

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/154084/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Wilkins, David and Meindl, Melissa 2025. Can child protection social workers forecast future actions, events and outcomes? A case study of long-term work with five families. *Child Care in Practice* 31 (2) , pp. 301-320. 10.1080/13575279.2022.2118674

Publishers page: <http://dx.doi.org/10.1080/13575279.2022.2118674>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Can Child Protection Social Workers Forecast Future Actions, Events and Outcomes?

A case study of long-term work with five families.

Abstract

Across the UK, child protection social workers are routinely called upon to assess the likelihood of future significant harm to children. Yet making consistently accurate judgements about what may or may not happen in future can be a difficult task. In a previous study, we tested social workers' abilities (n=283) to forecast the likelihood of different actions, events and outcomes following real-life referrals to social services. Aggregate group performance was only 6% better than you would expect by chance. As a result, we wondered whether social workers could make more accurate forecasts in relation to families they know well.

In this paper, we report the results of an in-depth case study, involving two social workers and five families. For eight-months, the social workers generated their own forecasting questions and provided estimates about the likelihood of different actions, events, and outcomes. One of the social workers, with more experience, made forecasts that were on average 12% more accurate than you would expect by chance. The other social worker, with less experience, made forecasts that were 6% less accurate than chance.

These findings suggest that simply having more information about the family may not make a consistent difference to the accuracy of social work judgements. What we do not know is the extent to which these findings might be replicated with a larger sample, or the nature of any potential relationship between more accurate judgements and better decision-making within the complex ecology of social work.

Key words: decision-making, accuracy, judgment, child protection, social work, case study

Word count (excluding abstract and references): 6,257

Authors:

1. Dr David Wilkins, Reader in Social, Cardiff University, WilkinsD3@cardiff.ac.uk (corresponding author)
2. Ms Melissa Meindl, Research Associate, Cardiff University.

Introduction

The Children Act 1989 places upon local authorities a ‘duty to investigate’ (section 47). This means that where a local authority “*have reasonable cause to suspect a child...is suffering, or is likely to suffer, significant harm*”, they must “*make...such enquiries as they consider necessary to enable them to decide whether they should take any action*”. According to Baroness Hale, former President of the UK Supreme Court, judging whether a child “*is likely to suffer significant harm*” involves making “*a prediction from existing facts...about what has happened in the past, about the characters and personalities of the people involved, [and] about the things they have said and done*” (Parliament. House of Lords, 2008). Former Lord Chief Justice Hoffman has similarly identified a link between making predictions and forming a judgement that the child is or is not likely to suffer significant harm (*ibid*). Indeed, it is hard to see how social workers could make *any* decision without making *some* judgement about the likelihood of different consequences, however implicitly or intuitively (Haran & Moore, 2014).

More broadly, we are interested in how you can measure and improve the quality of social work judgement. There are, according to Hood et al (2022), five ways in which the quality of a judgement (or a decision) can be defined:

1. Accuracy (the extent to which judgements or decisions are corroborated by other knowledge)
2. Consistency (the extent to which different professionals make similar decisions about the same or similar cases)
3. Outcomes (the consequences of a judgement or decision in terms of the child’s health and wellbeing)
4. Practice (the extent to which judgements and decisions are made according to sound principles and ethical ways of working)
5. Equity (the extent to which people from different socio-economic and demographic groups are treated the same).

In this paper, as in our previous work, we are interested in the question of *accuracy* (based on correspondence theory; Hammond, 1996). We do not focus on this measure because we think it is the only one that matters, or even the most important. Indeed, we believe strongly that all five of these criteria are essential for different reasons. We focus on it partly because it can be evaluated objectively, and partly because it has been a relatively

underused measure in the UK. According to Hood et al, out of thirteen studies included in their review, only three focused on accuracy, and none since 2014. We also focus on accuracy because it is ethically important (Keddell, 2022; Keddell et al., 2019). It matters whether social workers make similar or divergent judgements about similar cases (consistency). It matters whether decisions make a positive difference for children and families (outcomes). It matters whether and how social workers involve families in decision-making (practice). It matters that everyone receives a fair service (equity). And it matters whether social workers make judgements that are corroborated by what subsequently happens (accuracy).

A note on terminology

Before going any further, we want to be clear about how we are using various related terms. A **prediction** is a statement about what you think *will* happen in the future. If you predict that England will win a World Cup (in any sport) within the next 25 years, you will be right or wrong. A **forecast** is a statement about how likely you think something is to happen. If you forecast that England have a 30% chance of winning a World Cup (in any sport) within the next 25 years, you *cannot be right or wrong*. You can only be more or less accurate, and only then in relation to a large enough sample of forecasts. To be a good (accurate) forecaster, you need to align with reality. Things that you say have a 30% chance of happening should happen 30% of the time. A forecast is thus a special type of **judgement**, which has been defined in social work as “[when] a professional considers the evidence about the client or family situation in the light of professional knowledge to reach a conclusion or recommendation” (Taylor, 2013). **Making a decision** is a “commitment to action” (e.g., Wood, 2004), an agreement or determination to do something or that something should happen.

