, which may be redirected away from treatment. Without effective triage screening could also result in inappropriate treatment of people who are not ill through false positives. Martin *et al*. (2016) argue screening is likely to be ineffective in populations where there are low prevalence and prior detection rates, as few new cases are identified but the false positive rate will be high. Conversely, in populations with high prevalence rates and low prior detection rates (such as prisons), screening is likely to be effective as there are high numbers of cases to detect and improved specificity can reduce false positives. However, ultimately screening is only likely to be effective in terms of improving outcomes if coupled with available and effective interventions for those identified.

Screening within prisons also presents difficulties in terms of psychometric validity. . Being admitted to prison in itself is a departure from normality which is likely to be associated with increased distress for most individuals (Andersen *et al*., 2002; Hochstetler *et al*., 2004). Prisoners often enter prison with high levels of anxiety surrounding their legal, familial, financial, accommodation and parenting situation, as well as emotions such as guilt, shame and anger associated with their offending and hopelessness associated with being imprisoned (Dhami *et al*., 2007; Pratt *et al*., 2015). Additionally, individuals entering prison often present with a multitude of interrelated problems. For example, overlapping behavioural presentations associated with withdrawal from substances and mental health problems can create confusion and elevate scores on screening tools (Steadman, 2005). Low levels of literacy and poor communication skills can also make assessment difficult (Anthony & McFadyen, 2005). Consequently, screening for underlying mental health problems is affected by substantial confounding variables and situational stressors. As such the validity of many of the approaches and tools utilised in community settings cannot be assumed in the prison context (Martin *et al*. 2013).

Many psychometric tools are not well suited to the prison environment. Prisons receive high numbers of individuals daily and many tasks in addition to health assessments must be completed at reception (Ministry of Justice, 2011). Assessment must thus be brief; many existing screening tools (for example the SCL-90 R and the Brief Symptom Inventory) are too lengthy for this high churn environment. Many of the tools used in the community also require administration by a trained mental health professional, of which there is a paucity in prisons, as well as requiring requisite reading skills which many people entering prison do not possess (Little, 2013; Teplin & Swartz, 1989). A further problem in terms of language is the use of terms within tools that have different nuances within offender populations, for example 'guilt' which can result in over-identification with certain items amongst prisoners (Hewitt *et al*., 2011). As such the validity of tools considered reliable in the community cannot be assumed with prison populations.

Furthermore, problems of simulation or malingering and dissimulation generally associated with psychometric tests may be particularly salient in prison populations (Anderson *et al*., 2002). Prisoners may be motivated to feign symptoms in order to acquire medication to self-medicate or use as currency within the prison (Bowen *et al*., 2009; OHRN, 2010). Individuals may also seek to simulate or dissimulate symptoms in order to influence their legal or financial situation for example parole, risk classification and compensation claims (Higginson, 2005; Rogers, 2008). However, there is substantial consistency in the level of distress reported by prisoners across studies, which suggests prisoners' self-report of symptoms on the whole are likely to be truthful (Anthony & Mc Fadyen, 2005).

The complexity of prisoners' presentations as well as problems with psychometric tools can result in high levels of false positives or over identification. High false positive rates place a resource burden on already stretched services and have been cited as a potential drawback of prison based screening tools (Brooker *et al*., 2009; Ford, 2007; Martin *et al*., 2013). However, non-treatment resulting from false negatives is also costly in the long term (Steadman *et al*., 2005). When there are multiple stages to the

assessment process it has been suggested that it may be cost effective to initially prioritise sensitivity, whilst specificity can be emphasised at a later stage (Evans *et al*., 2010; Martin *et al*., 2013; Richardson *et al*., 2015).

A further issue concerning false positives is the potential negative impact this may have on individual prisoner's perception of themselves as well as staff's perception of prisoners (Martins *et al*., 2016). The issue of stigma however also applies to true positive cases. Stigmatisation (of oneself or by others) arising from being identified with mental health needs may affect perceptions of ability to cope, which in turn could affect individuals behaviour or how staff interact with them. Despite these difficulties, there is substantial evidence to support the use of standardised assessment procedures incorporating validated assessment instruments in prison screening processes (Watson *et al*., 2004: Birmingham, *et al*., 2000; Grubin *et al*., 2002; Teplin & Swartz, 1989). The following section presents the rational for using psychometric tools in prison mental health screening and outlines the strengths and weaknesses of two possible screening tools the GHQ-12 and CORE-10.

1.3.16 Psychometric screening tools

Psychometric tools assume that there are stable underlying characteristics (e.g. anxiety, depression, distress) that exist in everyone to varying degrees and can be assessed by measuring a range of items related to the underlying characteristic (e.g. behaviours, thoughts). As such scores on particular characteristics are standardised so that comparison can be made against a normative population in order to assess abnormal variation from the mean. The rationale for using psychometrics in screening for probable mental health problems is that valid tools with identified cut off points that signify abnormally high symptoms or distress can be used to improve the accuracy of predicting or identifying those with mental health problems. Predictive accuracy is, however, affected by the base rate of conditions in the population (in this case the prevalence of mental health problems). Improvement in predictive accuracy is greatest when the base rate is close to 50%: when base rates are extreme, improvement in predictive accuracy of a test can be negligible. As such, psychometric tools need to be validated within the population in which

they will be used. The base rate of mental health problems is lower amongst community populations as compared to prison populations, therefore the predictive accuracy of tools validated in the community will not be the same with prison populations. It is therefore imperative to validate possible mental health screening tools for prisoners amongst prisoners. Reports on base rates of probable mental health problems in male prisoners as measured by global self-report measures of psychiatric morbidity like the GHQ-12 have cited base rates of between 33% (Hassan *et al*., 2011) and 59% (Boothby *et al*., 2010), suggesting the non-extreme base rate will allow for validation to substantially improve tools' predictive accuracy amongst prisoners.

A range of psychometric tools are currently used for assessing anxiety and depression in prisoners, although the degree of use is variable across prisons (Little, 2013). Despite such tools being used in prisons, particularly within primary care mental health services (Adamson *et al*., 2015; Cox *et al*., 2015) many (including the PHQ9 and GHQ7) have not been validated for use within forensic or prison populations (Fizpatrick *et al*., 2010). Some generic tools which have been tested in prison populations, for example the BDI, have been found to have questionable construct and discriminant validity within this population (Boothby *et al*., 1999; Richter *et al*., 1991). As such, the tools being utilised currently are not necessarily providing clinically reliable data as they do in community populations. Their predictive accuracy is not known, they do not assess distress globally, and the breadth of tools used prevents comparison across prisons.

1.3.17 General Health Questionnaire (GHQ)

One tool, which assesses psychiatric morbidity more generally, is widely used within community primary care settings and has been validated within prisoner samples is the GHQ. The GHQ has been identified as the most widely used generic measure to assess depression in offenders and has been found to have relatively good reliability, sensitivity and specificity and discriminatory power (Boothby *et al*., 2010; Harding & Zimmerman, 1989; Hewitt *et al.,* 2011; Smith & Boorland, 1999). This research suggests the GHQ may be a useful

and reliable tool for identifying common and severe mental health problems in prisoner populations.

The GHQ is not however without its limitations with this population. Andersen (2003) concluded that the GHQ's overall performance as a screening tool was poor and noted that some items seemed absurd for a prison population. The GHQ focuses on deviation from normal functioning. Arguably, entry to prison is a departure from normality in itself, therefore scores on the GHQ will be higher for prisoner and many of the items will be less meaningful as prisoners by definition will have experienced change in their day to day activities and potential functioning. Hassan *et al*. (2011) also notes that the GHQ is subject to retest effects when re-administered over a short time period, and therefore may not be reliable for measuring adjustment or intervention outcome. Furthermore, the GHQ-12 items primarily focus on low mood and functioning without assessing psychotic symptoms or risk. Despite these problems, the GHQ-12 has been included in the current study for comparative purposes as it is probably the most widely validated brief measure of non-specific psychiatric distress in prisoner populations (Boothby *et al*., 2010).

1.3.18 Core Outcomes In Routine Evaluation (CORE)

The battery of CORE measures was developed in response to increasing demand on psychotherapeutic services for outcome measurement alongside the dearth of clinically usable measures available. Although many outcome measures were available at the time of the CORE development, many were designed for research purposes and thus prioritised theoretical constructs and fidelity over clinical relevance and utility in term of length, readability, referential normative data, generalizability and cost (Barkham *et al*., 2010). The original 34-item CORE-OM was thus developed as a clinically relevant global assessment and outcome measure which could be followed by problem specific measures. It was purported to tap four key domains consisting of related sub domains. These were subjective wellbeing, problems (including the subdomains anxiety, depression, physical and trauma), functioning (including the subdomains of general, close relationships and social relationships) and risk (including to self and others). Subsequent factorial

analysis has failed to consistently support this factor structure (Bedford *et al*., 2010; Evans *et al*., 2002; Lyne *et al*., 2006). Nevertheless, Bedford *et al*. (2010) found that a two factor solution for the CORE-OM consisting of a psychological distress and risk factor was supported, although the first factor had a great deal of item redundancy.

Reflecting such findings Connell & Barkham (2007) developed the shortened CORE-10 derived from the CORE-OM, for use in busy clinical settings (Barkham *et al*., 2010). The CORE-10 is a pan-theoretical measure to assess global distress. The CORE-10 consists of two anxiety items, two depression items, one physical item and one trauma item drawn from the psychological distress CORE-OM factor, as well as three functioning items and a risk item. The CORE-10 is thought to split items into high intensity and low intensity constructs, as well as measuring risk. Despite the authors hypothesising these factor structures for the CORE-10, they do not appear to have been confirmed using confirmatory factor analysis. The CORE-10 has however been found to be feasible and acceptable in general practice and primary care settings (Connell & Barkham, 2007; Barkham *et al*., 2013), but has not been validated in a prison context.

Although the CORE-10 has not been researched in the prison context, Perry *et al*. (2013) found the CORE-OM (the measure from which the CORE-10's items are drawn) to be acceptable and feasible in secure settings. McCloskey (2001) quantitatively explored the CORE-OM in a prison population and found that the measure had good internal consistency with mean distress score in the prison sample falling between the validation clinical and non-clinical sample. Additionally, the Engager project, an intervention for prisoner with common mental health problems has utilised the CORE-OM and CORE-10 and suggests these tools work best of all they have tried with this population (M. Maguire, personal communication, 1st April 2016). This information suggests that the CORE-10 may also be acceptable and valid as a screening tool in a prison population, although its psychometric proprieties and acceptability as a screening instrument need to be established.

A possible weakness of the CORE-10 is that although it differentiates between high and low intensity items, research with the CORE-OM has found minimal difference in scores between clients receiving primary and secondary care services, though those in secondary care had significantly higher risk scores (Barkham *et al*., 2005). This suggest that the CORE-10 may fail to differentiate between common and severe mental health problems, or that CORE scores do not increase incrementally with severity as assumed by service pathways. A related issue is that the CORE-10 focuses exclusively on current presentation and therefore prioritises the acuteness of symptoms (which tend to be rated higher amongst those in primary care), without considering the chronicity of problems - frequency characteristic of severe and enduring mental health problems. Nevertheless, its emphasis on psychological distress broadly, its brevity, its consideration of risk and its established robustness against other well established measures make it a potentially useful tool for screening prisoners for both common and severe mental health problems. For comparative purposes the following section presents a systematic review of existing validated tools designed globally to screen for mental health problems in offender populations.

1.4 Systematic Review

1.4.1 Aims

The present study aims to explore the psychometric properties of a self-report screening tool, the CORE-10 for use in screening for mental health problems in prisoners. As such, a systematic review was conducted to explore the quality of existing research concerning tools for screening mental health of prisoners.

Following the initial literature review, a systematic review was conducted to determine 'what tools have been developed and validated to screen for mental health difficulties in offender populations, what are their psychometric properties and clinical utility?'.

1.4.2 Search methodology

On 27th of October 2015 a review of research evidence from 1988[4] to 2015 was conducted. The following databases were searched: PsychINFO, Embase, Medline, PsychARTICLES, Web of Science, Scopus, ASSIA, PubMed, CINAHL and Emerald Insight. Grey literature was searched using Proquest dissertations, theses database and OPENGREY in addition to Google and Google Scholar. The following statutory sources were also searched: Ministry of Justice, Department of Health and Welsh Government.

The Boolean operator 'AND' was utilised to combine key search terms relating to 1) prisoner populations, 2) mental health, 3) assessment and 4) test development. The Boolean operator 'OR' was applied to search terms with a similar meaning within each category. Key search terms were:

1. **Prisoner population**: Topic = (detained OR detention OR prison* OR custody OR jail OR incarcerated OR forensic OR correction*)
   AND

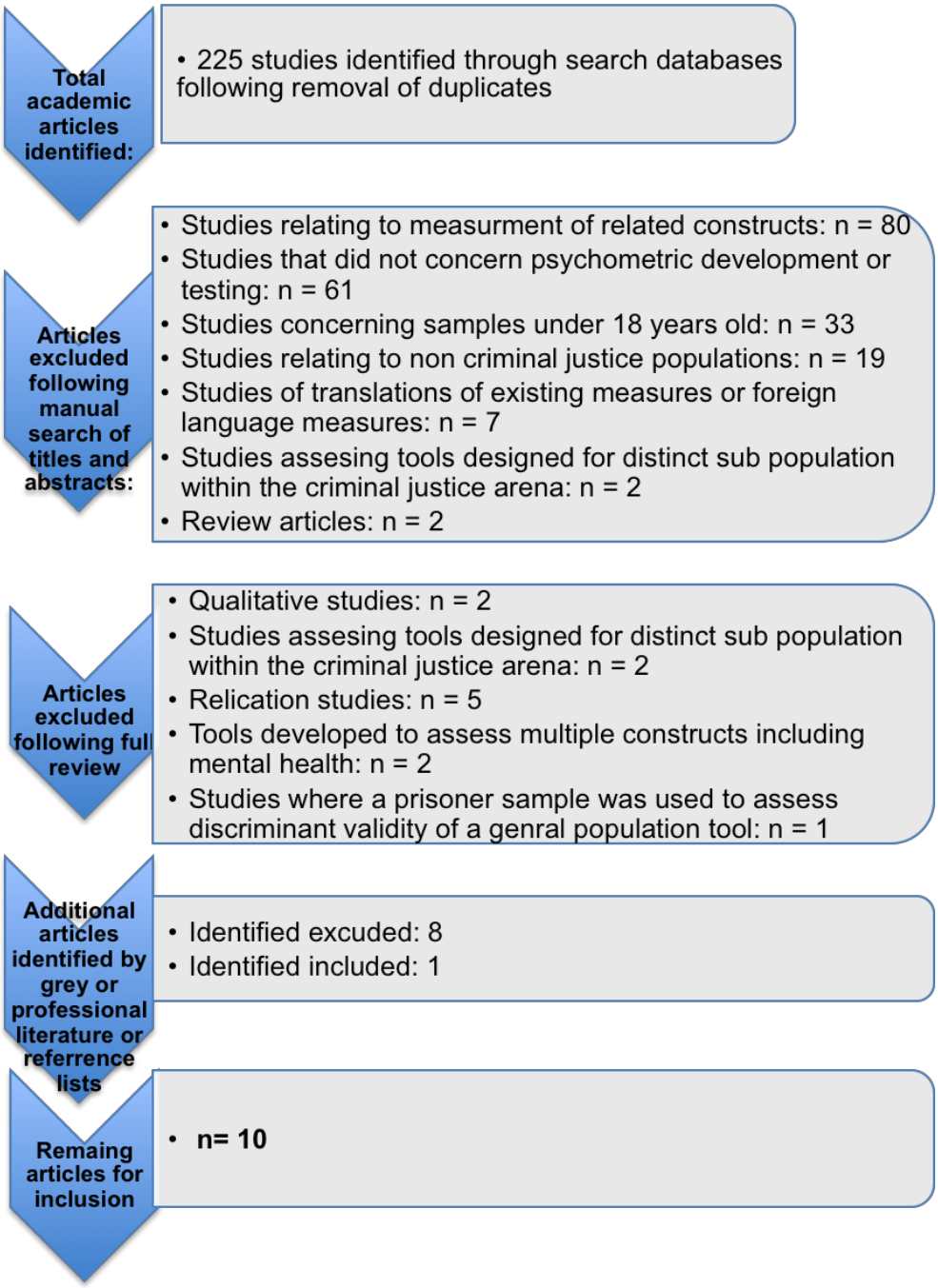
[4] 1988 was selected as the start date of the search as this is when literature around mental health screening in prisons began to emerge.

2. **Mental health**: Title = (mental health OR mental illness OR mental disorder OR psychiatric disorder OR wellbeing OR psychopathology) AND
3. **Assessment**: Title = (psychometric OR measure OR questionnaire OR tool OR instrument OR scale OR inventory OR assessment OR test) AND
4. **Test development:** Topic = (develop* OR valid* OR reliability OR ROC OR receiver operating curve OR sensitivity OR specificity OR item analysis OR factor analysis OR psychometric OR sensitivity OR specificity OR positive predicative value OR negative predicative value)

All titles and abstracts identified during these searches were reviewed (N= 222 following removal of duplicates). Full papers were reviewed when it was unclear if the study met the inclusion criteria from the abstract alone. The reference lists of all articles that met the inclusion criteria, key review papers, book chapters and meta-analyses were examined for relevant studies.

The search was repeated on 30th April 2016 to identify any further studies published between October 2015 and April 2016. A diagram of the systematic review search process is shown in figure 1.

Figure 1 Systematic review search process



1.4.3 Selection criteria

1.4.4 Inclusion criteria

- Studies concerning the development (including acceptability and feasibility studies) and or validation of (including studies of factor structure and diagnostic accuracy) screening tools for assessing prisoners mental health
- Original articles
- Peer reviewed papers
- Quantitative or mixed method studies
- Studies published between 1988-2015

1.4.5 Exclusion criteria

- Tools developed for people under 18 years old
- Studies published before 1988
- Studies of tools which have not been developed for or validated within a criminal justice context
- Studies of tools relating to assessment of related constructs (e.g. risk)
- Studies of tools designed for comprehensive assessment of mental health rather than screening tools[5]
- Studies which do not evaluate an assessment tool
- Studies assessing tools designed for distinct sub populations of the criminal justice populations (e.g. females only, indigenous populations)
- Studies validating translations of existing measures into different languages
- Replication studies aimed at validating existing screens (although these will be used for reference purposes)
- Studies not available in English
- Qualitative studies

[5] The Jail Assessment Tool (JSAT) (Nicholls *et al*., 2005) which has been described as a screening was excluded on the basis of its length, which is more a kin to that of a comprehensive assessment.

1.4.6 Publication status

Research concerning screening for mental health in prisoners published in peer reviewed journals, book chapters, conference papers as well as dissertations and theses were included in order to reduce the potential impact of publication bias. By prioritising studies with positive results publication bias may result in the overestimation of the psychometric properties of screening tools (Song *et al*., 2009).

1.4.7 Quality framework

Ten studies were included in the systematic review, a diagram of the selection process for these studies is provided in Figure 1. An outline of the clinical utility of each tool identified from the selected studies is shown in table 1. Table 2 summaries the articles included in the systematic review and the psychometric properties of each tool.

The selected studies were evaluated against a quality framework developed by the Critical Skills Appraisal Programme (CASP, 2013). The framework is designed specifically for evaluating diagnostic test studies (see appendix 1). A numerical scoring system was devised in addition to the existing descriptive quality framework. Studies were rated on each of the criterion[6] as good equating to a score of 2, mixed equating to a score of 1 or poor (e.g. not reported) equating to a score of 0. Scores across each criterion were then summed and divided by the number of criterion scored to create a mean quality score ranging between 0 and 2, with higher scores indicating better quality studies. Table 3 shows the quality assessment and score for each study included in the review.

1.4.8 Results

[6] Criterion 10 and 11 were combined thus forming 11 criteria in total.

Table 1 Clinical features of screening tools identified

Screening tool	Brief description of tool	Type of tool	Administration time	Administration training required	Costs
Referral Decision Scale (RDS) Teplin & Swartz (1989) (US)	The 14-item Referral Decision Scale (RDS) was derived from the Diagnostic Interview Schedule. It consists of sub-scales for depression (cut point 2), bipolar disorder (cut point 3), and schizophrenia (cut point 1) showing average sensitivity of 0.79 to 0.88 and average specificity of 0.99 for predicting full Diagnostic Interview Schedule diagnoses in offenders.	Self-report (Interviewer administered)	5-10 minutes	Yes	None
Core Outcomes in Routine Evaluation Outcome Measure (CORE-OM) McCloskey (2001) (UK)	The CORE-OM is a 34 item self-report measure comprising four domains: subjective wellbeing, symptoms, functioning and risk. Each item is scored on a scale from 0-4 with corresponding responses ranging from 'not at all' to 'all of the time'. Higher scores indicate more of the given construct. Clinical scores can be calculated by multiplying the mean score by 10, resulting in a score ranging from 0-40. Clinical scores can be divided into healthy (0-5), low level (6-9), mild (10-14), moderate (15-19), moderate- severe (20-24) and severe (over 25) categories.	Self-report	6 minutes	No	None

Grubin reception health screen (otherwise known as English Mental Health Screen) Grubin *et al*. (2002) (UK)	A tool administered at reception to prison The Grubin comprises of 15 basic screening questions relating to physical, mental health and withdrawal from substances. Four questions concerning a history of psychiatric treatment, a history of deliberate self-harm, being prescribed antidepressant or antipsychotic medication and murder/manslaughter index offence are used to screen for severe mental illness. A yes answer to any of these questions constitutes a positive screen and the tool includes protocols for action.	Self-report (interviewer administered)	5-10 minutes	Yes	None
General Health Questionnaire 28 (GHQ-28) Andersen *et al*. (2002) (Denmark)	The GHQ (Goldberg, 1978) is one of the most widely used questionnaires to screen for psychiatric morbidity, particularly in primary care. Individuals rate the extent to which their current state differs to their usual state over the last two weeks on 32 items. Different scoring methods are available all of which affect the total score.	Self-report	5 minutes	No	£1.10 + VAT per questionnaire

Prison Screening Questionnaire **(PriSnQuest)** Shaw *et al*. (2003) (UK)	A seven item measure which screens for serious mental illness. Items are a subset of items drawn from the General Health Questionnaire and Psychosis Screening Questionnaire (PSQ). Individuals are considered cases if they score positively on two or more of the GHQ items, either of the PSQ items or have previously received psychiatric treatment.	Self-report	< 5 minutes	No	None
Screening tool for identifying prisoners with severe mental illness Birmingham & Mullee (2005) (UK)	A seven item observational tool completed by prison officers for identifying prisoners with severe mental illness. Items focus on observations of behaviour and changes in behaviour.	Observational	< 5 minutes	No	None

Prisoner Mental Health Inventory (PMHI) Anthony & McFadyen (2005) (UK)	The PMHI is based on a subset of nine items from the Cardinal Needs Schedule. This self-report measure consists of items relating to mood swings, hearing voices, problems with thinking, depression and anxiety, alcohol, drugs, self-harm, aggression and sexual problems related to mental health. Respondents rate whether they consider each item to be a problem to them or not.	Self-report	<5 minutes	No	None
Brief Jail Mental Health Screen (BJMHS) Steadman *et al*. (2005) (US)	The BJMHS consists of eight items organised into two sections. The first section asks about the occurrence of mental health symptoms in the last six months and now. The second section asks if the individual has ever been hospitalised for emotional or mental health problems and if they are currently taking psychotropic medication. Answers are scored dichotomously yes/no. Individuals are considered cases if they endorse two or more current symptoms in section one and/or endorse either item in section two.	Self-report (Interviewer administered)	2.5 minutes	Yes	None

Correctional Mental Health Screen for men (CMHS M) Ford *et al.* (2007) (US)	A 12 item screen relating to a broad spectrum of DSM-IV axis I disorders including mood disorders, psychotic disorders, anxiety disorders, anxiety disorders, somatoform disorders, eating disorders and major features of Axis II personality disorders. Response categories are dichotomous yes/no. The validation study recommends a cut point of 6 or more items, equating to sensitivity of 74% and specificity of 75%.	Self-report (interviewer administered)	Approximately 5 – 10 minutes	Yes	None
K6 Louden *et al.* (2013) (US)	The K6 (Kessler *et al*., 2002) is a six item self-report screening tool designed to discern mental disorder from general distress. The six symptom items are rated in relation to the last 30 days on a scale from 0-4 (none of the time - all of the time). Possible scores range from 0 to 24. A score of 6 or more is recommended as indicative of a respondent being likely to have a mental disorder in offender populations.	Self-report	< 5 minutes	No	None

Table 2 Summary of systematic review studies

Screening tool, authors country and quality score	Design	Sample N and demographics	Exclusions	Comparator measures	Key findings (Sensitivity, specificity, reliability)	Limitations
Referral Decision Scale (RDS) Teplin & Swartz (1989) (US)	Cross sectional development and subsequent validation study	Development sample (S1) N = 728 Validation sample (S2) N = 1,149 All males Mean age 26 (S1) and 27 (S2) years 80% (S1), 51% (S2) black, 12% (S1), 45%	Incarcerated for safekeeping	Mental Health Diagnostic Interview Schedule (NIMH-DIS) (linked to DSMIII diagnostic categories)	• 14 item RDS for schizophrenia, manic and major depressive illness developed via discriminant analysis from the NIMH-DIS • Sensitivity 0.79 • Specificity 0.98 • Positive predictive value 0.79 • Negative predictive value 0.01 • Cut points: 2 for schizophrenia, 3 for manic depression and 2 for major depression	• Statistically derived from a diagnostic instrument rather than independently validated as a tool in its own right • Linked to specific diagnoses • Focus on lifetime disorder rather than current symptoms • Retest reliability not assessed • Veysey *et al.* (1998) questioned face validity of the RDS

		(S2) white, S1 remanded and short sentenced jail detainees (< 1 year) S2 convicted felons				
Core Outcomes in Routine Evaluation Outcome Measure (CORE-OM) McCloskey (2001) (UK)	Cross sectional with two week follow up	N = 53 All male Mean age 34 years 92% white, 4% black, 4% mixed Long term sentenced	Prisoners being transferred prior to data collection	CORE OM normative clinical and non-clinical data	• Mean score on the CORE-OM in the forensic population were slightly lower than those found in a normative clinical population but higher than those found in a non-clinical population • Internal consistency α = 0.95, α range 0.62-0.92 • Test-retest coefficient = .74 (range .62-.72) • The wellbeing, problem and function scores correlated highly with each other. The risk score	• Convergent validity against other measures was not tested • The measure was administered to the admissions wing only, and did not include the more stable wider population

		prisoners in a therapeutic community in HMP Grendon			correlation was much lower. • Acceptability of the measure was good with a refusal rate of less than 2%	
Grubin reception health screen (otherwise known as English Mental Health Screen) Grubin *et al*. (2002) (UK)	Measure compared against Schedule for Affective Disorders and Schizophrenia Lifetime Version (SADS L), audit of clinical use with 3 and 6 month follow up	N = 1360 59% adult male, 22% young offender a male and 19% female New admissions to prisons	Not specified	SADS-L administered to N = 150	• 28% of the whole sample screened positive from mental health problems • Amongst adult males 26% screened positive from mental health problems compared with 12% of male young offenders • 6% had open F2052SHs forms (the procedure for managing self-harm in the prison service at the time) • 3% reported currently feeling suicidal • Protocols were being followed appropriately in 99% of cases • When compared to the SADS-L one prisoner with severe mental illness had been missed (false	• The comparator measure was not administered to all participants • Primary level mental health need was not considered

					negative rate 1%) • The false positive rate for serious mental health problems was 13% • Sensitivity 97% • Specificity 84% • Positive predictive value 60% • Negative predictive vale 99% • Efficiency 86%	
General Health Questionnaire 28 (GHQ-28) Andersen *et al.* (2002) (Denmark)	Cross sectional with between groups and gold standard measure comparison using the Present State Examination version 10 (PSE-10)	N = 184 Age range 18-60 Men and women Remanded prisoners	Non Danish speaking Sentenced prisoners	PSE-10 (Based on ICD-10 criteria)	• Mean GHQ-28 scores were 9.96 • There was a high correlation between all subscales and total score • There was no correlation between GHQ score and IQ • The GHQ had low specificity at the normal cut-off of 4/5 • A cut off of 9/10 or 10/11 provides the best trade off in sensitivity (0.65) and specificity (0.69)	• Re test reliability was not explored • Details of participant demographics are not reported

Prison Screening Questionnaire (PriSnQuest) Shaw *et al*. (2003) (UK)	Cross sectional with comparison of measures against gold standard using the Schedule for Clinical Assessment in Neuropsychiatry (SCAN)	N = 2920 Male (86%) and female (14%) Mean age = 31 years Magistrates court attendants	Not reported	SCAN	• Logistic regression was used to create a 7 item measure including 4 GHQ and 2 Psychosis Screening Questionnaire (PSQ) items • Cases score 2 or more on the GHQ items, score on either of the PSQ items or have a history of psychiatric treatment • 40% of the sample were classed as cases across the GHQ and PSQ • Sensitivity 89% (combined measure) • Specificity 61% (combined measure) • The GHQ is good at distinguishing depression but not schizophrenia • Recommend a cut of 11 or more on the GHQ • Recommend a cut off of 1 on the PSQ	• Demographic information not reported • The final instrument was created though statistical analysis and has not yet been validated in its own right

Screening tool for identifying prisoners with severe mental illness Birmingham & Mullee (2005) (UK)	Within subjects measure descriptively compared with gold standard (SADS-L) and between subjects control group	N=100 Mean age 32 years (comparison) 33 years (control) Male prisoners at a local prison	Not specified	SADS-L	• Observational screening tool consisting of 6 items created • 19 prisoners in the 'case group' meet diagnostic criteria for severe mental illness compared with none in the control group	• Sensitivity, specificity and predictive values not explored • Tool developed based on qualitative information from a small number of prison officers • No blinding conducted
Prisoner Mental Health Inventory (PMHI) Anthony & McFadyen (2005) (UK)	Cross sectional self-report	N = 495 Male prisoners across five English prisons	Not specified	None	• Three quarters of prisoners were identified as having at least one symptom on the PMHI • The tool was assessed by staff to have face validity • Internal consistency α .83 • A two-factor structure was identified relating to mental health and substance misuse • PMHI identified higher level of need than the Camberwell	• Demographic information not reported • Convergent validity against another tool not carried out • Sensitivity and specificity not explored • Re test reliability not analysed

					Assessment of Need Short Version (CANSAS) which was previously trailed	
Brief Jail Mental Health Screen (BJMHS) Steadman *et al*. (2005) (US)	Cross sectional with comparison of measures against gold standard using the Structured Clinical Interview for DSM-IV (SCID)	N = 10,330 Inmates in four US jails Male and female 70% pre-trial detainees 58% African American Mean age 32 year	Not specified	SCID	• Twice as many women than men were classified as needing a mental health referral • 73.5% of males were correctly classified by the SCID • There was a false negative rate of 14.6% in males • Among women 63.6% were correctly classified, there was a false negative rate of 34.7% • For men sensitivity was .66 • For men specificity was .74	• Re test reliability was not explored • The SCID was only administer to 3% of the total sample • Proportion of males and females are not reported • Over sampled those with a positive screen on the diagnostic interview sample
Correctional Mental Health Screen for men (CMHS M)	Cross sectional with comparison of measures	N = 2,196 (1,526 men, 670 women)	Those on restricted units, with a high	SCID for axis I and II disorders Clinician	• 56% of men and 68% of women had a current psychiatric disorder • Internal consistency for the CMHS M was α .78	• The tool was administered as an embedded part of other tools therefore has not been validated on a stand-

Ford *et al.* (2007) (US)	against gold standard using the SCID	Mean age 32 years, age range 18-78 years 41% white, 38% black, 20% Hispanic New admissions to five US jails	bond security risk, admitted to medical or psychiatric units, did not speak English, were in court or were under 18 years old were excluded	Administered PTSD Scale	• Sensitivity .75 • Specificity .70 • The AUC for the CMHS M exceeded that of the RDS and BJMHS • CMHS M AUC 0.73-0.78 (95% CI = .72-.86) • CMHS M accuracy 75-80% • CMHS M false positive rate 22-29% < than the BJMHS • CMHS M false negative rate 18%-26% > than the BJMHS • The CMHS M showed good convergent, discriminant and criterion validity.	alone basis • The findings cannot be generalised beyond newly incarcerated detainees • Inmates with known mental health problems were excluded
K6 Louden *et al.* (2013)	Cross sectional with comparison of measures against gold	N = 4,670 probationers 72% male, 28% female	Those who could not speak English were	SCID	• K6 sensitivity at a cut point of six for men was 0.75 with 1-specificity at 0.36 • The K6 positive predictive value at a cut point of 6 was 0.26 and	• High attrition rate at follow up (66%) • Long delay between screening and diagnostic interview

(US)	standard using the SCID	Mean age 31 years Evenly distributed across Caucasian, African American and Hispanic ethnic groups	excluded		negative predictive value was 0.94 • The K6 demonstrated equivalent sensitivity to the BJMHS • Current legal problems did not predict mental disorder • BJMHS was found to have acceptable sensitivity and specificity in men and women contrary to previous research	

Table 3 Assessment of quality of screening tools identified via systematic review using the CASP (2013) framework

Mean quality score, author	1. Was there a clear question for the study to address?	2. Was there a comparison with an appropriate reference standard?	3. Did all patients get the diagnostic test and reference standard?	4. Could the results of the test have been influenced by the results of the reference standard?	5. Is the disease status of the tested population clearly described?	6. Were the methods for performing the test described in sufficient detail?	7. What are the results?	8. How sure are we about the results? consequences and cost of alternatives performed?	9 &10. Can the results/test be applied to your patients/ the population of interest?	11. Were all outcomes important to the individual or population considered?	12. What would be the impact of using this test on your patients/ population?

