British Journal of Social Work (2023) **54**, 1150–1169
https://doi.org/10.1093/bjsw/bcad234
Advance Access Publication November 8, 2023

Can Social Workers Estimate the Likelihood of Future Actions and Events? A Forecasting Accuracy Study

David Wilkins * and Melissa Meindl

School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK

*Correspondence to David Wilkins, School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK.
E-mail: *WilkinsD3@cardiff.ac.uk*

Abstract

Social workers routinely make decisions and formulate care plans in the course of their work with children and families. These decisions and care plans are based at least in part on the professional judgement of the worker. An important component of judgemental quality is the extent to which explicit or implicit forecasts about the likelihood of different future actions and events are made with accuracy. In this article, we report an analysis of 21,193 forecasts made by 581 child and family social workers in England about the likelihood of different future actions and events following referrals to children's services. We found that the more likely social workers said an action or event was to happen (as the forecast likelihood increased towards 100 per cent), the more often it occurred. However, we also found that social workers tend to over-estimate the likelihood of almost all specified future actions and events. Social workers were most accurate when forecasting something to be *very unlikely*, and less accurate when they forecast something to be *likely* or *very likely*. We consider these findings in relation to false positive and negative errors in child protection, and the theory of judgemental rationalism.

Keywords: accuracy, child and family social work, child protection, decision-making, judgement

Accepted: October 2023

Introduction

Social workers routinely make judgements and decisions as part of their day-to-day work with children and families. Such judgements are made whenever 'a professional considers the evidence about the client or family situation [and reaches] a conclusion or recommendation' (Taylor, 2013, p. 10). Social workers will 'make many judgements where [they] recommend a course of action even if [they] do not make the final decision' (Taylor, 2017, p. 105). Perhaps most significantly, child protection social workers make judgements about whether children are at risk of significant harm (Allen, 2005). Almost every decision involves some kind of judgement, however implicit, about the 'likelihood of various events' (Eysenck and Keane, 2015, p. 547), and the possible consequences of different choices (Haran and Moore, 2014, p. 5). For example, if a social worker thought that further incidents of domestic abuse were *very likely*, they would make different decisions than if they thought such incidents were *very unlikely*. Sometimes, social workers make individual judgements; more often they do so in collaboration with others, including supervisors, other professionals, children, parents and wider family members. The quality of social work judgements, and the accuracy of the implied or explicit forecasts they represent, are important not least because 'people act upon their beliefs—whether [they] are true or not (and it's far easier to be wrong than right)' (Boghossian and Lindsay, 2019, p. 5).

Judgemental rationalism

There are various criteria that can be used to evaluate the quality of social work judgements (see Hood *et al.*, 2022). For this study, the theory of judgemental rationalism provides a useful framework (Bhaskar, 2013). Judgemental rationalism suggests that whilst there can never be a 1:1 correspondence between our judgements, and the world-as-it-really-is, nonetheless our judgements 'generate expectations about the world and about the results of our actions' (Sayer, 2010, p. 69). Hence, one component of good judgement is the ability to forecast the likelihood of future events and actions. Not only so one can make (more) effective decisions, but also because of the moral imperative to act fairly, to balance the child's need for protection with the parent's right to a private family life (Masson, 2006).

The problem of negative and positive errors in child protection

In child protection, Munro (2004, 1999) has applied the concept of false negative and positive errors to help explain judgemental accuracy. As shown in Figure 1, if a social worker inaccurately judges a child to be

	Actual – child is 'safe'	Actual – child is 'not-safe'
Judgement – child is 'safe'	True negative	False negative
Judgement – child is 'not-safe'	False positive	True positive

Figure 1: Error types when judging whether a child is 'safe' or 'not-safe'.

'safe', they make a false negative error, whereas an inaccurate judgement that a child is 'not-safe' would be a false positive error (Figure 1).

Whilst these error types can have different consequences—for the child and family, and for the social workers and other professionals involved—one of the key features of many Serious Case Reviews (formal case enquiries held in England whenever a child has been seriously harmed or killed, and maltreatment is suspected to be a factor) is whether the outcome *was* or *should have been* foreseeable ahead of time (Sidebotham, 2012). Yet questions about whether social workers *should* be able to make accurate forecasts are practical-moral as well as technical-rational (Taylor and White, 2001, 2005). Ultimately, there can be no absolute standards for judgemental accuracy in social work—not least because the world is a complex and capricious place, and because people have free-will.

Can social workers make accurate judgements?

In a recent literature review, Hood *et al.* (2022) identified three UK and four international studies with a focus on the accuracy of social work judgements. They defined accuracy as 'the extent to which decisions are corroborated by other knowledge' (p. 5), including (but not limited to) subsequent actions and events following the decision.

Of the three UK studies, Dickens *et al.* (2005) examined care rate variations between English local authorities (LAs) and explored the relationship between relative need and the rate of children starting to be looked after. The authors concluded that 'it is hard to say whether there are children who need to become looked after, but who do not' (p. 607). Forrester (2008) collected data on 400 referrals, all of which were closed relatively quickly. Forrester analysed how many were subsequently re-referred because of actual or suspected maltreatment. The logic is that if the initial referral was closed and the child quickly re-referred, then the original judgement was potentially inaccurate. However, as Forrester

found that only 11 of the 400 children were re-referred, he concluded 'the level of accuracy for the identification of risk of serious abuse appears comparatively high' (p. 296). On the other hand, Farmer and Lutman (2014) analysed 138 cases in which a child returned home from care. Over a 5-year period, they found that 59 per cent of these children were subsequently maltreated and 65 per cent came back into care (p. 265), which they took to indicate a low level of judgemental accuracy.