In the current study, and in previous work, we have sought to evaluate the accuracy of social work *forecasting*, seeing this as one component of social work judgement more generally. It is also important to clarify that we see a difference between *forming a judgement* and *making a decision*, and while the judgement of an individual social worker is important, it does not (and should not) determine the decision. Not least because in many situations, the worker is not solely responsible for making it. For example, the decision to start or stop a child protection plan is made formally at a child protection conference, a meeting which must involve other professionals, and should involve the family. The decision about whether a child should come into care is made, in most cases, by a judge or by magistrates. Within

these fora, and others, the social worker's judgement can and should be an important influence, but it is not deterministic (Lipsky, 2010; Maynard-Moody & Portillo, 2010). A social worker may judge that a child is at high risk of significant harm, above the 'threshold for action' (figure 1). But what *should happen*, is a different question. For some such children, it will be right for them to come into care. For others, it will be right for them to live elsewhere within their own family. For still others, it will be right for them to remain living at home, with support. Thus, even allowing for the nature of multi-party decision-making in social work, there is an important difference between forming a judgement and making a decision.

<Insert figure 1 about here>

A summary of our previous work

In two previous studies, we explored how accurately social workers could make forecasts, using case vignettes based on real-life referrals to social services in England (Authors Own x 2). In the first study, social workers (n=283) read four vignettes and answered questions about the likelihood of different actions, events, and outcomes. In the second survey, a sub-set of the original sample (n=63) read two vignettes they had seen before, and four new ones. As before, they answered questions about the likelihood of different actions, events, and outcomes.

On average, we found that social workers in the first survey were 6% more accurate than you would expect by chance. Perhaps this is not surprising, given that making accurate judgements about the future is something that "*all humans, not just social workers, frequently get wrong*" (Coulthard et al., 2020, p. 2). Indeed, even widely acknowledged experts find it hard to make predictions with any meaningful degree of accuracy (Tetlock, 2017; Tetlock & Gardner, 2016). Yet we also found that some social workers were considerably more accurate.

In the second survey, we found similar levels of overall accuracy, but again a small group of workers who were more accurate, including some who had been more accurate in the first survey. We also found in both surveys that some of the vignettes were more difficult to forecast than others. A clear limitation in both studies was the artificial nature of the testing scenario, with respondents having only partial information about each referral. As such, it may be that even a modest ability to make forecasts more accurately than chance is

commendable. We wanted to find out whether social workers would be able to make more accurate forecasts in relation to families they knew well.

Information overlooked or information overload?

The idea that having more information can help you make better judgements and decisions is widespread in social work. In cases where children have been seriously hurt or killed, practitioners are often criticised for *over-looking* crucial information (Brandon et al., 2020). Social workers have been encouraged to “*hold detailed information about the child and family across different domains*”, while also collecting (and sharing) information with a range of partner agencies (The Independent Review of Children’s Social Care, 2021, p. 38 - 40). Yet it is not necessarily obvious how or why having more information would help you to make a more accurate forecast. In principle, there is almost no upper limit on the amount of information you *could* collect about a family, although there will always be a *practical* limit in terms of time and other resources. The fast-and-frugal model of decision-making (Gigerenzer and Kurzenhaeuser, 2005) suggests that accuracy can actually be improved when you limit the amount of time spent searching for (additional) information. For example, in a study of group-based decision-making, Reimer and Katsikopoulos (2004) found that having more information could lead to less accurate forecasts.

Cognitive Load Theory (CLT) helps us understand *why* more information may not always lead to more accurate forecasts, and more effective decision-making. Cognitive load refers to the work done when engaged in tasks such as forming a judgement and thinking about a decision (Sweller, 1988). This might involve the effort needed to think and express yourself out loud to others, the effort needed to understand what other people are saying to you, and the effort needed to store new information in your memory. Heavy cognitive load can have negative effects on task completion (Frein et al., 2013). In some fields, such as the aerospace industry, this theory has been applied to reduce the amount of information presented to pilots in their cockpits, to ensure they are able to make good and fast decisions, without becoming overwhelmed *and* without overlooking crucial information (Wilson et al., 1994). In social work, it is possible that accurate forecasting is made *more* difficult because of a surfeit of information, especially when this obscures more *critical* information. Nonetheless, relative to our previous survey results, we were interested to see how social workers would make forecasts in relation to families they were working with, and about whom they knew relatively a great deal, compared with those who completed our surveys with much more limited information.