Teplin & Swartz (1989)	To develop a severely truncated version of the Mental Health Diagnostic Interview Schedule (NIMH-DIS) suitable for screening purposes	NIMH-DIS	Yes, all prisoners completed the NIMH-DIS. The RDS was derived statistically and not administer separately.	Yes, the RDS was derived from results of the NIMH-DIS therefore the results would be influenced by the reference standard.	1.4% were diagnosed with schizophrenia, 1.5% with manic depression and 5% with major depression.	Participants were administered the NIMH-DIS at intake.	Sensitivity 0.79, specificity 0.98, positive predictive value 0.79, negative predictive value 0.01 against the NIMH-DIS.	No confidence intervals provided	Sample is comparative in terms of age and context, but over represents black ethnicity in comparison with the Welsh prison population.	Can't tell	The RDS could be used to screen prisoners in Wales however it is related to out of date diagnostic criteria and has not been validated in the UK.
1.6	2	2	2	0	2	2	2	1	2	n/a	1

McCloskey (2001)	Can the test characteristics of the CORE-OM be reproduced in a therapeutic forensic setting?	Normative CORE-OM data	n/a	n/a	Mean scores for wellbeing, problems functioning, risk and all items are presented with comparison against normative clinical and non-clinical samples.	All men admitted during a 4 month period were invited to complete the measure within two weeks of arrival and again two weeks later.	Mean score in the forensic population was between that found in the normative clinical and non-clinical samples. Internal consistency α = 0.95, Test-retest coefficient = .74. All scales other than risk were highly correlated Specificity and sensitivity were not explored.	Confidence intervals are not presented	The sample is largely comparable to that of the South Wales prison population although average sentence length is longer.	Yes	The CORE-OM could be usefully applied to prisoners in South Wales as it requires minimal training although for screening purposes the CORE-OM may be too long to be practical.
1.7	2	1	n/a	n/a	2	2	1	1	2	2	2

Grubin *et al.* (2002)	To develop and pilot a new physical and mental health screening procedure for prisons	SADS-L	SADS-L administered to N = 150 of the total sample of N = 1306	No – those administering the SADS-L were blind to the outcome of the screen	28% of the whole sample screened positive for mental health problems. 6% had open F2052SHs forms (self-harm). 3% reported currently feeling suicidal.	1306 reception case files were audited. A random sample of 15 prisoners from 10 different prisons took part in an interview 3 months after admission including administration of the SADS-L.	Sensitivity 97%. Specificity 84%. Positive predictive value 60%. Negative predictive vale 99%. Efficiency 86%.	Confidence intervals are not presented	The sample is comparable with that of the South Wales prison population	Yes	This brief screen is in use in the thesis sample population and helps to ensure accurate identification of those with severe mental health problems. The tool has been found to be both acceptable and feasible in British prisons.
1.8	2	2	1	2	2	2	2	1	2	2	2

Andersen *et al.* (2002)	To validate the GHQ-28 as a screening instrument in a random sample of prisoners	PSE-10	Yes	The GHQ-28 was filled out immediately after the diagnostic interview which may have had an impact on GHQ responses.	Disease status of the sample presented in a related paper.	Prisoners were interviewed using the PSE-10 within six days of imprisonment and diagnoses were agreed by two psychiatrists.	Mean GHQ-28 scores were 9.96. The GHQ had low specificity at the normal cut-off of 4/5. A cut off of 9/10 or 10/11 provides the best trade off in sensitivity (0.65) and specificity (0.69).	Confidence intervals were not presented	Insufficient demographic data is presented to make an accurate comparison to the prisoner population of South Wales.	Yes	The GHQ-28 could be applied to prisoners in South Wales as it requires minimal training although for screening purposes it may be too long to be practical, and has relatively poor sensitivity and specificity.
1.5	2	2	2	1	1	2	2	1	1	1	1

Shaw *et al.* (2003)	To develop a screening questionnaire for serious mental illness in magistrate's courts	SCAN	No, 1,306 of 2,920 were screened with the SCAN	Participants who completed the SCAN were a random sample of those identified as cases on the GHQ and PSQ. No blinding was put in place.	8.7% of the sample had a ICD 10 diagnosis	Participants were screened at magistrate's courts using the GHQ and PSQ. Second phase interviews using the SCAN was carried out although it is not specified when.	Logistic regression was used to create a 7 item tool. Sensitivity 89%. Specificity 61%. Cases scored 2 or more on the GHQ items, had a score on either of the PSQ items or have a history of psychiatric treatment. 40% of the sample were classed as cases.	Confidence intervals are not reported. Mean scores on the GHQ and PSQ are not reported	Insufficient demographic data is presented to make an accurate comparison to the prisoner population of South Wales	Can't tell	The tool has the potential to be used as a brief screen in prisons in South Wales although it is not clear what impact it would have as it has not be validated as a standalone tool.
1.4	2	2	1	1	2	1	2	1	1	1	1

Birmingham & Mullee (2005)	To develop and evaluate a screening tool based on the observational skills of prison officers to identify prisoner with severe mental illness	SADS-L	All participants underwent the observational screen and diagnostic interview.	No blinding was carried out, prison officer pre exiting knowledge could have influenced selection. The diagnostic interviewer was not blind to case status.	19% of the sample screened positively for severe mental disorder.	Prisoners identified by prison officers as meeting one of the criteria in the observational screen were approached for diagnostic interview. A random sample of those who screened negatively were also diagnostically interviewed.	Observational screening tool consisting of 6 items created. 19 prisoners in the 'case group' meet diagnostic criteria for severe mental illness compared with none in the control group.	The results may be confounded by the lack of blinding	The sample appears comparable to that of prisoners in South Wales.	Yes	The observational tool is short, tested on a similar population and uses the skills of prison officers therefore could have a positive and practical impact on identifying prisoners with severe mental illness in Wales.
1.6	2	2	2	1	2	2	1	1	1	2	2

Anthony & McFadyen (2005)	To develop a health needs assessment scale for prisons based on the Camberwell Assessment of Need tool	No	n/a	n/a	PMHI identified prevalence rates of: 47% depressed or anxious, 37% strong mood swings, 29% self-harm, 27% other problems with thinking, 19% hearing voices, 12% sexual problems related to mental health, 42% substance misuse.	The survey was administered as a self-report. How and when it was distributed is not however reported.	Three quarters of prisoners were identified as having at least one symptom on the PMHI. The tool was assessed by staff to have face validity. Internal consistency α .83. A two factor structure was identified relating to mental health and substance misuse. Sensitivity and specificity were not explored.	Confidence intervals are not reported.	The sample appears comparable to that of prisoners in South Wales although limited demographic data is presented to make appropriate comparisons.	Yes	The tool is short and developed with a similar population therefore could have a positive and practical impact on identifying prisoners mental health needs, although has not been widely validated.
1.3	2	0	n/a	n/a	2	1	1	1	1	2	2

Steadman *et al.* (2005)	Aimed to validate a revision of the Referral Decision Scale (RDS) – the Brief Jail Mental Health Screen (BJMHS) and assess its practicality, optimal scoring, sensitivity and specificity.	SCID	No, 357 participant of 10,330 were administered the SCID	Interviewers were blind to the outcome of screening.	11.3% of the overall sample were screened as needing further mental health assessment. Of the men clinically interviewed, 20% met criteria for a diagnosis of serious mental illness.	The BJMHS was administered by custodial staff at reception to jails. Approximately 90 participants from each jail were randomly selected for diagnostic interview conducted by a trained researcher.	73.5% of males were correctly classified by the SCID. There was a false negative rate of 14.6% in males. For men sensitivity was .66. For men specificity was .74.	Confidence intervals are not reported.	Sample is comparative in terms of age and context, but over represents black ethnicity in comparison the Welsh prison population.	Yes	The tool is short, tested on a similar population and requires little training therefore could have a positive and practical impact on identifying prisoners with severe mental illness.
1.8	2	2	1	2	2	2	2	1	2	2	2

Ford *et al.* (2007)	To develop a brief screening tool suitable for selecting individuals with clinically significant mental health problems for further assessment.	DSM-IV diagnostic categories SCID at follow up (20% of sample)	Yes	No, interviewers were blind to the outcome of screening.	56% of men and 68% of women had a current psychiatric disorder.	Self-report tools and structured interview administered within the first three days of admission by research assessors. The order of subscales was randomly varied. Follow up interviews were conducted 5 days later.	Internal consistency α .78. Sensitivity .75. Specificity .70. AUC exceeded RDS and BJMHS CMHS M AUC 0.73-0.78 (95% CI = .72-.86). CMHS M accuracy 75-80%. CMHS M false positive rate 22-29% < than the BJMHS. CMHS M false negative rate 18%-26% > than the BJMHS.	Confidence intervals reported.	The sample is comparative in terms of age and context, but over represents ethnicity minorities in comparison the Welsh prison population.	Yes	The tool is short, tested on a similar population and requires little training therefore could have a positive and practical impact on identifying prisoners with severe mental illness.
2.0	2	2	2	2	2	2	2	2	2	2	2

Louden *et al.* (2013)	To test the utility of the K6 and BJMHS for identifying probationers with lifetime DSM-IV Axis I mental disorders.	SCID at follow up	No, a sub sample of 149 participants were administered the structured clinical interview.	Can't tell if blinding was in place.	74% of probationers overall had a lifetime Axis I disorder. 65% of male and 90% of female probationers had a lifetime history of Axis I disorder.	Screening tools administered as self-reports as part of routine intake procedures at a probation agency. Follow up diagnostic interview on average 87 days post screening.	K6 sensitivity at a cut point of six for men was 0.75 with 1-specificity at 0.36. The K6 positive predictive value at a cut point of 6 was 0.26 and negative predictive value was 0.94. The K6 demonstrated equivalent sensitivity to the BJMHS.	Confidence intervals reported	Sample is comparative in terms of age but related to a community sample and over represents ethnicity minorities in comparison the Welsh prison population.	Yes	The tool is short, tested on a similar population and requires little training therefore could have a positive and practical impact on identifying prisoners with mental illness. However it does not appear to have been validated in UK prisons.
1.8	2	2	1	1	2	2	2	2	2	2	2

1.4.9 Synthesis of systematic review studies

Ten studies were included in the systematic review. The majority of studies originated from the UK (50%) or US (40%), with one study originating from Denmark. Four of the UK based studies and the Danish study were based in prisons (50%). Three of the US studies were based in jails (30%). Shaw *et al*. (2003) UK study was based in magistrates' courts and Louden *et al*. (2006) US study was set in a probation context. Table 4 shows the proportion of studies from each country and setting.

Table 4 Systematic review study origin and setting

Characteristic	n (%)	Study reference
Country of origin		
UK	5 (50%)	McCloskey (2001), Grubin *et al*. (2002), Shaw *et al*. (2003)., Birmingham & Mulle (2005), Anthony & McFayden (2005)
US	4 (40%)	Teplin & Swartz (1989), Steadman *et al*. (2005), Ford *et al*. (2007), Louden *et al*. (2013)
Denmark	1 (10%)	Andersen *et al*. (2002)
Study setting		
Jail	3 (30%)	Teplin & Swartz (1989), Steadman *et al*. (2005), Ford *et al*. (2007)
Prison	5 (50%)	McCloskey (2002), Grubin *et al*. (2002), Andersen *et al*. (2002), Birmingham & Mulle (2005), Anthony & McFayden (2005)
Court	1 (10%)	Shaw *et al*. (2003)
Probation	1 (10%)	Louden *et al*. (2013)

1.4.10 Mental health screening tools identified

1.4.10.1.1 Type of tools

In terms of the types of tools identified, half were self-completion questionnaires, 40% were interviewer administered self-report tools and one tool was observational. Table 5 provides a breakdown of tools by type and function.

1.4.10.1.2 Tool function

Six of the tools were designed specifically to detect likely severe mental illness, while two screen for mental health problems more generally and the other two assess psychological distress. Research has demonstrated that there is no consistency or operational definitions for severe mental illness (Ruggeri *et al*., 2000), and indeed definitions across the studies reviewed varied. The RDS, PriSNQuest, Birmingham and Mulle (2005) tool and the BJMHS defined severe mental illness as fulfilling diagnostic criteria for functional psychotic or major mood disorders. The K6 definition of severe mental illness additionally included fulfilment of diagnostic criteria for anxiety disorders. Although the function of the Grubin is also to screen for severe mental illness, Grubin *et al.* (2002) did not provide a definition for severe mental illness. Screening for severe mental illness is valuable given that it has been correlated to institutional violence and repeated incarceration (Baillargeon *et al*., 2009; Felson *et al*., 2012) as well as reflecting the typical referral criteria for secondary care mental health services such as in reach teams (Welsh Government, 2012b). However, such measures fail to differentiate between those with no need and those with low level need, which within a stepped care model can be treated at the primary care level (Layard, 2007; Welsh Government, 2012a).

Rather than screening for severe mental illness, both the CMHS-M and PMHI aim to screen more generally for mental health problems. The authors of the CMHS-M state that it screens for clinically significant mental health problems; specifically it provides dichotomous classifications for DSM-IV Axis I and Axis II mental disorders excluding Anti-Social Personality Disorder (ASPD). The PMHI is a broader screen, providing dichotomous

classifications for mood swings, hearing voices, problems with thinking, depression and anxiety, alcohol and drugs, self-harm, aggression and sexual problems related to mental health. The PMHI domains are drawn from the Cardinal Needs Schedule as oppose to being linked to diagnostic criteria. These tools may help to identify both those with lower level need as well as those with severe mental illness, but their dichotomous coding makes it difficult to identify the severity of need, which typically informs referral pathways.

The CORE-OM and GHQ-28 screen for the broader construct of psychological distress, but differ in terms of the subdomains that constitute this overall score. The CORE-OM measures four domains: subjective wellbeing, symptoms, functioning and risk. The CORE-OM also provides a continuous overall score of current global psychological distress, categorised within ranges from healthy to severe. The GHQ-28 provides scores for the subscales of depression, anxiety, social impairment and hypochondriasis as well as a continuous overall score for psychiatric distress. Cut offs on the GHQ-28 typically differentiate between 'caseness' for probable psychiatric morbidity (Goldberg, 1978). Such tools which focus on non-specific psychological distress are particularly useful in the context of stepped care models, where referral pathways are influenced by severity of need as well as diagnoses (Kessler *et al*., 2002).

Table 5 Systematic review screening tools by type and function

Characteristic	n (%)	Tool
Type of tool		
Self-report	5 (50%)	CORE-OM, GHQ-28,PriSnQuest, PMHI, K6
Interviewer administered	4 (40%)	RDS, Grubin, BJMHS, CMHS-M,
Observational	1 (10%)	Birmingham & Mulle (2005) tool
Screening function		
SMI	6 (60%)	RDS, Grubin, PriSnQuest, Birmingham & Mulle (2005) tool, BJMHS, K6
Mental health problems	2 (20%)	CMHS-M, PMHI
Psychological distress	2 (20%)	CORE-OM, GHQ-28

1.4.11 Clinical Utility

1.4.11.1.1 Practical considerations

Assessment of the clinical utility of the 10 screening tools identified (see table 1) indicated that all tools were appropriately brief for screening purposes. All tools could be completed within 10 minutes and half could be completed in less than five minutes. Tools taking longer than 5 minutes to administer may, however, be less practical in the high churn prison environment. The four interviewer administered tools (RDS, Grubin, BJMHS and CMHS-M) stipulated the need for staff training, thus would incur additional cost to implement. There was no indication training was needed to utilise the self-report or observational tools. All screening tools identified were freely available except the GHQ-28, which costs over a pound per administration. The cost of the GHQ-28 may decrease its clinical utility in high churn environments such as prisons with stringent budgets.

1.4.11.1.2 Service design considerations

In order for screening tools to have clinical utility, they must be consonant with the configuration of current mental health services. The oldest tool, the RDS, corresponds to the diagnostic categories schizophrenia, major depression and manic-depressive illness as defined by the out-dated Diagnostic and Statistical Manual for Mental Disorders third edition (DSM III), which was replaced by DSM IV in 1994 and subsequently DSM V in 2013 (American Psychiatric Association, 1980; 1994; 2013). More recently it has been acknowledged that diagnosis and distinction between disorders have little relevance at the screening phase, where identification of clinically significant mental health problems or distress in general is more useful (Ford *et al*., 2007; Division of Clinical Psychology, 2013; Kessler *et al*., 2002 Steadman *et al*., 2005). The RDS's strong diagnostic link impairs its utility in clinical practice today.

Similarly, although the PriSNQuest, Birmingham and Mulle (2005) tool, BJMHS and Grubin have reflected these trends to a degree and do not refer to specific diagnosis, all screen for SMI with implicit diagnostic definitions. Although consistent with referral criteria for many secondary care services, such measures' utility is diminished in light of stepped care models incorporating primary care services.

1.4.11.1.3 Temporal relevance

Eight out of the 10 tools identified screen for current symptoms. Two tools, the RDS and Grubin, however, focus on lifetime symptomology or historical indicators of psychiatric morbidity. Lifetime history of serious mental illness is important to assess since serious disorders can be episodic in nature and re-emerge (Teplin *et al.,* 1997; Teplin & Swartz, 1989). Nevertheless, current symptoms should arguably take precedence in a clinical context where there are limited resources and immediate treatment needs must be identified. As such, tools which solely focus on lifetime or historical factors such as the RDS and Grubin may be limited in their clinical utility.

1.4.12 Psychometric properties

All studies except Birmigham and Mullees (2005) reported some kind of psychometric properties associated with the tool in question as shown in Table 6. Only three original validation studies reported internal reliability (Ford *et al*., 2007; McCloskey *et al*., 2001; Teplin & Swartz, 1989). Where internal reliability was reported it was generally good with the lowest alpha reported for the CMHS-M at .78 (Ford *et al*., 2007) and the highest for the CORE-OM at .95 (McCloskey, 2001). Psychometric properties in terms of sensitivity and specificity were reported for seven out of the ten tools (Andersen *et al*., 2002; Ford *et al*., 2007; Grubin *et al*., 2002; Louden *et al*., 2013; Shaw *et al*., 2003; Steadman *et al*., 2005; Teplin & Swartz, 1989).

There is no consensus regarding what constitutes acceptable performance for a mental health screening tools in custody (Martin *et al*., 2013). Screening can improve identification of people with mental health problems, but often results in high false positive rates (Brooker *et al*., 2009). Some argue for screening purposes a high false positive rate is preferable (i.e. prioritising sensitivity) (Evans *et al*., 2010). On the other hand very high false positive rates can result in inefficient use of scarce resources (Hart *et al*., 1993; Steadman *et al.,* 2005), thus good specificity is also important. Martin *et al*. (2013) puts forward four different options in terms of evaluating performance, one of which is maximising overall accuracy with no priority given to sensitivity or specificity, which for sake of simplicity will be used as a point of comparison herein.

The Grubin tool had the highest sensitivity at .97 (Grubin *et al*., 2002), followed by the PriSnQuest with a sensitivity of .89 (Shaw *et al*., 2003) and the RDS with a sensitivity of .79 (Teplin & Swartz *et al*., 1989). The GHQ-28 had the poorest sensitivity at .65 (Andersen *et al*., 2002). In terms of specificity the RDS was the most specific, with a specificity of .98 (Teplin & Swartz, 1989) followed by the Grubin (.84) (Grubin *et al*., 2002) and BJMHS (.74) (Steadman *et al*., 2005). The K6 had by far the poorest specificity at .36 (Louden *et al*., 2013). Overall it appeared the Grubin tool followed by the RDS offered the best trade-off between sensitivity and specificity. The

BJMHS was developed to address limitations in the validity of the RDS and is recommended over the RDS (Martin *et al*., 2013). The BJMHS however has substantially lower sensitivity (.66) and specificity than the RDS (Steadman *et al*., 2005). As such, overall it seems the Grubin tool followed by the CMHS-M (Ford *et al*., 2007) (sensitivity .75 and specificity .70) currently offer the best overall accuracy in terms of mental health screening tools in prisons. Further research is needed to explore the accuracy of the CORE-OM (McCloskey, 2001), the Brimigham and Mulle (2005) tool and the PMHI (Anthony & Mc Fadyen, 2005).

Of the eight studies that reported tool accuracy, four had replication studies (Ford *et al*., 2007; Grubin *et al.,* 2002; Steadman *et al*., 2005; Teplin & Swartz, 1989). The RDS has several replication studies (e.g. Hart *et al*., 1993; Rogers *et al*., 1995). Hart *et al*., (1993) in contrast to the validation study, reported good sensitivity with poor specificity, while Rogers *et al*. (1995) questioned the discriminant validity of the RDS but confirmed good internal reliability. Birmingham *et al*. (2000) tested very similar questions to the Grubin and reported reasonable sensitivity (.76) and specificity (.71). However in a replication study of the Grubin and BJMHS with prisoners in New Zealand (Evans *et al*., 2010), both tools had around half the sensitivity reported in their respective validation study sample (Grubin *et al*., 2002; Steadman *et al*., 2005). Similarly, Gagnon's (2009) evaluation of BJMHS in a remand population found poorer accuracy in screening for SMI when compared to the original validation study. Baksheev *et al.* (2012) however reported better sensitivity for the BJMHS when screening for any Axis I disorder in police custody with the MINI as a criterion measure. With regard to the CMHS-M, Ford's (2009) validation study reported sensitivity was substantially lower than previously found. Overall, replication studies across tools largely failed to reproduce strong predictive properties across samples.

Table 6 Psychometric properties of systematic review tools

Screening tool	Study	Sensitivity	Specificity	PPV	NPV	α
RDS	Teplin & Swartz (1989)	.79	.98	.79	.01	-
	Hart *et al.* (1993)	-	-	.33	.89	-
	Rogers *et al.* (1995)	-	-	-	-	.78
Grubin	Birmigham *et al.* (2000)	.76	.71	.39	.93	-
	Grubin *et al*. (2002)	.97	.84	.60	.99	-
	Evans *et al*. (2010)	.42	.75	-	-	-
PriSnQuest	Shaw *et al*. (2003)	.89	.61	-	-	-
BJMHS	Steadman *et al*. (2005)	.66	.74	-	-	-
	Evans *et al*. (2010)	.34	.86	-	-	-
	Gagnon (2009) (SMI)	.67	.59	-	-	-
	Baksheev *et al.* (2012)	.82	.64	.68	.78	-
CMHS-M	Ford *et al.* (2007)	.75	.70	-	-	.78
	Ford *et al.* (2009)	.64	-	.66	.62	.78
K6	Louden *et al.* (2013)	.75	.36	-	-	-
GHQ-28	Andersen *et al.* (2002)	.65	.69	-	-	-
CORE-OM	McCloskey (2001)	n/a	n/a	n/a	n/a	.95
	Perry et al. (2009)	n/a	n/a	n/a	n/a	.81
Birmingham & Mullee tool	Birmingham & Mullee (2005)	n/a	n/a	n/a	n/a	n/a
PMHI	Anthony & McFadyen (2005)	n/a	n/a	n/a	n/a	.83

Note: replication studies excluded from the systematic review are shown in grey.

1.4.13 Methodological issues

1.4.14 Study Design

All studies were cross sectional in design. Eight of the studies employed a within subjects design to test development tools against an established gold standard (Andersen *et al*., 2002; Birmingham & Mullee, 2005; Ford *et al*., 2007; Grubin *et al*., 2002; Louden *et al*., 2013; Shaw *et al*., 2003; Steadman *et al*., 2005; Teplin & Swartz, 1989). Birmingham & Mulle (2005) also utilised a control group. Neither the CORE-OM nor the PMHI tools were tested against gold standard tools, which consequently makes establishing their predictive validity and thus utility as screening tools difficult.

1.4.15 Recruitment and sampling

All studies except Anthony & McFayden's (2005) specified at what point participants were recruited. As expected, eight of the studies recruited participants at point of reception to either court, jail, prison or probation (Andersen *et al*., 2002; Ford *et al*., 2007; Grubin *et al*., 2002; Louden *et al*., 2013; McCloskey, 2001; Shaw *et al*., 2003; Steadman *et al*., 2005; Teplin & Swartz, 1989). All of the studies which recruited participants from point of reception administered measures within the first week. McCloskey (2001) administered measures within two weeks of reception. Given the vast majority of studies recruited participants at point of reception, the validity of the screening tools for prisoners who have been in custody longer is not clear. Three studies administered diagnostic interviews as follow-ups. Ford *et al*. (2007) conducted a follow up five days after initial administration of tools, Louden *et al*., (2013) on average 87 days later and Shaw *et al*. (2003) did not report when the follow up was conducted. Conducting diagnostic interviews at later time points potentially reduces the reliability of findings in these studies, since psychological symptoms can vary over time. Birmigham & Mullee (2005) recruited a stratified sample of prisoners from wings who would have been there for varying time periods; average length of stay for participants was not reported making establishing the tools utility for screening at different time points in custody difficult.

Four studies employed a randomised sampling method (Anderson *et al.*, 2002; Ford *et al*., 2007; Grubin *et al*., 2002; Teplin & Swartz, 1989). Five studies systematically sampled all new admissions within a fixed time period (Anthony & McFayden, 2005; McCloskey *et al*., 2001; Shaw *et al*., 2003; Steadman *et al*., 2005; Louden *et al*., 2013). Brimigham & Mullee (2005) sample was selected based on prison officers identifying prisoners who met observational screening criteria, with random selection of a control group. The lack of randomisation in these six studies limits the generalizability of the findings since results may be influenced by selection bias.

Five studies reported both the number of participants who refused to consent and were excluded (Andersen *et al.*, 2002; Ford *et al*., 2007; Louden *et al*., 2013; McCloskey, 2001; Teplin & Swartz, 1989). Two studies reported figures relating to the decline rate but did not report exclusions (Birmigham & Mullee, 2005; Shaw *et al*., 2007). Three studies did not report figures for those who were excluded or declined (Anthony & McFayden, 2005; Grubin *et al*., 2002; Steadman *et al*., 2005). Failure to report declination rates and exclusion criteria reduce the validity of findings for the PMHI, Grubin and BJMS since it is not possible to ascertain if those who declined or were excluded had distinct characteristics from the study samples (for example higher levels of mental health problems).

1.4.16 Sample size

Sample size is a key indicator to assess when appraising quantitative studies since power to detect a statistically significant effect is depended upon sample size (Field, 2009). Cohen (1992) recommends a sample size of 783 to detect a small effect size, 85 to detect a medium effect size and 28 to detect a large effect size at a α level of .05 with recommended power of .8. Sample size varied considerably across the studies reviewed ranging from N=53 (McCloskey, 2001) to N=10,330 (Steadman *et al*., 2005). Two of the studies had large sample sizes capable of detecting even small effect sizes increasing the validity and reliability of their results (Shaw *et al*., 2003; Teplin & Swartz, 1989). Seven studies had medium sample sizes (Anthony & McFadyen, 2005; Andersen *et al.*, 2002; Birmigham & Mullee, 2005; Ford

et al., 2007; Grubin *et al*., 2002; Louden *et al*., 2013; Steadman *et al*., 2005) capable of detecting medium to large effect sizes and one study had a small ample size (McCloskey, 2001). The smaller sample sizes of these studies may have resulted in an underestimation of the tools performance, as it there would have been insufficient power to detect small effects.

1.4.17 Sample demographics

1.4.17.1.1 Age

Participant age was reported in eight out of the ten studies. Two studies did not report age thus reducing confidence in the applicability of their results to other samples (Anthony & McFadyen, 2005; Grubin *et al*., 2002). Andersen *et al*. (2002) reported age ranges from 18 to 60 years, but did not specify a mean age. All other studies provided the samples mean age with means ranging from 26 (Teplin & Swartz, 1989) to 34 years (McCloskey, 2001). The five studies reporting mean ages between 26 and 34 years all reported mean age as either 31 to 32 years (Birmigham & Mullee, 2005; Ford *et al*., 2007; Louden *et al*., 2013; Steadman *et al*., 2005) suggesting relative homogeneity in age across studies and increasing the relevance of these results to this study population, which is of a similar mean age.

1.4.17.1.2 Sex

Six of the studies included both male and female participants (Andersen *et al*., 2002; Ford *et al*., 2007; Grubin *et al*., 2002; Louden *et al*., 2013; Steadman *et al*., 2005; Shaw *et al*., 2003). Percentages of females within samples ranged from 14% (Shaw *et al*., 2003) to 31% (Ford *et al*., 2007). For the purposes of the current review, where findings were reported separately for males and females, the figures for males were extracted given the current emphasis on male prisoners. Two studies which included males and females did not report the proportions of each gender negating ability to identify differences in psychometric properties of the tools by sex (Andersen *et al.*, 2002; Steadman *et al*., 2005). Four studies had all male samples and therefore the tools validity with female prisoners cannot be assumed (Anthony & McFadyen, 2005; Birmigham & Mullee, 2005; McCloskey, 2001; Teplin & Swartz, 1989).

1.4.17.1.3 Ethnicity

Six of the ten studies reported sample distribution by ethnicity (Birmigham & Mullee, 2005; Ford *et al*., 2007; Louden *et al*., 2013; McCloskey, 2001; Steadman *et al*., 2005; Teplin & Swartz, 1989). Notably, the four US studies reported much larger proportions of black or African American (between 33% (Louden *et al*., 2013) and 58% (Steadman *et al*., 2005)) and Hispanic 20% (Ford *et al*., 2007) participants compared to the UK studies thus reducing reliability of their findings in UK samples. Both McCloskey (2001) and Birmingham & Mullee's (2005) UK samples consisted of 92% white participants in kin with the South Wales prisoner population. Four studies did not report the ethnic composition of their samples (Andersen *et al*., 2002; Anthony & McFadyen, 2005; Grubin *et al*., 2002; Shaw *et al*., 2003) limiting their generalizability.

1.4.18 Management of confounding variables

The potential impact of confounding variables was considered by all studies to varying degrees. Three tools, the RDS, PriSnQuest and CMHS-M were derived statistically from longer composite measures (Ford *et al*., 2007; Shaw *et al*., 2003; Teplin & Swartz, 1989). Additionally, in Andersen *et al's*. (2002) study the GHQ-28 was embedded in a longer assessment. As such the assessment of the independent validity and accuracy of these tools is compromised and cannot be assumed.

Selection procedures also have the potential to produce bias in some studies. Randomisation was only employed in four studies. One study offered financial incentive for participation resulting in high participation rates, but potentially introducing contamination bias into the results (Teplin & Swartz, 1989). Another study had a high drop out rate (> 40%) (Louden *et al*., 2013) although possible differences across those who dropped out and participated were assessed and did not differ. Furthermore, Ford *et al*. (2007) excluded those with known mental health problems from the study entirely significantly reducing the relevance of their result to populations with high rates of mental health problems. Steadman *et al*. (2005) on the other hand over sampled those who screened positive for mental health problems

in their follow up without statistically adjusting for this, thus reducing the relevance of their findings to non-clinical samples. Almost all studies (n=8) also focused on new receptions. Such sampling biases decrease the generalizability of findings and consequent validity of tools across samples.

Personnel administrating screening tools also varied between researchers and operational staff across studies. In half of the studies researchers administered the measures (Andersen *et al*., 2002; Birmingham & Mullee, 2005; Ford *et al*., 2007; Shaw *et al*., 2003; Teplin & Swartz, 1989). In three studies operational staff administered screening tools (Grubin *et al*., 2002; Louden *et al*., 2013; Steadman *et al*., 2005) and two studies did not report who the tools were administered by (Anthony & McFadyen, 2005; McCloskey, 2001). It is possible that disclosures of mental health difficulties may have been higher in studies where tools were administered by researchers since research has shown that offenders are less likely to disclose health problems to operational staff (Steadman *et al*., 2005). However, studies where operational staff administered tools (Grubin *et al*., 2002; Louden *et al*., 2013; Steadman *et al*., 2005) may well have greater ecological validity.