Of the international studies, Cross and Casanueva (2009) considered a sample of 4,000 child maltreatment substantiation decisions in America. They examined whether caseworker judgements of harm, risk and strength of evidence could predict the outcome, finding that these decisions did vary depending on these criteria. They also found that allegations of maltreatment towards girls were more likely to be substantiated than for boys. By using a logistic regression, they estimated a predictive accuracy of between 77.5 per cent and 87.2 per cent for substantiated, and between 73.5 per cent and 79.4 per cent for unsubstantiated cases. However, when DePanfilis and Girvin (2005) analysed a similar sample of 129 decisions from New Jersey, they found that only 42 per cent were accurate. Cyr *et al.* (2022) evaluated an assessment tool and established that it could improve the accuracy of judgements about subsequent child maltreatment, whilst Gillingham and Humphries (2010) found that practitioners would selectively use a Structured Decision-Making tool to ensure it confirmed their own judgements. It is worth noting that these studies used different conceptions of accuracy and different methods for measuring it. This makes it difficult to compare them directly. It is also worth noting that the UK studies collected data on decision outcomes and *inferred* the underlying judgements. By contrast, the American studies collected data on judgements and decisions separately.

In our study, we set out to measure the accuracy of social work forecasts more directly. We did so by asking practitioners to read a series of anonymised referrals, and answer questions about the likelihood of subsequent actions and events. We then examined how many of these actions and events happened or did not happen. Our research question was as follows:

> When social workers make forecasts about the likelihood of specific actions and events, how often do they occur?

A word on definitions

Before describing our methods, it is important to define some key terms. We start with Taylor's definition of a 'judgement' as something formed whenever a social worker 'considers the evidence about the client or family situation [and reaches] a conclusion or recommendation' (Taylor,

2013). Social workers make judgements to inform their own decisions, and to inform decisions made by others. We define a decision as being a 'commitment to action' (Wood, 2004), an agreement or determination to do something or that something should happen. An *effective decision* is one that achieves its intended outcome (Drucker, 1967). A decision to keep a child safe by placing them in foster care is *effective* if the child is kept safe, without necessarily being *right* (perhaps the child would have been safe at home with more intensive support services). The judgements made by individual workers may inform but do not dictate decision-making. If a social worker says that a child is at risk of significant harm, *some* response is called for but there is no one right thing to do. The action taken should be commensurate with the judgement, whilst considering other factors, especially the child and parent's wishes and feelings, the legal framework, equitable use of resources and professional ethics (see Wilkins and Boahen, 2013).

We also refer to *forecasts* (or *probability estimates*) as one component of judgemental quality. A forecast is a statement about how likely something is to happen. It is not the same as a prediction (unless you make a forecast of 0 per cent or 100 per cent). A prediction is a statement about what you think *will* happen. If you predict a 100 per cent chance of snow tomorrow, you will be right or wrong. If you forecast an 80 per cent chance of snow, you cannot be right or wrong—although you can be more or less accurate (if it does snow, an 80 per cent forecast is more accurate than a 60 per cent forecast). You can only obtain a meaningful measure of accuracy in relation to a sufficiently large set of forecasts. Making a 'lucky guess' that it is going to snow tomorrow does not make you a good weather forecaster. It is also important to consider the nature of the forecasting environment. Forecasting that it is going to be sunny in Hawaii tomorrow does not make you a good weather forecaster either (it is almost always sunny in Hawaii). To measure the accuracy of a forecast, you need to know what happened. For any given set of forecasts, you can calculate the observed frequency of the specified actions and events (the *outcome*). For example, if a perfectly accurate weather forecaster says there is an 80 per cent chance of snow on a given set of days, it would snow on precisely 80 per cent of those days, giving an observed frequency of 0.8. If it snows on more or less than 80 per cent of days, the observed frequency would be more or less than 0.8. The greater the difference between the forecast probability (for a given set of forecasts) and the observed frequency (ditto), the less accurate the forecasts (and by implication, the forecaster).

Methods

In a series of online surveys hosted on Qualtrics (copyright © 2022, Provo, Utah; https://www.qualtrics.com), respondents were presented

Table 1. An overview of the referrals used in our survey

Pseudonym for the child	Brief details
1. Aadesh	Male pre-school child referred by community mental health team in relation to concerns about domestic abuse in the home.
2. Ava	Female pre-school child referred by nursery in relation to concerns about neglect.
3. Clark	Unborn child referred by midwife team in relation to concerns about mother's presentation and lack of engagement.
4. Emelia	Female baby referred by CAFCASS in relation to concerns about domestic abuse and physical abuse.
5. Malalai	Female teenager referred by hospital emergency department in relation to concerns about self-harm and domestic abuse.
6. Omar	Male primary-school child referred by another LA in relation to concerns about parental learning difficulties and domestic abuse.
7. Poppy	Female pre-school child referred by police in relation to concerns about parental alcohol misuse and domestic abuse.
8. Salma	Female teenager, referred by the police in relation to concerns about possible sexual exploitation, having shared explicit images of herself with an adult.
9. Stephanie	Female primary-school child referred by CAFCASS due to allegations made by the mother in relation to father's partner during private family court proceedings.
10. Taryn	Male teenager referred by mother in relation to concerns about emotional and behavioural difficulties.
11. William	Male teenager referred by police in relation to concerns about domestic abuse.
12. Wooten	Unborn child referred by another LA in relation to concerns about maternal mental health problems and learning difficulties.

with anonymised referrals to social services and asked to forecast the likelihood of different subsequent actions and events (Meindl and Wilkins, 2021). For example, whether there would be a social work assessment, whether the LA would arrange a strategy meeting, whether the child would come into care and whether the mother or father would attend a future appointment (see Table 1 for an overview of the referrals and Table 2 for a more detailed example.). For each possible action or event, respondents provided a forecast using a numerical scale from 0 to 100, in which 0 = will definitely not happen, and 100 = will definitely happen. We could then compare respondents' forecasts with what actually happened, as derived from the case file records subsequent to each referral. Thus, our dataset consisted of (i) a collection of numerical forecasts (0–100 per cent) made by social workers in relation to twelve referrals and associated questions and (ii) for each forecast whether the specified event or action (the outcome) happened or did not happen (1, 0).