A brief review of judgement and decision-making in social work

The literature on social work judgement and decision-making is extensive, and it is beyond the scope of this article to provide a comprehensive review. However, it is worth making four broad points. First, many studies in the UK provide insightful descriptions of existing decision-making practices. They do so by using methods such as interviews and observations. These studies provide rich insights into important aspects of judgement and decision-making including: how social workers use their emotions during home visits (Cook, 2016); the factors that social workers say they rely on when forming a judgement, such as the level of risk, parental accountability, severity, and corroboration (Platt, 2006); and how intuitive reasoning can take place between people and across teams (Helm, 2016).

Second, many studies in the UK focus on case- or worker-related factors, rather than the external and organisational factors of the wider decision-making ecology (figure 2). This approach, which we have also adopted in our work, can generate important knowledge about how and why decisions are made by individuals, and in relation to specific cases. Yet it is important to acknowledge that focusing on case- and worker-related factors means focusing on only a part of the decision-making environment. A more complete understanding relies on our ability to integrate knowledge and understanding from a variety of perspectives.

<Insert Figure 2 about here>

Third, there is a growing body of literature about the socioeconomic and demographic inequalities of the UK social work system (Bywaters et al., 2016; Bywaters et al., 2015, Webb & Bywaters, 2018; Webb et al., 2021), and about the growth in the proportion of child protection investigations, compared to numbers of referrals and assessments (Bilson, 2021; Bilson & Martin, 2017; Devine & Parker, 2015). Children living in the most deprived circumstances are more likely to experience social work involvement and intervention, especially those living in poor families within more affluent areas (Bywaters et al., 2015). This situation reflects poorly, not on individual social workers per se, but on the investigative stance adopted by the systems as a whole, as well as the chronic underfunding of services.

Finally, the quality of social work judgement and decision-making in the UK has been criticised. It is suggested both that too many children are in care (Tickle, 2016), and that too many children are left in dangerous situations at home without proper support (Stevenson, 2015). According to the Independent Review of Children's Social Care in England, "*decision*

making and risk assessment related to harm is too inconsistent and often isn't good enough" (2021, p. 37). Where similar concerns have been expressed in other places, one response has been the introduction of Structured Decision-Making tools (Kim et al., 2008). However, much debate exists as to the benefits and unintended consequences of these tools (Bosk & Feely, 2020), and about the risks of de-professionalising the workforce (Bastian, 2017). In the UK, such tools are not in widespread use, although they have been trialled (Broadhurst et al., 2010a; Broadhurst et al., 2010b).

Methods

For this project, we used a case study design, described as "*an in-depth, multifaceted investigation, using qualitative research methods, of a single...phenomenon*" (Feagin et al, 1991, p. 2). Two social workers from one inner London local authority met regularly with the lead author over a period of eight-months. During these meetings, they discussed anonymously a small number of families from their caseloads and answered questions about the likelihood of different future actions, events, and outcomes. Their responses were scored with the aim of determining i) each social worker's overall level of forecasting accuracy, ii) the relationship between forecasting accuracy and the length of time between the question being forecast and resolved, and iii) their level of accuracy in relation to different families.

An example of the questions generated by the social workers can be seen in table 1. For each one, a suitable timescale was identified, and the worker provided a forecast, using a scale from 0 (definitely will not happen) to 100 (definitely will happen).

<Insert table 1 about here>

Sampling

In December 2020, the lead author contacted a Head of Service from one inner London local authority, to ask permission to circulate an information sheet about the study to social workers employed within their child protection teams. Two social workers volunteered to take part, referred to subsequently as Social Worker A and Social Worker B. Social Worker A was newly qualified, having graduated in 2020, and worked in the local authority ever since. Social Worker B qualified in 2016 and worked in two other local authorities before this one. The local authority in question uses the Signs of Safety practice model (Turnell & Edwards, 1997, 1999). To what extent this may have influenced the approach

taken by the two social workers, or the accuracy of their forecasts, we are not in any position to comment.

The meetings with the social workers took place via video-call, while they were either working from home or in their socially distanced offices. In the initial meetings, the social workers identified two or three families each, about whom they would provide regular anonymised case updates, and answer forecasting questions. The families identified by Social Worker A are referred to as Family 1 and 2. The families identified by Social Worker B are referred to as Families, 3, 4 and 5. The families were selected at the discretion of the social workers, with the only criterion being that they were likely to remain open for the duration of the study – a forecast which turned out to be accurate in every case (table 2).