Of the seven studies which utilised independent diagnostic interviews as criterion measures, only three ensured blinding to the outcome of other measures (Grubin *et al*., 2002; Ford *et al*., 2007; Steadman *et al.,* 2005). Blinding was not carried out in Birmingham & Mullee (2005) or Shaw *et al*'s. (2003) studies and it was not clear if blinding was implemented or not in Louden *et al's*. (2013) study. Failure to blind interviewers to either the outcome of screening or diagnostic assessment (depending on the order of administration) can result in biased assessments, which may result in overestimation of tools performance (Karanicolas *et al*., 2010).

Similarly, although common practice Whiting *et al*. (2011) argue that selecting a post-hoc cut point to optimise sensitivity and specificity results in overestimation of test performance, whereby the test is likely to perform worse in an independent sample where the same cut off is applied. In four

studies the test cut off was selected to optimise sensitivity and specificity (Andersen *et al*., 2002; Ford *et al*., 2007; Shaw *et al*., 2003; Teplin & Swartz *et al*., 1989). For tools that had replication studies including the RDS (Hart *et al*., 1993), Grubin (2002), and CMHS-M (Ford *et al*., 2009), all except the BJMHS (Evans *et al*., 2010) had lower sensitivity and specificity in replication studies compared to the original study reducing confidence in the generalizability of their validity.

Although descriptive demographic information such as age was reported in eight studies and ethnicity in six studies, none of the studies explored these variables with regard to the study results. Grubin *et al*. (2002) did explore differences between male young offenders and adults, although age differences within the adult category were not explored further. Potential confounding factors such as sentence length, index offence and detention status were also not explored making it difficult to establish if the tools had similar psychometric properties or not across differing groups of prisoners.

1.4.19 Criterion measures

Eight of the ten studies used criterion measures as 'gold standards' to evaluate screening tools against. The most commonly cited criterion measure was the SCID utilised in three studies (Ford *et al*., 2007; Louden *et al*., 2013; Steadman *et al*., 2005). The SADS-L was adopted in two studies (Birmingham & Mullee, 2005; Grubin *et al*., 2002). The NIMH-DIS (Teplin & Swartz, 1989), PSE-10 (Andersen *et al*., 2002) and SCAN (Shaw *et al*., 2003) were also utilised in a study each. Two studies did not assess tools against a criterion measure (Anthony & McFadyen, 2005; McCloskey, 2001). Half of the studies administered the criterion measure to all participants (Andersen *et al*., 2002; Birmingham & Mullee, 2005; Ford *et al*., 2007; Tepli & Swartz, 1989), while the other half administered the criterion measure to a sub sample only substantially weakening power to establish predictive validity. In Louden *et al's*., (2013) study there was a substantial delay (more than a month) between administration of the screening measure and criterion measure, decreasing the meaningfulness of comparison. Use of criterion measures allows for the estimation of

sensitivity and specificity as well as assessing construct validity. The effectiveness of tools in a screening context is not clear if meaningful comparison has not been made with a criterion measure to allow for estimation of sensitivity and specificity.

1.4.20 Quality of reports

The quality of reporting within the ten studies included in the review was variable. Based on the quantitative scoring system applied to the studies, Ford *et al's*. (2007) study scored highest scoring two out of two followed by Grubin *et al's*. (2002), Steadmean *et al*. (2005) and Louden *et al's*. (2013) studies each scoring 1.8 out a possible 2. Anthony and McFayden's (2005) study was the weakest (score of 1.3) in terms of quality closely followed by Shaw *et al's.* (2005) study (score of 1.4). All studies included in the systematic review included and introduction drawing on relevant literature and provided a rationale for the study. One study did not have an abstract (Andersen *et al*., 2002). Only two studies clearly specified hypotheses (Louden *et al*., 2013; Steadman *et al*., 2005). Failure to specify a priori hypothesis increases the possibility of false-positive results from post hoc testing (Corner & Kendall, 2013), thus reducing confidence in the results of these studies.

With regard to methodology, most studies reported the dates of data collection, three did not specify a time period making it difficult to establish their contemporary relevance (Ford *et al*., 2007; Louden *et al*., 2013; Anthony & Mc Fadyen, 2005). Half of the studies specified inclusion and exclusion criteria, half did not (Anthony & McFadyen, 2005; Birmigham & Mullee, 2005; Grubin *et al*., 2002; Shaw *et al*., 2003; Steadman *et al*., 2005). All studies except Anthony & McFayden's (2005) provided a description of how their sample was arrived at increasing their reliability, although none were based on power calculations.

Statistical procedures utilised to deduce results were presented in all papers and summaries of key findings were provided in each paper's discussion. All papers discussed the potential utility and generalisability of the tools in

question. To varying degrees, all but one paper (McCloskey, 2002) considered limitations of the study. Six of the studies provided detail of how the research was funded; four did not which may introduce bias into their results based on the funding source (Andersen *et al*., 2002; Anthony & McFadyen, 2005; McCloskey, 2002; Shaw *et al*., 2007).

1.4.21 Implications for clinical practice

Screening for common and severe mental health problems using evidence-based tools is stipulated within the Policy Implementation Guidance as a core function of prison mental health services in Wales (Welsh Government, 2014); which screening tools are utilised to do so, however, is not stipulated. The current systematic review has identified a range of brief screening tools validated for identifying mental health needs amongst offenders, which are mostly freely available and can be quickly administered. However substantial variation in methodological rigour and approach including sample characteristics, size, criterion measures and psychometric evaluation make selecting appropriate measures for the clinical context in Welsh prisons difficult. Furthermore, existing research has failed to reliably replicate strong psychometric properties of tools in new samples, thus the utility and accuracy of the existing tools for the Welsh prison population cannot be assumed.

Systematic analysis of the studies included in this review indicated that four tools, the Grubin, BJMHS, CMHS-M and K6 had validation studies of good quality, which support their use in clinical practice. However, three of those tools the BJMHS, CMHS-M and K6 were developed and validated in the US. This is problematic since the samples they were validated on are not representative of the British prisoner population, particularly in terms of ethnic composition. The sensitivity and specificity of these three tools were also all below .80 in their validation studies. Furthermore, the BJMHS and CMHS-M both had lower sensitivity in their replication study samples (Evans *et al*., 2010: Ford *et al*., 2009) and no studies further validating the K6 with prisoners were identified. This brings into question their utility for clinical practice in the UK, given that their predictive accuracy is likely to differ within

a British prisoner population. A further issue is that none of these three studies have been validated across a spectrum of prisoners. Those with mental health problems were excluded from Ford *et al*. (2007) study, Steadman *et al*. (2005) oversampled those identified with a mental health problem and Louden *et al*. (2013) only sampled probationers. As such their utility for screening prisoners across the range from those with no mental health problems through to those with severe mental health problems cannot be assumed. Despite the BJMHS, CMHS-M and K6 demonstrating some promise in terms of being freely available short well validated tools, they would need to be further validated with British prisoners in order to establish their utility in clinical practice in the UK.

The Grubin (Grubin *et al*., 2002) screening tool, which this review also identified as having a high quality validation study, is the only tool to have been adopted in the UK. It is not however used in its original validated form in the prisons sampled herein, and adapted versions may not possess the same psychometric properties. Moreover, the Grubin was developed prior to introduction of the stepped care model, thus once again is designed to screen for severe mental illness based solely on historic factors. As such it is not appropriate to address screening for current primary care needs as stipulated by Part 1 of the Mental Health Measure (2010). The few tools identified that do screen more broadly for mental health problems and psychological distress have shown poor psychometric properties (CMHS-M; GHQ-28) within prison populations or have not been sufficiently validated (CORE-OM; PMHI). In sum, the current review has highlighted the lack of sufficiently validated screening tools for assessing the mental health of British prisoners in line with current clinical practice using a stepped care model as required by the Mental Health Measure (2010). Consequently, there is a need to validate screening tools which may be psychometrically robust, be feasible and reflect the needs of current service pathways in British prisons.

1.4.22 Limitations of the systematic review

Firstly, while a systematic approach was adopted in searching the literature, it is possible that some tools may not have been identified – particularly those that performed poorly which may have been subject to publication bias (Martin *et al*., 2013). Secondly, the review is dependent on the quality of papers included. Studies varied dramatically in their approach to and reporting of samples and psychometric properties thus rendering comparison difficult at times. Many of the studies were conducted in countries outside of the UK where differing legal frameworks and population demography alter the composition of the offender populations. Additionally, cross cultural differentiation in the construct of mental health influences base rates, which in turn affects psychometric properties.

Thirdly, comparison of positive and negative predictive values was beyond the scope of this review. Given that estimates of base rates for mental health problems in the prison population are relatively high, however, predictive values are not likely to be heavily affected (Glaros & Kline, 1988). Fourthly, most studies were based on self-reports and as such could be subject to biases such as social desirability (van de Mortel, 2008) as well as malingering (Rogers *et al*., 1996). None of the studies checked content validity through seeking prisoner's views. Fifthly, the populations included all volunteered to take part. It could be argued that those that are most unwell would be the least likely to take part, as well as those with known mental health problems being specifically included or excluded from some studies reducing the generalisability of findings. Finally, the current review focused specifically on screening tools validated for use with offenders and thus excluded existing tools used in the community, which may be useful for screening for mental health problems but have yet to be validated in this population.

1.4.23 Summary and rationale

In sum, this review has identified ten screening tools for assessing the mental health of prisoners. Four tools had high quality validation studies (BJMHS; CMHS-M; K6; Grubin), though only one tool, the Grubin was

validated in the UK. Though the Grubin is utilised in current screening processes in British prisons it is now outdated, focusing on historic predictors of mental health problems as opposed to current distress and treatment need. Similarly most other tools identified typically focused on identifying severe mental illness therefore do not meet the policy and service requirement to also screen for common mental health problems. Identified tools which have the potential to screen more broadly are poorly validated in prisons, do not have established cut point and are not used routinely in British prisons. As such, the systematic review has highlighted a gap in the literature in term of establishing effective screening tools and identifying cut off points for common and severe mental health needs. Addressing this gap by validating appropriate tools will allow for faster and more accurate assessment of need and thus appropriate referral and targeting of limited resources. Consequently, the current research aims to compare the utility of the CORE-10 as screening tool for common and severe mental health problem in comparison to the GHQ-12 and existing practice in terms of referral decision.

1.5 Research Aims

Specifically the current research aims to improve screening for common and severe mental health problems in line with current policy by:

- Identifying appropriate cut off points for common and severe mental health needs in prisoners on brief standardised screening measures (CORE 10 and GHQ-12), comparing against a gold standard diagnostic interview (Mini International Neuropsychiatric Interview 6.0 (MINI 6.0)).
- Comparing the sensitivity and specificity of differing tools and methods (CORE 10, GHQ-12 and referral decisions) against a gold standard (MINI 6.0) to establish which gives the most accurate prediction of common and severe mental health problems in prisoners.
- Assessing the psychometric reliability and validity of the CORE-10 within a prison population.

1.5.1 Hypotheses

Based on existing research and the study aims the following hypotheses will form the basis of this study:

H[1] CORE-10 scores will be higher amongst those meeting MINI 6.0 screening criteria for common or severe mental illness than amongst those with no disorder.

H[2] GHQ-12 scores will be higher amongst those meeting MINI 6.0 screening criteria for common or severe mental illness than amongst those with no disorder.

H[3] Variables identified by research as predictive of mental health problems in custody will be correlated with measures of mental health in the current study.

H[4] There will be a strong positive correlation between CORE-10 and GHQ-12 scores.

H[5] There will be a positive correlation between CORE-10 and clinical referral decisions.

H[6] CORE-10 score and MINI 6.0 SMI, any current and lifetime mental health condition status will be correlated.

H[7] CORE-10 and MINI 6.0 suicidality scores will be correlated.

H[8] The CORE-10 will have discriminant validity in that scores will not be correlated with variables not directly associated with mental health (e.g. age, ethnicity, prison).

H[9] The CORE-10 will have superior sensitivity and specificity than the GHQ-12 and referral decisions for identifying prisoners with common mental health needs at the primary care level.

H[10] The CORE-10 will have superior sensitivity and specificity than the GHQ-12 and referral decisions for identifying prisoners with severe mental health needs at the secondary care level.

H[11] The CORE-10 will have high internal and re-test reliability in the prison population

H[12] Confirmatory Factor Analysis (CFA) will confirm the original six factor structure of the CORE-10 in a prison population.

2 Method

2.1 Design

A cross-sectional design was adopted in order to explore the construct validity, sensitivity and specificity of the CORE-10 as a screening instrument for assessing common (primary care) and severe (secondary care) mental health needs. The CORE-10 was compared against the GHQ-12 and referral to mental health services decisions, with the MINI 6.0 diagnostic interview utilised as the gold standard benchmark. Two-week re-test reliability of the CORE-10 and GHQ-12 was assessed via a longitudinal design. Data was collected by self-completion of questionnaires and analyses used correlational approaches, ROC and confirmatory factor analysis.

2.2 Participants

2.2.1 Power analysis

Participants were recruited from two prisons in south Wales. The sample size was based on a power calculation for ROC analysis using MedCalc version 15.11 (MedCalc Software, Ostend, Belgium). The four ROC analyses would require a total minimum sample of 116 to distinguish a typical area under the curve of 0.8 from an area of 0.5 (no prediction) at power of 0.8 with alpha set at .05. Given high attrition rates in-prison based research over-sampling is recommended (Trestman *et al*., 2015). A total sample of *N* =150 was recruited to maximise statistical power.

2.2.2 Inclusion and exclusion criteria

All male prisoners aged 18 years and above who had been received into the establishments in the six months preceding data collection were eligible for inclusion. Prisoners who did not speak English, who were deemed ‘unsafe

to see' based on prison risk assessment, or who the researcher assessed as not having capacity to consent[7] were excluded.

2.3 Procedure

2.3.1 Ethical and research approval

The study was approved by Cardiff University Psychology Research Ethics Committee (see appendix 2) and the National Offender Management Service (NOMS) research committee (see appendix 3). Permission to recruit participants was granted by the respective governor and director of the two establishments following consultation with key stakeholders (see appendix 4 for details of stakeholder engagement).

2.3.2 Recruitment

An invitation explaining the nature of the research, inclusion criteria and inviting prisoners to participate in the research was sent out to every prisoner a week before the researchers were due to visit the establishments. Wings were then chosen at random and visited by the researcher[8]. All prisoners who had shown an interest in taking part on that wing were approached by the researcher who provided further information and sought informed consent to participate. On wings where few prisoners had shown an interest, all prisoners on that wing at the time were approached and offered the opportunity to participate. On wings where many prisoners had volunteered to participate participants were randomly selected by pulling prisoner numbers out of a hat based on the number of participants it was possible to see during that period.

2.3.3 Consent

Ethically concerns have been raised regarding prisoners being coerced into research (Fazel & Lubbe, 2005). Moser *et al*. (2004) formally assessed a sample of 40 prisoners' competence to consent to participate in research and concluded that normal procedures for ensuring informed consent are

[7] Capacity was informally assessed by the researcher based on judgment of whether individuals could understand information provided regarding the study and the implications and potential consequences of taking part.

[8] Wings specifically for prisoners staying longer durations at the prisons were excluded due to the exclusion of individuals who had been at establishments for six months or more.

sufficient for prisoner populations.

As such, information outlining the purpose of the study, including that participation was voluntary, was provided in the information sheet for prisoners (see appendix 5). The researcher gave prisoners time to read the information sheet and also summarised the information verbally for all prisoners to account for the low level of literacy within the prisoner population (Morgan & Kett, 2003). Prisoners provided informed consent to participate by signing the consent form (see appendix 6). A debrief sheet was provided for all prisoners who participated (see appendix 7).

2.3.4 Confidentiality

Participants were informed that all information they provided would be treated confidentially unless they disclosed information to suggest a risk to themselves, others, security or a breach of prison rules. Where risks were identified, prisoners were informed that confidentiality would be breached and prison staff were informed (see appendix 8 for a description of the risk management protocol).

Data was stored anonymously to maintain confidentiality. Participants were assigned a research ID number in order to link data from the two time points. Front sheets of the interview battery with prisoner numbers on were removed and consent forms were stored separately to maintain confidentiality.

2.3.5 Data collection and storage

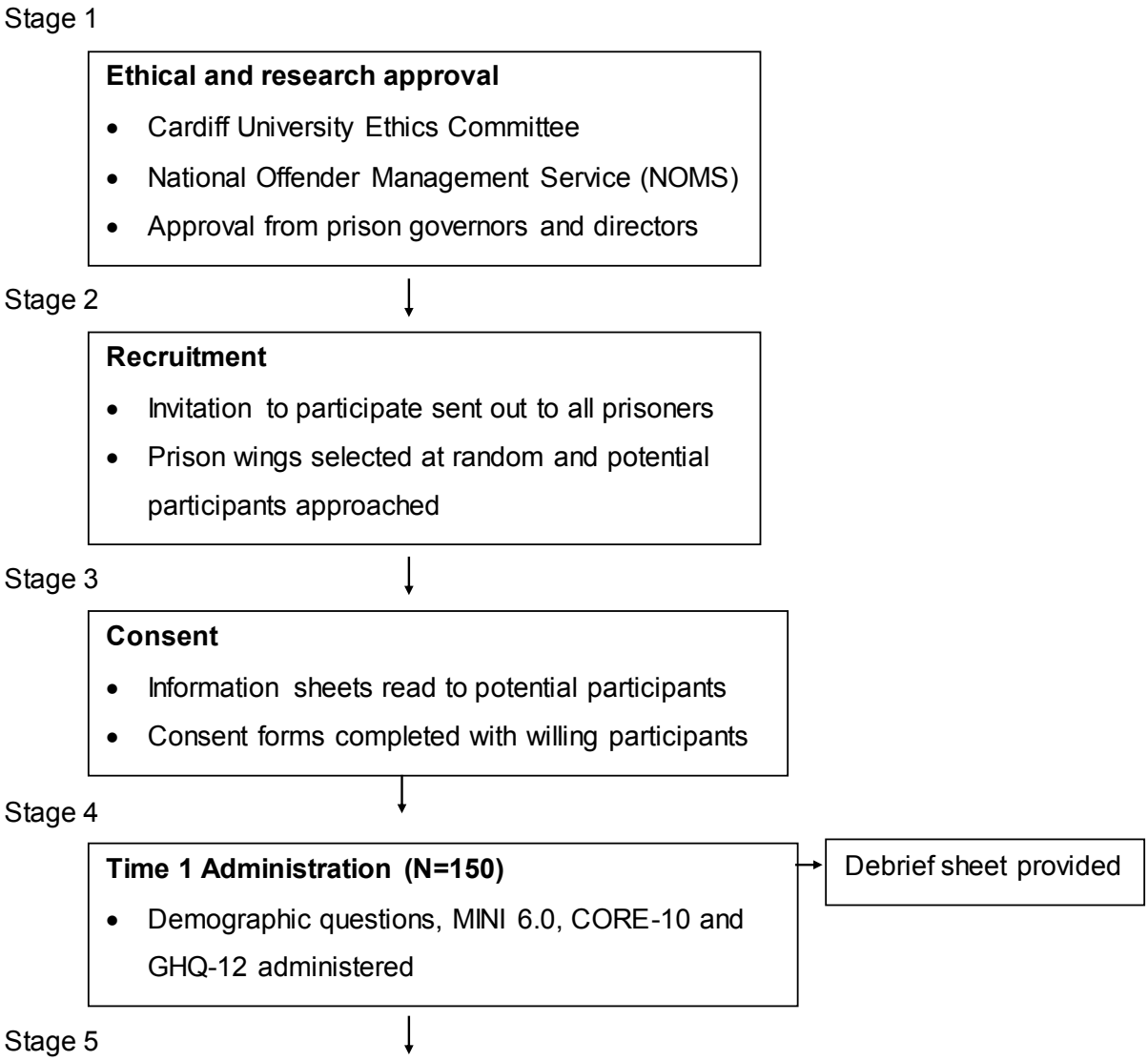
Data collection took place during July and August 2015. Interviews were conducted by the author (*n*=101) and three research assistants (*n*=49) all of whom had been trained in administration of the measures and were supervised by the author. In line with the Data Protection Act 1998 data was stored anonymously. All data was stored in a locked filing cabinet.

2.3.6 Administration

Interviews took place in interview rooms or at tables in secluded areas of wings. Researchers were in view of, but out of earshot of, custodial staff during interviewing. Interviews lasted between 15 and 50 minutes. At time

one measures were administered in the following order: demographic questions, CORE-10, GHQ-12, MINI 6.0 (appendix 9). At time two the GHQ-12 was administered first followed by the CORE-10 (appendix 10). Fifty four per cent (*n*=81) of the sample completed time two interviews approximately two weeks (range 11-16 days) following the initial interview. An overview of the research procedure is provided in figure 1.

Figure 2 Research procedure



Time 2 Follow-up administration (N=81)

- Sub group of prisoners administered the GHQ-12 and CORE 10 (11-16 days post time 1)

→ Debrief sheet provided

2.4 Measures

2.4.1 Demographic measures

Demographic information regarding: age, ethnicity, if the person was a foreign national, what their first language was, if it was their first time in prison sentence length, length of time in current prison, index offence and if they were a veteran of the armed forces was collated (see appendix 9). Index offence was categorised according to Home Office counting rules for recorded crime (Home Office, 2015). Where several index offences were reported the most serious (based on sentence length) was categorised. Items included were derived from the existing literature and discussions with commissioners.

2.4.2 Predictors of mental health problems in prisons

Research has identified a number of factors that are correlated with metal health difficulties in prison (Birmingham *et al*., 2000; Greenberg & Rosenheck, 2008). Variables consistently identified are a history of self-harm, past psychiatric care, prescription of antidepressants before prison (Brimigham *et al*., 2000; Grubin *et al*., 2002), homelessness, alcohol and drug problems in the in the year before prison (Greenberg & Rosenheck, 2008; McNeil *et al*., 2005). As such, six dichotomous questions relating to these factors were included in the interview battery.

2.4.3 Core Outcomes in Routine Evaluation 10 (CORE-10)

The CORE-10 (Connell & Barkham, 2007) is a brief measure derived from the CORE-OM, a 34 item assessment and outcome measure. The CORE-10 is a pan theoretical measure to assess global distress. It is suitable for use as screening tool and outcome measure. The CORE-10 includes two anxiety items, two depression items, three functioning items and an item each for trauma, physical symptoms and risk. These items are split across six high intensity and four low intensity items. Individuals rate how much

they have felt a certain way in the last week on a five point scale ranging from ‘not at all’ (scored 0) to ‘most of the time’ (scored 3). The CORE-10 contains two reverse scored items. The measure was presented in its original form.

A total clinical score ranging between 0 and 40 was derived by summing scores from the 10 items. Higher scores indicate the individual is reporting more problems/distress. Scores above 10 are considered to be in the clinical range, with severity levels ranging from mild to severe defined by different cut of points as shown in table 7 (Connell & Barkham, 2007). At a cut off of 13 the CORE-10 has sensitivity of .92 and specificity of .72 in community populations when compared against a diagnosis of DSM-IV depression using the SCID. In community samples there are also good correlations between CORE-10 scores and scores on other measures, including the SCL-90-R, Brief Symptom Inventory and the Beck Depression Inventory (BDI). The CORE-10 has very good internal reliability (α .82) and is sensitive to change, with a recommended change score of 6 being reliable (Connell & Barkham, 2007). The CORE-10 has yet to be validated in prison populations, although items from its parent measure the CORE-OM were found to be acceptable and feasible with hospitalised forensic inpatients (Perry *et al*., 2013).

Table 7 Normative cut off scores for the CORE-10 (Connell & Barkham, 2007)

	Severity	**CORE-10 cut off score**
Non clinical	Healthy	≤ 5
	Low level	≤ 10
Clinical	Mild	11-14
	Moderate	15-19
	Moderate to severe	20-24
	Severe	≥25

2.4.4 General Health Questionnaire 12 (GHQ-12)

Derived from the longer GHQ-60, the GHQ-12 is one of the most widely used screens of psychiatric morbidity. The tool was designed to be a unidimensional measure of psychiatric morbidity, although there has been contention that it may have a two or three factor structure (Kalliath *et al*., 2004; Shevlin & Adamson, 2005; Werneke *et al*., 2000). Hawkins (2008) concluded that the GHQ-12 is undimensional, but is affected by response bias to negatively framed items.

The GHQ-12 aims to distinguish 'psychiatric cases' from non-cases, thus assessing deviation from 'normal functioning' (Goldberg, 1972). The shorter 12 item version is thought to be just as good, or better, than longer versions at detecting psychiatric morbidity (Goldberg *et al*., 1997).

Participants were asked to rate the GHQ-12 items in terms of whether they had felt that way in the preceding few weeks more or less than usual on a four point scale in line with the original format. Various scoring options are available for the GHQ-12, the 0-0-1-1 method was adopted for the current study. This method of scoring is preferred since it has the advantage of eliminating errors caused by 'end' and 'middle' users (Goldberg, 1978; Goldberg *et al*., 1997) and has been utilised in previous studies with prison populations (Andersen *et al*., 2002; Boothby *et al*., 2010; Hassan *et al*., 2011). Using this scoring method, a total GHQ-12 score ranging between 0-12 was created by summing responses to the 12 items. Higher scores indicate greater psychiatric morbidity.

Multiple thresholds for identifying cases on the GHQ-12 have been proposed. In community samples thresholds as low as one or more (Gureje & Obikoya, 1990) and as high as six or more (Goldberg *et al*., 1997) have been proposed, although three or more is the typical cut-off utilised in the community (Hassan *et al*., 2011). Higher cut points are however thought to offer a better balance between sensitivity and specificity in prison populations (Andersen *et al*., 2002), although there is no established cut off point for forensic populations. Cut-off scores of four (McGilloway & Donnelly,

2004) five (Boothby *et a*l., 2010; Smith & Borland, 1999) and seven or more (Hassan *et al*., 2011; OHRN, 2010) have also been used as threshold for caseness in offender populations.

2.4.5 Mini Neuro Psychiatric Interview Version 6 (MINI 6.0)

The MINI 6.0 is considered a gold standard short diagnostic interview for major Axis I disorders consistent with DSM-IV and ICD-10 classifications. It has similar reliability and validity to the SCID, but is faster to administer (Sheehan *et al*., 1997; 2010). The MINI 6.0 produces dichotomous yes/no classifications for presence of disorders both currently and over the person's lifetime. The MINI 6.0 has previously been utilised as a gold standard against which to validate mental health screening tools in prison populations (Baksheev *et al*., 2012; Evans *et al*., 2010), as well as being widely used in epidemiological research within custodial settings (Black *et al*., 2004; Borrill *et al*., 2003; Gunter *et al.,* 2008; Falissard *et al*., 2006; Fotiadou *et al*., 2006; Rivlin *et al*., 2010; Westmoreland *et al*., 2010).

In order to reduce demands on participants, only the major depressive, suicidality, manic and hypomanic, panic disorder, agoraphobia, social phobia, obsessive compulsive disorder, post-traumatic stress disorder, generalised anxiety disorder and psychotic disorders sections of the MINI 6.0 were administered. The MINI screens for antisocial personality disorder[9], anorexia nervosa, bulimia nervosa and substance dependence and abuse were excluded for sake of brevity. These disorders were considered of less interest as they are not typically treated by prison mental health services.

[9] Personality disorders were specifically excluded from consideration for a number of reasons. Firstly, personality disorders have not typically been included in mental health screening tools previously. Secondly, assessment of personality disorder is complex and would not be captured by CORE-10 screening items. Finally, the very high prevalence rate of personality disorders in male prison populations (Fazel & Seewald, 2012) meant that inclusion would have resulted in the vast majority of prisoners screening positive, thus rendering the utility of the screening tool in practice redundant.

Three dichotomous variables, any current serious mental illness (SMI), any current mental health disorder and any lifetime mental health disorder were created based on MINI 6 classification. In line with previous research (Birmingham & Mullee, 2005; Shaw *et al*., 2003; Steadman *et al*., 2005; Teplin & Swartz, 1989), the definition of a current SMI was based on a positive screen in the last month for current major depressive disorder, bipolar (i), bipolar (ii), bipolar disorder not otherwise specified, mood disorder with psychosis or psychotic disorder. The 'any current mental health disorder' variable was based on a positive screen for any of the following in the last month: major depression episode, suicidality, manic episode, major depressive disorder and/or bipolar (i) and/or bipolar (ii) and/or bipolar disorder not otherwise specified and/or mood disorder with psychosis and/or psychotic disorder, panic disorder (with and without agoraphobia), agoraphobia, social phobia, obsessive compulsive disorder, post traumatic distress disorder and generalised anxiety disorder. Similarly, any lifetime mental health disorder was based on a positive screen for any of the above disorders over the life course. For suicidality, the MINI 6.0 produces a score ranging from 0 to 68, with scores from 0 to 8 considered low, 9 to 16 considered moderate and 17 or more considered high in term of current suicidality.

2.5 Data analysis

2.5.1 Descriptive analysis

Data analysis was carried out using SPSS version 23 (IBM corporation, 2015). Preliminary analysis was conducted to check for errors in the data, outliers and test assumptions for parametric tests in continuous variables. Correlational analysis was carried out to test hypothesised associations between differing mental health variables, as well as exploring relationships with demographic variables. Alpha was set at 0.05 for all two-tailed tests of statistical significance.

2.5.2 Receiver Operating Curve (ROC) analysis

The sensitivity, specificity, negative and positive predictive power of the CORE-10 and GHQ-12 for identifying primary care and secondary care level

need were assessed against MINI any current mental health disorder (primary) and MINI current SMI (secondary) classification using ROC analysis with MedCalc version 15.11 (Medcalc Software, 2015).

2.5.3 Confirmatory Factor analysis

Confirmatory Factor Analysis was conducted on the time 1 CORE-10 data (*N*=150) using AMOS version 20 (Arbuckle, 2011). Structural Equation Modelling estimation was conducted using the variance-covariance matrix with Maximum Likelihood (ML) method. Bollen-Stine bootstrapping was applied to assess model fit due to the data not conforming to the assumption of multivariate normality (Bollen & Stine, 1992). With non-normal data particularly in small samples the ML Chi Square statistic can be inflated, therefore assessment of the Bollen-Stine bootstrap p-value is recommended to ascertain overall model fit. A non-significant Bollen-Stine bootstrap p value supports the null hypothesis that the model is a good fit (Ghofar & Islam, 2015). Three factor solutions were compared, the original six factor solution[10], the two factor solution representing high and low intensity factors – both proposed by Connell & Barkham (2007), and a single factor solution including all CORE-10 items (see appendix 12 for hypothesised factor solutions). Model fit was assessed using the goodness-of-fit index (GFI); comparative fit index (CFI); normed fit index (NFI); root-mean-square error approximation (RMSEA) and Akaike's information criterion (AIC) with assessment of recommended acceptable values.

[10] Single item factors were excluded from the six factor model since the observed item represents the latent factor and it is not possible to establish an identified model when single item factors are included (McDonald, 1985). As such the six-factor model contained the three factors with more than one item: anxiety, depression and functioning.

3 Results

3.1 Chapter outline

The current chapter presents the results of the study with reference to the stated hypotheses. Firstly, results of the preliminary data analysis to assure data quality and assess appropriateness of statistical tests are presented. Secondly, descriptive results for demographic and mental health variables are presented, with an analysis of the relationship between variables. Thirdly, ROC analysis results are presented to compare the performance of each tool in screening for mental health problems in prisoners. Finally, results of confirmatory factor analysis are presented to confirm the factor structure of the CORE-10 with prison populations.

3.2 Preliminary data analysis

3.2.1 Error analysis

Minimum, maximum and frequency values for each variable were calculated to identify if any point fell outside the defined range or category. One data point was identified as entered incorrectly and two as not entered, these were corrected by referring back to the raw data. Total scores were calculated electronically using SPSS to avoid human error.

3.2.2 Missing data

There was only one missing data point from the CORE-10 and GHQ-12 continuous variables. These missing data points were replaced with the individual's mean for that scale. Categorical data was complete with the exception of two participants. Categorical MINI 6.0 data was incomplete for one participant due to his interview being terminated prematurely for operational reasons. His positive classification on earlier MINI 6.0 indices, however, allowed his data to be utilised in the overall analysis. Referral decision data was not available for one other participant, who consequently was excluded from the ROC analyses.

3.2.3 Parametric test assumptions

3.2.4 Normality

None of the continuous variables (age, suicidality score, CORE-10 and GHQ-12 at times one and two) were normally distributed as indicated by significant Kolmogrov-Sirmov (*K-S*) test statistics ($p > .05$). All variables were significantly negatively skewed as indicated by significant skew z scores ($p > .05$).

3.2.5 Linearity

Parametric tests require a linear relationship between variables (Field, 2009). Analysis of scatter plots of each continuous variable against each other confirmed linear distributions.

3.2.6 Homoscedasticity

Homoscedasticity assumes similar variances for residuals across the range of the predictor variables (Field, 2009). Visual analysis of standardised scatterplots suggested the data was homoscedastic.