Sampling

Respondents were recruited to complete the survey via social media requests and by direct email invitation (circulated within a small number

Table 2. A more detailed example of one of the referrals used in our survey studies, and a selection of the associated questions

'I am concerned about Aadesh (aged 11) and his siblings due to domestic abuse at home. Aadesh's uncle is drug taking and selling within the family home, this is having a grave effect upon his mental health. There is domestic violence between mother and her brother-in-law. He is known to use cannabis, heroin and crack-cocaine. He was previously admitted to a mental health unit under section. Diagnosed with schizoaffective disorder and a personality disorder. The children must be witnessing domestic violence and I think they would be scared of their uncle. I've spoken to the mother, and she said she is scared of him and they cannot ask him to leave the home in case he reacts aggressively. I think he may also be drug dealing within the home. He is currently under a community mental health treatment order'. (Referral received from a community mental health nurse.)
Within the next 12 weeks, how likely is it that Aadesh will become the subject of a child protection plan?
Within the next 12 weeks, will the LA convene a strategy meeting?
Within the next 6 months, will there be another referral about Aadesh?
Within the next 6 months, will Aadesh come into care?

of LAs in England). All respondents self-declared that they were (i) a qualified social worker and (ii) currently practicing with children and families in England. Reflecting the wider workforce, most of our respondents were female (85.2 per cent); approximately one-third were in the age range 25–34 (34.9 per cent), and another approximately one-third aged between 35 and 44 (30.7 per cent). Most identified themselves as White English, Welsh, Scottish, Northern Irish or British (86.1 per cent). Nearly one-quarter said they had been qualified as a social worker for between 1 and 3 years (23.6 per cent), whilst approximately one-third said they had more than 10 years' experience (35.1 per cent). Finally, just over a quarter worked in child in need or child protection teams (26.6 per cent), and around one in ten worked either in a Referral and Assessment team (11.4 per cent) or with Looked After Children (12.7 per cent). As with all non-probability sampling, we relied on the availability and interest of respondents to take part, resulting in a non-representative sample.

Data analysis

In 1950, Glenn Brier, working for the federal Weather Bureau in America, published a paper on the verification of forecast outcomes. He collected eighty-five rain forecasts, expressed in numerical terms, and organised them into five categories from least to most likely. He then observed the proportion of cases in which it had rained. For an ideally calibrated forecaster, there would be a linear relationship between the forecast and the observed frequency of outcomes. Whilst this degree of accuracy is impossible to achieve in real-life, you might still expect that as the forecast increased, so the observed frequency of outcomes also increased. As shown in Figure 2 and Table 3, this is what Brier (1950)

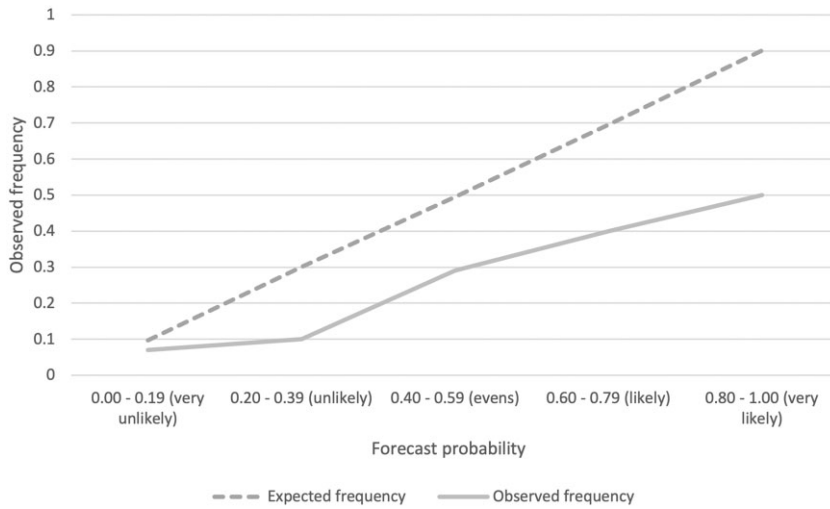

Figure 2: A comparison of expected frequencies and observed frequencies for rain forecasts from the early 1950s (Brier, 1950). The observed frequency is *below* the expected frequency indicating overestimation (it rained less than forecast). If the observed frequency was *above* the expected frequency, this would have indicated underestimation (it rained more than forecast).

Table 3. A comparison of forecasted probability, expected frequency and observed frequency in relation to rain forecasts from the early 1950s (Brier, 1950)

Forecasting probability (category)	Expected frequency (for a perfectly accurate forecaster)	Observed frequency	Difference between expected and observed outcomes
0.00–0.19 (very unlikely)	0.10	0.07	−0.03
0.20–0.39 (unlikely)	0.30	0.10	−0.20
0.40–0.59 (evens)	0.50	0.29	−0.21
0.60–0.79 (likely)	0.70	0.40	−0.30
0.80–1.00 (very likely)	0.90	0.50	−0.40

found—although he also discovered that weather forecasters routinely *overestimated* the chances of rain, and increasingly so as their forecasts approached 100 per cent.

We replicated this approach in our analysis. Using Microsoft Excel for Mac (version 16.69), we grouped our forecasts together using the same categories as Brier (1950) and labelled them—*very unlikely* (for forecasts of 0–19 per cent), *unlikely* (20–39 per cent), *evens* (40–59 per cent), *likely* (60–79 per cent) and *very likely* (80–100 per cent). The observed frequency of outcomes was then calculated for each category by counting the number of events and actions that happened, comparing this with the overall number of forecasts, and calculating the frequency. For example,

assume a set of 100 forecasts, and that the specified actions and events occurred ninety times. This would result in an observed frequency of 0.90. We applied this approach to the entire dataset ($n = 21{,}193$), by sorting all the forecasts into the categories described above, and then for each category counting how many of the forecast events happened or did not happen. We then did the same again in relation to each individual referral (i.e., we sort the forecasts made in relation to each referral into these same categories and for each category counted how many of the forecast events happened or did not happen). Throughout, we used descriptive statistics only, and did not test to see if the differences between forecast and observed outcomes were statistically significant.