<Insert table 2 about here>

Data analysis

Data were analysed using Microsoft Excel (version 16.55 for Mac) and SPSS (version 26 for Mac). For each forecast, Brier scores were calculated to measure accuracy. Brier scores range from 0 to 2, where 2 indicates complete inaccuracy and 0 indicates complete accuracy (see table 3 for some worked examples). The closer the Brier score is to zero, the more accurate the forecast. Random guessing in relation to binary outcomes will produce an aggregate score of 0.5, being equal to chance (Tetlock & Gardner, 2016, p. 64). Brier scores are calculated using the following formula, whereby x = the forecast for the outcome that does occur and y = the forecast for the outcome that does not occur:

$$Z = (1 - x)^2 + (0 - y)^2$$

Aggregate Brier scores were calculated for each family, and for each worker (using a mean average of all related questions).

<Insert table 3 about here>

Ethics

The study was approved by the [x] University School of Social Sciences' ethics committee in October 2020. The two social workers were provided with an information sheet

and gave verbal and written consent to take part. The social workers discussed their casework without sharing any identifying information about the families.

As part of our agreement with the local authority, if there had been any serious concerns about the social workers' practice, these would have been notified to the Head of Service (fortunately, this did not occur). As part of the study design, family consent was not sought, as we were not aware in advance which families would be discussed, and because no identifiable, non-anonymised information was collected. In the meetings, we asked the social workers some basic questions about their case work (e.g., 'how would you describe the current situation?') and for them to make forecasts about the likelihood of different actions, events, and outcomes. The forecasting questions were generated by the social workers. We did not seek to challenge the social worker's conclusions or suggest other ways of thinking about their work. While it is possible that by meeting with the lead author on a regular basis, the social workers may have thought slightly differently about the families, or their casework, this is also true of many other data collection methods. It is important to emphasise that all decisions were made about the families subject to the same local authority and multi-agency processes as normal.

Findings

In total, the social workers answered 249 questions in relation to five families. Tables 4 and 5 show how many questions were answered i) by each social worker, ii) in relation to which family, and iii) at what point during the study. A final session with each worker took place in September 2021 to finalise the outcome of remaining questions.

<Insert tables 4 and 5 about here>

Overall levels and range of forecasting accuracy

Social Worker A's overall level of accuracy, measured using Brier scores, was 0.53 (6% worse than chance), with a range between 0.00 (perfect accuracy) and 2.00 (perfect inaccuracy). Social Worker B's overall level of accuracy was 0.44 (12% better than chance), with a range also between 0.00 and 2.00.

We categorised each of the 249 forecasts based on whether they were more accurate than you would expect by chance, or not (a Brier score of 0.50; figure 3).

<Insert Figure 3 about here>

Forecasting accuracy and the length of time between forecast and resolution of the question

For every forecast, we recorded the date it was made and the date it resolved (i.e., when the outcome became known), and calculated the number of days between the two. Figures 4 and 5 show the relationship between this variable and forecasting accuracy. The dotted lines of best fit indicate a clear positive correlation for Social Worker B, and a very small positive correlation for Social Worker A.

<Insert Figures 4 and 5 about here>

Table 6 shows a series of correlations between these two variables (accuracy and length of time), for the two social workers combined, for each one separately, and for each family.

<Insert table 6 about here>

These data show that for Social Worker B, the correlation is weakly positive and statistically significant. For Family 5, the relationship is strongly positive and statistically significant. The forecasts for Family 1 are noteworthy, as they are the only ones for which there is a negative correlation, albeit the relationship is very weak, and non-significant.

Forecasting accuracy for different families

Of the five families, the most accurate forecasts were made in relation to Family 4. The least accurate forecasts were made in relation to Family 1. Overall, the two social workers made forecasts more accurately than you would expect by chance in relation to three out of the five families (table 7).

<Insert table 7 about here>

Using the same categories as in figure 3, we explored what percentage of forecasts for each family were made more or less accurately than chance (figures 6 and 7).

<Insert Figures 6 and 7 about here>

Forecasting accuracy and question updating

We then considered whether individual questions had been forecast only once, or more than once (whether the forecasts had been updated). Table 8 shows the mean Brier scores for each of these categories, indicating that updated forecasts were more accurate, on average, than non-updated ones.

<Insert table 8 about here>

Predictors of forecasting accuracy

Finally, a direct logistic regression was performed to assess the impact of several variables on Brier scores (whether they were higher or lower than you would expect by chance). The model contained five independent variables (social worker, family, the month the forecast was made, the length of time between forecast and outcome, and whether the forecast was updated or not). The full model containing all predictors was statistically significant, $\chi^2(12, n=249) = 34.40, p < .001$, indicating that the model was able to distinguish between Brier scores equal to or worse than chance, and those better than chance. The model explained between 12.9% (Cox and Snell R square) and 17.2% (Nagelkerke R squared) of the variance in Brier score status, and correctly classified 65.5% of cases. As shown in table 9, two of the independent variables made a unique statistically significant contribution to the model (social worker and timeframe). The strongest single predictor – timeframe – had an odds ratio of 1.014. This indicates that the shorter the timeframe, the more likely it was for the Brier score to fall below 0.5 (indicating a forecast more accurate than chance), controlling for the other variables in the model.