3.2.7 Independence and outliers

Participants represented independent cases. Outlier analysis was carried out using box plots and z score analysis. Z scores did not exceed the critical value of 3.29 for any variable. For all variables less than 1% of Z scores exceed the critical value of 2.57 and less than 5% exceeded the critical value of 1.96, suggesting outliers would not unduly bias parametric models fitted to the data (Field, 2009).

3.2.8 Summary

The data met the assumptions of linearity, homoscedasticity, independence and had minimal outliers. However, the assumption of normality was violated for all continuous variables. As such, bootstrapping was employed when calculating correlations and confirmatory factor analysis, a method of statistical inference that does not rely on the normal distribution of data (Efron & Tibshirani, 1997). Bootstrapping uses the sample data to empirically build a picture of the sampling distribution of the sample mean by ‘resampling’ the data with replacement many times, rather than relying on central limit theorem, which states the sample mean has a normal distribution when

assumptions are met (Mooney & Duval, 1993). Percentile bootstrap confidence intervals are produced based on the values between which 95% of the bootstrap sample estimates lie (Field, 2013). Pearson's with bootstrapping was used to calculate correlations between continuous variable[11]. Cramer's V was utilised for calculating correlations between categorical variables (Field, 2009). Eta with bootstrapping was used for calculating correlations between continuous and categorical variables.

3.3 Descriptive analysis

3.3.1 Sample

One hundred a fifty male prisoners from two prisons in South Wales participated. Just over half of the participants were recruited from a state run urban local prison (52.7%) and just under half were drawn from a privately run resettlement prison.

3.3.2 Demographics

Age ranged from 18 to 81 years (M=31.7, SD=10.8). As illustrated in Table 8, participants were predominantly White (92%). English was the first language of 96% of the sample. For just over 30% of the sample this was their first time in custody. Ten per cent of the sample were classified as vulnerable prisoners[12] (all vulnerable prisoners were recruited from the wings specifically for sex offenders at the resettlement prison). Ten participants (6.7%) were veterans who had previously served in the armed forces.

Only prisoners who had been received into the establishment in the six months before data collection were recruited. Data was not collected regarding individuals' duration in custody across establishments.

[11] Field (2009) has also recommended the use of Kendall's tau for calculating correlations in non-parametric data. Kendall's tau correlations between continuous variables and bootstrapped Person correlations produced identical results; Pearson's correlations are reported.

[12] Vulnerable prisoners are those at risk of bullying (including sex offenders, police informants and ex-officials) and those at risk of suicide or self-harm (Gov.UK, 2015).

Table 8 Demographics

Characteristic	***n***	**%**
Ethnicity		
White British	134	89.3
White other	4	2.7
Black British	1	0.7
Black other	2	1.3
Asian British	4	2.7
Asian other	1	0.7
Mixed	2	1.3
Other	2	1.3
Prison		
Local urban prison	79	47.3
Resettlement prison	71	52.7
Foreign nationals	4	2.7
First time in custody	46	30.7
Vulnerable prisoners	15	10.0
Veterans	10	6.7
Length of stay		
Less than 2 days	4	2.7
2 days less than 5 days	17	11.3
5 days less than 7 days	13	8.7
7 days less than 14 days	17	11.3
14 days less than 1 month	20	13.3
1 month – less than 3 months	44	29.3
3 – 6 months	35	22.6

3.3.3 Sample offence and custodial profile

The offence and custodial profile of the sample is presented in table 9. More than half of the sample (55.3%) were sentenced prisoners, 28.7% of participants were on remand. The sample captured individuals across a range of sentence lengths, including a small number of individuals serving life

sentences. The most commonly reported index offence category was violence against the person (28%), followed by burglary (20%) and drugs offences (14%). Eight per cent had been convicted for sexual offences.

Table 9 Status, sentence length and index offence

Characteristic	*n*	%
Status		
Remanded	43	28.7
Sentenced	83	55.3
Convicted un-sentenced	2	1.3
Recalled	22	14.7
Sentence length[13]		
Less than 6 months	16	19.3
6 months – less than 1 year	13	15.7
1 year – less than 2 years	9	10.8
2 years – less than 4 years	20	24.1
4 years to less than 10 years	16	19.3
10 years or more	5	6.0
Life	3	3.6
Indefinite Public Protection	1	1.2
Index offence[14]		
Violence against the person	42	28.0
Burglary	30	20.0
Drugs offences	21	14.0
Sexual offences	12	8.0
Theft or handing stolen goods	12	8.0
Other	11	7.3
Robbery	9	6.0
Fraud and forgery	5	3.3
Arson and criminal damage	3	2.0
Possession of weapons	3	2.0
Public order offences	1	0.7
Civil offences	1	0.7

[13] For sentenced participants only

[14] Where more than one index offence was reported (*n*= 9) only the most serious was coded.

3.3.4 Predictors of mental health problems in prison

Research has identified a number of factors that are correlated with mental health difficulties in prison (Birmingham *et al*., 2000; Greenberg & Rosenheck, 2008). Self-report data concerning these factors were collected in the current study and are presented in table 10.

Table 10 Historic predictors of mental health problems in prison

Predictor variable	**n**	**%**
Homeless in the year before prison	28	18.7
Drug problem in the year before prison	63	42
Alcohol problem in the year before prison	39	26
Ever had contact with mental health services	70	46.7
Being prescribed an antidepressant before prison	63	42.0
Ever self-harmed	40	26.7

Correlational analysis (see table 17) indicated that ever having had contact with mental health services, being prescribed an antidepressant before prison and having ever self-harmed (Birmingham *et al*., 2000) were all significantly correlated with all measures of current mental health and distress supporting hypothesis three. Specifically, there were significant correlations between mental health service contact, antidepressants and self-harm on the one hand and MINI 6.0 SMI (*V*=.30, .24, .17 respectively), any current mental health disorder (*V*=.41, .26, .27 respectively), and lifetime mental health disorder (*V*=.35, .27, .33 respectively), CORE-10 score (*ε*.= .40,.28,.42 respectively), GHQ-12 (*ε*.=.32, .21, .33 respectively) score and referral decision (*V*=.49, .35, .41 respectively) on the other hand (all *p*'s. <.05). Being homeless and having substance misuse problems in the year before prison (Greenberg & Rosenheck, 2008) were correlated with some, but not all, mental health and distress variables measured. Homelessness was significantly correlated with MINI 6.0 current SMI classification (*V*=.21) and referral decision (*V*=.21) (*p's*. <05), but was not significantly correlated with CORE-10, GHQ-12 scores, MINI 6.0 any or lifetime current disorder. Drug problems were correlated with

CORE-10 score (ε. =.21) and MINI 6.0 any current mental health disorder (*V*= .23), but not GHQ-12 score, MINI 6.0 current SMI or mental health service referral. Alcohol problems were significantly correlated with MINI 6.0 lifetime mental health disorder (*V*= .18) and referral decision (*V*= .25) (*p's* <.05), but were not significantly correlated with CORE-10, GHQ-12 scores and MINI current SMI or any current disorder.

3.3.5 Test scores

3.3.6 CORE-10

CORE-10 scores were available for all 150 participants. CORE-10 scores ranged between 0 and 36 with a mean of 12.4 (SD 8.7, 95% CI 11.0-13.8) as shown in Table 11. When compared to Connell & Barkham's (2007) CORE-10 general population validation study, the current sample's mean was significantly lower (*M*=12.4, SD=8.8) than that of males in primary care (*M*=18.5, SD=8.1)(*t*(226)=-7.61, *p* <.05) but significantly higher than that of males in the general population (*M*=4.8, SD=4.6) (*t*(195)=9.85, *p*<.05).

Table 11 CORE-10 mean scores

Sample	CORE-10 Mean (SD)	Mean 95% CI
Current prison population (n =150)	12.4 (8.8)	11.0-13.8
Male Primary care (*N*=516) (Connell & Barkham, 2007)	18.5 (8.1)	17.8-19.2
Male General population (*N*=268) (Connell & Barkham, 2007)	4.8 (4.6)	4.3-5.4

Forty-nine per cent of participants' CORE-10 scores fell within the non-clinical range (≤ 10) based on traditional cut points proposed by Connell & Barkham (2007). Of those in the non-clinical range, 28% of scores fell into the healthy range (≤ 5) and 21% fell into the low level range (6-≤ 10). Fifty one per cent of CORE-10 scores fell within the clinical range (11-40). Within the clinical range, 11% of scores were classified as mild (11-14), 18% moderate (15-19), 11% moderate-severe (20-24), and 11% fell in the severe range (≥25), as shown in Table 12.

Table 12 CORE-10 score by severity

	Severity	CORE-10 cut off score	*n* (%)
Non clinical	Healthy	≤ 5	42 (28%)
	Low level	≤ 10	31 (21%)
Clinical	Mild	11-14	17 (11%)
	Moderate	15-19	27 (18%)
	Moderate to severe	20-24	17 (11%)
	Severe	≥25	16 (11%)
	Total		**150 (100%)**

Mean CORE-10 scores according to grouping by MINI 6.0 classification of current SMI, any current mental health disorder or no disorder are presented in Table 13. Confirming hypothesis one, mean CORE-10 scores were positively correlated with severity of mental health need (ε. =.57, p <.05). CORE-10 scores were highest amongst those with a current SMI (M=17.8, SD 7.5) and lowest amongst those with no current disorder (M=6.3, SD 5.4). Although mean CORE-10 scores amongst those with a SMI were slightly higher than mean scores amongst those with any current disorder (excluding an SMI) the difference was not significant (F (1, 90) =2.52, p >.05). However there was a significant difference in mean CORE-10 scores between those with any current disorder and those with no current disorder (F (1,106) =42.4, p <.0001). These findings suggest the CORE-10 has good criterion validity in distinguishing between individuals with, and individuals without, a disorder. However, the CORE-10 is less sensitive to difference between differing levels of mental health need amongst those with a disorder.

Table 13 Mean CORE-10 score by MINI diagnosis classification

Sample	CORE-10 Mean (SD)	Mean 95% CI
MINI 6.0 current SMI (n=41)	17.8 (7.5)	15.4-20.2
MINI 6.0 any current disorder[15] (n=51)	15.1 (8.7)	12.6-17.5
MINI 6.0 no current disorder (n=58)	6.3 (5.4)	4.9-7.7

3.3.7 GHQ-12

GHQ-12 scores were available for all 150 participants. GHQ-12 scores ranged between 0 and 12 with a mean of 4.06 (SD 3.6). Mean GHQ-12 scores according grouping by MINI 6.0 classification of current SMI, any current mental health disorder or no disorder are presented in Table 14. Confirming hypothesis two, mean GHQ-12 scores were positively correlated with severity of mental health need (ε. =.56, p <.05). Mean GHQ-12 scores were highest amongst those with a current SMI (M=6.0, SD 3.7) and lowest amongst those with no current disorder (M=1.5, SD 2.1). Although mean GHQ-12 scores amongst those with a SMI were slightly higher than mean scores amongst those with any current disorder (excluding an SMI) the difference was not significant (F (1, 90) =.79, p >.05). However, there was a significant difference in mean GHQ-12 scores between those with any current disorder and those with no current disorder (F (1,106) =54.6, p <.0001). These findings suggest the GHQ-12, like the CORE-10, has good criterion validity in distinguishing between individuals with and individuals without a disorder. However, the GHQ-12 is also less sensitive to difference between differing level of mental health need within the disorder group.

[15] Any current mental health disorder excluding current SMI

Table 14 Mean GHQ-12 score by MINI diagnosis classification[16]

Sample	GHQ-12 Mean (SD)	Mean 95% CI
Grand mean (n=150)	4.06 (3.6)	3.5-4.6
MINI 6.0 current SMI (n=41)	6.0 (3.7)	4.8-7.2
MINI 6.0 any current disorder[17] (n=51)	5.4 (3.3)	4.4-6.3
MINI 6.0 no current disorder (n=58)	1.5 (2.1)	0.9-2.0

3.3.8 MINI 6.0 classifications

MINI 6.0 classifications were examined in terms of those screening positively for:

- Current SMI: a positive screen in the last month for current major depression and/or bipolar (i) and/or bipolar (ii) and/or bipolar disorder not otherwise specified and/or mood disorder with psychosis and/or psychotic disorder.
- Any current mental health disorder: a positive screen for any of the following in the last month - major depression episode, suicidality, manic episode, major depressive disorder and/or bipolar (i) and/or bipolar (ii) and/or bipolar disorder not otherwise specified and/or mood disorder with psychosis and/or psychotic disorder, panic disorder (with and without agoraphobia), agoraphobia, social phobia, obsessive compulsive disorder, post traumatic distress disorder and generalised anxiety disorder.
- Any lifetime mental health disorder – a positive screen for any of the above at any time point

Table 15 indicates that according to the MINI 6.0, 27.3% of the sample screened positive for a current SMI, 61.3% screened positive for any current

[16] The samples scores are not compared against community population data since studies using the GHQ-12 typically only report the proportion of people falling above a given cut off without reporting means and the cut-offs applied are highly variable preventing meaningful comparison.

[17] Any current mental health disorder excluding current SMI

mental health disorder and 77.3% screened positively for a mental health disorder in their lifetime.

Table 15 MINI 6.0 screening results

MINI 6.0 screen	***n* (%)**
Current SMI	41 (27.3%)
Any current mental health disorder	92 (61.3%)
Any lifetime mental health disorder	116 (77.3%)

In terms of MINI sucidality scores, 40.7% (*n*=61) did not score on the suicidality scale. Thirty four per cent (*n*=51) fell within the low MINI suicidality range (score 1-8), 8% (*n*=12) scored within the moderate MINI suicidality range (score 9-16) and 17.3% scored within the high MINI suicidality range (*n*=26).

3.3.9 Referral decisions

Referrals for mental health services in the prisons can be made by health care staff during reception screening or subsequently by custodial or other prison staff. During their current stay at the respective establishments, 61.3% of the sample had not been referred for any mental health service, 34.7% had been referred for a primary care mental health service and 3.3% had been referred for a secondary care mental health service. Data was not available for one individual as indicated in table 16.

Table 16 Proportions of participants referred for mental health services

	***n* (%)**
No referral	92 (61.3%)
Primary care referral	52 (34.7%)
Secondary care referral	5 (3.3%)
Unknown	1 (0.7%)

3.3.10 Correlational analysis

In order to test correlational hypotheses, correlations between all key variables were calculated as shown in Table 17. Supporting hypothesis four, there was a significant positive correlation between CORE-10 and GHQ-12 scores (r= .728, p <.001).

Supporting hypotheses five, six and seven there were significant relationships between CORE-10 scores and referral decisions (ε=.363), MINI SMI (ε=.385), MINI any current disorder (ε=.557) and MINI lifetime disorder classification (ε=.456). These significant correlations suggest good convergent validity of the CORE-10 with existing clinical decision-making, as well as other established structured assessments of mental health.

There was also a significant correlation between CORE-10 scores and MINI suicidality score (r=.559; all p's <.001). Further exploration of this relationship was undertaken using two between subject analyses of variance (ANOVA). The first between subjects analysis of variance (ANOVA)[18] with the CORE-10 risk item as the dependant variable revealed a significant difference in scores on the CORE risk item between subjects in different MINI suicidality risk categories (F (3,146)=18.043, p <.001). Bonferroni adjusted post hoc tests confirmed that scores on the CORE risk item differed significantly between those with a high MINI suicidality categorisation and those in any of the other categories (none, low and moderate suicide risk). A second between subjects analysis of variance (ANOVA)[19] with the CORE-10 sum score as the dependant variable also revealed a significant difference in CORE-10 scores between subjects in different MINI suicidality risk categories (F (3,146)=36.32, p <.001). Bonferroni adjusted post hoc tests confirmed that CORE-10 scores were significantly lower amongst those in the no risk category compared to any other MINI category of suicide risk. CORE-10 scores did not differ significantly between those in the low and medium or between those in the

[18] Bootstrapping was applied to account for the non normal distribution of the CORE risk item scored and the CORE-10 scores.

[19] Bootstrapping was applied to account for the non normal distribution of the CORE risk item scored and the CORE-10 scores.

medium and high MINI suicidality categories. CORE-10 scores for those in the high MINI suicidality category differed significantly from those in the none and low categories. These findings further support the convergent validity of the CORE-10 risk item with structured assessments of suicidal risk.

The CORE-10 also appeared to have good discriminant validity (hypothesis eight), as there were no significant correlations between CORE-10 scores and variables not typically linked to mental health or psychological distress. Specifically, CORE-10 scores were not correlated with age, ethnicity, prison establishment, sentencing status, sentence length or offence.

Table 17 Bivariate correlations between main study variables

	GHQ-12	MINI SMI	MINI current	MINI lifetime	MH Referral	Suicidality score	Homeless	Drugs	Alcohol	MH service	Antidepressants	Self-harm	Veteran	1st sentence	Age	Ethnicity	Prison	Status	Sentence length	Offence
CORE-10	.728**	.385**	.557**	.456**	.363**	.559**	.093	.213*	.158	.395**	.283*	.419**	.129	.025	-.036	.328	.158	.178	.234	.256
GHQ-12	1	.335**	.551**	.414**	.233**	.490**	.053	.152	.036	.329**	.209**	.328**	.122	.009	.031	.256	.137	.162	.242	.178
MINI SMI		1	.487**	.332**	.195	.274**	.205*	.145	.148	.296**	.236**	.171*	.076	.181*	.055	.210	.132	.205	.191	.277
MINI current			1	.682**	.379**	.420**	.134	.232**	.159	.413**	.260**	.293**	.212**	.184*	.012	.222	.097	.221	.247	.247
MINI lifetime				1	.363**	.323**	.096	.170*	.176*	.347**	.267**	.326**	.174	.227**	.149	.215	0.93	.235*	.301	.282
MH Referral					1	.388**	.213*	.115	.254**	.488**	.345**	.413**	.215*	.206**	.057	.153	.115	.174	.305**	.288
Suicidality score						1	.269**	.245**	.208*	.440**	.416**	.570**	.236**	.072	.081	.261	.137	.174	.239	.270
Homeless							1	.216**	.223**	.066	.216**	.137	.146	.244**	.113	.174	.077	.180	.250	.351
Drugs								1	.389**	.206*	.234*	.220**	.119	.185*	.082	.186	.049	.161	.260	.315
Alcohol									1	.177*	.327**	.261**	.085	.229**	.009	.211	.014	.078	.153	.268
MH service										1	.395**	.403**	.232**	.158	.054	.275	.164	.241	.258	.184
Antidepressants											1	.312**	.097	.185*	.002	.243	.005	.139	.139	.239
Self-harm												1	.141	.172*	.025	.230	.032	.134	.368	.292
Veteran													1	.004	.129	.337*	.093	.311**	.259	.251
1st sentence														1	.013	.234	.080	.330**	.308	.438**
Age															1	.155	.118	.177	.408**	.360*
Ethnicity																1	.232	.153	.195	.307
Prison																	1	.371**	.677**	.521**
Status																		1	.584**	.360**
Sentence length																			1	.358**

Note: * p<.05, **p<.001

3.4 Statistical analysis

3.4.1 ROC analysis

There is no consensus regarding what constitutes acceptable performance for mental health screening tools in custody (Martin *et al*., 2013). Screening can improve identification of people with mental health problems, but often results in high false positive rates (Brooker *et al*., 2009). Some argue that for screening purposes a high false positive rate is preferable (Evans *et al*., 2010), however very high false positive rates can result in inefficient use of scarce resources (Hart *et al*., 1993; Martin *et al*., 2013; Steadman *et al.,* 2005).

A series of ROC analyses were conducted in order to measure the predictive validity of the CORE-10 and GHQ-12 in distinguishing between those with no current mental health problem, MINI 6.0 any current mental health disorder (primary level) and those with a MINI 6.0 classification of severe mental illness (secondary level).

The tools' sensitivity (rate of true positive-predictions) was plotted against its specificity (the rate of true negatives) across a range of scores in order to produce a ROC curve. An area under the curve (AUC) of 0.5 signifies a model for which prediction is no better than chance, whereas an AUC of 1.0 signifies a perfect predictive model. Sensitivity and specificity across a range of scores on the tools were inspected to establish optimal cut points (based initially on maximising overall accuracy rather than prioritising sensitivity or specificity) for predicting primary and secondary level need.

3.4.2 Primary level need

3.4.2.1.1 Psychometric tools

In order to test hypothesis nine that the CORE-10 will have superior sensitivity and specificity than the GHQ-12 for identifying prisoners with mental health needs at the primary care level, two ROC analyses were

conducted against MINI 6.0 any current mental health diagnosis classification.

The ROC curve of the CORE-10 against the MINI 6.0 any current mental health diagnosis produced a significant AUC of 0.85 (95% CI, 0.78-0.90; z=10.58, *p* <.0001). This is comparative to the significant GHQ-12 AUC of 0.84 (95% CI, 0.77-0.89; z=10.54, *p* <.0001). The CORE-10 produced a slightly larger AUC than the GHQ-12 for identifying those at the primary care level although the difference was not significant (z=0.30, SE=0.03, *p* >.05) thus partially confirming hypothesis nine as shown in Table 18 and Figure 2.

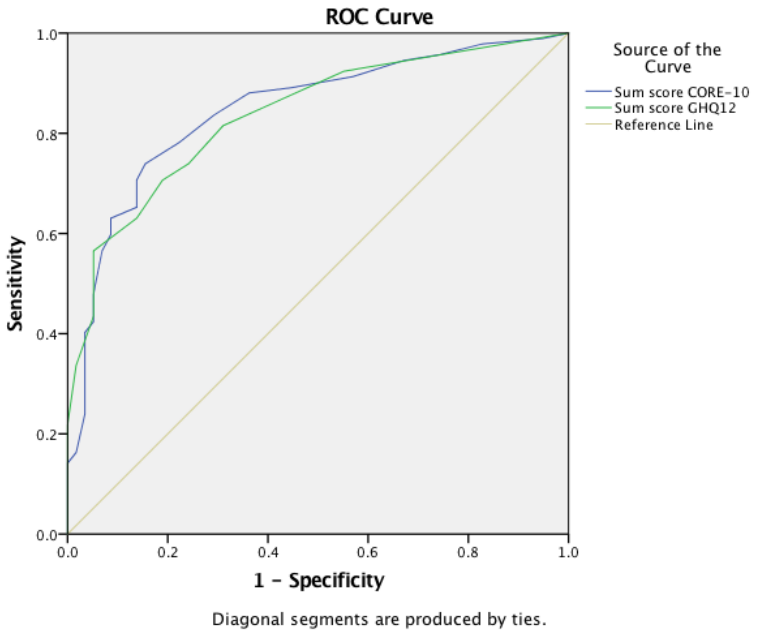
Statistically, the optimal cut point for identifying primary care level mental health problems on the CORE-10 was >6 (see appendix 11), producing a sensitivity of 88 (95% CI, 80-94) and a specificity of 64 (95% CI, 50-76). The optimal cut point for identifying primary care level mental health problems on the GHQ12 was >1, giving a sensitivity of 82 (95% CI, 72-89) and specificity of 69 (95% CI, 57-82). Overall the CORE-10 was more sensitive but less specific in identifying those with a current mental health problem at primary care level of need than the GHQ-12.

The CORE-10 positive predictive value (PPV) (the proportion of those who screen positively who have a current disorder) was 79 (95% CI, 70-87) and the negative predictive value (NPV) (the proportion of those who screen negatively who do not have a current disorder) was 77 (95% CI, 63-88). The PPV and NPV of the GHQ-12 were 81 (95% CI, 71-88) and 70 (95% CI, 57-82) respectively. Both tools had higher PPV than NPV thus most true positive cases are detected.

Table 18 ROC for the CORE-10 and GHQ-12 against MINI 6.0 any current mental health diagnosis

Measure	AUC	Sensitivity (CI)	Specificity (CI)	PPV (CI)	NPV (CI)
CORE-10 (Cut >6)	0.85*** (0.78-0.90)	88.04 (79.6-93.9)	63.79 (50.1-76.0)	79.4 (70.2-86.8)	77.1 (62.7-88.0)
GHQ-12 (Cut >1)	0.84*** (0.77-0.89)	81.52 (72.1-88.9)	68.97 (55.5-80.5)	80.6 (71.1-88.1)	70.2 (56.6-81.6)

Figure 3 ROC for the CORE-10 and GHQ-12 against MINI 6.0 any current mental health diagnosis



3.4.2.1.2 Referral decisions

In order to test the hypothesis that psychometric tools will offer superior accuracy in screening for primary level mental health problems in prisoners than referral decisions, referral decisions were compared against the MINI 6.0 any current mental health disorder variable. Of those with a MINI 6.0 current mental health disorder, 52.7% were referred for a mental health service, however 47.3%, despite screening positive for a current disorder, were not referred for a service as shown in Table 19. This indicates that

while most true negatives were correctly not referred, many true positives were missed. Comparatively, the sensitivity of referral decisions (53%) was substantially lower than the sensitivity of the CORE-10 (88%) and GHQ-12 (82%). Nevertheless, there was a statistically significant association between referral decisions and MINI 6.0 any current disorder (*V* (1) =.37, p <.05).

Table 19 Comparison of referral decisions against MINI 6.0 any current disorder

		MINI 6.0 any current mental health disorder	
		No	Yes
Referred for a mental health service[20]	No	49 (84.5%)	43 (47.3%)
	Yes	9 (15.5%)	48 (52.7%)

3.4.3 Secondary care level need

3.4.3.1.1 Psychometric tools

In order to test hypothesis ten, that the CORE-10 will have superior sensitivity and specificity than the GHQ-12 for identifying prisoners with mental health needs at the secondary level (SMI) two ROC analyses were conducted against MINI 6.0 SMI classification.

The ROC curve of the CORE-10 against the MINI 6.0 SMI classification produced a significant AUC of 0.76 (95% CI, 0.68-0.83; z=6.34, *p* <.0001). The GHQ-12 produced a significant AUC of 0.70 (95% CI, 0.62-0.78; z=4.23, *p* <.0001). The CORE-10 produced a slightly larger AUC than the GHQ-12 for identifying those at the secondary care level (SMI) although the difference was not statistically significant (z=1.49, SE=0.04, *p* >.05), thus partially confirming hypothesis ten as shown in table 20.

[20] This includes referrals for primary and secondary care service as the MINI 6.0 any current mental health disorder variable included current SMI.

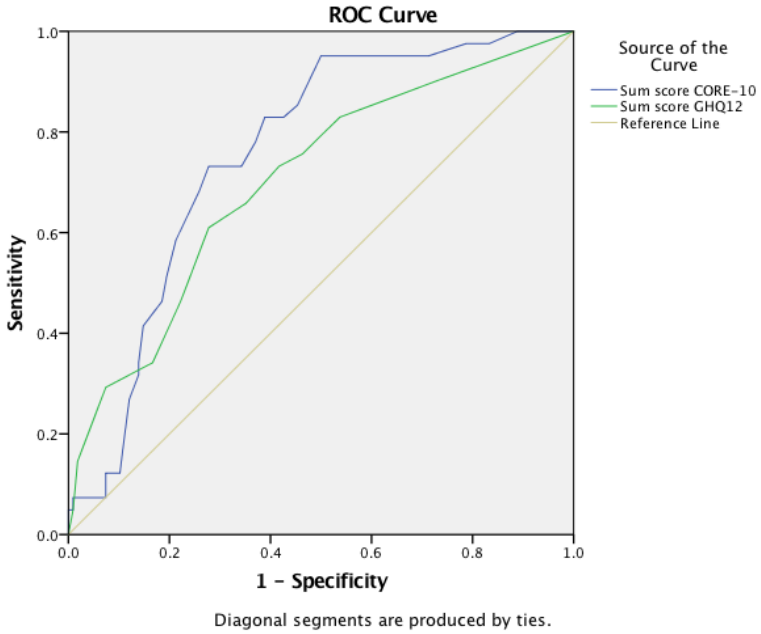
Statistically, the optimal cut point for identifying secondary care level SMI on the CORE-10 was >10 (see appendix 11), producing a sensitivity of 83 (95% CI, 68-93) and a specificity of 61 (95% CI, 51-70). The optimal cut point for identifying secondary care level SMI on the GHQ12 was >3, giving a sensitivity of 73 (95% CI, 57-86) and specificity of 58 (95% CI, 48-67) as shown in figure 3. Overall the CORE-10 had better sensitivity and specificity than the GHQ-12 for identifying those with secondary care level SMI.

The CORE-10 positive predictive value (PPV) (the proportion of those who screen positively who have a current SMI) was 44 (95% CI, 33-56) and the negative predictive value (NPV) (the proportion of those who screen negatively who do not have a current SMI) was 90 (95% CI, 81-96). The PPV and NPV of the GHQ-12 were 39 (95% CI, 28-51) and 85 (95% CI, 75-92) respectively. Both tools had higher NPV than PPV on this occasion, reflecting their lower specificity.

Table 20 ROC for the CORE-10, GHQ-12 and referral decisions against MINI 6.0 SMI classification

Measure	AUC	Sensitivity (CI)	Specificity (CI)	PPV (CI)	NPV (CI)
CORE-10 (Cut >10)	0.76*** (0.68-0.83)	82.93 (67.9-92.80)	60.55 (50.7-69.8)	44.1 (32.8-55.9)	90.4 (81.3-96.1)
GHQ-12 (Cut >3)	0.70*** (0.62-0.78)	73.17 (57.1-85.8)	57.80 (48.0-67.2)	39.4 (28.4-51.3)	85.2 (75.0-92.4)

Figure 4 ROC for the CORE-10 and GHQ-12 against MINI SMI classification



3.4.3.1.2 Referral decisions

In order to test the hypothesis that psychometric tools will offer superior accuracy in screening for secondary care level mental health problems in prisoners than referral decisions, referral decisions were compared against the MINI 6.0 current SMI variable.

Of those with a MINI 6.0 current SMI, 5% were referred for a secondary care mental health service, 95% of cases that screened positive for a SMI were not referred for a secondary care mental health service as shown in Table 21. The Table demonstrates that while most true negatives were correctly not referred, a very high numbers of true positives were missed. There was no statistically significant association between referral to secondary care and the MINI 6.0 current SMI variable (V (1) =.05, p >.05). Comparatively, the sensitivity of secondary care referral decisions (5%) was substantially lower than the sensitivity of the CORE-10 (83%) and GHQ-12 (73%).

Table 21 Comparison of referral decisions against MINI current SMI

		MINI 6.0 current SMI	
		No	Yes
Referred for a mental health service[21]	No	104 (97.2%)	39 (95.1%)
	Yes	3 (2.8%)	2 (4.9%)

3.5 Reliability

3.5.1 Internal reliability

Internal reliability of the CORE-10 and GHQ-12 at time one and two was calculated using Cronbach's alpha (α). Internal reliability was good (Kline, 1999) for both the CORE-10 (time 1 α.84, time 2 α.89) and GHQ-12 (time 1 α.87, time 2 α .89).

3.5.2 Test re test reliability

Two-week test re-test reliability was calculated using Interclass Correlation Coefficients (ICC) since this accounts for both consistency in performance from test to re-test as well as average performance over time (Vaz et *al*., 2013). As such ICC is considered more appropriate than Pearson's correlation coefficient (which measures linear association) when looking at fluid constructs like psychological distress which vary somewhat over time and can be open to systematic error (Mitogen & Li-Hua, 2002). Two week test-re-test ICC's for the CORE-10 (ICC=.83, p <.001) was moderate and for the GHQ-12 (ICC=.71 p <.001) was low, suggesting relative stability in responding over time (Vincent, 1999). The test re test correlations using Pearson's correlation coefficient, as anticipated, suggested lower test retest reliability (CORE-10 r=.71 (p <.001); GHQ-12 r=.56 (p <.001)).

[21] This includes referrals for primary and secondary care service as the MINI 6.0 any current mental health disorder variable included current SMI.

3.6 CORE-10 Confirmatory Factor analysis

A Confirmatory Factor Analysis (CFA) was conducted to test the final hypothesis that the factor structure of CORE-10 within a prison population would conform to the original six factor structure proposed by Connell & Barkham (2007). The original six factor solution was contrasted with the two factor solution (high vs. low intensity items) and a single factor solution as show in Table 22 (see appendix 12 for modelled factor solutions). All three models demonstrated good fit with non-significant Bollen-Stine Bootstrap p-values, supporting the null hypothesis that the hypothesised models fits the data. The six factor solution however provided a significantly better fit (as indicated by significantly better X^2 statistics) with the data than both the two and single factor solutions.

The six factor model showed good fit with a non-significant Bollen-Stine bootstrap p-value. The six factor model also had an acceptable Goodness of Fit Index (GFI), acceptable Comparative Fit Index (CFI) and acceptable Normed Fit Index (NFI) all of which were above .95, as recommended by Bryne (2010) to represent good model fit. The six factor model had an RMSEA of .06 indicating a good fit between the hypothesised model and the data (Hu & Bentler, 1999). The two factor and single factor models on the other hand despite having non-significant Bollen-Stine bootstrap p-values, failed to meet acceptable levels of model fit in terms of CFI, NFI and RMSEA. Furthermore the six factor model has the lowest Akaike's information criterion value, suggesting a better fit with the six factor hypothesised model. The results support hypothesis 12 and support the construct validity of the CORE-10 within prison populations.