Ethics

Approval for data collection was provided by the School of Social Sciences' ethics Committee at Cardiff University. As part of the informed consent process, we asked respondents for their permission to reuse their data in future secondary analyses.

Results

As social workers' forecasts increased towards 100 per cent, so did the observed frequency of specified actions and events. When social workers judged an action or event to be *very unlikely* (a forecast of 0–19 per cent) it happened less often than when they judged the action or event to be *unlikely* (20–39 per cent). When they judged the action or event to be *unlikely*, it happened less often than when they judged the action or event to be *likely* (60–79 per cent)—and so on. We also found that social workers routinely overestimated the likelihood of the specified action or event, unless they were forecasting that it was *very unlikely* (0–19 per cent)—in which case, as shown in Figure 3 and Table 4, the specified actions or events happened slightly more often than anticipated.

We also considered forecasts in which social workers were certain (when they forecast either 0 per cent = will definitely not happen, or 100 per cent = will definitely happen). As shown in Figure 4, when social workers said the specified action or event *would definitely not happen*, it occurred 9 per cent of the time; when they said the specified action or event *would definitely happen*, it occurred 65 per cent of the time.

We then considered the difference between the expected frequency and the observed frequency in relation to the forecasting categories. As shown in Table 5, we found that for most of them, the difference between the minimum and maximum observed frequencies was between 0.35 and 0.44, apart from the *most likely* category (80–100 per cent),

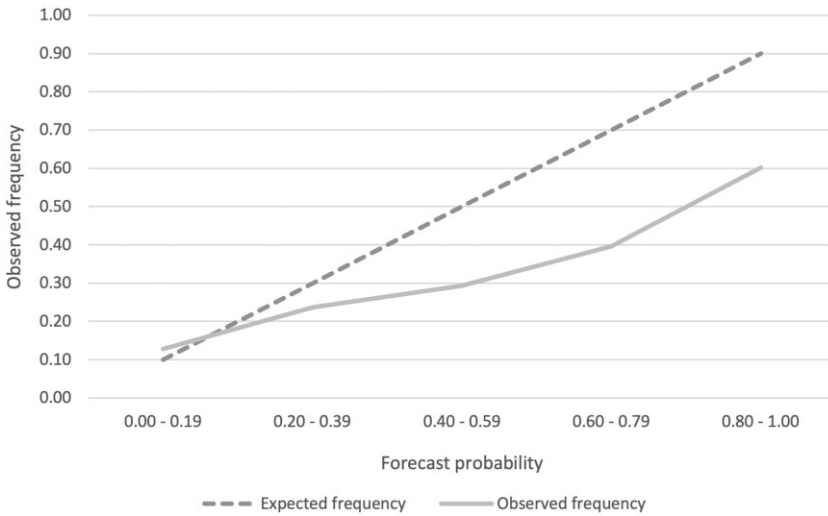

Figure 3: A comparison of expected frequencies and observed frequencies for social worker forecasts. The observed frequency is *below* the expected frequency indicating overestimation (outcomes mostly occurred less often than forecast).

Table 4. A comparison of expected outcomes (for an idealised forecaster) and verified outcomes in relation to social work forecasts

Forecast range	Number of forecasts	Expected frequencies (for a perfectly accurate forecaster)	Observed frequency of outcomes	Difference between expected and observed frequencies
0.00–0.19 (very unlikely)	6,849	0.10	0.128	+0.028
0.20–0.39 (unlikely)	2,385	0.30	0.237	−0.063
0.40–0.59 (evens)	4,053	0.50	0.293	−0.207
0.60–0.79 (likely)	3,763	0.70	0.396	−0.304
0.80–1.00 (very likely)	4,143	0.90	0.602	−0.298
Total	21,193	–	–	–

where the difference was 0.70. When social workers forecast that an action or event was *very unlikely* (0–19 per cent), the proportion of observed outcomes fell between 0.02 and 0.38. When social workers forecast the action or event was *unlikely* (20–39 per cent), the proportion of observed outcomes fell between 0.00 and 0.43. However, when social workers forecast the action or event was very likely (80–100 per cent), the proportion of observed outcomes fell between 0.09 and 0.79. Thus, if a social worker said that an action or event was *very unlikely* or *unlikely*, it would probably happen no more than four times in every ten. However, if a social worker said that an action or event was *very likely*, it could happen as often as eight times in ten, or less than one time in ten.

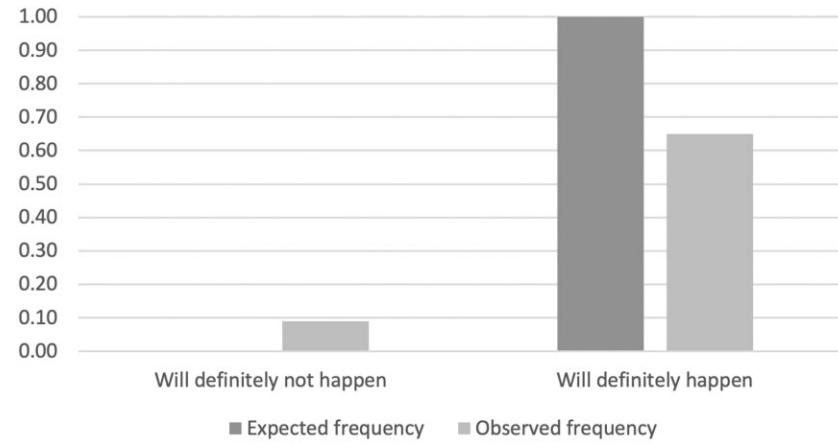

Figure 4: A comparison of expected frequencies and observed frequencies for social work forecasts made with absolute certainty (either 0 = definitely will not happen or 100 = definitely will happen).