<Insert table 9 about here>

Summary

Social Worker A's overall level of forecasting accuracy was slightly worse than chance (by 6%), and Social Worker B's was better than chance (by 12%). Forecasts made in relation to Families 2, 4, and 5 were better than chance (by 6%, 34%, and 16% respectively). Forecasts made in relation to Families 1 and 3 were worse than chance (by 16% and 10% respectively). We found a weak relationship overall between forecasting accuracy and length of time, with forecasts made in relation to shorter timeframes being more accurate on average than those made in relation to longer timeframes.

Discussion

Making accurate forecasts about the future is hard, at least where human behaviour is concerned. This we knew already. We also know that in England and Wales, child protection social workers are routinely required to make judgements about the likelihood of future significant harm to children. Yet compared with other fields, we know little about the nature of forecasting in social work, or how the relationship between forecasting accuracy, judgement, and decision-making might operate in practice.

In a previous survey, we found that a large group of social workers (n=283), when reading real-life referrals, made forecasts only slightly more accurate than chance (by 6%). Based on this finding, we asked ourselves whether social workers would make more accurate forecasts when the questions were “*about the families they are working with and...based on [a] greater availability of information*” (Authors Own). Albeit on a small scale, our results here imply that *simply* having more information is not necessarily a good predictor of more accurate forecasts. For at least one family each, the two workers in this study outperformed the benchmark set in our previous surveys, but they also fell short of the same benchmark for at least one other family. Cognitive Load Theory, as discussed above, suggests that having more information can make decision-making more difficult, and the same may well be true for making accurate forecasts too.

We also found in our previous surveys that some of the referrals had lower mean Brier scores than others. Some of them also had a narrower range of scores (indicating greater agreement between respondents). This could suggest that some cases are more easily ‘forecastable’ than others. In this study, Family 4 had the lowest mean Brier score (25% more accurate than chance), while Family 5 had the second lowest (19% more accurate than chance). While it is impossible to know for sure whether or how these cases differed from the others – and logically one of the families had to be the easiest to forecast – it is interesting to note something they had in common. In both Families 4 and 5, the decision had already been taken to remove the child/ren from the family home, and for them to live permanently elsewhere. For Families 1, 2 and 3, the children were either living at home or living elsewhere temporarily, while the local authority considered or undertook care proceedings. Perhaps the family court’s final decision to grant a care order provides a degree of stability in relation to social work decision-making such that making accurate forecasts post-care proceedings is easier than beforehand. This is a testable hypothesis.

We also found a correlation in this study between forecasting accuracy and the length of time between forecast and resolution. The longer the time, the less accurate the forecast. This may seem like an obvious finding. After all, one-day weather forecasts (of temperature) are accurate more than 80 per cent of the time, and five-day forecasts only 60 per cent of the time (Rose & Floehr, 2017). Yet if you try and predict the level of the FTSE 100 stock market tomorrow, you will probably get it wrong (and you certainly will if you keep doing so day after day). On the other hand, it is relatively easy to predict that in 10- or 20-years' time, the FTSE 100 will be higher than it is today. For weather forecasting, and at least some forecasts in social work, it is easier to be more accurate over the short-term. For the stock-market, and perhaps some other forecasts in social work that we have not yet considered, it is easier to be more accurate over the long-term. For example, you may not know what will happen to a child tomorrow, if they are living with domestic violence in the home, but you can say with some confidence that over the longer-term, living with domestic violence is very likely to be very harmful.

More generally, and however implicitly, forecasting must play at least some part in social work decision-making. Because “*every decision depends, to some degree, on [making] a forecast of the consequences*” (Haran & Moore, 2014, p. 5), the question is not whether social workers should make forecasts about future events and outcomes. They already do, albeit perhaps not so explicitly as in our studies, and without needing to characterise their own practice in such terms. Any kind of plan implies a set of expectations about the future. A child protection plan which says that parents should attend a parenting programme makes a series of (implicit) forecasts – that the parenting programme will still be available, that if the parents attend, the programme has a fair chance of being helpful, and that any subsequent changes in parenting behaviour will be of benefit to the child. These expectations need not be the *only* reasons for making the plan. Other factors will also be considered, and often much more explicitly, such as the parents' own views, and those of the child, and the availability (or lack) of other resources. Nonetheless, if you expected (or forecast) that the parenting programme would make things worse, you should not include it on the plan. As such, it is possible to ask – when social workers *forecast* that a parenting programme will help, how often are they right or wrong? Are they right more often than they are wrong? Are they better than chance, or worse than chance? And what does this suggest about the way in which decisions are made in social work?