Table 22 CORE-10 Confirmatory Factor Analysis fit indices

Fit indices									Comparison with six factor model	
	B-S *p* value	χ^2	df	GFI	CFI	NFI	RMSEA	AIC	$\Delta\chi^2$	df
Six factor model	.251	15.90	11	.97	.98	.95	.06	49.90		
Two factor model	.162	58.70	35	.92	.94	.88	.07	100.70	42.80*	24
Single factor model	.126	63.94	34	.91	.94	.87	.07	103.9	48.04*	23

Note. B-S p value= Bollen-Stine p value; df=degrees of freedom; GFI=goodness-of-fit index; CFI=comparative fit index; NFI=normed fit index; RMSEA=root-mean-square error approximation; AIC=Akaike's information criterion

3.7 Summary of results

As predicted, mean CORE-10 and GHQ-12 scores were highest amongst those with a current SMI and lowest amongst those with no current disorder. The difference between mean scores for those with a SMI (secondary care level) and those with any other current disorder (primary care level) was however non-significant. Scores on both the CORE-10 and GHQ-12 did however differ significantly between those with any current disorder and those with no current disorder, suggesting good criterion validity for both tools.

The CORE-10 demonstrated good convergent validity with established structured assessments of mental health and clinical decision-making. Significant positive correlations were identified between the CORE-10, GHQ-12, all MINI variables including suicidality, referral decisions and variables predictive of mental health problems in prison. The CORE-10 also appeared to have good discriminant validity as demonstrated by non-significant correlations between CORE-10 scores and variables not typically linked to mental health or psychological distress.

ROC analysis on the CORE-10 against the MINI 6.0 revealed significant areas under the curve, which were comparable to those produced by the GHQ-12 for predicting both primary (AUC .85) and secondary care (AUC .76) levels of mental health need. The areas under the curve suggested that the CORE-10 has moderate accuracy in distinguishing between groups of prisoner with no mental health needs, primary care and secondary care level needs (Steiner & Cairney, 2007). At a cut point of >6 for primary care level need the CORE-10 had sensitivity of .88 and specificity of .64. With a cut point of >10 for secondary care level need, the CORE-10 had sensitivity of .83 and specificity of .61. The sensitivity of the CORE-10 substantially exceeded that of exiting referral decisions for both primary and secondary level needs. These findings support the use of the CORE-10 as a screening tool for mental health problems in prisoners.

The CORE-10 had good internal reliability (α .84-.89) and moderate two-week re-test reliability (ICC=.83) within the sample of prisoners. Confirmatory factor analysis confirmed the CORE-10's original six-factor model to be a good fit, with acceptable fit across all indices. The six factor model was a significantly better fit than the two factor and one factor models. Taken together, these findings suggest the CORE-10 is a reliable and valid tool for use with prisoners.

As such it can be concluded that all hypotheses were confirmed, with the exception of the CORE-10 having comparable rather than superior predictive accuracy compared to the GHQ-12 in screening for primary and secondary care level mental health needs in prisoners.

4 Discussion

4.1 Overview

The current study explored the utility of the CORE-10 as a screening tool for common and severe mental health problems in prisoners. This was achieved by comparing the psychometric properties of the CORE-10 to another brief screening tool validated with prisoners - the GHQ-12, a gold standard diagnostic interview – the MINI 6.0 and existing practice in terms of referral decisions. Effective screening for mental health problems to identify level of need is imperative for accurately targeting resources in line with a stepped care or matched care model of mental health service provision. The availability of valid and reliable instruments to measure prisoners' mental health is essential for effective identification and treatment, which in turn can support wellbeing and rehabilitation as well as broader institutional outcomes. The current chapter discusses the key findings of the study in the context of previous research, outlines the clinical and service implications, and considers the limitations of the research and areas for further development.

4.2 Summary of findings

This is the first study to explore the utility of the CORE-10 as a screening tool for common and severe mental health needs in prisoner populations. The primary aim of the study was to establish and compare the sensitivity, specificity and predictive value of the CORE-10 in screening for common (primary care level) and severe (secondary care level) mental health needs in prisoners. ROC analysis on the CORE-10 against the MINI 6.0 revealed significant areas under the curve, which were comparable to those produced by the GHQ-12 for predicting both primary and secondary care level of mental health need. The areas under the curve suggested that the CORE-10 has moderate accuracy (AUC .70-.90) in distinguishing between groups of prisoners with no mental health needs, primary care and secondary care level needs (Steiner & Cairney, 2007).

Identified cut-offs on the CORE-10 at primary and secondary care levels of need gave sensitivities of 88% and 83% respectively, with specificities of 64% and 61% respectively. This was not significantly different to the sensitivity and specificity of the GHQ-12. Pintea & Moldovan (2009), however, note that difference which is not significant does not equate to equivalence of two tests and suggest comparison of partial AUC (AUC at a specific value of sensitivity or specificity) is more useful. For screening purposes it is often thought to be advantageous to prioritise sensitivity (Evans *et al*., 2010; Ford *et al*., 2007) because those who meet the screening threshold will normally undergo further detailed assessment before any intervention is applied, and in this way false positives can be detected and eliminated. A sensitivity of at least 80% is suggested for screening purposes (Barnes, 1982; Sharifi *et al*., 2008). Using sensitivity of 80% as a benchmark for comparison would suggest that the CORE-10 is more useful than the GHQ-12 for screening at both primary (sensitivity .88 versus .82) and secondary (sensitivity.83 versus .73) care level.

For primary care level need, the CORE-10 has a positive predictive value of 79%, which was substantially better than accuracy of existing referral decisions (53%). The positive predictive value of the CORE-10 for secondary care level need was lower at 44%, although this was likely affected by the low base rate for SMI (Elwood, 1993). Nevertheless, accuracy of the CORE-10 at secondary care level was again substantially better than existing referral decisions (5%).

A secondary aim of the study was to establish the validity and reliability of the CORE-10 in a prison population. As hypothesised, the CORE-10 demonstrated construct validity with severity of mental health need positively correlated with CORE-10 score. The CORE-10 significantly correlated with other clinical and psychometric indices of mental health suggesting good convergent validity. Furthermore, CORE-10 scores were not significantly correlated with variables not traditionally thought be associated with mental health (for example age, ethnicity, prison establishment and sentencing status), indicating the tool has discriminate

validity within this population. Internal reliability of the CORE-10 was also good, and comparable to that of the GHQ-12.

Confirmatory factor analysis is integral in ensuring tools' proper measurement of proposed constructs within specific populations (Jackson *et al*., 2009). Confirmatory factor analysis confirmed the CORE-10's original suggested six-factor model proposed by Connell & Barkham (2007) to be a good fit within a prisoner population. The six-factor model had acceptable fit across all indices and was a significantly better fit than both the two-factor and one-factor models. This is promising since it confirms the items included in the CORE-10 do reflect latent variables concerning mental health including, depression, anxiety, functioning, physical symptoms, trauma and risk in prisoners.

Overall these findings demonstrate that the CORE-10 is a valid and reliable psychometric tool within prisoner populations, which can reliably be used as a screening tool for assessing common and severe mental health needs and offers improved accuracy compared to existing practice.

4.3 Study findings in the context of past research

4.3.1 Prevalence of mental health problems

4.3.2 CORE-10

Overall 51% of prisoners' CORE-10 scores fell within the clinical range when applying the community clinical cut of score of more than 10 (Connell & Barkham, 2007). The proportion of individuals falling within the clinical range within the community general population and primary care validation study samples was not reported, thus preventing comparison. However, previous research validating the CORE-34 within a prison population found that prisoners formed a distinct group in terms of mean levels of distress, which fell between those of the validation clinical and non-clinical samples (McCloskey, 2001). Consistent with this research, mean scores on the CORE-10 amongst prisoners in the current study were significantly higher than those of the normative general population sample, but significantly

lower than the average of males in primary care counselling in the validation study (Connell & Barkham, 2007). This is not surprising given that imprisonment is thought to be associated with elevated distress (Gibbs, 1991; Haney, 2001; Linquist & Linquist, 1997), therefore it is expected that scores amongst prisoners would exceed those found in the general population. However, reflecting research that has demonstrated that imprisonment does not necessarily result in clinically significant mental health problems (Andersen *et al*., 2000; 2003; Hassan *et al*., 2011; Taylor *at al*., 2010), prisoners' scores remained distinct from those of a clinical sample.

4.3.3 GHQ-12

Previous research using the GHQ-12 in prison populations has not reported mean scores and has applied heterogeneous cut off scores for identifying cases in terms of probable psychiatric morbidity which makes drawing clear comparisons difficult (Boothby *et al*., 2010; Hassan *et al*., 2011; OHRN, 2010; Smith & Borland, 1999). Nevertheless, the prevalence of GHQ-12 caseness in the current sample is lower than previously identified amongst male offenders. At a cut point of four or more, 44% of the current sample would be defined as cases, compared with 68% in McGilloway & Donnelly's (2004) study amongst offenders entering police custody and 59% amongst Boothby *et al's*. (2010) sample entering a local prison. At the higher cut point of seven or more, 29% of the current sample would meet caseness, compared to 33% in Hassan *et al*. (2011) sample of men entering prison. The lower prevalence of those meeting GHQ-12 caseness is not surprising given that the current sample's inclusion criteria was men who had been received into the prison within the preceding six months, whereas all previous studies focused on immediate reception to custody. Indeed, Hassan *et al.* (2011) reported a significant decrease in the number of men meeting GHQ-12 caseness at two month follow up following reception to prison, and several studies have reported improvements in symptoms following entry to prison (Andersen *et al*., 2000; 2003; Taylor *at al*., 2010). As such the lower prevalence of psychiatric morbidity in the current study is likely attributable to the inclusion of men who have been in prison for longer

durations who may have lower levels of distress following time to adjust to the institutional environment.

4.3.4 MINI 6.0

Consistent with previous research, prevalence of mental health disorders as assessed by diagnostic clinical interview was high within the current prisoner sample. Seventy-seven per cent of prisoners screened positively on the MINI for a mental health disorder in their lifetime. Using the MINI, Black *et al*. (2004) and Gunter (2008) found that 81% and 90% of their US prisoner samples respectively had a lifetime disorder. Both studies included substance dependence disorders that were excluded from the current study, which may explain the lower prevalence rate reported here. Prevalence of any current mental health disorder in the current sample was 61%. The OHRN (2010) reported a comparable prevalence for any current mental health disorder of 54% in British prisoners, while Ford *et al*. (2007) similarly reported a prevalence rate of 56% for any current disorder in US prisoners.

In terms of SMI, 27% of the sample screened positively for a current SMI. This is consistent with Grubin's (2002) findings that 28% of male British prisoners screened positive for an SMI using the same diagnostic definition with the Schedule for Schizophrenia and Affective Disorders (SADS) clinical interview. Birmingham & Mulle (2005), and the OHRN (2010) have however reported higher rates of 38% and 41% respectively for current SMI using the SADS with British prisoners. Studies in the US have reported prevalence of SMI in prisoners ranging between 14.5% (Steadman *et al*., 2009) using the SCID, and 54% using the MINI (Black *et al*., 2004), while Evans *et al*. (2010) reported that 54% of prisoners in a New Zealand sample screened positively for an SMI on the MINI. Making meaningful comparison beyond similar British studies is however confounded by use of differing definitions of SMI and temporal periods explored across studies. Nevertheless, the current study supports the consistent finding that rates of SMI are substantially higher within prison populations compared to community populations (Fazel & Danesh, 2002; Fazel & Seewald, 2012).

Within the current sample, 59% of prisoners gained a score on the MINI suicidatily scale, with 34% scoring in the low risk range, 8% scoring in the moderate risk range and 17% scoring in the high risk range. Rates of suicidality were thus considerably higher than those identified by Black *et al.* (2004) who found 31% of their US prisoner sample gained a suicidality score on the MINI, with 7% scoring in the high risk range. Similarly, the rate of prisoners gaining a score for suicidality in the current study was substantially higher than that identified by Grubin *et al.* (2002) who found that 3% of British prisoners in their sample reported feeling suicidal. It is not surprising that the rate of suicidality identified herein is higher than that reported by Grubin given that their study relied on one item, whereas the MINI suicidality scale used in the current study comprises of 12 items which can trigger a score for suicidality, including items relating to hopelessness and suicidal ideation. However, the discrepancy with Black *et al's*. (2004) findings is surprising given that they utilised the same measure. The high rate of suicidality identified warrants further investigation beyond the scope of the current study.

4.3.5 Construct validity of the CORE-10

An important aspect of the construct validity of an instrument within a population is establishing its ability to discriminate between clinical and non-clinical populations (Connell & Barkham, 2007). Consistent with the validation study, amongst prisoners CORE-10 scores were significantly different between those without a current mental health condition and those who screened positively for a current mental health condition. Importantly, scores on the CORE-10 also differed significantly between those with clinically significant suicidality and non-suicidal prisoners. These findings support the first hypotheses and suggest good construct validity for the CORE-10 in a prison population.

Scores did not, however, differ significantly between prisoners with a SMI and those with less severe but clinically significant mental health conditions. This is consistent with research with community samples which found that scores on the CORE-OM did not differ significantly between those in primary

and secondary mental health services, although those in secondary care did score significantly higher on the risk subscale (Barkham *et al*., 2005; Leach *et al*., 2005). Interestingly, the current study replicated the same findings with the GHQ-12, which has already been validated in prison populations, suggesting the CORE-10 has similar discriminatory power. Mean GHQ-12 scores were also highest amongst prisoners with a current SMI and lowest amongst those with no current mental health disorder. GHQ-12 scores differed significantly between those with no disorder and those with any disorder, but did differ significantly between those with an SMI and those with other mental health disorders.

It has been suggested that due to the nature of SMI (which can impact on insight, cognition and affect), patients with an SMI may have difficulty with the process of self-reporting symptoms and therefore scores on self-report measures may not increase in a linear fashion between those with less and more severe illness (Atkinson *et al*., 1997; Barkham *et al*., 2005; Wilde, 1972). Furthermore, Barkham *et al*. (2005) found that self-reported severity ratings on the CORE-OM reflected the immediacy of problems experienced by those in primary care, whereas those in secondary care had problems which were characterised more in term of chronicity, which was better captured by clinicians' assessment. Both the CORE-10 and GHQ-12 focus on immediate symptoms and thus both seem less sensitive in differentiating SMI, although previous research suggest the inclusion of a risk item on the CORE-10 may help to identify those with an SMI. Talking this into account it may be that the CORE-10 is useful in identifying those with clinically significant mental health problems; however combining it with questions pertaining to historical problems, for example those already included in the Grubin tool (Grubin *et al*., 2002) may support differentiation between those with primary care needs and those requiring support for chronic problems through secondary care.

4.3.6 Convergent validly of the CORE-10

As hypothesised the CORE-10 had good convergent validity with other established measures mental health and distress. CORE-10 scores correlated significantly with GHQ-12 score and all MINI diagnostic screening variables. These findings extend previous research validating the CORE-10 in community settings, which have demonstrated strong correlations with established generic measures of mental health including the SCL-90-R, the Brief Symptom Inventory as well as depression specific measures such as the BDI (Connell & Barkham, 2007). Previous research has also demonstrated that the CORE-10's parent measure, the CORE-OM, has comparable accuracy to the PHQ-9 in screening for depression (Gibody *et al*., 2007). This suggests the CORE-10 may also be valuably utilised in prison instead of the PHQ-9, which is available in Welsh prisons but used sporadically (Little, 2013). Strong correlations with all MINI diagnostic screening variables further suggest the CORE-10 is valid in screening for distress across diagnostic categories with prisoners, extending findings of the validation study which demonstrated its convergence with SCID diagnoses of depression (Connell & Barkham, 2007). Importantly, CORE-10 scores also correlated significantly with MINI suicidality score. Given the known association between mental ill health and suicide attempts in prison (Jenkins *et al*., 2005; Rivlin *et al*., 2010; Shaw *et al*., 2004) and the ethical imperative to reduce suicide risk in custody, these findings support the use of the CORE-10 as a brief screen which can support further assessment of risk to self.

In addition to demonstrating good convergent validity with other psychometric measures, the CORE-10 also had good convergence with existing practice in British prison mental health screening. CORE-10 scores correlated significantly with referral decisions made at reception screening suggesting good convergence with clinical judgment. Furthermore, supporting Birmingham *et al*. (2000) and Grubin *et al*. (2002) research, CORE-10 scores correlated significantly with historic items concerning mental health service contact, antidepressant use and self-harm included in the existing prison mental health screening tool. These findings support the

real world validity of the CORE-10 as a screening tool with prisoners, in addition to its psychometric convergence.

Large-scale correlational studies in the US have shown that homelessness and drug and alcohol use are associated with SMI on entry to prison (Greenberg & Rosenheck, 2008; McNeil *et al*., 2005). CORE-10 scores were significantly correlated with drug use, but surprisingly not alcohol use and homelessness. It may be that drug use prior to prison contributes to increased distress scores due to the process of detoxification on entry to prison, or distress as a result of on-going attempts to maintain drug habits once in custody, although the same hypothesis may be expected with regard to alcohol use. Distress as a result of homelessness prior to custody on the other hand may be ameliorated by entry to custody. Interestingly, homelessness and alcohol problems were significantly associated with current SMI and lifetime mental health disorders respectively, and as such may be better predictors of chronic problems over the life course as oppose to current distress as assessed by measures like the CORE-10 and GHQ12. Nevertheless, the CORE-10 demonstrated good convergent validity with both established psychometric and diagnostic tools as well as existing practice in mental health screening in prisons, suggesting it may be a valid tool for mental health screening in prison populations.

4.3.7 Discriminant validity of the CORE-10

Discriminant validity refers to whether measurements that are conceptually unrelated are in fact unrelated (Campell & Fiske, 1959). As hypothesised, the CORE-10 showed good discriminant validity. Non-significant correlations were found between CORE-10 scores and variables not conceptually expected to be associated with psychological distress including age, ethnicity, prison establishment, sentencing status, sentence length or offence. Establishing discriminant validity is important for assessing the usefulness of mental health screens in prisons (Ford *et al*., 2007) and historically screening tools such as the RDS have failed to achieve discriminate validity (Rogers *et al*., 1995). As such the current findings are

promising in terms of suggesting the utility of the CORE-10 as mental health screening tool in prisons.

4.3.8 Predictive validity of the CORE-10 as a screening tool

4.3.9 Primary care level mental health

The current study is the first of its kind to explore the predictive validity of the CORE-10 for identifying prisoners with both primary and secondary care level mental health needs. Findings confirmed the hypothesis that the CORE-10 would have adequate sensitivity and specificity for identifying those with primary care needs, although sensitivity and specificity was only marginally better than the GHQ-12 at identified cut-offs. Despite there being no established consensus regarding acceptable performance for a mental health screening tool in custody (Martin *et al*., 2013), the CORE-10 compared well at primary care level with existing tools. The CORE-10 achieved a sensitivity of 88% and specificity of 64% for identifying those with primary care level needs. When compared to the only other tool designed to broadly screen for mental health problems, the CMHS-M (Ford *et al*., 2007), the CORE-10 had substantially better sensitivity (CORE-10 88% vs. CMHS-M .64-.75), but slightly lower specificity (CORE-10 64% vs. CMHS-M .70). Furthermore, the CORE-10 achieved far better sensitivity and slightly lower specificity than the GHQ-28 (sensitivity .65, specificity .69: Andersen *et al*., 2002), the only other tool measuring psychological distress broadly which appears to have been validated with a prison population. Compared to existing referral decisions at primary care level the CORE-10 also compared favourably with a positive predictive value of 79% compared to the accuracy of 53% for existing referral decisions. As such, these findings suggest that the CORE-10 offers better sensitivity for identifying those with primary care level needs than existing available tools and current practice. However the CORE-10 has poorer specificity, resulting in high false positives, which can place a resource burden in services (Hart *et al*., 1993; Steadman *et al.,* 2005). In the Welsh prison context where initial screens are often not carried out by mental health professionals, but later followed by multi-disciplinary mental health team discussion or further assessment, initial prioritisation of sensitivity may be preferable (Evans *et al*., 2010).

4.3.10 Secondary care mental health

Consistent with hypothesis nine, the CORE-10 had better predictive validity than referral decisions in identifying prisoners with secondary care level mental health needs. For identifying prisoners at secondary care level the CORE-10 had sensitivity of 83% with specificity of 61%, which marginally exceeded the sensitivity and specificity of the GHQ-12. Compared alongside existing tools designed to identify prisoners with SMI, the CORE-10 performed relatively well. The CORE-10 had superior sensitivity than the RDS (.79) (Teplin & Swartz, 1989), the BJMHS (.34-.66) (Evans *et al*., 2010; Steadman *et al*., 2005) and the K6 (Louden et al., 2013), although it was less sensitive than the Grubin (.97) (Grubin *et al*., 2002) and PriSnQuest (.89) (Shaw *et al*., 2003) in their validation studies. In terms of specificity, the CORE-10 was less specific than the RDS (.98) (Teplin & Swartz, 1989), the Grubin (.84) (Grubin *et al*., 2002) and the BJMHS (.74-.86) (Evans *et al*., 2010; Steadman *et al*., 2005), as specific as the PriSnQuest (.61) (Shaw *et al*., 2002) and more specific than the K6 (.36) (Louden *et al*., 2013). The CORE-10 had lower positive predictive value (PPV .44) than the Grubin (.60) for identifying SMI, although was substantially more accurate than existing referral decisions which accurately identified only 5% of those with an SMI. The CORE-10 thus offers adequate sensitivity in identifying SMI when compared to existing available tools although lacks specificity.

The CORE-10's poorer predictive value in predicting SMI compared to the Grubin may reflect its emphasis on current distress as opposed to historic indicators of mental health problems. SMIs generally treated in secondary care are typically characterised in terms of chronicity as oppose to the immediacy of problems (Barkham *et al*., 2005), which the CORE-10 does not account for. As such the CORE-10 may be most useful used in conjunction with the historic indicators contained within the Grubin.

4.3.11 Reliability of the CORE-10

The reliability of the CORE-10 with prisoner populations has not previously been explored. The current findings supported the internal and test-re test reliability of the CORE-10 in this population. Confirming hypothesis ten, the

CORE-10 demonstrated good internal reliability (Cronbach's alpha=.84-.89), which was comparable with that of the GHQ-12. Alphas exceeded those found it its community validation study (.84) (Connell & Barkham, 2007), and supported previous findings demonstrating high internal reliability of its parent measure the CORE-OM in a prison population (Andersen *et al*., 2002).

The CORE-10 also demonstrated moderate retest reliability (Vincent, 1999) with an ICC above .8 suggesting consistency in responding. The retest reliability of the CORE-10 exceeded that of the GHQ-12. Re-test reliability of the CORE-10 in community sample does not appear to have been explored thus preventing comparison. Early in custody prisoners experience significant and variable stressors (Gagliardi *et al*., 2004) that may decrease the re test reliability of screening tests (Ford *et al*., 2007). Inclusion of prisoners in the first stages of custody may thus contribute to findings of moderate re-test reliability for the CORE-10. Nevertheless, Connell & Barkham (2007) propose a change score of six as signifying clinically significant change on the CORE-10: mean scores from time one to time two changed by less than one point lending support to the relative stability of the CORE-10 in this sample[22]. The findings of the current study suggest that the CORE-10 is a reliable psychometric tool for use with prisoners.

4.3.12 Factor structure of the CORE-10

To date the factor structure of the CORE-10 does not appear to have been scrutinised despite the authors hypothesising a six factor and two factor model (Connell & Barkham, 2007). Furthermore, previous research has failed to confirm the original proposed factor structure of its parent measure, the CORE-OM (Bedford *et al*., 2010; Evans *et al*., 2002; Lyne *et al*., 2006). Interestingly, results of the current study supported the original six-factor model for the CORE-10 with good fit indicated across all indices. The six factor model was a significantly better fit than both the two factor and single factor models also tested. This conflicts with the view of some authors that

[22] Sensitivity to change was not assessed in the current study as predictive validity was prioritized due to the research questions.

the notion of domains and sub domains (or factors) is severely threatened when scales are very short (Bedford *et al*., 2010). However, given that the CORE-10 was developed by choosing items based on clinical utility, coverage and statistical procedure (namely selecting items using regression that were most predictive of each CORE-OM domain) it is not surprising that the factor structure of the CORE-10 is more robust than its parent measure.

Some support was lent to Connell and Barkham's (2007) proposal that the CORE-10 items can be divided into high and low intensity domains, as this two factor model did demonstrate acceptable levels of overall fit, but scrutiny of the CFI, NFI and RMSEA measures of fit for this model did not meet acceptable thresholds (Bryne, 2010). Whilst lending some to support to items being dividable into low and high intensity factors, it does not appear that this is the best structure to explain the data in the current population. This is not surprising, given that that CORE-10 scores did not differ significantly between prisoners with primary and secondary care level needs. Furthermore, inadequate fit of the high/low intensity model is consistent with previous research finding that while the CORE-OM adequately reflected low intensity problems in primary care, it was not as good as clinician's judgment in assessing those with severe problems (Barkham *et al*., 2005).

Again the single factor model showed adequate overall fit, but CFI, NFI and RMSEA indices did not meet acceptable thresholds (Bryne, 2010). It may be that this is due to the combination of conceptually distinct theoretical constructs – namely problems/symptoms, functioning and risk (Bedford *et al*., 2010). In clinical practice the combination of these constructs is helpful as the authors intended; however, theoretically they are not unified and therefore loadings onto a unitary factor will be decreased. The CORE-10 does summate items into an overall clinical score which whilst helpful in practice, is not of theoretical value.

Overall, the findings of the CFA suggest the CORE-10 has a robust and distinguishable six-factor structure. Theoretically the findings support the

notion of distinguishable clusters of symptoms relating to specific problems like anxiety, depression, trauma and physical symptoms. As such, the factor structure supports the theoretical model of classifiable psychiatric disorders. It is not surprising that the two factor structure of high and low intensity items was less robust given that the notion of 'high' and 'low' intensity symptoms is derived from ideas surrounding offering services to those with more severe presentations as oppose to an epistemological understanding of mental health. It is more surprising however that the single factor structure proved to be less robust than the six factor structure. The single factor structure is based on the theoretical premise that people with a heterogeneous mental health problems will similarly have elevated scores across a range of common cognitive, emotional and behavioural symptoms which could be described as 'non-specific distress' (Kessler *et al*., 2002). As such this single factor idea assumes a commonality in the experience of mental health problems, which should supersede the notion of specific disorders. The current findings although showing adequate fit for a single factor structure, do not however entirely support this idea since the six discrete factors provide a preferable fit suggesting some notable distinction in experience across these factors. In clinical practice this suggests that while looking at total scores may be initially helpful, scrutinising distinct sub-domains scores is also important when considering what type of intervention may be required or not.

4.4 Clinical and service implications

The primary purpose of validating mental health screening tools for prisoners is to support accurate identification and description of individuals' mental health needs (Grisso *et al*., 2005). Effective screening should enable those with identified needs to access appropriate services (Teplin, 1990) and thus should ultimately promote improved wellbeing amongst screened populations relative to those who are not assessed (Sackett & Haynes, 2002). Despite the promotion of use of structured screening tools in policy (NICE, 2011b; Welsh Government, 2014), tools which can delineate severity of needs in line with current service models have not been validated and are

not widely used in Welsh prisons (Little, 2013). In addressing this gap, the current study aimed to assess the validity and consequent clinical utility of the CORE-10 for screening prisoners' common and severe mental health needs.

4.4.1 Establishing clinically useful CORE-10 cut points

In achieving this aim, the statistical and clinical utility of the CORE-10 at differing cut points to detect primary and secondary care level mental health needs was assessed. While there is no consensus regarding acceptable performance for mental health screening tools in custody (Martin *et al*., 2013), authors writing on screening in other domains have typically suggested tools need to have a minimum sensitivity of 0.8 and a minimum specificity of 0.6 (Barnes, 1982; Bennett & Lincon, 2006; Sharifi *et al*., 2008). Prioritising sensitivity is typically preferred (Evans *et al*., 2010; Ford *et al*., 2007) because those who meet the screening threshold will normally undergo further detailed assessment whereby false positives can be detected and eliminated.

For detecting primary care level need, a cut point of >6 on the CORE-10 offered the best trade of between sensitivity and specificity in statistical terms. The CORE-10s' specificity was relatively poor (64%) compared to the existing practice - referral decisions, which despite having poor sensitivity, were relatively specific (84%). Poor specificity may be problematic in this context since referring high numbers of prisoner who screen positive but do not have a disorder (false positives) may delay treatment of those in greatest need given limited mental health service provision in the prisons (Hart *et al*., 1993; Steadman *et al.,* 2005). However, given that an initial brief screen (such as the CORE-10) should be followed further clinical assessment (which has been shown to be relatively specific) false positives could be filtered out at this point.

In the community the CORE-10 clinical cut-off recommended by Connell & Barkham (2007) is 10. Increasing the CORE-10 primary cut point with prisoners from >6 to >10 improves specificity to 85% (95% CI, 73-93) which

is comparative to the specificity of current referral decisions. However sensitivity reduces to 74 (95% CI, 64-83) at a cut point of >10, which despite still being significantly better than the sensitivity of current referral decisions (53, 95% CI, 42-63) is unacceptably low for a screening tool (Bennett & Lincon, 2006). If a cut point of 10 were applied some 26% of prisoners with common mental health problems would still be missed. As such a clinically useful cut point for identifying prisoners with primary level mental health need of >6 is recommended, with further assessment to filter out false positives.

For detecting secondary care level need, statistically a cut point of >10 on the CORE-10 offered the best trade of between sensitivity (>.8) and specificity (>.6) in statistical terms. However again the CORE-10s' specificity was relatively poor (61%) compared to the existing practice referral decisions, which despite having very poor sensitivity (5%) and therefore limited utility in detection, were very specific (97%). If the cut-point for secondary care was increased to >14 the cut off between mild and moderate clinical ranges in a community context (Connell & Barkham, 2007), specificity would improve to 72% (95% CI, 63-81); however sensitivity would be unacceptably low at 73 (95% CI, 57-86). As such the most clinically useful cut point for identifying prisoners with secondary level need (SMI) on the CORE-10 is likely to be >10. Alternatively, the CORE-10 could be used to identify clinically significant distress without a secondary care cut point, as a primary care cut off would encapsulate individuals with secondary care needs also, with clinical judgment alone informing referral to secondary care. This approach would reflect the findings that there is not a significant difference in CORE-10 scores between those with primary and secondary level needs.

Clinically, administering the CORE-10 with the Grubin items (Grubin *et al*., 2002) which have been found to be highly predictive of SMI may be particularly valuable, although further research is required to establish the predictive validity of the CORE-10 and Grubin combined. From a clinical perspective, attaining information pertaining to historic factors including

previous mental health service contact and medication use is imperative in order to promote continuity of care (Byng *et al*., 2012) and meet legal requirements in terms of Care and Treatment Planning and re-assessment imposed by parts 2 and 3 of the Mental Health Measure (OHRN, 2008; Welsh Government, 2010; 2014). At the same time, introducing a measure focusing on presenting problems like the CORE-10 addresses criticisms raised regarding the solely historical emphasis of the Grubin (OHRN, 2008). As such, the CORE-10 should be used alongside other evidence based assessment questions and not in isolation.

The results of the current study suggest that the CORE-10 may be a clinically valid screening tool for use with prisoners in conjunction with existing processes. A cut point of above 6 may be most clinically suitable for identifying those with primary care level need. A CORE-10 cut point of 10 or more may be most appropriate for differentiating those with secondary care level needs, with inclusion of historic items having the potential to improve predictive validity particularly at secondary care level.

4.4.2 Considering risk

Within a custodial context, in addition to detection, screening should also allow for identification of those at elevated risk of self-harm and suicide and support risk management processes given the elevated risk of suicide in custody (Dooley *et al*., 1990; Partton & Borrill, 2004; Slade & Edelman, 2014). Most brief screens of mental health such as the GHQ-12 do not include risk based items. Evaluation of the prison reception screening process has also revealed that staff generally perceived the current standardised tool to be least effective for identifying risk of suicide, self-harm and mental health problems (OHRN, 2008). A strength of the CORE-10 is its inclusion of a risk item (*'I made plans to end my life'*). Importantly, scores on the CORE risk item differed significantly between those with a high MINI suicidality categorisation and those in any of the other suicidality categories with overall CORE-10 scores also differing significantly between those with clinically significant suicidality and non-suicidal prisoners. This suggests that assessment of response to the CORE-10 risk item could be clinically useful

in terms of identifying prisoner who may need further assessment of suicidal risk. Clear protocols would need to be developed to support practitioners in decision making surrounding responses to the risk item. Furthermore, items relating to self-harm are not included in the CORE-10 and thus an additional validated item(s) would need to be considered to identify risk of self-harm.