Table 5. The range, maximum and minimum for the proportion of observed outcomes in each forecasting category

Forecast category	Observed frequencies (mean)	Min. observed frequency	Max. observed frequency	Difference between min. and max. observed frequencies
0.00–0.19 (very unlikely)	0.14	0.02	0.38	0.36
0.20–0.39 (unlikely)	0.27	0.00	0.43	0.43
0.40–0.59 (evens)	0.28	0.11	0.55	0.44
0.60–0.79 (likely)	0.38	0.22	0.57	0.35
0.80–1.00 (very likely)	0.59	0.09	0.79	0.70

Finally, we considered the frequency of observed outcomes for the forecasting categories in relation to each referral. Figure 5 shows that (i) as the forecasts increased towards 100 per cent , the specified actions and events happened more often and (ii) the general trend of over-estimation. We also observed some apparent similarities and differences between the referrals. For referrals 1 and 12, the forecasts made in the *very unlikely* category (0–19 per cent) trended towards over-estimation (the specified actions or events almost never happened, when you would expect a frequency of 0.1). This included a mixture of possible outcomes, including the likelihood of another referral, of the child coming into care, or the LA holding a strategy meeting. On the other hand, for referrals 3, 9 and 11, the proportion of observed outcomes in the *very likely* category (80–100 per cent) was very low (no higher than 0.20). As indicated in Table 1 (above), referral 1 involved a pre-school child and

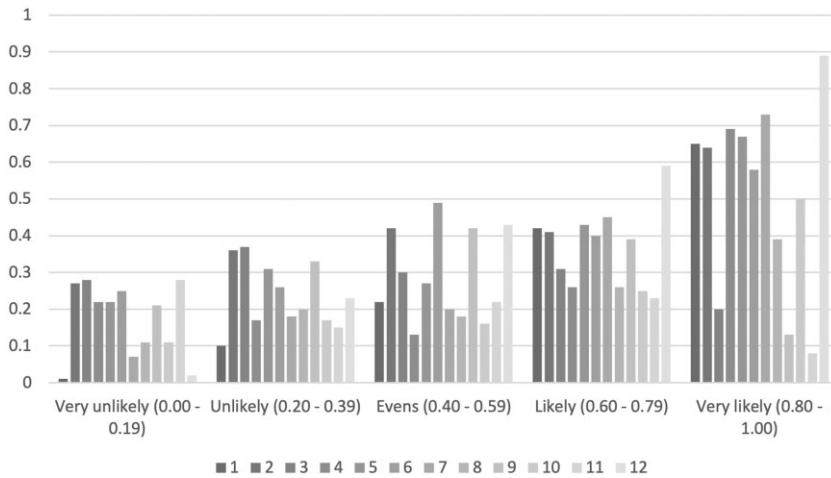

Figure 5: A comparison of observed frequencies for social work forecasts made in relation to each of the referrals individually.

concerns about domestic abuse concerns. Referral 12 involved an unborn child and concerns about maternal mental health problems and learning disability. What these two referrals might have in common to explain the similar pattern of over-estimating *very unlikely* events is hard to say. Likewise, referral 3 involved an unborn child and a perceived lack of maternal engagement, whilst referral 9 involved a primary-school age child and problems in private family proceedings. Referral 11 involved a male teenager and concerns about domestic abuse. Again, what might link these referrals such that respondents significantly over-estimated the likelihood of *very likely* actions and events is difficult to discern. It may be that these similarities are less explicable by reference to the nature of the referrals, and more influenced by the specificity of the information (Platt, 2006).

Discussion

Based on more than 20,000 forecasts made by nearly 600 social workers in relation to twelve referrals, we found that for each step-increase in forecasting category, the specified action or event was more likely to happen. When social workers judged that an action or event was *unlikely*, it occurred on average less often than when they judged an action or event to be *likely*. Thus, the general trend was in the right (i.e., more accurate) direction. However, much like American weather forecasters in the 1950s, social workers consistently overestimated the likelihood of specified actions or events, unless they judged them to be *very unlikely*

(in which case, they slightly underestimated them). When social workers were certain that something *would not* happen, it happened on average around one time in ten. When social workers were certain that something *would* happen, it happened on average less than seven times in ten. The range of observed outcomes suggests that these judgements would in many cases lack 'practical adequacy', such that the workers' expectations about the world would often go unrealised (Sayer, 2010). This is not the same as saying that social workers are good at judging when something will *not* happen, and less good at judging when it *will*. It is the task of estimating the specified outcome, whether defined positively *or* negatively, that seems to pose the challenge, and not the nature of the event itself. Of course, one can frame forecasting questions so that anyone can accurately estimate the outcome. For example, how likely is it that the sun will rise tomorrow? So likely as to be practically certain. Or, how likely is it that you will win millions of pounds on the national lottery in the next week? So unlikely as to be practically impossible. There are not many judgements in social work that would be similarly certain or uncertain and yet still worth making. Instead, it may be that social workers routinely operate outside the bounds of the genuinely foreseeable (at least, without the dubious benefits of hindsight). Thus—to reiterate—we have not found that social workers *can* judge when something *will not* happen (e.g., a re-referral will not be made) whilst being less good at judging *when it will* (e.g., a re-referral will be made). Rather, both 'types' of outcome (positive and negative) are inherently challenging to foresee and based on our data, we cannot say that one is any easier to forecast than the other. In future, we want to explore in more detail the issue of whether and what differences there are in forecasting accuracy in relation to different question types, the characteristics of the individual worker (e.g., levels of experience) and the organisational context. For example, can social workers make more accurate forecasts about procedural responses than they can about the behaviour of parents and young people?