Yet it is *relatively* unimportant whether a parent attends or even benefits from a parenting programme or not. What really matters is whether the child suffers significant

harm. We did not include this exact question in our study, or others like it, because you cannot objectively measure the outcome. Whether a child is made subject to a child protection plan, or comes into care, are questions with definitive outcomes – they happen, or not. The question of whether a child has suffered significant harm is a more subjective judgement, albeit one based in part on a set of objective facts. It is a question about which reasonable people could disagree. In at least some cases, we may never be able to say for sure. It might be possible to use other events, such as the child coming into care, as a proxy indicator of significant harm. Yet no proxy measure is perfect, and many children come into care without having been significantly harmed (indeed, coming into care may protect them from such harm).

This highlights the complexity of the social work decision-making ecology, and the recursive nature of social work judgement. Whether you think it is going to rain or not has no bearing on what happens. As a social worker, what you think is going to happen will influence what happens. The social worker's judgement operates as a form of commentary about the situation, *and* as a feature of the situation itself. If the social worker judges the child to be at risk of significant harm, this increases the chances of a child protection conference, all else being equal. But it also influences the parent's view of the social worker, and their willingness to 'engage' with services. This in turn influences the social worker's judgement about the need for a child protection conference. The social worker can also do things to affect the future, based on their own judgements about what is likely or not likely to happen. If a social worker judges a high likelihood of domestic abuse, they will work with the family to ensure this does not happen, thus helping to protect the child and others, while 'disproving' their own initial judgement. All of which serves to highlight the need to take a measured and comprehensive view of what accuracy means in relation to social work judgements, and which questions might be most relevant to consider.

Strengths and limitations

The most significant limitation of this study is the small sample size. The findings cannot be generalised to the wider social work population in England, or even to other social workers in the same authority. It is also important to acknowledge that we have not studied how social workers actually form their judgements and make decisions in day-to-day practice. Ideally, they will do so in collaboration and partnership with families, their supervisors, and other professionals. As such, another limitation is our focus on individual forecasting abilities, separate from the wider practice setting and organisational context. We

fully acknowledge that this is not how social workers form judgements and make decisions. However, it was not our aim to explore how judgements and decisions are made in practice (there are many excellent studies of this already in the UK) but to continue laying the foundations for a programme of study based on the importance of accurate forecasts, including how you can measure accuracy, what a reasonable level of accuracy might be for social work, and what kinds of questions it is helpful to consider. We also fully acknowledge that making good judgements and good decisions in social work involves much more than making accurate judgements (or forecasts) about the future.

Yet the case study method has advantages of its own. It helps provide a more detailed understanding than can be obtained via larger-scale studies. Because the case study is embedded within a more natural context, it allows for this understanding to be developed over time, rather than relying on a snapshot. It also allows respondents to become more involved in the study design. As a tangible example, in our previous work, the forecasting questions were provided by us. In this study, the social workers were able to set and answer their own questions.

Conclusions

For child protection social workers, the absolute worst-case scenario – of a child dying – happens only very rarely (Munro, 2009, 2010a, 2010b), yet far too often. The very rareness of the event makes it difficult to predict or to forecast – and impossible in many cases. There is a limit to our knowledge, however much information we have, and epistemic humility is an important component of good judgement in any setting (Mellers et al., 2019). We also have much to learn from effective interventions in other fields, including the benefits of cognitive debiasing training (Jenkins & Youngstrom, 2016), the importance of regularly updating your forecasts (Atanasov et al., 2020), and the power of working together in structured group formats to improve accuracy (Horowitz et al., 2019).

We do not (yet) have a good understanding in social work of what might be possible when it comes to forecasting accuracy, or the extent to which increased accuracy is a worthwhile goal. Part of our motivation for undertaking this work is to explore the limits of what might be possible in social work, the extent to which forecasting accuracy might be improved, and the relationship between forecasting accuracy, professional judgement more broadly, and decision-making in real-life practice. In principle, it seems reasonable to think that making more accurate forecasts will allow you to make more effective decisions. Yet we also know, from Baumann et al's (2014) ecology of social work decision-making (see figure

2, above), that individual factors are only one component. Perhaps the benefits of more accurate forecasts will be outweighed by the time and resources they take to achieve. Perhaps there is a level of 'good enough' accuracy that social workers already achieve. Or perhaps the influence of individual factors is so relatively minor, that improvements in other parts of the system would be more beneficial. We do not know - but hope to find out.

Acknowledgements

We are very grateful to the two social workers who took part in the study, when they were busy enough with their own casework already. They both responded with patience and good humour to the lead author's many naïve questions about the nature of contemporary social work practice with children and families. We are also grateful to the local authority for supporting the study.

Funding

The CASCADE partnership receives infrastructure funding from Health and Care Research Wales.