4.4.3 Practical utility

Practicality of psychometric tools in terms of ease of administration, administration time, competency in using tools, perceived relevance of tools, costs and organisational support for structured assessment are key predictors of the whether tools are used in practice (Belazi & Goldfarb, 2002; Jensen-Doss & Hawley, 2011; Trauer *et al*., 2006). The CORE-10, like several existing prison mental health screening tools (Ford *et al*., 2007; Grubin *et al*., 2002; Steadman *et al*., 2007) is very brief with 10 items on one page and was administered to prisoners in approximately two to three minutes in this study. Its brevity supports its use in busy prison environments, where there is limited time to make clinical assessments. Unlike the GHQ-12 the CORE-10 is also freely available and can be downloaded online with ease, thus reducing service costs and increasing its accessibility to staff.

The CORE-10 was administered by researchers in the current study and thus its ease of administration and perceived relevance by staff was not directly assessed. Nevertheless, previous research concerning the acceptability and feasibility of the CORE-10s parent measure the CORE-OM found that nurses reported it was simple to administer, relevant in terms of the variety of items included and useful in terms of clinical decision making (Perry *et al*., 2013). These findings combined with informal discussion with prison staff suggest that the CORE-10 may be feasibly administered by staff. Further research piloting the CORE-10 within staff and prisoners in the clinical screening context is however required to confirm its feasibility in terms of practical application.

No specific training is recommended for staff to develop competence in administering the CORE-10 and paper based scoring is relatively straightforward. However, implementation and routine use of measures has been found to be more successful when organisations provide appropriate training and administrative support, particularly for less qualified staff (Duncan & Murray, 2012; Jensen-Doss & Hawley, 2011). Given that screening is often conducted by Health Care Assistants in prisons, and that at times it may be appropriate for Prison Officers to contribute to assessment, both of whom have limited mental health, training in administration and interpretation of the CORE-10 would be imperative to ensure effective application.

Furthermore, the CORE-10 is designed to be a component of assessment and no one tool should be relied upon for screening purposes (Bryne *et al*., 2010; Connell & Barkham, 2007; Perry *et al*., 2013). As such it would be necessary to further explore how the CORE-10 could be integrated with existing practices and measures and develop clear protocols of responses dependent on screening outcome and services to ensure effective clinical use.

The current results which indicate high prevalence rates and low levels of previous detection indicate that introducing the CORE-10 into mental health screening is likely to be beneficial (Martins *et al*., 2016). However, ultimately the utility of the CORE-10 and screening more generally in terms of improving outcomes is dependant on the provision of effective services and treatments post identification. Primary care mental health services are currently under developed in Welsh prison (Little, 2013). Furthermore, the demands placed on MHIRTs at secondary care level have historically outstripped their ability to provide good quality care (Brooker *et al.,* 2005; OHRN, 2009; Pratt *et al*., 2015). Given this context of limited existing service provision, careful consideration would need to be given to implementing the CORE-10 as this would likely result in a substantial increase in detection rates, which would then require a treatment response from prison mental health services. That said, not all prisoners detected will

require treatment. For example, research has shown that depression and anxiety symptoms at intake resolve naturally for up to half of prisoners (Hassan *et al*., 2011; Taylor *et al*., 2010), and not all are likely to consent to treatment. Nevertheless, careful service planning, particularly at the primary care level would need to take place to ensure effective and ethical implementation of the CORE-10 as a screening tool.

In addition to its potential utility as a screening tool the CORE-10 has also been validated as an outcome measure (Connell & Barkham, 2007) which implies that it is sensitive to change. Although not the emphasis of the current research, validation of the CORE-10 with prisoners herein suggests it may also be usefully applied as an outcome measure for prison based mental health interventions. The Offender Positive Practice Guidance (NHS England, 2013), Policy Implementation Guidance (Welsh Government, 2014) in addition to NICE (2011b) guidance recommend evidence based interventions for common mental health problems across the stepped care framework for prisoners. Recommended interventions include at step two guided self-help, psycho-education and Cognitive Behavioural Therapy (CBT), at step three CBT, Interpersonal therapy and Eye Movement Desensitisation and Reprocessing and at step four third wave CBT approaches, Acceptance and Commitment Therapy, Integrative Therapy and Cognitive Analytic Therapy (NHS England, 2013). The CORE-10 has the potential to inform what type of interventions may be helpful as well as being utilised as an outcome measure in assessing the effectiveness of such interventions for prisoners. Further research is required however to establish the reliability of the CORE-10 in terms of its sensitivity to change in prisoner populations.

The findings of the current study provide preliminary support for the practical utility of the CORE-10 as a mental health screening tool for common and severe mental health problems in prisoners. Furthermore, the findings suggest the CORE-10 may be a valid and reliable outcome measure for use with prisoners. Further research is however needed to corroborate these initial findings.

4.4.4 Acceptability

Acceptability of measures can be judged in part by rates of completion and non-completion of items (Connell & Barkham, 2007). There was an excellent completion rate in the current study with only one CORE-10 item missing from 230 completed CORE-10s consistent with Perry *et al's*. (2013) findings regarding the CORE-OM in a forensic setting. The high completion rate may be in part attributable to the CORE-10 being interviewer administered in the current study. Administering the CORE-10 verbally likely increases the ability of many prisoners with low literacy levels (Morgan & Kett, 2003) to complete it successfully. Indeed, Perry *et al.* (2013) reported that patients and staff showed a preference for completing the CORE-OM together since this allowed for emotional and skills support. Based on the current and previous findings it is recommended that the CORE-10 be administered in interview format to prisoners in order to maximise its acceptability and consequent clinical utility.

4.4.5 Timing of administration

Acceptability may also be affected by when the tool is administered. Initial reception processes into prison already incorporate a number of tasks (Ministry of Justice, 2011) and reception is a time associated with elevated distress (Gigliardi *et al*., 2004; Perry *et al*., 2013). Perry *et al's*. (2013) study indicated that some prisoners felt that administration of the CORE-OM at reception would not provide an accurate assessment, despite staff perceiving this would be valuable. Although scores may be skewed by elevated distress at reception, research has also shown that identifying symptoms at reception is correlated with subsequent access to mental health services (Teplin, 1990). Furthermore, reception offers one of the few systematic opportunities to engage with all individuals entering the prison system. As such it seems that administering the CORE-10 either at initial health screening during the first 24 hours or during the general health assessment in the first week would maximise its utility. Given that the CORE-10 is thought to be sensitive to change and brief it is possible that it could be re administered by mental health professional upon service contact or initial implications drawn from it logically adjusted in light of new clinical

information. Establishing the most appropriate time to administer the CORE-10 requires piloting within the existing prison health screening process. The results of the current study provide preliminary support for the acceptability of the CORE-10 amongst prisoners, although further research is needed to establish how administration can be tailored to maximise acceptability.

4.5 Limitations

While the current study provides provisional support for the validity, reliability and utility of the CORE-10 as a screening tool for assessing common and severe mental health needs, the findings are confounded by a number of methodological limitations.

4.5.1 Study Design

4.5.2 Sampling

The current study's findings are limited by the non-random selection of the sample. Although random selection was partially achieved by randomly selecting wings, the sample was not entirely random as it was only possible to interview those who volunteered and were available on the days the researchers visited the prisons. Voluntary participation meant that participants self-selected to participate. Self-selection bias may threaten the internal validity of the findings since those who volunteer may differ from the population as a whole. Research has suggested that individuals who volunteer for research may be those who are more extroverted or have a particular interest in the topic (Rosenthal & Rosnow, 1975), in this case mental health. It is possible that such bias may exaggerated results, for example it is possible those with experience of mental health problems may have been more motivated to participate which may inflate prevalence rates and correlations between measures of mental health. However, comparison of the study sample and the whole populations of the studied prisons suggest minimal difference between participants and the entire population, suggesting the sample was representative minimising the possible effects of self-selection bias (HMIP, 2013a; 2013b). Furthermore, large-scale research concerning self-selection bias in population based surveys has found

minimal difference in demographic and health characteristics of responders and non-responders (Soggard *et al*., 2004).

In terms of sample size, although the sample size had adequate power to detect main effects, the relatively small sample size prevented exploration of differences in the validity of the CORE-10 between subgroups of prisoners. As such, it was not possible to explore differences in terms of age, ethnicity or sentencing status. Previous research assessing the validity of mental health screening tools for prisoner has found mixed results with regard to the effect of race. In Prins *et al.* (2012) evaluation of the BJMH, Black and Latino prisoners were found to have lower odds of screening positive compared to white prisoners, although this was thought to reflect known racial disparities rather than the tool's properties. Similarly, in terms of age, Grubin *et al*. (2002) reported higher rates of positive screens in adult prisoners compared to young offenders reflecting known differences in prevalence of mental health problems across age groups, although the measure appeared equally valid across age groups. Previous studies like the current study have not explored the effect of custodial factors on screening tools, despite research suggesting offence type (Birmingham *et al*., 2000) and sentencing status (Hassan *et al*., 2011) may impact upon distress, which may affect the validity of screening tools between subgroups of prisoners. Research utilising the CORE-10 with larger samples of prisoners is required to assess its validity and reliability across subgroups of prisoners.

4.5.3 Administration

A number of issues related to the administration of the measures in the current study may limit the validity of findings. Firstly, the measures were administered to prisoners who have been in the prisons up to six months as opposed to at point of reception as typically done in previous prison screening tool validation research (Andersen *et al*., 2002; Ford *et al*., 2007; Grubin *et al*., 2002; McCloskey, 2001; Steadman *et al*., 2005; Teplin & Swartz, 1989). This approach was necessary due to the operational impracticality of administration at reception in the sampled prisons and the

aim to consider the broader applicability of the CORE-10 in mental health assessment in prisons. However, the approach decreases the external validity of the CORE-10 with regard to its clinical utility at point of reception health screening.

Secondly, measures were also administered by researcher in the current study potentially decreasing the external validity of findings since, in practice, measures would be administered by health care staff. Previous research has suggested that prisoners are more likely to disclose symptoms to researchers than prison personnel (Grubin *et al*., 2002). As such the administration by researchers may have inflated mean scores and prevalence results. Nevertheless, the findings of the current study are consonant with those of Grubin *et al*. (2002) who utilised operational staff for administration suggesting the findings may be externally valid.

Finally, the CORE-10 was administered as part of a battery of tests as such fatigue may have had an impact on responses although the CORE-10 was administered near the beginning of the battery. Moreover, the CORE-10 in practice is likely to be used alongside other questions and tools related to mental health, therefore the design in part had a degree of ecological validity. In order to establish the validity and utility of the CORE-10 in clinical practice further research piloting its use at reception screening, alongside other standard questions by prison healthcare staff is required.

4.5.4 Criterion measure

Utilisation of a diagnostic instrument in terms of the MINI presents a possible limitation to the current study design. Importantly, diagnostic criteria and eligibility or need for a mental health service in prison do not necessarily correspond (Gangon, 2009). For example, not all of those with a current mental health diagnosis will need a mental health service and conversely prisoners without a diagnosis displaying a risk to themselves may need imminent intervention. Referral decisions may also be driven by service factors such as availability of resources, which are not taken into consideration using diagnostic tools (Gangon, 2009). As such benchmarking

the CORE-10's predictive validity against the MINI fails to account for broader procedural and systemic issues such as staffing levels, local practices and policy affecting access to mental health services in prisons. However, evidence based treatments recommended by policy are typically developed based on diagnostic criteria and use of structured tools, although agreement between diagnostic tools and clinical decision-making remains poor (Jensesn-Doss & Hawley, 2011). As such, promoting use of structured tools associated with diagnostic criteria, particularly in screening, has been suggested as a way of improving treatment outcome (Jensesn-Doss & Hawley, 2011).

Despite these suggested long-term benefits, a related issue is testing the predictive validity of a tool for measuring distress (the CORE-10) against a diagnostic interview for disorders (the MINI), given the two are not necessarily synonymous. However, testing a screening instrument against a gold standard criterion is an important first step in establishing a tools psychometric validity, and validating distress measures against diagnostic criteria is common practice (Weiner & Graham, 2003). Further research is required to explore the relationship between CORE-10 scores and mental health diagnoses.

Additionally, it has been argued that diagnostic clinical interviews can produce arbitrarily high epidemiological estimates for mental health disorder since they classify general difficulties in adjusting in terms of mental disorders (Evans *et al*., 2010). This is particularly problematic in the prison context, where high level of distress and situational stressors are known to exist. Furthermore, items contained in the MINI associated with some disorders, for example mania[23] have the potential for over identification amongst offenders, who may by the nature of their offending relate to characteristics like impulsivity or grandiosity which overlap with, but do not necessarily construe a psychiatric diagnosis. As such, the MINI may

[23] For example the MINI 6.0 mania screen contains the following items C1 a. *'have you ever had a period of time where you were feeling so high, hyper or so full of energy or full of yourself you got into trouble'* and C3 a. *'felt you could do things others couldn't do, or that you were an especially important person'*.

overestimate mental health disorders, although the prevalence rates herein are largely consonant with those identified in other British prison-based research (Grubin, 2002; OHRN, 2010). Evans *et al*. (2010) argue that over estimation of mental health disorders is not necessarily problematic in the context of evaluating screening tools. Screening tools would need to detect mental illness in those with questionable diagnoses, therefore artificially high rates of mental disorder would result in an underestimation of tools sensitivity rather than an inflation of its predictive power (Evans *et al*., 2010). Additionally, the MINI has been validated against longer gold standard clinical interviews including the SCID and CIDI and has been widely utilised in prison based research (Baksheev *et al.*, 2012; Black *et al*., 2004; Evans *et al*., 2010; Gunter *et al.,* 2008) supporting its validity in this context. Further research evaluating the CORE-10 against other diagnostic interviews such as the SCID or SADS is warranted to confirm its predictive validity with prisoners.

A further limitation with regard to the MINI is the use of a subset of its screens focusing on Axis I mental health problems in the current study, with the exclusion of eating disorders, personality disorders and substance use disorders. This approach was adopted to reduce burden on participants, to reflect key presentations treated in prison mental health services and to promote consistency between the gold standard measure and the psychometric tool under validation. However, exclusion of personality disorder and substance use disorders meant it was not possible to consider the predictive validity of the CORE-10 in screening for comorbidity, which is highly prevalent in prison populations (Fazel & Seewald, 2012). Furthermore, it was not possible to assess the impact of personality disorder on CORE-10 scores. Further research is required to explore whether a diagnosis of personality disorder impacts upon distress scores. With regard to substance misuse, Andersen (2002) argues if dependence disorders have a significant impact on mental health, disorders in the affective, psychosomatic and anxiety domains may be identified. Indeed the positive correlation between drug use in the year before prison and CORE-10 scores support this assertion. No brief screening tool is likely to adequately screen

across the whole spectrum of Axis I, Axis II disorders and substance misuse disorders, with contemporary research suggesting screening tools focusing on non-specific distress are more appropriate in the context of stepped care models (Kessler *et al*., 2002). As such, diagnostically screening for all disorders was perceived to be superfluous.

A final limitation associated with use of the MINI was that it was not possible to calculate inter-rater reliability between researchers due to time and resource limitations. Future research using the MINI should ensure that a proportion of MINI interviews are scored by at least two researchers blind to each other's scoring outcome in order to assess inter-rater reliability.

4.5.5 Malingering and social desirability

A possible limitation of the current study is the failure to control for malingering and 'super normality' or social desirability effects. Malingering is the exaggeration of symptoms or 'faking bad' motivated by external incentives (Cima *et al*., 2003), which has been highlighted as an issue within forensic populations (Rogers, 2008). Resnick & Knoll (2008) argue that malingering is difficult to accurately assess in prisoners since it can be an adaptive response to the stressful and potentially dangerous prison environment. As such they recommend not denying treatment based on malingering of some symptoms amongst prisoners. Malingering has been discounted in previous research identifying high rates of mental illness in prisoners as records confirmed may were previously known to community psychiatric services (Linehan *et al*., 2005). It is possible that prevalence rates of disorders and psychometric scores in the current study were elevated by malingered symptoms. However, given the complexity in assessing malingering in prisoners it was deemed beyond the scope of the current small-scale study.

In contrast to malingering, social desirability (or 'feigning good') in terms of creating a positive impression of oneself by denying minor faults or difficulties has also been found to bias self-report measures (van de Mortel

& Thea, 2008). Some authors closely associate social desirable and 'super normality' type responses (Hall & Poirer, 2001). Others suggest 'super normality' goes beyond social desirable responses since individuals deny all negative symptoms, even those commonly reported in the general population (Cinma, 2003). The phenomenon of 'super normal' responding is thought to be prevalent in forensic population where individuals may be motivated to demonstrate progress for parole purposes or to avoid hospitalisation (Cinma, 2003). Unlike malingering, social desirability may bias the current result in term of underreporting of symptoms. The consistency between the current findings and result of previous research with similar samples (Grubin *et al*., 2002; McCloskey, 2001; OHRN, 2010), however suggests biases created as a result of social desirability and malingering did not have an undue effect on the result, or at least had an equitable effect as to other studies. Furthermore, studies have identified that clinicians are not very accurate in identifying feigning (Cinma *et al*., 2002; Faust *et al*., 1988). The exclusion of feigning measures is consistent with the clinical context in which the measure would be administered as health care staff do not systematically assess feigning when assessing prisoners mental health. Excluding feigning measures also served to reduced burden on participants. Further research is needed to explore the effect of socially desirable responding and malingered symptoms on the CORE-10.

4.5.6 Limitations of the CORE-10

Despite the findings of the current study suggesting the CORE-10 may be a valid and reliable tool for screening prisoner distress, it is not without its limitations. The CORE-10 focuses on current distress and does not consider historical factors predicting mental health problems or the chronicity of existing problems. Emphasis on current symptoms measured at entry to prison when increased stress is expected due situational factors may skew accurate assessment of clinical need, since distress at this time may not predict subsequent adjustment within the prison environment (Ford *et al*., 2007). Indeed, previous feasibility research amongst prisoners and prison staff with the CORE-OM indicated that some individuals felt that the symptom-based nature of questions may pathologise normal contextually

contingent experiences (Perry *et al.*, 2013). Assessment of current distress also fails to acknowledge the often chronic nature of SMI and does not account for individuals who may externally be displaying debilitating symptoms, but do not recognise these as distressing or problematic (Barkham, 2005). Additionally, the CORE-10 does not explore self-harm, which while not being defined as mental health problem per se, often triggers the involvement of mental health services in prisons. As such it is important that in practice the CORE-10 is utilised alongside holistic assessment, clinical judgement and other validated questions concerning historic indicators of mental health problems (Grubin *et al.*, 2002).

4.6 Recommendations for future research

The findings of the current study are promising in terms of supporting the use of the CORE-10 in screening for common and severe mental health needs of prisoners in line with contemporary stepped care service models. The current study supports the psychometric validity of the CORE-10 for screening purposes. Replication studies with larger samples of prisoners and using alternative criterion measures are however required to confirm the positive results found herein, and to explore validity and reliability of the CORE-10 within subgroups of prisoners. Further research should also seek to assess the impact of biases in responding, substance misuse and personality disorder on CORE-10 scores.

Further research piloting the CORE-10 in practice with staff and prisoners in the clinical screening context is imperative to confirm its feasibility in terms of practical application, acceptability and to establish optimum timing for administration (for example at first or second reception health assessment). Exploration of how the CORE-10 could be integrated with existing practices and historical measures such as the Grubin (2002) is also required to develop clear protocols to inform referral pathways dependent on screening outcome.

A particularly valuable avenue of further research may be assessing the predictive validity of the CORE-10 when combined with questions pertaining to historical problems. For example, assessing predictive validity when combined with items already included in the Grubin tool may support improved differentiation between those with primary care needs and those requiring support for chronic problems through secondary care.

Furthermore, the current study focused specifically on screening therefore exploration of the utility of the CORE-10 for other aspects of mental health provision in prisons such as outcome measurement was not considered. Given the strong psychometric properties of the CORE-10 within this population, assessment of its sensitivity to change by testing it in relation to interventions may be beneficial in terms of establishing its utility for measuring outcomes of mental health service interventions in prisons.

Despite substantial scope for further research, the current study represents the first in Wales to specifically assess a screening tool focused on current distress in line with existing mental health service models in prisons.

4.7 Conclusions

Psychiatric morbidity within prison populations is considerably higher than in community populations (Singleton *et al*., 1998). The prison environment can have a detrimental impact upon mental health and untreated mental health problems amongst prisoners have been linked to increased institutional violence, self-harm, suicide and reoffending (Ballargeon *et al*., 2009; Martin *et al*., 2013). In improving access to treatment for common mental health problems in Wales, the Mental Health Measure (2010) stipulates that Primary Care Mental Health Services should be available to all, including prisoners. Although such services have been embedded within the community, their implementation in Welsh prisons is in its infancy (Little, 2013). Accurately targeting such services to those in need is dependent on effective screening and identification of mental health problems. The current study is the first of its kind to explore the predictive validity of the CORE-10

in screening for common and severe mental health problems amongst prisoners.

The results of the current study provide preliminary evidence that the CORE-10 is a valid and reliable tool for screening for common and severe mental health problems in prisoners. The CORE-10 demonstrated adequate predictive validity against diagnostic criteria and its predictive validity was superior to existing clinical practice in terms of referral decisions. The CORE-10 had construct, convergent and discriminant validity as well as good reliability amongst this population. The CORE-10 also conformed to its originally proposed clinically useful six factor structure within this population. Importantly, the CORE-10 also appeared to be both acceptable to prisoners and feasible in terms of administration, although further research is warranted to establish its utility in clinical practice. The current study also found preliminary evidence for the validity and reliability of the GHQ-12, which has comparable psychometric properties to the CORE-10 in screening for mental health problems in prisoners.

The CORE-10 offers prison staff a structured and reliable method for screening for mental health problems and suicidal risk in prisoners, which could be used to support referral decisions across mental health pathways. Integrating the CORE-10 with holistic assessment of historic indicators of mental health difficulties may offer the most accurate approach to screening. Specifically, combining the CORE-10 with items from the Grubin contained in the existing reception screen may improve accuracy and warrants further research.

In sum, the current research supports the use of the CORE-10 in screening for common and severe mental health problems in prisoners consistent with current stepped care pathways of mental health care advocated in policy and practice (Little, 2013; Welsh Government, 2010; 2014). Improving accuracy of screening offers the potential to improve patient care as well as offer long term cost savings by supporting accurate timely allocation of resources, thus potentially ameliorating the negative effects associated with

undiagnosed mental health problem in prisoners. The initial findings are promising and indicate the need for further research piloting the CORE-10 for screening in clinical practice in Welsh prisons.

5 References

Abramson, M. F. (1972). The criminalization of mentally disordered behavior: possible side effect of a new commitment law. *Hospital and Community Psychiatry*, 23, 101–107.

Adams, K. (1986). The disciplinary experiences of mentally disordered inmates. *Criminal Justice and Behaviour*, 13, 297-316.

Adamson, V., Gibbs, M. & McLaughlin, D. (2015). Evaluation of the improving access to psychological therapies for offenders programme at HMP Lincon: A three-year prospective study. *The Journal of Forensic Psychology and Psychiatry*, 26, 185-201.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd. edn). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). Washington, DC: Author.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th edn). Washington, DC: Author.

Andersen, H. S., Sestoft, D., Lillebaek, T., Gabrielsen, G. & Hemmingsen, R. (2002). Validity of the General Health Questionnaire (GHQ-28) in a prison population: Data from a randomized sample of prisoners on remand. *International Journal of Law and Psychiatry*, 25, 573-580.

Andersen, H. S., Sestoft, D., Lillebaek, T., Gabrielsen, G., Hemmingsen, R. A. (2003). Longitudinal study of prisoners on remand: Repeated measures of confinement. *International Journal of Law and Psychiatry*, 26, 165-177.

Andersen, H. S., Sestoft, D., Lillebæk, T., Gabrielsen, G., Hemmingsen, R., & Kramp, P. (2000). A longitudinal study of prisoners on remand. Prevalence, incidence and psychopathology in solitary vs. non-solitary confinement. *Acta Psychiatrica Scandinavica*, 102, 19–25.

Aneshensel, C. S. (1992). Social stress: theory and research. *Annual Review of Sociology*, *18*, 15-38.

Anthony, D. & McFayden, J. (2005). Mental health needs assessment of prisoners. *Clinical effectiveness in Nursing*, 9, 26-36.

Appleby, L. (2010). *Prison mental health: vision and reality*. London: Royal College for Nursing, NACRO & Centre for Mental Health.

Arbuckle, J. (2011). *IBM SPSS AMOS version 20*. IBM Corp.

Armitage, C., Fitzgerald, C. & Cheong, P. (2003). Prison in-reach mental health nursing. *Nursing Standard,* 17, 40-42.

Atkinson, M., Zibin, S. & Chuang, H. (1997). Characterizing quality of life among patients with chronic mental illness: A critical examination of the self-report methodology. *American Journal of Psychiatry*, 154, 99-105.

Baillargeon, J., Binswanger, I. A., Penn, J. V., Williams, B. A., & Murray, O. J. (2009). Psychiatric disorders and repeat incarcerations: The revolving prison door. *American Journal of Psychiatry*, 166, 103–109.

Baillargoeon, J., Hoge, S. K. & Penn, J. V. (2010). Addressing the challenge of community reentry among released inmates with serious mental illness. *American Journal of Community Psychology*, 46, 361-375.

Baksheev, G., Ogloff, J. & Thomas, S. (2012). Identification of mental illness in police cells: A comparison of police processes, the Brief Jail Mental Health Screen and the Jail Screening Assessment Tool. *Psychology, Crime & Law*, 18, 529-542.

Ballargeon, J., Binswanger, I., Penn, J., Williams, B. & Murray, O. (2009). Psychiatric disorders and repeated incarcerations: The revolving prison door. *American Journal of Psychiatry*, 166, 103-109.

Bennett, H. & Lincon, N. (2006). Screening for depression and anxiety after stroke: Potential measures for routine screening. *International Journal of Therapy and Rehabilitation*, 13, 401-406.

Barkham, M., Bewick, B., Mullin, T., Gilbody, S., Connell, J. & Cahill, J. *et al*. (2013). The CORE-10: a short measure of psychological distress for routine use in psychological therapies. *Counseling and Psychotherapy Research*, 13, 3-13.

Barkham, M., Gilbert, N., Connell, J., Marshall, C. & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry*, 186, 239-246.

Barkham, M., Mellor-Clark, J., Connell, J., Evans, C. & Margison, F. (2010). Clinical Outcomes in Routine Evaluation (CORE) – The CORE measures and system: Measuring, monitoring and managing quality evaluation in the psychological therapies. In Barkham, M. & Hardy, G. & Mellor-Clarke, J. (Eds.) *Developing and delivering practice based evidence: A guide for the psychological therapies* (pp. 175-219). London: John Wiley and Sons Ltd.

Barnes, K. E. (1982). *Preschool screening: The measurement and prediction of children at-risk.* Springfield: Charles C. Thomas.

Bebbington, P. E., Minot, S., Cooper, C., Dennis, M., Meltzer, H. & Jenkins, R. *et al*. (2010). Suicidal ideation, self-harm and attempted suicide: results from the British psychiatric morbidity survey 2000. *European Psychiatry*, 25, 427-431.

Beck, A. T., Steer, R. A. & Brown, G. K. (1996). *Manual for the Beck Depression Inventory–II.* San Antonio, TX: Psychological Corporation.

Bedford, A., Watson, R., Lyne, J., Tibbles, J., Davies, F. & Deary, J. (2010). Mokken scaling and principal component analyses of the CORE-OM in a large clinical sample. *Clinical Psychology and Psychotherapy*, 17, 51-62.

Belazi, D., Goldfarb, N., He, H. (2002). Measuring health-related quality of life in the clinical setting. *Expert Review of Pharmacoeconomics Outcomes Research,* 2, 109–117.

Benedek, E., Ash, P. & Scott, C. (2009). *Principle and practice of child and adolescent mental health*. Washington, London: American Psychiatric Publishing Inc.

Birmingham, L. (1999). Between prison and the community. The 'revolving door psychiatric patient' of the nineties. *British Journal of Psychiatry*, 174, 378-379.

Birmingham, L. & Mullee, M. (2005). Development and evaluation of a screening tool for identifying prisoners with severe mental illness. *The Psychiatric Bulletin*, 26, 334-338.

Birmingham, L., Mason, D. & Grubin, D. (1996). Prevalence of mental disorder in remand prisoners: Consecutive case study. *British*

Medical Journal, 313. Retrieved 13th November 2015 from doi: http://dx.doi.org/10.1136/bmj.313.7071.1521

Birmingham, L., Mason, D. & Grubin, D. (1997). Health screening at first reception into prison. *The Journal of Forensic Psychiatry*, 8, 435-439.

Birmingham. L., Gray, J., Mason, D. & Grubin, D. (2000). Mental illness at reception into prison, *Criminal Behaviour and Mental Health*, 10, 77-87.

Blaauw, E., Roozen. H. & Val Marie, H. (2007). Saved by structure? The course of psychosis within a prison population. *International Journal of Prison Health*, 3, 248-256.

Blaauw, E., Winkel, F. W., & Kerkhof, J. F. M. (2001). Bullying and suicidal behavior in jails. *Criminal Justice and Behavior,* 28, 279-299.

Black, D., Ardnt, S., Hale, N. & Rogerson, B. (2004). Use of the Mini Neuropsychiatric Interview (MINI) as a screening tool in prisons: Results of a preliminary study. *Journal of American Academy of Psychiatric Law*, 32, 158-162.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods and Research*, 21, 205-229.

Borrill, J., Burnett, R., Atkins, R., Miller, S., Briggs, D., Weaver, T., *et al.* (2003). Patterns of self-harm and attempted suicide among white and black/mixed race female prisoners. *Criminal Behaviour and Mental Health*, 13, 229–240.

Boothby, J. & Durham, T. W. (1999). Screening for depression in prisoners using the Beck Depression Inventory. *Criminal Justice and Behavior,* 26, 107-124.

Boothby, M., Mulholland, I., Cases, A., Carrington, K. & Bolger, T. (2010). Towards mental health promotion in prisons: The role of screening emotional distress. *Procedia Social and Behavioural Science*, 5, 90-94.

Botna, J. & Gendreau, P. (1990). Re-examining the cruel and unusual punishment of prison life. *Law and Human Behaviour*, 14, 347-372.

Bowen, R. A., Rogers, A., Shaw, J. (2009). Medication management and practices for people in prison with mental health problems: A

qualitative study. *International Journal of Mental Health*, 3. Retrieved 22nd November 2015 from doi: http://dx.doi.org/10.1186%2F1752-4458-3-24.

Bower, P., Gilbody, S., Richards, D., Fletcher, J. & Sutton, A. (2006). Collaborative care for depression in primary care. *British Journal of Psychiatry*, 189, 484-493.

Boyle, M. & Johnstone, L. (2014). Alternatives to psychiatric diagnosis. *Lancet Psychiatry*, 14, 70359-70361.

Bradley, L. (2009). *The Bradley Report. Lord Bradleys review of people with mental health problems or learning disabilities in the criminal justice system.* London: Department of Health.

Brooker, C., Repper, J., Siedifield, C. & Gojkovic, D. (2009). Review of service delivery and organisational research focused on prisoners with mental disorders. *Journal of Forensic Psychiatry and Psychology*, 20, S102-S123.

Brooker, C., Rickets, T., Lemme, F. & Dent-Brown, K. (2005). *The evaluation of the mental health in-reach collaborative*. Retrieved 23rd August 2015 from: http://www.nfmhp.org.uk

Bryne, B. M. (2010). *Structural equation modeling with AMOS. Basic concepts, applications and programming* (2nd Edn.). New York: Routledge.

Burgha, T., Singleton, N. Meltzer, H., Bebbington, P., Farrell, M., Jenkins, R. *et al*. (2005). Psychosis in the community and in prisons: A report from the British National Survey of Psychiatric Morbidity. *American Journal of Psychiatry*, 162, 774-780.

Burns, T. (2001). Generic versus specialist mental health teams. In G. Thornicroft & G. Szmukler (Eds.) *Textbook of Community Psychiatry* (pp. 231 -241). Oxford: Oxford University Press.

Byng, R., Quinn, C., Sheaff, R., Samele, C., Duggan, S. & Harrison, D. *et al*. (2012). COCA: *Care for offenders' continuity of access. Final Report*. NIHR Service Delivery and Organizational Programme.

Cameron, I., Crawford, J., Lawton, K. & Reid, I. (2008). Psychometric comparison of the PHQ-9 and HADS for measuring depression

severity in primary care. *British Journal of General Practice*. Retrieved 3rd August 2015 from DOI: 10.3399/bjgp08X263794

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Cavanagh, J., Carson, A., Sharpe, M. & Lawrie, S. (2003). Psychological autopsy studies of suicide: A systematic review. *Psychological Medicine*, 3, 395-405.

Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R. & Muris, P. (2003). The other side of malingering: Supernormality. *The Clinical Neuropsychologist,* 00, 1-10.