Perhaps the most pressing question is whether these results are reflective of actual practice. It is important to acknowledge that these judgements were made under study conditions, far removed from the ways in which social workers would normally respond to referrals. However, whether implicitly or explicitly, social workers make professional judgements in their day-to-day work all the time, and one aspect of this is whether they can anticipate the likelihood of future actions and events. In many forecasting studies, it has been found that what predicts greater judgemental accuracy is not subject expertise, but a more diffuse set of abilities, such as pattern recognition, critical reflection and open-mindedness (Mellers *et al.*, 2014, 2015a,b, 2017, 2019). In fact, subject expertise has been associated with *less* accuracy, perhaps because it can engender over-confidence (Arkes, 2001) Thus, whilst the study conditions were

artificial, this does not imply that judgements made by social workers in practice conditions would necessarily be more accurate (although they could be).

Hindsight bias and the influence of organisational conditions

Munro (2019) has noted that when we review social work judgements and decisions, we are often prone to hindsight bias—thinking that what seems obvious *now*, should have been obvious *at the time*. In day-to-day practice, it has also been suggested that social workers tend to be over-optimistic (Child Safeguarding Practice Review Panel, 2020, p. 9), susceptible to the availability heuristic (how easy it is to think of examples), confirmation bias (looking for evidence to confirm an existing view), groupthink (striving for consensus), decision avoidance (postponing for as long as possible) and emotional bias (Department for Education, 2014, pp. 22–27). Yet it would be meaningless to consider human judgement separately from context (Rose, 2016). Whilst individual cognitive debiasing training may result in a ~10 per cent improvement in forecasting accuracy (Morewedge *et al.*, 2015), the real challenge is to recognise that 'situations can be more or less error-provoking [and to] create [conditions] in which errors are less likely' (Reason, 2009, p. 32). Other factors associated with forecasting accuracy include—having more deliberation time (Mellers *et al.*, 2015a) and structured collaboration in diverse groups (Önkal *et al.*, 2011). In social work, several authors have noted the significant influence of organisational culture. Munro and Hubbard (2011) identified the influence of factors such as alignment between organisational and individual goals, the negative impact of stress resulting from excessive workloads, and whether errors are treated as opportunities for learning or not. Similarly, Platt and Turney (2014) explained how individual judgements are mediated via policy and organisational contexts.

Conditions of uncertainty: false positive and false negative errors

Thus, it would be a mistake to focus solely on the individual and overlook the complex ecology and conditions of uncertainty within which social work happens (Baumann *et al.*, 2014). There is an unavoidable chance of error when making judgements and not primarily because of individual deficits (Munro, 2020). As it is impossible to make entirely accurate judgements in social work, there must always be a balance between false positives and false negatives (Wilkins and Meindl, 2023). The degree to which these error types are acceptable is determined not by the individual, or by social work organisations, but by wider society

(Munro, 2020). One interpretation of our results is that when asked to make *any* kind of judgement, social workers have 'learned' not to make the mistake of underestimation. In other words, it is 'safer' (psychologically, professionally, organisationally) to say something is more likely to occur than less. If so, this would be a problematic basis on which to make decisions. Yet it would also be understandable, as it is for making false negative errors that social workers are most often criticised. In a national review into the tragic deaths of Star Hobson and Arthur Labinjo-Hughes (Child Safeguarding Practice Review Panel, 2022), it was suggested that workers demonstrated 'a lack of critical thinking' (p. 8). The report said we 'need … sharper specialist child protection skills and expertise, especially in relation to complex risk assessment and decision-making' (p. 8). When such awful events are reported in the media, it is common to see social workers castigated for missing 'obvious' signs of maltreatment (Cavanagh and Sims, 2021)—for making false negative errors. Formal case reviews are only rarely held because of false positive errors (Cousin *et al.*, 1991), and whilst there is some reporting of 'unnecessary' child protection investigations and interventions, these are usually less high-profile and employ less incendiary language. Social workers are spat at in the street when they 'miss' child abuse, not when they place children into care 'unnecessarily' (BBC, 2021).

Services have tended to respond to this challenge 'by lowering the threshold for intervention to minimise the chances of missing another child in such extreme danger' (Munro, 2020, p. 87). All else being equal, this results in more families being subject to child protection investigations, a growing proportion of which result in no further action (Bilson and Martin, 2016). And whilst it would be easy to criticise organisations for lowering thresholds, it is a reasonable response to these wider pressures. When you are repeatedly disparaged for missing 'obvious' signs of maltreatment, 'the easiest coping response is the crudest form of organisational learning: whatever you do next time, don't make the last mistake. In signal detection terms, you … shift your response threshold for 'crying wolf'' (Tetlock and Mellers, 2014, p. 11575).

Limitations of the study

There are three main limitations to our study. First, we recruited a non-probability sample. All such samples are skewed in more and less obvious ways. As such, we do not make any claims of generalisability. Second, we could not consider systematically the differences between question types and referrals. This is something to explore in future studies, alongside the influence of factors such as individual experience, workloads and organisational values. Finally, we asked respondents to make forecasts by reading anonymised referrals and completing a survey.

This is not how social workers make judgements in real-life. Hence, these findings may not reflect the accuracy that social workers achieve in practice. It is also possible that by asking social workers to express their judgements using a numerical scale, they were biased towards over-estimation (feeling more confident that they knew what was going to happen). On the other hand, over-confidence is a common finding in the literature on judgement and decision-making and in many different fields, and numerical scales may actually provide one way of *reducing* over-estimation (Arkes, 2001; Russo and Schoemaker, 1992).

Conclusion

There can be no objective standard as to what level of accuracy social workers 'should' be able to achieve, and in relation to what kinds of questions. Using a baseline of common sense, one would expect that social workers should make forecasts more accurately than chance (e.g., compared to flipping a coin) and our results suggest that they often can—even when presented only with very limited referral information, and in the absence of other contextual information that would be available in real-life practice settings. Thus, in many ways the respondents in our study performed admirably in achieving the reported level of accuracy.