References

Authors Own x 2

- Atanasov, P., Witkowski, J., Ungar, L., Mellers, B. & Tetlock, P. 2020. Small steps to accuracy: Incremental belief updaters are better forecasters. *Organizational Behavior and Human Decision Processes*, 160, 19-35.
- Bastian, P. 2017. Negotiations with a risk assessment tool: Standardized decision-making in the United States and the deprofessionalization thesis. *Transnational Social Review*, 7, 206-218.
- Baumann, D.J., Fluke, J.D., Dalglish, L. & Kern, H., 2014. The decision making ecology. *From evidence to outcomes in child welfare: An international reader*, 24-40.
- Benbenishty, R. & Fluke, J. D. (2020). Frameworks and models in decision-making and judgement in child welfare and protection. In *Decision-making and judgement in child welfare and protection: Theory, research, and practice* (pp. 3-26). Oxford University Press.
- Bilson, A. (2021). Child Protection Investigations in Scotland: A 33 Per Cent Increase in Two Years. *Child Abuse Review*.

- Bilson, A., & Martin, K. E. (2017). Referrals and child protection in England: One in five children referred to children's services and one in nineteen investigated before the age of five. *British Journal of Social Work*, 47(3), 793-811.
- Bosk, E. & Feely, M. 2020. The goldilocks problem: Tensions between actuarially based and clinical judgment in child welfare decision making. *Social Service Review*, 94, 659-692.
- Brandon, M., Sidebotham, P., Belderson, P., Cleaver, H., Dickens, J., Garstang, J., . . . Wate, R. (2020). *Complexity and challenge: a triennial analysis of SCRs 2014-2017*.
- Broadhurst, K., Hall, C., Wastell, D., White, S. & Pithouse, A. 2010a. Risk, instrumentalism and the humane project in social work: Identifying the informal logics of risk management in children's statutory services. *British Journal of Social Work*, 40, 1046-1064.
- Broadhurst, K., Wastell, D., White, S., Hall, C., Peckover, S., Thompson, K., Pithouse, A. & Davey, D. 2010b. Performing 'initial assessment': identifying the latent conditions for error at the front-door of local authority children's services. *British journal of social work*, 40, 352-370.
- Bywaters, P., Brady, G., Sparks, T. & Bos, E. 2016. Inequalities in child welfare intervention rates: The intersection of deprivation and identity. *Child & Family Social Work*, 21, 452-463.
- Bywaters, P., Brady, G., Sparks, T., Bos, E., Bunting, L., Daniel, B., Featherstone, B., Morris, K. & Scourfield, J. 2015. Exploring inequities in child welfare and child protection services: Explaining the 'inverse intervention law'. *Children and Youth Services Review*, 57, 98-105.
- Cook, L. 2016. *Professional judgement in social work: making sense of the initial home visit*. University of East Anglia.
- Coulthard, B., Mallett, J., & Taylor, B. (2020). Better Decisions for Children with "Big Data": Can Algorithms Promote Fairness, Transparency and Parental Engagement? *Societies*, 10(4), 97.
- Dalgleish, L.I., 1988. Decision making in child abuse cases: Applications of social judgment theory and signal detection theory. In *Advances in psychology* (Vol. 54, 317-360). North-Holland.
- Deveans, T., & Kewley, R. H. (2009). *Overcoming information overload in the cockpit*.
- Devine, L., & Parker, S. (2015). Rethinking child protection strategy: Learning from trends.

- Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. *International journal of information management*, 20(1), 17-28.
- Feagin, J. R., Orum, A. M., & Sjoberg, G. (1991). *A case for the case study*. UNC Press Books.
- Featherstone, B., Gupta, A., Morris, K., & Warner, J. (2018). Let's stop feeding the risk monster: towards a social model of child protection'. *Families, Relationships and Societies*, 7(1), 7-22.
- Frein, S. T., Jones, S. L. & Gerow, J. E. 2013. When it comes to Facebook there may be more to bad memory than just multitasking. *Computers in Human Behavior*, 29, 2179-2182.
- Gigerenzer, G. & Kurzenhaeuser, S. 2005. Fast and frugal heuristics in medical decision making. *Science and medicine in dialogue: Thinking through particulars and universals*, 3-15.
- Hammond, K. (1996) *Human Judgement and Social Policy*. Oxford University Press, USA.
- Haran, U., & Moore, D. A. (2014). A better way to forecast. *California Management Review*, 57(1), 5-15.
- Helm, D. 2016. Sense-making in a social work office: an ethnographic study of safeguarding judgements. *Child & Family Social Work*, 21, 26-35.
- Hood, R., Abbott, S., Coughlan, B., Nilsson, D., Duschinsky, R., Parker, P. & Mannes, J. 2022. Improving the quality of decision making and risk assessment in children's social care: a rapid evidence review.
- Horowitz, M., Stewart, B. M., Tingley, D., Bishop, M., Resnick Samotin, L., Roberts, M., Chang, W., Mellers, B. & Tetlock, P. 2019. What makes foreign policy teams tick: Explaining variation in group performance at geopolitical forecasting. *The Journal of Politics*, 81, 1388-1404.
- Independent Review of Children's Social Care. (2021). *The Case for Change*. Available at: <https://childrensocialcare.independent-review.uk/wp-content/uploads/2021/06/case-for-change.pdf> [Accessed April 2022]
- Jenkins, M. M. & Youngstrom, E. A. 2016. A randomized controlled trial of cognitive debiasing improves assessment and treatment selection for pediatric bipolar disorder. *Journal of Consulting and Clinical Psychology*, 84, 323.
- Keddell, E. 2022. On Decision-Variability in Child Protection: Respect, Interactive Universalism and Ethics of Care. *Ethics and Social Welfare*, 1-16.