Cima, M., Merckelbach, H., Nijman, H., Knauer, E., & Hollnack, S. (2002). I can't remember Your Honor: Offenders who claim amnesia. *German Journal of Psychiatry*, 5, 24–34.

Clear, T. R. & Sumter, M. T. (2002). Prisoners, prison and religion. *Journal of Offender Rehabilitation*, 35, 125-156

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.

Connell, J. & Barkham, M. (2007). *CORE-10 User Manual, Version 1.1.* CORE System Trust & CORE Information Management Systems Ltd.

Cooper, C. & Berwick, S. (2001). Factors affecting psychological well-being of three groups of suicide prone prisoners. *Current Psychology*, 20, 269-182.

Corner, J. & Kendall, P. (2013). *The Oxford handbook of research strategies for clinical psychology*. London: Oxford University Press.

Cox, R. M., Lenz, S. & James, R. K. (2015). A pilot evaluation of the ARRAY programme with offenders with mental illness. *Journal of Counselling & Development*, 93, 471-480.

Critical Skills Appraisal Programme (CASP) (2013). *CASP diagnostic checklist.12 questions to help you make sense of a diagnostic test study*. Retrieved on 30th October 2015 from: http://www.casp-uk.net/#!casp-tools-checklists/c18f8

Davis, S. (1992). Assessing the "criminalization" of the mentally ill in Canada. *Canadian Journal of Psychiatry,* 37, 532-538.

Department of Health (2002). *Health Promoting Prisons: A Shared*

Approach. London: Department of Health.

Department of Health (2008). *Improving access to psychological therapies implementation plan: National guidelines for regional delivery*. London: Department of Health.

Department of Health (DoH) and HM Prison Service (HMPS) (2001). *Changing the outlook: A strategy for developing and modernizing mental health services in prisons.* London: Department of Health.

Dhami, M., Ayton, P. & Loewenstein, G. (2007). Adaptation to imprisonment: Indigenous or imported? *Criminal Justice and Behavior*, 18. Retrieved June 9th 2015 from doi: 10.1177/0093854807302002

DiCataldo, F., Greer, A., & Profit, W. (1995). Screening prison inmates for mental disorders: An examination of the relationship between mental disorder and prison adjustment. *Bulletin of the American Academy of Psychiatry and Law,* 23, 573-585.

Ditton, P. M. (1999). Mental health treatment of inmates and probationers. *Bureau of Justice Statistics Special Report*, 1-12.

Division of Clinical Psychology (2013). *Position statement on the classification of behaviour and experience in relation to functional psychiatric diagnosis.* Leicester: British Psychological Society.

Dooley E. (1990). Prison suicide in England & Wales 1972-87. *British Journal of Psychiatry,* 150, 40-45.

Duncan, E. & Murray, J. (2012). The barriers and facilitators to routine outcome measurement by allied health professionals in practice: A systematic review. *Health Services Research*, 12-96.

Efron, B. Tibshirani, R, (1997). *An introduction to the bootstrap*. Dordrecht: Springer Science & Business Media.

Elbogen, E., & Johnson, S. (2009). The intricate link between mental disorder and violence: Results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry*, *66*, 152-161.

Elwood, R. W. (1993). Psychological tests and clinical discriminations: Beginning to address the base rate problem. *Clinical Psychology Review,* 13, 409–419.

Evans, C., Brinded, P., Simpson, A., Frampton, C. & Mulder, R. (2010). Validation of a brief screening tools for mental disorders among New Zealand prisoners. *Psychiatric Services*, 61, 923-928.

Evans, C. E., Cornell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J. *et al.* (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, *180*, 51–60.

Evans, G. W. (2003). The built environment and mental health. *Journal of Urban Health*, 80, 536-555.

Falissard, B., Loze, J-Y., Gasquet, I., Duburc, A., de Beaurepaire, C., Fagnani, F. *et al*. (2006). Prevalence of mental disorders in French prisons for men. *BMC Psychiatry*, 6-33.

Faust, D., Hart, K., Guilmette, T.J., & Arkes, H.R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, 19, 508–515.

Fazel, S. & Seewald, K. (2012). Severe mental illness in 33,588 prisoners worldwide: Systematic review and meta-analysis. *British Journal of Psychiatry*, 200, 364-373.

Fazel, S., Benning, R., & Danesh, J. (2005). Suicides in male prisoners in England and Wales 1978-2003. *Lancet*, 366, 1301-1302.

Fazel, S., Cartwright, J., Norman-Nott, A. & Hawton, K. (2008). Suicide in prisoners: A systematic review of risk factors. *Journal of Clinical Psychiatry*, 69, 1721-1731.

Fazel, S., Cartwright, J., Norman-Nott, A. & Hawton, K. (2008). Suicide in prisoners: A systematic review. *The Journal of Clinical Psychiatry*, 4, e1-e11.

Fazel, S., Hope, T., O'Donnell, I., Jacoby, R. (2001). Hidden psychiatric morbidity in elderly prisoners. *British Journal of Psychiatry*, 179, 535–539.

Fazel, S. & Danesh, J. (2002). Serious mental disorder in 23,000 prisoners. A systematic review of 62 surveys. *Lancet,* 359, 545-550.

Fazel, S. & Lubbe, S. (2005). Prevalence and characteristics of mental disorders in jails and prisons. *Current Opinion in Psychiatry*, 18, 550-554.

Feder, L. (1991). A comparison of the community adjustment of mentally ill offenders with those from the general prison population: An 18-month follow-up. *Law and Human Behavior*, 15, 477–493.

Fellner, J. (2006). A corrections quandary: Mental illness and prison rules. *Harvard Civil Rights-Civil Liberties Law Review*, 41, 391-412.

Felson, R., Silver, E., Remster, B. (2012). Mental disorder and offending in prison. *Criminal Justice and Behavior*, 39, 125-143.

Feron, J. M., Paulus, D., Tonglet, R., Lorant, V. & Pestiaux, D. (2005). Substantial use of primary health care by prisoners: Epidemiological description and possible explanations. *Journal of Epidemiological Community Health*, 59, 651-655.

Field, A. (2009). *Discovering statistics using SPSS*. 2nd Edn. London: SAGE Publications Ltd.

Field, A. (2013). *Discovering statistics using IBM SPSS*. 4th Edn. London: SAGE Publications Ltd.

Fisher, W., Silver, E. & Wolff, N. (2006). Beyond criminalization: towards a criminologically informed framework for mental health policy and service research. *Administration and Policy in Mental Health*, 33, 544-557.

Fizpatrick, E., Chambers, J., Burns, T., Doll, H., Fazel, S., Jennkinson, C. K. *et al*. (2010). A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. *Health Technology Assessment*, 14, 9-27.

Ford, J. D., Trestman, R. L., Weisbrock, V. & Zhang, W. (2007). Development and validation of a brief mental health screening instrument for newly incarcerated adults. *Assessment,* 14, 279-299.

Ford, J., Trestman, R., Weisbrock, V. & Zhang, W. (2009). Validation of a brief screening instrument for identifying psychiatric disorders among newly incarcerated adults. *Psychiatric Services*, 60, 842-846.

Fotiadou, M., Livaditis, M., Manou, I., Kaniotou, E. & Xenitidis, K. (2006). Prevalence of mental disorders and deliberate self-harm in Greek male prisoners. *International Journal of Law and Psychiatry*, 29, 68–73.

Fruehwald, S., Matschnig, T., Koenig, F., Bauer, P. & Frottier, P. (2004).

Suicide in custody: case–control study. *British Journal of Psychiatry*, 185, 494–498.

Gagnon, N. C. (2009). *Mental health screening in jails.* Unpublished PhD Thesis: Simon Fraser University, Psychology Department.

Ghofar, A. & Islam, S. (2015). *Corporate governance and contingency theory: a structural equation modelling approach and accounting risk implications.* Switzerland: Springer International.

Gibbs, J. (1991). Environmental congruence and symptoms of psychopathology: A further exploration of the effects of exposure to the jail environment. *Criminal Justice and Behavior*, 18, 351-374.

Gigliardi, G., Lovell, D., Peterson, P. & Jemelka, R. (2004). Forecasting recidivism in mentally ill offenders released from prison. *Law and Human Behavior*, 28, 133-155.

Gilbody, S., Richards, D. & Barkaham, M. (2007). Diagnosing depression in primary care using self-completion instruments. *British Journal of General Practice*, August, 650-652.

Glaros, A. G. & Kline, R. B. (1998). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity and predictive value model. *Journal of Clinical Psychology*, 44, 1033-1023.

Goldberg, D. (1978). *Manual of the general health questionnaire*. Windsor: NFER-Nelson.

Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire*. London: London University Press.

Goldberg, D. P. (1978). *Manual of the General Health Questionnaire*. Sussex: DJS Spools.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O. *et al*. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, 27, 191-197.

Goodman, L. A., Salyers, M. P., Mueser, K. T., Rosenberg, S. D., Swartz, M., Essock, S. M., & Vidaver, R. M. (2001). Recent victimization in women and men with severe mental illness: Prevalence and correlates. *Journal of Traumatic Stress*, *14*, 615-632.

Gov. UK (2015). *Vulnerable prisoners*. Retrieved on 12th November 2015 from: https://www.gov.uk/life-in-prison/vulnerable-prisoners

Grann, M., Danesh, J & Fazel, S. (2008). The association between psychiatric diagnosis and violent re-offending in adult offenders in the community. *BMC Psychiatry,* 8, 92.

Greenberg, G. A. & Rosenheck, R. A. (2008). Jail, incarceration, homelessness and mental health: A national study. *Psychiatric Services*, 59, 170-177.

Greenhalagh, J., Long, A. F. & Lynn, R. (2005). The use of patient reported outcomes measures in routine clinical practice: Lack of impact or theory? *Social Sciences & Medicine*, 60, 833-843.

Grey, N., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. & Snowden, R. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of the HCR-20, PCL-R, and psychiatric symptomology. *Journal of Counselling and Clinical Psychology*, 71, 443-451.

Grisso, T. Vincent, G. & Seagrave, D. (2005). *Mental health screening and assessment in juvenile justice.* New York & London: The Guilford Press.

Grubin, D., Carson, D. & Parsons, S. (2002). *Report on new prison reception health screening arrangements: the results of a pilot study in 10 prisons.* London: Department of Health.

Grubin, D., Parsons, S., Hopkins, C. (1999). *Report on the evaluation of a new reception health questionnaire and associated training*. Report for Her Majesty's Prison Service.

Gunn, J., Maden, A. & Swinton, J. (1991). Treatment needs of prisoners with psychiatric disorders. *British Medical Journal*, 303, 338-341.

Gunter, T. D., Arndt, S., Wenman, G., Allen, J. Loveless, P., Sieleni, B. *et al.* (2008). Frequency of mental and addictive disorders among 320 men and women entering the Iowa prion system: Use of the MINI-Plus. *Journal of American Academy of Psychiatry and Law*, 36, 27-34.

Gureje, O. & Obikoya, B. (1990). The GHQ as a screening tool in a primary care setting. *Social Psychiatry and Psychiatric Epidemiology,* 25, 276–280.

Hall, H.V. & Poirier, J.G. (2001). *Detecting malingering and deception: Forensic distortion analysis*. Boca Raton, FL: CRC Press.

Haney, C. (2001). *The psychological impact of incarceration: Implications for post-prison adjustment.* ASPE, US Department of Health & Human Services. Retrieved on 20th November 2015 from: http://aspe.hhs.gov/basic-report/psychological-impact-incarceration

Haney, C. (2003). Mental health issues in long term solitary and 'supermax' confinement. *Crime and Delinquency*, 49, 124-156.

Harding, T., Zimmerman, E. (1989). Psychiatric symptoms, cognitive stress and vulnerability factors. A study in a remand prison. *British Journal of Psychiatry* 155, 36–43.

Hart, S., Roesch, R., Corrado, R. & Cox, D. (1993). The Referral Decision Scale: A validation study. *Law and Human Behaviour*, 17, 611-623.

Harty, M., Tighe, J., Leese, M., Parrott, J. & Thornicroft, G. (2003). Inverse care for mentally ill prisoners: Unmet needs in forensic mental health services. *The Journal of Forensic Psychiatry and Psychology,* 14, 600–614.

Hassan, L., Birmingham, L., Harty, M., Jarrett, M., Jones, P., King, C. *et al*. (2011). Prospective cohort study of mental health during imprisonment. *British Journal of Psychiatry*, 198, 37-42.

Hawkins, M. (2008). The factor structure of the twelve item General Health Questionnaire: The results of negative phrasing? *Clinical Practice and Epidemiology in Mental Health*, 4, 1-8.

Hawtton, K., Linseel, L., Adeniji, T., Sariaslan, A. & Fazel, S. (2014). Self-harm in prisons in England and Wales: An epidemiological study of prevalence, risk factors, clustering and subsequent suicide. *Lancet*, 383, 1147-1154.

Hemmens, C. & Marquart, J. W. (1999). Straight time: Inmates perceptions of violence and victimization in the prison environment. *Journal of Offender Rehabilitation*, 28 1-21.

Her Majesty's Inspectorate of Prisons (HMIP) (2007). *The mental health of prisoners: A thematic review of care and support of prisoners with mental health needs*. London; HMIP.

Her Majesty's Inspectorate of Prisons (HMIP) (2013a). *Report on an*

announced inspection of HMP Cardiff 18-22 March 2013. London; HMIP.

Her Majesty's Inspectorate of Prisons (HMIP) (2013b). *Report on an unannounced inspection of HMP/YOI Parc 9-19 July 2013*. London; HMIP.

Her Majesty's Inspectorate of Prisons for England and Wales (HMIP) (1996). *Patient or prisoner? A new strategy for health care in prisons*. London: Home Office.

Hewitt, C., Perry, A., Adams, B & Gilbody, S. (2011). Screening and case finding for depression in offender populations: A systematic review of diagnostic properties. *Journal of affective disorders*, 128, 72-82.

Higginson, S. (2005). Malingering psychosis among prisoners: Part 1. *Mental Health Practice,* 9, 22-25.

Hillbrand, M., Krystal, J., Sharpe, K., Foster, H. (1994). Clinical predictors of self-mutilation in hospitalized forensic patients. *Journal of Nervous & Mental Disease*, 182 (1), no pagination.

HM Prison Service (2006). *Prison Service Order 3050. Continuity of healthcare for prisoners*. HM Prison Service.

Hochestetler, A., Murphy, D. & Simons, R. (2004). Damaged goods: Exploring predictors of distress in prison inmates. *Crime & Delinquency*, 50, 436-457.

Home Office (2015). *Counting rules for recorded crime*. Retrieved on 10th January 2015 from: https://www.gov.uk/government/publications/counting-rules-for-recorded-crime

House of Commons (2013). *Older prisoners*. Fifth report in session 2013-14. London: The Stationary Office.

Hu, L. T. & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,* 6, 1-55.

Hughes, R. (2000). Health, place and British prisons. *Health & Place*, 6, 57-62.

Humber, N., Emseley, R., Pratt, D. & Tarrier, T. (2013). Anger as a predictor of psychological distress and self-harm-ideation: A structured self-

assessment diary study. *Psychiatry Research*, 210,166-173.

IMB Corporation (2015). *SPSS for Macintosh, version 23.0*. Armonk, New York: IBM Corporation.

Jackson, D., Gillaspy, J. A. & Purc-Stephenson, R. (2009). Reporting practices in Confirmatory Factor Analysis: An overview and some recommendations. *Psychological Methods*, 14, 6-23.

Jenkins, R., Bhugra, D., Meltzer, H., Singleton, N., Bebbington, P., Burgha, T. *et al*. (2005). Psychiatric and social aspects of suicidal behavior in prisons. *Psychological Medicine*, 35, 257-269.

Jensen-Doss, A. & Hawley, K. M. (2011). Understanding clinicians' diagnosis practices: Attitudes towards the utility of diagnosis and standardized diagnostic tests. *Administration and Policy in Mental Health Services Research*, 38, 476-485.

Jiang, S. & Winfree, L. T. (2006). Social support and inmate adjustment to prison life. *The Prison Journal*, 86, 32-55.

Kalliath, T. J., O'Driscoll, M. P. & Brough, P. (2004). A confirmatory factor analysis of the General Health Questionnaire-12. *Stress and Health: Journal of the International Society for the Investigation of Stress,* 20, 1-20.

Karanicolas, P. J., Farrokhyar, F. & Bhandari, M. (2010). Blinding: Who, what, when, why how? *Canadian Journal of Surgery*, 53, 345-348.

Kendrick, T., King, F., Albertella, L., & Smith, P. W. F (2005). GP treatment decisions for patients with depression. *British Journal of General Practice,* 55, 280–286.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. *et al.* (2002). Short screening to monitor population prevalence's and trends in non-specific psychological distress. *Psychological Medicine*, 32, 959-976.

Kline, P. (1999). *The handbook of psychological testing* (2nd Edn). London: Routledge.

Lamb, R. & Weinberg, L. (1998). Persons with severe mental illness in jails and prisons a review. *Psychiatric Services*, 49, 483-492.

Layard, R., Clark, D.M., Knapp, M. & Mayraz, G. (2007). Cost– benefit analysis of psychological therapy. *National Institute Economic Review*, 202, 90–98.

Leach, C., Lucock, M., Barkham, M., Nobel, R., Clarke, L. & Iveson, S. (2005). Assessing risk and emotional disturbance using the CORE-OMand HoNOS outcome measures at the interface between primary and secondary mental healthcare. *The Psychiatric Bulletin*, 29, 419-422.

Lewis, G. & Meek, R. (2013). *A comprehensive health needs assessment of prisons and IRCs on behalf of NHS England.* London: NHS England.

Lindquist, C. (2000). Social integration and mental well-being among jail inmates. *Sociological Forum*, 15, 431-455.

Linehan, S. A., Duffy, D., M., Wright, B., Curtin, K., Monks, S. & Kennedy, G. (2005). Psychiatric morbidity in a cross-sectional sample of male remand prisoners. *Irish Journal of Psychiatric Medicine*, 22, 128-123.

Link, B., Phelan, J., Bresnahan, M. Stueve, A. & Pescoslido, B. (1999). Public conceptions of mental illness: Labels, cause, dangerousness and social distance. *American Journal of Public Health*, 89, 1329-1333.

Linquist, C. & Linquist, C. (1997). Gender differences in distress: Mental health consequences of environmental stress among jail inmates. *Behavioural Sciences and the Law*, 15, 303-523.

Little, K. (2013). *Prison Health Needs Assessment*. *Thematic review 2013: Mental health needs and provision across the Welsh prison estate.* NHS Wales: Public Health Wales.

Lohner, J. & Konrad, N. (2006). Deliberate self-harm and suicide attempts in custody: Distinguishing features in male inmate self-injurious behaviour. *International Journal of Law and Psychiatry*, 29, 370-385.

Louden, J. E., Skeem, J. L. & Blevins, A. (2013). Comparing the psychometric utility of two screening tools for mental disorder among prisoners. *Psychological Assessment*, 25, 405-415.

Lovell, D., Gagliardi, G. J., & Peterson, P. D. (2002). Recidivism and use of services among persons with mental illness after release from prison. *Psychiatric Services*, 53, 1290–1296.

Lyne, K. J., Barrett, P., Evans, C., & Barkham, M. (2006). Dimensions of variation on the CORE-OM. *British Journal of Clinical Psychology*, *45*, 185–203.

Maniglio, R. (2009). Severe mental illness and criminal victimization: A systematic review. *Acta Psychiatrica Scandinavica*, *119*, 180-191.

Martin, S. M., Colman, I., Simpson, A. I. F. & McKenzie, K. (2013). Mental health screening tools in correctional facilities: A systematic review. *BMC Psychiatry*, 13, 275-285.

Martin, S. M., Potter, B. K., Crocker, A. G., Wells, G. A. & Coleman, I. (2016). Yield and efficiency of mental health screening: a comparison of screening protocols at intake to prison. *PLOSone*, 11, doi:10.1371/journal.pone.0154106

Marshall, T., Simpson, S. & Stevens, A. (2001). Use of health services by prison inmates: Comparisons with the community. *Journal of Epidemiology and Community Health,* 55, 364–65

McCloskey, H. (2001). Evaluation of the CORE outcome measure in a therapeutic forensic setting. *The British Journal of Forensic Practice*, 3, 22-28.

McCorkle, R. C. (1995). Gender, psychopathology, and institutional behavior: A comparison of male and female mentally ill prison inmates. *Journal of Criminal Justice,* 23, 53-61.

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M. & Lawson-Smith, S. (2008). *Paying the price. The cost of mental health care in England to 2026. Kings Fund*. London.

McDonald, R. P. (1985). *Factor analysis and related methods*. New York: Erlbaum.

McGilloway, S. & Donnelly, M. (2004). Mental illness in the UK criminal justice system: A police liaison scheme for mentally disordered offenders in Belfast. *Journal of Mental Health*, 13, 263-275.

McNeil, D., Binder, R. & Robinson, J. (2005). Incarceration associated with homelessness, mental disorder and co-occurring substance abuse. *Psychiatric Services*, 56, 840-846.

McReynolds, L. S. & Wasserman, G. A. (2008). Risk for disciplinary infractions among incarcerated male youths. *Criminal Justice and*

Behavior, 35, 1147-1185.

MedCalc Software (2015). *MedCalc version 15.11*. Ostend, Belgium: MedCac

Medical Dictionary (2007). *Screening*. Retrieved on 4th February 2016 from http://medical-dictionary.thefreedictionary.com/screening

Meltzer, H., Lader, D., Corbin, I., Singleton, N., Jenkins, R. & Brugha, T. (2002). *Non-fatal suicide behaviour among adults aged 16–74 in Great Britain*. The Stationery Office: ONS, London.

Miaofen, Y. & Li-Hua, L. (2002). Examining re-test reliability: An interclass correlation approach. *Nursing Research*, 51, 59-62.

Mickeljohn, C., Hodges, K. & Capon, D. (2004). In-reach work with prisoners. *Mental Health Nursing*, 24, 8-10.

Ministry of Justice (2011*). Early days in custody – reception in, first night and induction to custody. Prison Service Instruction 74/2011*. Ministry of Justice: National Offender Management Service.

Ministry of Justice (2014a). *Offender management statistics prison population 2014*. London: Ministry of Justice.

Ministry of Justice (2014b). *Safety in custody statistics England and Wales. Deaths in custody to September 2014. Assaults and self-harm to June 2014*. Ministry of Justice statistics bulletin 30th October 2014. London: Ministry of Justice.

Ministry of Justice (2015a). *Population bulletin: weekly 18th March 2016.* London: Ministry of Justice and National Offender Management Service.

Ministry of Justice (2015b). *Proven re-offending statistics quarterly bulletin. January to December 2013, England and Wales*. London: Ministry of Justice.

Ministry of Justice (2015c). *Safety in custody statistics quarterly update to December 2014*. London: Ministry of Justice.

Mooney, C. & Duval, R. (1993). *Bootstrapping: A non-parametric approach to statistical inference*. Newbury Park: Sage publications.

Morgan, M. & Kett, M. (2003). *The prison adult literacy survey. Results and implications*. Dublin: Irish Prison Service.

Moser D., Arndt, S., Kanz ,J., Benjamin, M. L., Bayless, J. D., Reese, R. L.

et al. (2004). Coercion and informed consent in research involving prisoners. *Comprehensive Psychiatry*, 45, 1-9.

National Institute for Clinical Excellence (2011a). *Common mental health disorders: identification and pathways to care. QS11*. London: NICE.

National Institute for Health and Clinical Excellence (NICE) (2011b). *Self-Harm: Long term management CG133*. Manchester: National Institute for Health and Clinical Excellence.

NHS England (2009). *Improving Access to Psychological Therapies (IAPT). Offenders positive practice guidance*. London: NHS England.

NHS England (2013). *Improving Access to Psychological Therapies (IAPT). Offenders positive practice guidance*. London: NHS England.

Nicholls, R. R., Roesch, M., Olley, J. & Hemphill, J. F. (2005). *Jail Assessment Tool (JSAT) Manual*. Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

Nurse, J., Woodcock, P. & Ormsby J. (2003). Influence of environmental factors on mental health within prisons: Focus group study. *British Medical Journal*, 327, 480– 485.

Offender Health Research Network (OHRN) (2008). *An evaluation of the reception screening process used within prisons in England and Wales.* Manchester: OHRN.

Offender Health Research Network (OHRN) (2009). *A national evaluation of prison mental health in reach services. Report for the National Institute of Health Research*. Offender Health Research Network: University of Manchester.

Offender Health Research Network (OHRN) (2010). *The pathway of prisoners with mental health problems through prison health services and the effect of the prison environment on the mental health of prisoners. A report to the National Institute of Health Research.* Manchester: OHRN.

Ogloff, J. R. P., Roesch, R., & Hart, S. D. (1994). Mental health services in jails and prisons: Legal, clinical, and policy issues. *Law & Psychology Review,* 18, 109-136.

Owens, D., Horrocks, J., & House, A. (2002). Fatal and non-fatal repetition of self-harm. *British Journal of Psychiatry*, 181, 193–199.

Paton, J., & Borrill, J. (2004). Prisons. In D. Duffy & T. Ryan (Eds.), *New approaches to preventing suicide: A manual for practitioners* (pp. 114–130). London: Jessica Kingsley.

Perry, T., Barkham, M. & Evans, C. (2013). The CORE-OM (SV) in secure settings: A template analysis of the experiences of male patients and their staff. *Journal of Forensic Practice*, 15, 32-43.

Pintea, S., Moldovan, R. (2009). The Receiver-Operating Characteristic (ROC) analysis: Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies*, 1, 49-66.

Pratt, D., Piper, L., Appleby. L., Webb, R. & Shaw, J. (2006). Suicide in recently released prisoners: A population-based cohort study. *Lancet*, 368, 119-123.

Pratt, D., Goodling, P., Eccles, S., Awenat, Y. & Tarrier, N. (2015). Cognitive behavioural suicide prevention for male prisoners: case examples. *Cognitive Behavioural Practice*, DOI: http://dx.doi.org/10.1016/j.cbpra.2015.09.006

Priebe, S., Badesconyi, A., Fioritti, A., Hansson, L., Kilian, R. & Torres-Gonzales, F. *et al.* (2005). Reinstitutionalization in mental health care: Comparison of data on service provision from six European countries. *British Medical Journal*, 330, 123-126.

Prins, S., Osher, F., Steadmanm, H., Robbins, P. & Case, B. (2012). Exploring racial disparities in the brief jail mental health screen. *Criminal Justice and Behavior*, 39, 635-645.

Reed, J. (2003). Mental health care in prisons. *British Journal of Psychiatry*, 182, 287-288.

Resnick, P. J. & Knoll, J. L. (2008). Malingered psychosis. In Rogers, R. (Ed.) *Clinical assessment of malingering and deception* (3rd edn., pp. 51-68). New York: Guilford Press.

Richardson, R., Tepel, D., Perry, A., Ali, S., Duffy, S., Gabe, R. *et al*. (2015). Screening for psychological and mental health difficulties in young people who offend: A systematic review and decision model. *Health Technology and Assessment*, 19. Retrieved 2nd June 2015 from http://dx.doi.org/10.3310/hta19010

Richter, P., Scheurer, H. & Krober, H. (1991). Measuring depression of prisoners. *Zeitschrift fur Klinische Psychologie, Psychopathologie und Psychotherapie*, 39, 382-388.

Rivlin, A., Hawton, K., Marzano, L. & Fazel, S. (2010). Psychiatric disorders in male prisoners who made near-lethal suicide attempts: A case control study. *The British Journal of Psychiatry*, 197, 313-319.

Rivlin, A., Hawton, K., Marzano, L. & Fazzel, S. (2013). Psychosocial characteristics and social networks of suicidal prisoners: Toward a model of suicidal behaviour in detention. *PLOS ONE*, 8, e68944.

Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd edn.). New York, London: Guilford Press.

Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, *67,* 629–640.

Rogers, R., Sewell, K., Ustad, K., Reinhardt, V & Edwards, W. (1995). A preliminary study of convergent and discriminant validly. *Law and Human Behavior*, 19, 481-492.

Rosenthal, R. & Rosnow, R. L. (1975). *The volunteer subject*. New York: Wiley.

Ruggeri, M., Leese, M., Thornicroft, G., Bisoffi, G. & Tansella. (2000). Definition and prevalence of severe and persistent mental illness. *British Journal of Psychiatry*, 177, 149-155.

Sackett, D. L. & Haynes, R. B. (2002). The architecture of diagnostic research. In J. Knottnerus (Ed.) *The evidence base of clinical diagnosis* (pp.19-38). London: BMJ Books.

Scott, S. (2004). Opening a can of worms? Counseling for survivors in UK women's prisons. *Feminist and Psychology*, 14, 605-608.

Sharifi, F., Mousavinasab, N., Mazloomzedeh, S., Jaberi, Y., Saeini, M., Dinmohammadi, M. *et al*. (2008). Cut off point of waist circumference for the diagnosis of metabolic syndrome in an Iranian population. *Obesity Research & Clinical Practice*, 2, 1-11.

Shaw, J., Baker, D., Hunt, I., Maloney, A. & Appleby, L. (2004). Suicide by prisoners: National clinical survey. *British Journal of Psychiatry*, 184,

263–267.

Shaw, J., Tomenson, B. & Creed, F. (2003). A screening questionnaire for the detection of serious mental illness in the criminal justice system. *The Journal of Forensic Psychiatry and Psychology*, 14, 138-150.

Sheehan, D., Lecrubier, Y., Harnett-Sheehan, K., Janavas, J., Weiller, E., Bonara, L. I. *et al* (1997). Reliability and validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, 12, 232–241.

Shevlin, M. & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: A large sample analysis using confirmatory factor analysis. *Psychological Assessment* 17, 231-236.

Silver, E., Felson, R., Vaneseltine, M. (2008). The relationship between mental health problems and violence among criminal offenders. *Criminal Justice and Behavior*, 35, 405-426.

Silver, S. B., Cohen, M. I., & Spodak, M. K. (1989). Follow-up after release of insanity acquittees, mentally disordered offenders, and convicted felons. *Bulletin of the American Academy of Psychiatry and the Law*, 17, 387–400.

Sim, J. (1994). Prison medicine and social justice. *Prison Service Journal*, 95, 30-38.

Simmonds, S., Coid, J., Joseph, P., Mariott, S. & Tyrer, P. (2001). Community mental health team management in severe mental illness: A systematic review. *British Journal of Psychiatry*, 178, 497 - 502.

Singleton, N., Meltzer, H. & Gatward, R. (1998*). Psychiatric morbidity among prisoners*. London: Office for National Statistics.

Slade, K. & Edelman, R. (2014). Can theory predict the process of suicide on entry to prison? Predicting dynamic risk factors for suicide ideation in a high-risk population. *Crisis*, 35, 82-89.

Smith, C., & Borland, J. (1999). Minor psychiatric disturbance in women serving a prison sentence: The use of the General Health Questionnaire in the estimation of the prevalence of non-psychotic disturbance in women prisoners. *Legal and Criminological Psychology*, 4, 273-384.

Sogaard, A., Selmer, R., Bjertness, E. & Thelle, D. (2004). The Oslo health study: the impact of self-selection in a large, population based survey. *International Journal for Equity in Health*, 3. Retrieved 7th September 2015 from DOI: 10.1186/1475-9276-3-3.

Song, F., Parekh-Bhurke, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. *et al*. (2009). Extent of publication bias in different categories of research cohorts: A meta-analysis of empirical studies. *BMC Medical Research Methodology*, 9. Retrieved 24th July 2015 from doi: 10.1186/1471-2288-9-79.

Steadman, H. J., Mulvey, E. P., Monahan, J., Robbins, P. C., Appelbaum, P. S. & Grisso, T. *et al*. (1998). Violence by people discharged from acute psychiatric inpatient facilities by others in the same neighborhoods. *Archives of General Psychiatry*, *55*, 393-401.

Steadman, H. J., Scott, J. E., Osher, F., Agnese, T. K. & Robbins, P. C. (2005). Validation of the Brief Jail Mental Health Screen. *Psychiatric Services*, 56, 816-822.

Steadman, H., Osher, F., Robbins, P., Case, B. & Samuels, S. (2009). Prevalence of serious mental illness among jail inmates. *Psychiatric services*, 60, 761- 765.

Steadman, H., Osher, F., Robbins, P., Case, B. & Samuels, S. (2009). Prevalence of serious mental illness among jail inmates, *Psychiatric Services*, 60, 761-765.

Steel, J., Thornicroft, G., Birmingham, L., Mills, A., Harty, M. & Shaw, J. (2007). Prison mental health in-reach services. *British Journal of Psychiatry*, 190, 373-374.

Steiner, B. & Wooldredge, J. (2008). Inmate versus environmental effects on prison rule violations. *Criminal Justice and Behaviour*, 35, 438-456.

Streiner, D. L. & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry*, 52, 121-128.