In a more perfect world, social workers would never 'miss' a case of child abuse or neglect—in a truly perfect world, there would be no child abuse or neglect to miss—but we do not live in a perfect world. Given this unfortunate reality, what can we do, other than accept an apparently continual lowering of intervention thresholds? If we believe that social work judgements are already being made at the limits of the forecasting frontier (Tetlock *et al.*, 2012), there may not be much else we can do. Yet if we can 'facilitate higher-order forms of learning that push out performance frontiers, not just shift response thresholds' (Tetlock and Mellers, 2014, p. 11575) then we might be able to reduce positive and negative errors of judgement at the same time. The welcome news is that 'good judgement can be learned, honed, and sharpened' (Mellers *et al.*, 2017, p. 379).

Making judgements under conditions of uncertainty 'is one of the most pervasive and difficult aspects of life. Uncertainty … makes error inevitable, and error makes injustice unavoidable' (Hammond, 2000, p. 35). Thus, we must not be surprised that social workers make judgemental errors. After all, when 'irreducible uncertainty exists in the environment, then all our knowledge will not completely eliminate all errors from our actions' (Hammond, 2000, p. 20). However uncomfortable it may be, we must (learn to) tolerate such uncertainty. But we make it even harder for social workers to make accurate forecasts when they are castigated

and attacked for making false-negative errors. Within this context, it is understandable to over-estimate the likelihood of almost all actions and events, with the aim of 'not missing' anything. Yet there are also costs—personal, social and financial—to this approach. Whilst it is an understandable and absolute priority to protect children from abuse and neglect, 'shifting the decision criterion … does not affect the total amount of injustice; it [only] changes its allocation' (ibid, p. 45).

Acknowledgements

Thank you to all the social workers and other members of staff who took part, to the LAs that supported the study and to Professor Donald Forrester for his helpful comments and feedback on earlier drafts.

Funding

This study was partially funded by What Works for Children's Social Care (now called Foundations) and supported by infrastructure funding from Health and Care Research Wales.

Conflict of interest statement

None declared.

References

Allen, N. (2005) *Making Sense of the Children Act 1989*, Hoboken, NJ, John Wiley & Sons.

Arkes, H. R. (2001) 'Overconfidence in judgmental forecasting', in J. Armstrong (ed), *Principles of Forecasting*, International Series in Operations Research & Management Science, pp. 495–515, Berlin, Germany, Springer.

BBC. (2021) 'Star Hobson: Social Workers 'spat at on the street', available online at: https://www.bbc.co.uk/news/uk-england-leeds-59684207 (accessed January 2021).

Baumann, D. J., Fluke, J. D., Dalgleish, L. and Kern, H. (2014) 'The decision making ecology', in A. Shlonsky and R. Benbenishty (eds), *From Evidence to Outcomes in Child Welfare: An International Reader*, pp. 24–40. New York, NY, OUP USA.

Bilson, A. and Martin, K. E. (2016) '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), pp. 793–811.

Bhaskar, R. (2013) *A Realist Theory of Science*, London, Routledge.

Boghossian, P. and Lindsay, J. (2019) *How to Have Impossible Conversations: A Very Practical Guide*, Lebanon, IN, Da Capo Lifelong Books.

Brier, G. W. (1950) 'Verification of forecasts expressed in terms of probability', *Monthly Weather Review*, **78**(1), pp. 1–3.

Cavanagh, N. and Sims, P. (2021) 'Left to die' Star Hobson: Shocking failure of social workers to spot horrific abuse of TWO murdered toddlers', available online at: https://www.thesun.co.uk/news/17042982/star-hobson-failure-social-workers-two-toddlers/ (accessed January 2023).

Child Safeguarding Practice Review Panel. (2020) 'Annual report 2018 to 2019', available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/870033/Child_Safeguarding_Practice_Review_Panel__Annual_Report_2018_2019.pdf (accessed January 2023).

Child Safeguarding Practice Review Panel. (2022) 'Child protection in England', available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1078488/ALH_SH_National_Review_26-5-22.pdf (accessed January 2023).

Cousin, J., Cousin, G., McGrath, S. and Fine, R. (1991) 'System abuse of children: The Orkney case', *Probability Journal*, **38**, 121.

Cross, T. P. and Casanueva, C. (2009) 'Caseworker judgments and substantiation', *Child Maltreatment*, **14**(1), pp. 38–52.

Cyr, C., Dubois-Comtois, K., Paquette, D., Lopez, L. and Bigras, M. (2022) 'An attachment-based parental capacity assessment to orient decision-making in child protection cases: A randomized control trial', *Child Maltreatment*, **27**(1), pp. 66–77.

DePanfilis, D. and Girvin, H. (2005) 'Investigating child maltreatment in out-of-home care: Barriers to effective decision-making', *Children and Youth Services Review*, **27**(4), pp. 353–74.

Dickens, J., Howell, D., Thoburn, J. and Schofield, G. (2005) 'Children starting to be looked after by local authorities in England: An analysis of inter-authority variation and case-centred decision making', *British Journal of Social Work*, **37**(4), pp. 597–617.

Drucker, P. F. (1967) *The Effective Decision*, Cambridge, MA, Harvard University. Graduate School of Business Administration.

Eysenck, M. W. and Keane, M. T. (2015) *Cognitive Psychology: A Student' Handbook*, London, Psychology Press.

Farmer, E. and Lutman, E. (2014) 'Working effectively with neglected children and their families—What needs to change', *Child Abuse Review*, **23**(4), pp. 262–73.

Forrester, D. (2008) 'Child protection and re-referrals involving serious concerns: a follow-up study of 400 referrals closed by Social Services Departments', *Child & Family Social Work*, **13**(3), pp. 286–99.