- Keddell, E., Davie, G. & Barson, D. 2019. Child protection inequalities in Aotearoa New Zealand: Social gradient and the 'inverse intervention law'. *Children and Youth Services Review*, 104, 104383.
- Kim, A. K., Brooks, D., Kim, H. & Nissly, J. 2008. Structured Decision Making® and child welfare service delivery project. *California: California Social Work Education Centre*.
- Lipsky, M. 2010. *Street-level bureaucracy: Dilemmas of the individual in public service*, Russell Sage Foundation.
- Maynard-Moody, S. & Portillo, S. 2010. Street-level bureaucracy theory. *The Oxford handbook of American bureaucracy*.
- Mellers, B., Tetlock, P. & Arkes, H. R. 2019. Forecasting tournaments, epistemic humility and attitude depolarization. *Cognition*, 188, 19-26.
- Munro, E. (2009). Managing societal and institutional risk in child protection. *Risk Analysis: An International Journal*, 29(7), 1015-1023.
- Munro, E. (2010a). Conflating risks: Implications for accurate risk prediction in child welfare services. *Health, Risk & Society*, 12(2), 119-130.
- Munro, E. (2010b). Learning to reduce risk in child protection. *British Journal of Social Work*, 40(4), 1135-1151.
- Parliament. House of Lords. (2008). *Opinions of the lords of appeal for judgement in the cause – In Re B (Children) (Fc)*. London: The Stationary Office.
- Platt, D. 2006. Threshold decisions: How social workers prioritize referrals of child concern. *Child Abuse Review: Journal of the British Association for the Study and Prevention of Child Abuse and Neglect*, 15, 4-18.
- Reimer, T. & Katsikopoulos, K. V. 2004. The use of recognition in group decision-making. *Cognitive Science*, 28, 1009-1029.
- Rose, B., & Floehr, E. (2017). Analysis of high-temperature forecast accuracy of consumer weather forecasts from 2005–2016. *Forecast Watch Report*. (Dublin, OH, USA).
- Sbaffi, L., Walton, J., Blenkinsopp, J., & Walton, G. (2020). Information overload in emergency medicine physicians: a multisite case study exploring the causes, impact, and solutions in four North England National Health Service Trusts. *Journal of Medical Internet Research*, 22(7), e19126.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285.
- Taylor, B. 2013. *Professional decision making and risk in social work*, Learning Matters.

- Tetlock, P. E. (2017). *Expert political judgment: How good is it? How can we know?*-New edition. Princeton University Press.
- Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: The art and science of prediction*. Random House.
- Tickle, L. (2018). Number of children in care for emotional abuse soars. *The Guardian*. Available at: <https://www.theguardian.com/society/2018/sep/14/number-of-children-in-care-for-emotional-abuse-soars> [Accessed November 2021]
- Turnell, A., & Edwards, S. (1997). Aspiring to Partnership. The Signs of Safety approach to child protection. *Child Abuse Review: Journal of the British Association for the Study and Prevention of Child Abuse and Neglect*, 6(3), 179-190.
- Turnell, A., & Edwards, S. (1999). Signs of safety. *A solution oriented approach to child protection casework*.
- Webb, C. J. & Bywaters, P. 2018. Austerity, rationing and inequity: trends in children's and young peoples' services expenditure in England between 2010 and 2015. *Local Government Studies*, 44, 391-415.
- Webb, C. J. R., Bywaters, P., Elliott, M. & Scourfield, J. 2021. Income inequality and child welfare interventions in England and Wales. *J Epidemiol Community Health*, 75, 251-257.
- Wilson, G. F., Fullenkamp, P. & Davis, I. 1994. Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, space, and environmental medicine*.
- Wood, M. A. 2004. *A study of the perception of the impact of modeling on the development of commitment to action in Decision Conferencing*. Curtin University.