Swartz, M., Swanson, J., Hiday, V., Borum, R., Wagner, R. & Burns, J. (1998). Violence and severe mental illness: The effects of substance abuse and non-adherence to medication. *American Journal of Psychiatry*, 155, 226-231.

Taylor, P. J., Walker, J., Dunn, E., Kissell, A., Williams, A. & Amos, T. (2010). Improving mental state early in imprisonment. *Criminal Behavior and Mental Health,* 20, 215-231.

Teplin, L. A. & Swartz, J. (1989). Screening for severe mental disorder in jails: the development of the Referral Decision Scale. *Law and human Behaviour*, 13, 1-18.

Teplin, L. Abram, K & McLelland, G. (1997). Mentally disordered women in jail: Who receives services? *American Journal of Public Health*, 87, 604-609.

Teplin, L, A. (1994). Psychiatric and substance abuse disorders among male urban jail detainees. *American Journal of Public Health*, 84, 290–293.

Topp, D. O. (1979). Suicide in prison. *British Journal of Psychiatry* 134, 24-27.

Torch, H. & Adams, K. (1986). Pathology and disruptiveness among prison inmates. *Criminal Justice and Behaviour*, 14, 93-109.

Trauer, T., Gill, L., Pedwell, G., & Slattery, P. (2006). Routine outcome measurement in public mental health – what do clinicians think? *Australian Health Review,* 30,144–147.

Trestman, R., Appelbaum, K. & Metzner, J. (2015). *Oxford textbook of correctional psychiatry*. Oxford & New York: Oxford University Press.

Tyrer, P., Coid, J., Simmonds, S., Joseph, P. & Marroitt, S. (2003). *Community mental health teams (CMHTs) for people with severe mental illnesses and disordered personality (Cochrane Review)*. Cochrane Library, issue 2. Oxford: Update Software.

van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, 25, 40-48.

Vaz, S. Falkmer, T., Passmore, A., Parsons, R & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One*, 8, e73990.

Veysey, B. M., Steadman, H. J., Morrissey, J. P., Johnsen, M. & Beckstead, J. W. (1998). Using the Referral Decision Scale to screen mentally ill

jail detainees: Validity and implementation issues. *Law and Human Behaviour*, 22, 305-315.

Vincent, W. J. (1999). *Statistics in Kinesiology.* Champaign, IL: Human Kinetic.

Walters, G. D. (2011). Criminal thinking as a mediator of the mental illness-prion violence relationship: A path analysis study and causal mediation analysis. *Psychological Services*, 8, 189-199.

Watson, R., Stimpson, A. & Hostick, T. (2004). Prison health care: a review of the literature. *International Journal of Nursing Studies*, 41, 119-128.

Weiner, I. B. & Graham, J. R. (2003). *Handbook of psychology, assessment psychology*. London: John Wiley and Sons.

Welsh Government (2011). *Mental health (Wales) measure 2010: National service model for local primary mental health support services*. Cardiff: Welsh Government.

Welsh Government (2012a). *Together for Mental Health: A Strategy for Mental Health and Wellbeing in Wales.* Cardiff: Welsh Government.

Welsh Government (2012b). *Government Policy Implementation Guidance on Local Primary Mental Health Support Services and Secondary Mental Health Services for the purposes of the Mental Health (Wales) Measure 2010 and related subordinated legislation.* Cardiff: Welsh Government.

Welsh Government (2012c). *The Psychological Therapies in Wales Implementation Guidance.* Cardiff: Welsh Government.

Welsh Government (2014). *Policy implementation guidance. Mental health services for prisoners*. Cardiff: Welsh Government.

Werneke, U., Goldberg, D. P., Yalcin I. & Ustun, B. T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine,* 30, 823-829.

Werner, R. E. & Keys, C. (1988). The effects of changes in jail population densities on crowding, sick call and spatial behavior. *Journal of Applied Social Psychology*, 18, 852-866.

Westmoreland, P., Gunter, T., Loveless, P., Allen, J., Sieleni, B., & Black, D. W. (2010). Attention deficit hyperactivity disorder in men and women

newly committed to prison. *International Journal of Offender Therapy and Comparative Criminology*, 54, 361–377.

Whiting, P., Rutjes, A., Westwood, M., Mallett, S., Deeks, J., Reistman, J *et al.* (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155, 529-537.

Wilde, G. J. S. (1972). Trait description and measurement by personality questionnaires. In Cattell, R. (Ed.) *Handbook of modern personality theory* (pp. 93-103). Chicago: Aldine.

Wilper, A., Woolhandler, J., Boyd, W., Lasser, K., McCormick, D., Bor, D. *et al*. (2009). The health and health care of US prisoners: results of a nationwide survey. *American Journal of Public Health*, 99, 666-672.

Wilson, J. M. G. & Jungner G. (1968). *Principles and practice of screening for disease*. Geneva: WHO. Retrieved on 20th January 2016 from: http://apps.who.int/iris/bitstream/10665/37650/1/WHO_PHP_34.pdf

Wolff, N., & Shi, J. (2009). Victimisation and feelings of safety among male and female inmates with behavioural health problems. *Journal of Forensic Psychiatry & Psychology,* 20*,* S56-S77.

Wolff, N., Blitz, C. L., & Shi, J. (2007). Rates of sexual victimization in prison for inmates with and without mental disorders. *Psychiatric Services*, *58*, 1087-1094.

Wood, S. R. & Buttaro, A. (2013). Co-occurring severe mental illness and substance abuse disorders as predictors of state prison inmate assaults. *Crime and Delinquency*, 59, 510-535.

Wright, K. (1991). Environmental and interactive effects in explaining adjustment to prison. *Justice Quarterly*, 8, 217-239.

6 Appendixes

6.1 Appendix 1. Critical Skills Appraisal Programme (2013) framework for assessing diagnostic test studies

6.2 Appendix 2. Ethical Approval

Ethics Feedback - EC.15.05.12.4136R2

Wed 24/06/2015 10:00

Dear Gwen,

The Chair of the Ethics Committee has considered your revised postgraduate project proposal: Screening for common mental health problems in prisoners (EC.15.05.12.4136R2).

The project has now been approved.

Please note that if any changes are made to the above project then you must notify the Ethics Committee.

Best wishes,
Natalie

6.3 Appendix 3. NOMS National Research Committee Approval

Miss Gwen Lewis
Doctoral Programme in Clinical Psychology
11th Floor, School of Psychology
Tower Building
70 Park Place
Cardiff CF10 3AT
lewisge@cardiff.ac.uk

National Offender Management Service
National Research Committee
Email: National.Research@noms.gsi.gov.uk

13th April 2015

National Offender Management Service

APPROVED SUBJECT TO MODIFICATIONS – NOMS RESEARCH

Ref: 2015-056
Title: Screening for Common Mental Health Problems in Prisoners in Wales

Dear Gwen,

Further to your application to undertake research across NOMS, the National Research Committee (NRC) is pleased to grant approval in principle for your research. The Committee has requested the following modifications:

This image has been removed by the author for copyright reasons.

Yours sincerely,
National Research Committee

6.4 Appendix 4. Stakeholder consultation process

Pre-implementation

The following stakeholders were consulted with through six meetings prior to implementing the research project:

1. **Welsh Government and National Offender Management Service commissioners**
    - Discussion of the political and strategic relevance of the proposed research project.
    - Discussion of possible related areas to consider and broader research streams to link with.
    - Sign posting from government advisors and NOMS commissioners to other interested parties.
2. **Prison senior management teams and research boards**
    - Negotiation of access.
    - Discussion of operational practicalities including risk management plans (see appendix 8).
    - Explanation of possible measures given and feedback sought.
3. **Prison health care teams**
    - In depth discussion of existing practice and the strengths and weaknesses of this.
    - Consultation and advice regarding choice of measures.
    - Planning for implementation of research design.

Dissemination of research findings post completion

The research findings were disseminated to stakeholders through the following means in order to support implementation of the research findings in practice:

1. Presentation at prison research fair (attended by prison senior management, Welsh Government representatives, prison staff and academics).
2. Presentation and discussion with managers and healthcare teams at each prison.
3. Presentation and discussion with Welsh Government and NOMS commissioners.
4. Presentation and discussion with the Prison Health Improvement Network.
5. Research summary submitted to NOMS National Research Committee.
6. Plans to disseminate via peer reviewed publication.

6.5 Appendix 5. Participant Information Sheet

PARTICIPANT INFORMATION SHEET

Version 2.0 April 2015

Title of study: **Improving screening for Common Mental Health Problems in Prisoners**

Researchers: External research, Cardiff University

We work for Cardiff University and would like to invite you to take part in this study to find out how screening for common mental health problems can be improved for prisoners. The study looks at different screening questionnaires to see which would be best for use in prisons.

What does it involve?

Taking part in the study involves answering some brief questions about how you are feeling now and then again in two weeks' time. It will take about 30 minutes the first time and about 10 minutes the second time.

Why have I been chosen?

Everyone who has recently come into the prison while the research is taking place and can speak English is being offered the chance to take part.

Do I have to take part?

No! You only take part if you want to. Even if you begin to complete the questionnaire you can stop at any time without giving a reason. If you don't take part or decide to stop, it will not affect any of the services you receive or your Incentives and Earned Privileges (IEPs) status.

Potential benefits and disadvantages of taking part

We hope you will find it interesting, and by taking part you will be helping to improve knowledge of how prisons can better screen for common mental health problems in future. But if you find any of the questions upsetting you can stop completing the questionnaire and. if you want me to, I can let a member of prison staff know so that you can get support.

If you are upset after taking part you can talk to your personal officer, a listener or make an

appointment with healthcare yourself.

Is it confidential?

Yes, all the information you provide will be kept confidential unless you say something that means there is a risk to you or someone else, or if you tell me about an illegal act or behaviour that breaks prison rules and can be adjudicated against. If you did say something like this, I would have to let prison staff know and prison rules would apply.

Your questionnaire will be kept securely in a locked cabinet for 5 years and the front page with your prison number on will be kept separately so that your answers can't be traced back to you (it will be anonymous). You can pull out of the study at any point up until your questionnaire is made anonymous by letting [NAMED CONTACT AT EACH PRISON] know.

Why do we need your prisoner number?

We need your prisoner number so that:

- We can find out if you are offered a mental health service in this prison in the next month.
- We can get in touch with you in two weeks to complete the questionnaire again.
- We can link the information you give us the first time with the questionnaire you fill in two weeks later.

What will happen to the results of the study?

When everyone's information has been put together so that it is anonymous (it can't be linked to anyone who took part) it will be analyzed by the researchers and written up as part of a project for Cardiff University.

Who has said that the study is OK to go ahead?

The research study has been reviewed and approved by the School of Psychology Research Ethics Committee at Cardiff University. If you have any concerns or complaints about the research you can contact [NAMED CONTACT AT EACH PRISON] through the complaints system, who will pass on your concern to the Ethics Committee.

If you would like more information about the project, please feel free to ask me now or ask [NAMED CONTACT AT EACH PRISON] if you have a question afterwards.

Thank you for reading this!

6.6 Appendix 6. Consent form

CONSENT FORM

Version 2.0 April 2015

Title of study: **Improving screening for Common Mental Health Problems in Prisoners**

Researchers: External research Cardiff University

I understand that my participation in this study will involve answering some brief questions about how I feel now and then again in two weeks' time and will take about 30 minutes the first time and about 10 minutes the second time. ☐

I have read and understood the information sheet and have been able to ask any questions. ☐

I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason. This will not affect my access to services or my Incentives and Earned Privileges (IEPs) status.

I understand that I am free to ask any questions at any time. I can discuss any concerns with [NAMED CONTACT AT EACH PRISON].

I understand that the information provided by me will be kept securely and anonymously and will be held for 5 years.

I understand that the information I give will be confidential (only seen by the researchers) unless I give information that means there is a risk to me or someone else, or if I tell you about an illegal act or behaviour that breaks prison rules and can be adjudicated against; I understand that you will tell prison staff if I give such information.

I understand that the researchers are external to the prison and that taking part in this study WILL NOT make any difference to me accessing health services in prison. I know that if I am concened about my mental health I need to make a healthcare application.

I understand that my prisoner number will be used by the researcher:

a) To see if I am offered a mental health service in this prison in the next month

b) To get in touch with me to complete the questionnaire again in two weeks' time

c) To link the information I give now with my questionnaire that I will fill in two weeks' time

8. I understand that the researchers will NOT access any other information held about me.

I, _____(NAME) consent to participate in the study conducted by Cardiff University

Signed:……

Date……

6.7 Appendix 7. Debrief sheet

This study was about improving screening for common mental health problems in prisoners and was carried out by Cardiff University.

Why is this important to study?

Very little research has been carried out in the past looking at screening tools for identifying common mental health problems in prisoners, but many prisoners do experience common mental health problems like anxiety and depression. It is therefore important that screening tools are tested so that people's needs can be identified and support offered in future.

How was this done?

You completed two different screening questionnaires, which will be compared to the clinical interview you took part in and whether you get offered a mental health service in the next month. This comparison will allow us to see which screening tool is best at identifying mental health needs of people in prison.

Main questions in this study

- Which screening tool is best at identifying people with common mental health problems?
- Which screening tool is best at telling the difference between people who have common mental health problems and severe mental health problems?
- Which screening tool is most reliable over time?

What will happen to the information I have given?

When everyone's information has been put together so that it is anonymous (it can't be linked to anyone who took part) it will be analyzed by the researchers and written up as part of a project for Cardiff University.

Your questionnaire will be kept securely and the front page with your prison number on will be kept separately so that your answers can't be traced back to you (it will be anonymous). You can pull out of the study at any point up until your questionnaire is made anonymous by letting [NAMED CONTACT AT EACH PRISON] know.

What if I feel I need help for my mental health?

If you are concerned about your mental health or feel you would like help with an emotional problem please put in a health care application or talk to your personal officer. We are external researchers and do not work for the prison or healthcare so taking part in this study will not mean you are or are not offered services.

What if I want to know more?

If you would like more information about the project, please feel free to ask me now or ask [NAMED CONTACT AT EACH PRISON] if you have a question afterwards and they will get in touch with me.

Thank you for taking part!

6.8 Appendix 8. Risk management protocol

The following risk management protocol was agreed with prisons to manage disclosure of risk to self, others, security or significant distress as these issues were anticipated to arise in the course of the research:

1. **Risk to self**

   Action would be taken if risk to self in terms of the following were identified:

   - Disclosure of current suicidal plans and or preparation (through CORE-10, MINI or general conversation)
   - Disclosure of plans to hurt self (through MINI suicidality interview or in general conversation)
   - High suicidality score on the MINI

   ACTIONS:

   - Inform prisoner of need to break confidentiality (unless this presents risk to the researcher)
   - Ask prisoner if they are under Assessment Care in Custody and Teamwork (ACCT) procedures used to monitor and support prisoners who pose a risk to themselves.
   - If risk to self appears imminent remain with the prisoner and call over prison staff to observe the prisoner.
   - Check with wing staff if the prisoner is under ACCT procedures, if no ACCT is open, open an ACCT.
   - Inform wings staff of concerns identified and ensure staff record this on the computerised system P-NOMIS.
   - Notify nominated research lead at the prison if any issues arise in applying the above protocol.

2. **Risk to others/security**

   Action would be taken if risk to others or security of the following nature were identified:

   - Threats of violence/abuse to others
   - Previously undisclosed illegal acts
   - Breaches of prison rules (e.g. possession of drugs, weapons, escape plans)

   ACTIONS:

   - Inform wings staff of concerns identified and ensure staff record this on the computerised system P-NOMIS

- With wing staff submit a Security Information Report.
- Notify nominated research lead at the prison of prisoner details, risk issue and action taken.

3. Significant distress

Action would be taken if a prisoner was displaying significant distress requiring immediate support:

- If safe to do so tell prisoner that confidentiality will be breached in order for them to gain support
- Notify wing staff of prisoners distress and ask them to provide support and record this on P-NOMIS
- If appropriate notify Prison Mental Health team (N.B. if concerns are regarding florid psychosis always notify mental health team in addition to wing staff).
- Advise prisoner to submit health care application for on-going support.

6.9 Appendix 9 Time one interviewer administered questionnaire

N.B. Instructions for the interviewer are in CAPITALS, what is to be read to the participant is in lower case.

Improving screening for Common Mental Health Problems in Prisoners

- PROVIDE INFORMATION SHEET (READ TO PARTICIPANT IF NECESSARY)
- ANSWER ANY QUESTIONS
- IF HAPPY TO PARTICIPATE PROVIDE CONSENT FORM TO READ (READ TO PARTICIPANT IF NECESSARY) AND SIGN

Thank you for choosing to take part in this study!

Today involves spending about 30 minutes answering some questions about how you feel. I will then come and see you again in two weeks' time with a similar very short questionnaire for you to fill in. As part of the study we would also like to find out if you are offered any mental health service in this prison.

Please provide your prisoner number so we can:

- Get in touch with you in two weeks
- See if you are offered a mental health service here in the next month

Your prisoner number will not be used to access any other information about you.

Your questionnaire will be kept separately from your prisoner number so that you can't be identified.

IF HAPPY TO PROVIDE ASK FOR PRISONER NUMBER

Prisoner number:

I am an external researcher; prison staff will not see what you have written, and taking part in this study will not help you or stop you from getting healthcare services.

We will only let prison staff know if you say or write something that means there is a risk to you or someone else, or if you tell us about an illegal act or behaviour that breaks prison rules and can be adjudicated against.

Please answer all the questions honestly even if some seem very similar to each other. We are interested in your own experiences; there are no right or wrong answers.

About you

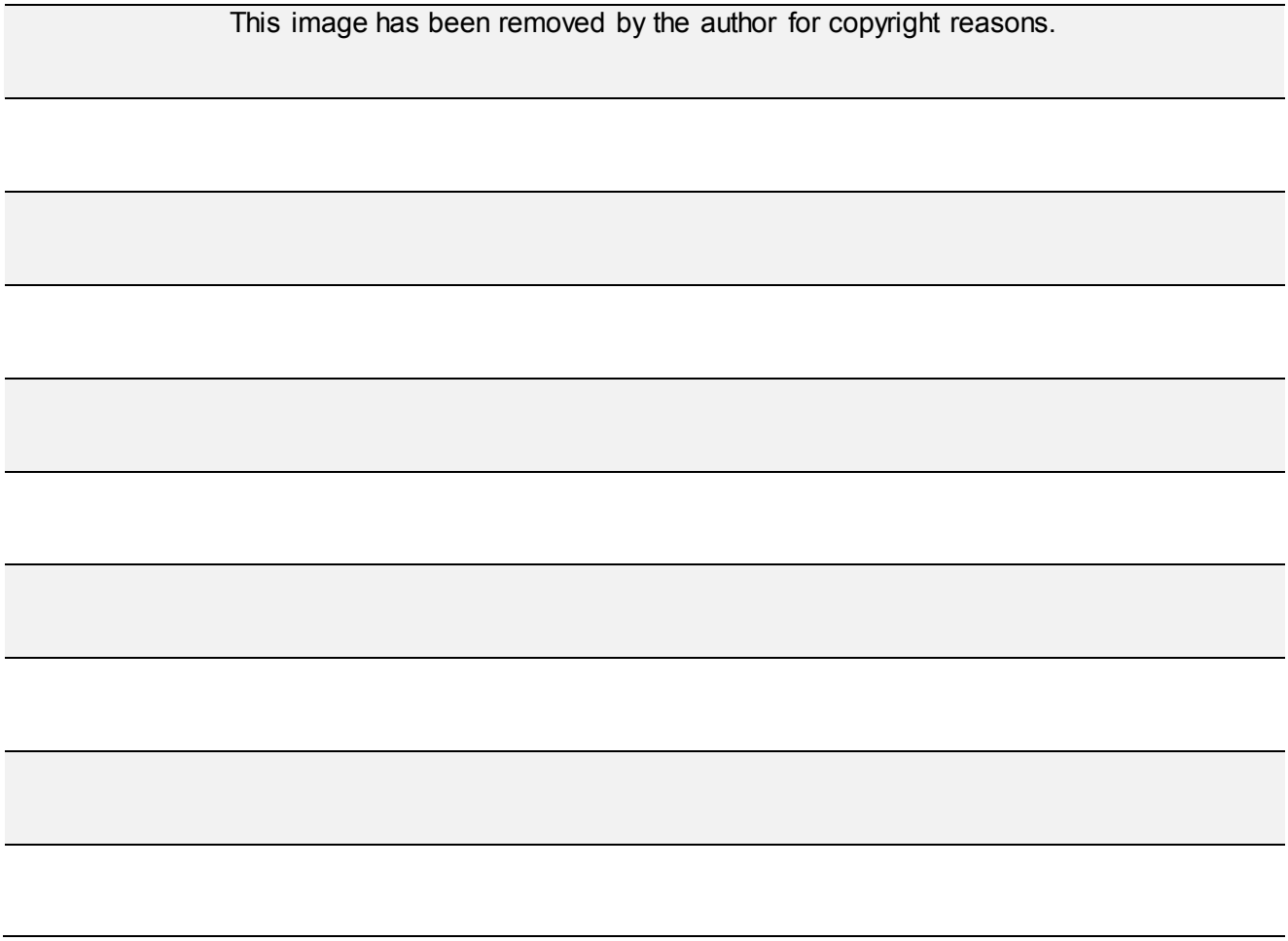
How old are you (in years)?		
What is your ethnicity:	○ White British or Irish ○ Other White background ○ Black British ○ Other Black background	○ Asian British ○ Other Asian background ○ Mixed ○ Other (please specify) .....
Are you a foreign national?	○ Yes ○ No	
Is English your first language?	○ Yes ○ No	
Is this your first time in prison?	○ Yes ○ No	
Are you:	○ Remand ○ Sentenced ○ Convicted un sentenced	○ Recall ○ Detainee
How long is your sentence?	○ Not sentenced ○ Less than 6 months ○ 1 year to less than 2 years ○ 2 year to less than 4 years	○ 4 years to less than 10 year ○ 10 years or more ○ IPP ○ Life
How long have you been in this prison?	○ Less than 2 days ○ 2 to less than 5 days ○ 5 to less than 7 days ○ 7 days to less than 14 days ○ 14 days to –1 month ○ 1 to less than 3 months	○ 3 to less than 6 months ○ 6 to less than 12 months ○ 12 month to less than 2 years ○ 2 to less than 4 years ○ More than 4 years
What was you index offence?	○ Violence against the person ○ Sexual offences ○ Burglary ○ Robbery ○ Theft and handling ○ Arson and criminal damage	○ Drugs offences ○ Possession of weapons ○ Public order offences ○ Civil offences ○ Fraud and forgery ○ Other
Are you a veteran (ex-military)?		○ Yes ○ No
Were you homeless in the year before coming to prison?		○ Yes ○ No
Did you have a problem with drugs in the year before coming to prison?		○ Yes ○ No
Did you have a problem with alcohol in the year before coming to prison?		○ Yes ○ No
Have you ever had contact with mental health services?		○ Yes ○ No
Were you being prescribed an antidepressant before you came to prison?		○ Yes ○ No
Have you ever self-harmed?		○ Yes ○ No

PLACE FORM IN FRONT OF PARTICIPANTS SO THEY CAN SEE RESPONSE CATEGORIES AND TICK

This form has 10 statements about how you have been over the last week.

Read or listen to each statement and think how often you felt that way last week.

Then tick the box which is closest to this.



PLACE FORM IN FRONT OF PARTICIPANTS SO THEY CAN SEE RESPONSE CATEGORIES AND TICK This form has 12 statements about how you have been over the last few weeks. Read or listen to each statement and think how often you felt that way over the last few weeks. Then circle which is closest to this.

Have you recently

ADMINISTER MINI INTERNATIONAL NEUROPSYCHIATRIC INTERVIEW (MINI 6.0) MODULES:

EXPLAIN:

- The next section asks some similar questions to help us to see if the screening questionnaires we just did are any good.
- The questions are quite structured and to keep it short- I just need yes/no answers from you.
- You will have the chance to ask more questions at the end.

Thank you for taking part! Do you have any questions?

It would be really helpful if you are happy to do a similar but much shorter questionnaire in about two weeks' time when we will come in to see you again. Thank you.

MINI Score Sheet

Disorder	Yes	No
Major Depression episode		
Current		
Past		
Recurrent		
Major Depressive Disorder (MDD)		
Current		
Past		
MDD with psychotic features		
Current		
Past		
Suicidality		
Low		
Moderate		
High		
Manic episode		
Current		
Past		
Hypomanic episode		
Current		
Past		
Bipolar I		
Current		
Past		
Bipolar II		
Current		
Past		
Bipolar NOS		
Current		
Past		
Panic Disorder		
Current		
Past		
Panic Disorder with agoraphobia (current)		
Panic Disorder without agoraphobia (current)		
Agoraphobia without panic disorder		
Social phobia (current)		
Obsessive Compulsive Disorder (current)		
Post-Traumatic Stress Disorder (current)		
Mood disorder with psychosis		
Current		
Lifetime		
Psychotic disorder		
Current		
Lifetime		
Generalised Anxiety Disorder (current)		
ANY CONDITION		

6.10 Appendix 10. Time two interviewer administered questionnaire

Improving screening for Common Mental Health Problems in Prisoners

Thank you for taking part in the first part of this study and for agreeing to complete another very short questionnaire.

Detailed information about the study is on the information sheet (which is the same as before) for you to keep.

Please put your prisoner number so that:

- We can match this questionnaire to the information you gave me before
- See if you are offered a mental health service here in the next month

Your prisoner number will not be used to access any other information about you.

Your questionnaire will be kept separately from your prisoner number so that you can't be identified.

I am an external researcher, prison staff will not see what you have written and taking part in this study will not help you or stop you from getting healthcare services.

We will only let prison staff know if you say or write something that means there is a risk to you or someone else, or if you tell us about an illegal act or behaviour that breaks prison rules and can be adjudicated against.

Please answer all the questions honestly even if some seem very similar to each other. We are interested in your own experiences; there are no right or wrong answers.

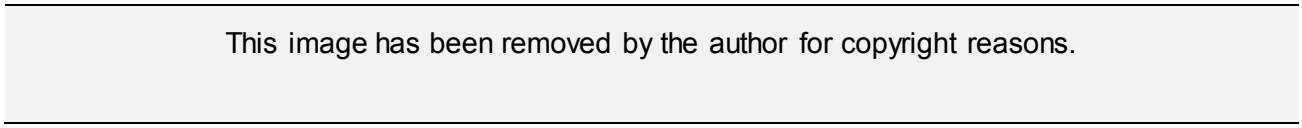
Thank you!

This form has 12 statements about how you have been over the last few weeks.

Read each statement and think how often you felt that way over the last few weeks.

Then circle which is closest to this.

Have you recently



Please turn over

This form has 10 statements about how you have been over the last week.

Read each statement and think how often you felt that way last week.

Then tick the box which is closest to this.

Over the last week	Not at all	Only occasionally	Sometimes	Often	Most or all of the time
This image has been removed by the author for copyright reasons.					

Thank you!

6.11 Appendix 11. Sensitivity and specificity at differing cut points on the CORE-10

CORE-10 sensitivity and specificity across cut points for primary care need

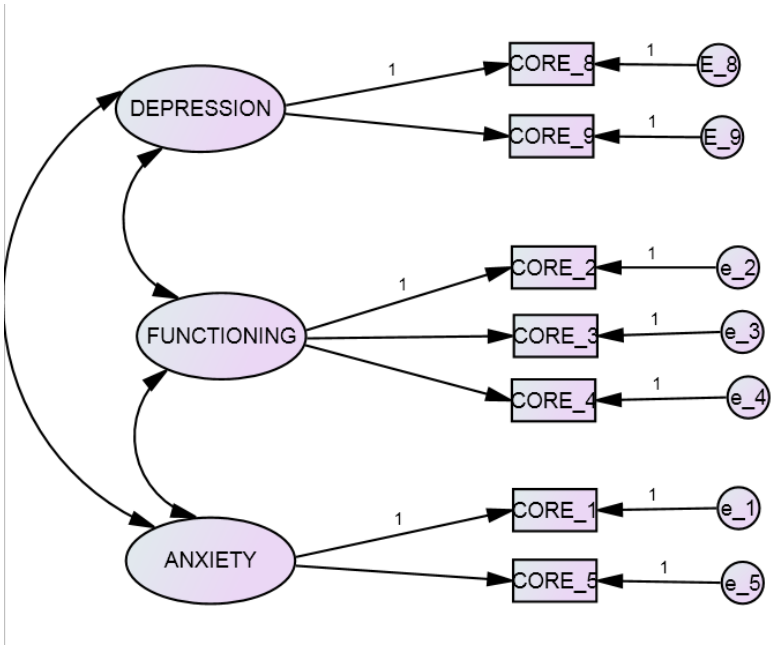
CORE-10 cut point	Sensitivity	95% CI	Specificity	95% CI
>4	91.30	83.6 - 96.2	43.10	30.2 - 56.8
>5	89.13	80.9 - 94.7	55.17	41.5 - 68.3
>6	88.04	79.6 - 93.9	63.79	50.1 - 76.0
>7	83.70	74.5 - 90.6	70.69	57.3 - 81.9
>8	78.26	68.4 - 86.2	77.59	64.7 - 87.5
>9	76.09	66.1 - 84.4	81.03	68.6 - 90.1
>10	73.91	63.7 - 82.5	84.48	72.6 - 92.7
>11	70.65	60.2 - 79.7	86.21	74.6 - 93.9
>12	65.22	54.6 - 74.9	86.21	74.6 - 93.9
>13	63.04	52.3 - 72.9	91.38	81.0 - 97.1
>14	59.78	49.0 - 69.9	91.38	81.0 - 97.1
>15	56.52	45.8 - 66.8	93.10	83.3 - 98.1

CORE-10 sensitivity and specificity across cut points for secondary care need

CORE-10 cut point	Sensitivity	95% CI	Specificity	95% CI
>7	95.12	83.5 - 99.4	49.54	39.8 - 59.3
>8	85.37	70.8 - 94.4	54.13	44.3 - 63.7
>9	82.93	67.9 - 92.8	56.88	47.0 - 66.3
>10	82.93	67.9 - 92.8	60.55	50.7 - 69.8
>11	78.05	62.4 - 89.4	62.39	52.6 - 71.5
>12	73.17	57.1 - 85.8	65.14	55.4 - 74.0
>14	73.17	57.1 - 85.8	72.48	63.1 - 80.6
>15	68.29	51.9 - 81.9	74.31	65.1 - 82.2
>16	58.54	42.1 - 73.7	78.90	70.0 - 86.1
>17	51.22	35.1 - 67.1	80.73	72.1 - 87.7

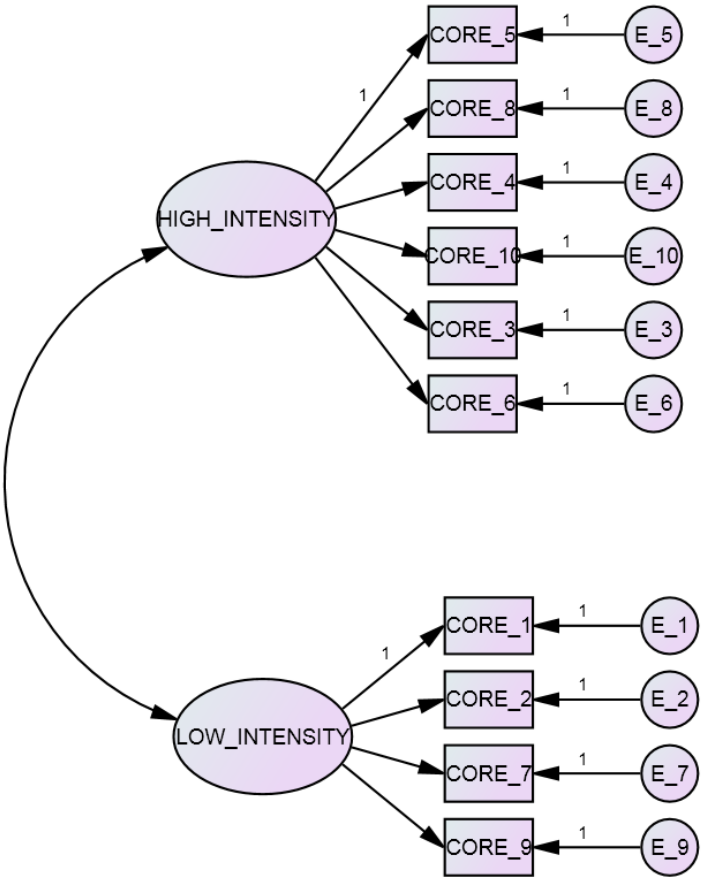
6.12 Appendix 12. Hypothesised factor solutions

Six factor model[24]



[24] Single item factors were excluded from the six factor model since the observed item represents the latent factor and it is not possible to establish an identified model when single item factors are included (McDonald, 1985). As such the six-factor model contained the three factors with more than one item: anxiety, depression and functioning.

Two factor model



Single factor model