Gillingham, P. and Humphreys, C. (2010) 'Child protection practitioners and decision-making tools: Observations and reflections from the front line', *British Journal of Social Work*, **40**(8), pp. 2598–616.

Hammond, K. R. (2000) *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*, Oxford, Oxford University Press on Demand.

Haran, U. and Moore, D. A. (2014) 'A better way to forecast', *California Management Review*, **57**(1), pp. 5–15.

Hood, R., Abbott, S., Coughlan, B., Nilsson, D., Duschinsky, R., Parker, P. and Mannes, J. (2022) Improving the quality of decision making and risk assessment in children's social care: A rapid evidence review. What Work's for Children's Social Care, London.

Masson, J. (2006) 'Fair trials in child protection', *Journal of Social Welfare and Family Law*, **28**(1), pp. 15–30.

Meindl, M. and Wilkins, D. (2021) 'Can social workers forecast future actions, events, and outcomes? A study of referrals to Children's Services in England', *Child Care in Practice*. https://doi.org/10.1080/13575279.2021.2001434

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E. and Tetlock, P. (2015a) 'The psychology of intelligence analysis: Drivers of prediction accuracy in world politics', *Journal of Experimental Psychology: Applied*, **21**(1), pp. **1**, 1–14.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., Ungar, L. and Tetlock, P. (2015b) 'Identifying and cultivating superforecasters as a method of improving probabilistic predictions', *Perspectives on Psychological Science*, **10**(3), pp. 267–81.

Mellers, B., Tetlock, P. and Arkes, H. R. (2019) 'Forecasting tournaments, epistemic humility and attitude depolarization', *Cognition*, **188**, 19–26.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E. and Tetlock, P. E. (2014) 'Psychological strategies for winning a geopolitical forecasting tournament', *Psychological Science*, **25**(5), pp. 1106–15.

Mellers, B. A., Baker, J. D., Chen, E., Mandel, D. R. and Tetlock, P. E. (2017) 'How generalizable is good judgment? A multi-task, multi-benchmark study', *Judgment & Decision Making*, **12**(4), pp. 369–81.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H. and Kassam, K. S. (2015) 'Debiasing decisions: Improved decision making with a single training intervention', *Policy Insights from the Behavioral and Brain Sciences*, **2**(1), pp. 129–40.

Munro, E. (1999) 'Protecting children in an anxious society', *Health, Risk & Society*, **1**(1), pp. 117–27.

Munro, E. (2004) 'A simpler way to understand the results of risk assessment instruments', *Children and Youth Services Review*, **26**(9), pp. 873–83.

Munro, E. (2019) 'Decision-making under uncertainty in child protection: Creating a just and learning culture', *Child & Family Social Work*, **24**(1), pp. 123–30.

Munro, E. (2020) Causality and its implications for Theories of Change and evaluations of complex systems. CHESS Working Paper No. 2020-04. Durham University, Durham.

Munro, E. and Hubbard, A. (2011) 'A systems approach to evaluating organisational change in children's social care', *British Journal of Social Work*, **41**(4), pp. 726–43.

Önkal, D., Lawrence, M. and Sayım, K. Z. (2011) 'Influence of differentiated roles on group forecasting accuracy', *International Journal of Forecasting*, **27**(1), pp. 50–68.

Platt, D. (2006) 'Threshold decisions: How social workers prioritize referrals of child concerns', *Child Abuse Review*, **15**(1), pp. 4–18.

Platt, D. and Turney, D. (2014) 'Making threshold decisions in child protection: A conceptual analysis', *British Journal of Social Work*, **44**(6), pp. 1472–90.

Reason, J. (2009) 'From little slips to big disasters: An error quest', in P. Rabbitt (ed), *Inside Psychology: A Science over 50 Years*, pp. 233–42. Oxford University Press, Oxford.

Rose, T. (2016) *The End of Average: How to Succeed in a World That Values Sameness*, London, Penguin UK.

Russo, J. E. and Schoemaker, P. J. (1992) 'Managing overconfidence', *Sloan Management Review*, **33**(2), pp. 7–17.

Sayer, A. (2010) *Method in Social Science: Revised*, 2nd edn, Abingdon, Routledge.

Sidebotham, P. (2012) 'What do serious case reviews achieve', *Archives of Disease in Childhood*, **97**(3), pp. 189–92.

Taylor, B. (2013) *Professional Decision Making and Risk in Social Work*, Exeter, Learning Matters.

Taylor, B. J. (2017) 'Heuristics in professional judgement: A psycho-social rationality model', *The British Journal of Social Work*, **47**(4), pp. 1043–60.

Taylor, C. and White, S. (2001) 'Knowledge, truth and reflexivity: The problem of judgement in social work', *Journal of Social Work*, **1**(1), pp. 37–59.

Taylor, C. and White, S. (2005) 'Knowledge and reasoning in social work: Educating for humane judgement', *British Journal of Social Work*, **36**(6), pp. 937–54.

Tetlock, P. and Mellers, B. (2014) 'Judging political judgment', *Proceedings of the National Academy of Sciences of the United States of America*, **111**(32), pp. 11574–5.

Tetlock, P. E., Horowitz, M. C. and Herrmann, R. (2012) 'Should "systems thinkers" accept the limits on political forecasting or push the limits', *Critical Review*, **24**(3), pp. 375–91.

Wilkins, D. and Boahen, G. (2013) *Critical Analysis Skills for Social Workers*, London, McGraw Hill.

Wilkins, D. and Meindl, M. (2023) 'Measuring the ratio of true-positive to false-positive judgements made by child and family social workers in England: A case vignette study', *Child & Family Social Work*. https://doi.org/10.1111/cfs.13086

Wood, M. A. (2004) A study of the perception of the impact of modeling on the development of commitment to action in Decision Conferencing. Unpublished PhD thesis, Curtin University.