Can Social Workers Forecast Future Actions, Events, and Outcomes? A Study of Referrals to Children's Services in England

Child protection social workers in England are required to make many decisions in their day-to-day work, including whether to accept a referral, undertake a child protection investigation, pursue care proceedings, or close the case. Many of these decisions involve implicit or explicit predictions about the likelihood of future actions, events, and outcomes. This paper presents the results of a study in which social workers and social work students in England were asked at two time points, six to eight months apart, to read a series of case vignettes and make forecasts about the likelihood of different actions, events, and outcomes. The accuracy of these forecasts was measured to determine 1) the aggregate level and range of forecasting accuracy, 2) whether forecasting accuracy is stable over time, 3) whether different vignettes are harder or easier to forecast and 4) whether personal or professional factors are predictors of forecasting accuracy. On average, respondent's forecasts were 6% better than you would expect by chance, although there was significant variation within the sample. Nearly three-quarters of the respondents who made forecasts more accurately than chance at Time 1 did so again at Time 2. Four of the vignettes were found to be consistently easier to forecast, and four were consistently more difficult. No personal or professional characteristics were found to be significant predictors of forecasting accuracy. There are few straight-forward decisions in social work and the question of how best to support practitioners as they undertake this critical aspect of their role will continue to be an important focus for research.

Introduction

Child and family social workers make potentially life-changing decisions almost every day of their working lives. Many of these decisions require some element of forecasting – making predictions about possible future actions, events, or outcomes (Mumford et al., 2015), whether implicitly or explicitly. Social workers use relevant information at hand to "arrive at a judgement and to predict what might happen to the child and family" (Benbenishty & Fluke, 2020, p. 12). Most significantly, social workers in England (and Wales) are required by the 1989 Children Act to make judgements about the likelihood of children experiencing significant harm. Such judgements involve making "a prediction from existing facts, often from a multitude of such facts, about what has happened in the past, about the characters and personalities of the people involved, [and] about the things which they have said and done" (Parliament. House of Lords, 2008). The very definition of an 'at-risk' child implies a judgement about the likelihood of future harm (Campbell & Humphreys, 1993), and there is a strong association between social work assessments of risk, and the decision to investigate or intervene in private family life (King et al., 2021).

Social workers are not the only professionals required to make predictions about complex problems. Probation officers assess the likelihood of future offending behaviour (Lancaster & Lumb, 2006; Vincent et al., 2016), economists make predictions about the timing and extent of future recessions (Fratesi, 2009), and meteorologists make regular and public forecasts about the weather (Ehrendorfer & Murphy, 1988). However, human beings, in general, are poor at making accurate predictions about the future (Buturovic, 2010; Stenmark et al., 2010). Tetlock (2017) reports, in relation to political forecasting, that even widely acknowledged experts are only marginally more accurate, on average, than a group of 'dart-tossing chimps.' Yet, the ability to make high quality, accurate forecasts remains a critically

important component of decision-making. Stenmark has found that higher quality forecasting – for example, examining a wider range of possible outcomes - is associated with more ethical decisions (Stenmark, 2013; Stenmark et al., 2010). This finding may be especially important for fields, such as social work, where decisions are likely to have substantial consequences for individuals involved (Stenmark, 2013).

There are many complexities to consider when studying social work forecasting, judgement, and decision-making (Helm, 2016). It is not simply the individual characteristics and abilities of the worker that are important, let alone their ability to forecast the future. Organisational context matters, as do the views and wishes of the child and family. Social work decision-making is also constrained and underpinned by the law, by statutory guidance (Braye & Preston-Shoot, 1998), and by personal and professional values (Landau, 1999). The Baumann et al. (2014) well-known model provides a helpful outline of the key factors involved (Figure 1).

[Figure 1 to be inserted here]

Despite the influence of these factors, the role of the individual social worker, their beliefs, and their judgements about the family, are still important. Rodrigues et al. (2020) suggest that "professional and individual decisions are...the most elementary unit of the system" (p. 150, emphasis added). This is reminiscent of Evans and Harris' (2004) conception of social workers as 'street-level bureaucrats', interacting directly with the public to interpret and implement the law. While individual social workers do not usually make significant decisions by themselves, their judgement about the likelihood of future actions, events, and outcomes is an influential part of the process. For example, social workers refer people to

services in part because they expect – or forecast – that the service will be beneficial. They visit a family at a particular time in part because they expect someone will be home. They seek to remove a child from their parent's care in part because they expect the child to be unsafe otherwise.

As described by Benbenishty and Fluke (2020), "[social workers] use case information to arrive at a judgement and to predict what might happen to the child and family…the worker's prediction may be accurate, but it may also prove to be wrong" (p. 13 – 14). In social work, this argument is often presented in terms of false-positives and false-negatives (Mansell et al., 2011; table 1). If a worker believes a child is at significant risk of abuse, they may seek to remove the child from home. If the child was not at risk, the worker's belief about the child was a false-positive. On the other hand, if a worker believes a child is safe at home, they may seek to close the case. If the child was at risk of abuse, the worker's belief was a false-negative (Forrester & Harwin, 2006). Munro (2004, 2009, 2010) has previously noted the consequences of making false-negative errors may be far more consequential for the worker than false-positive errors – primarily, a child may end up being abused or killed – even though both types of errors are sub-optima and harmful in different ways.

[Table 1 to be inserted here]

The question we examine in this paper is – to what extent can social workers accurately forecast different actions, events, and outcomes following a referral to social services? We propose that the ability to accurately predict the likelihood of different actions, events, and outcomes can serve as a useful proxy measure for the ability to judge other important outcomes that may be less tangible (such as whether a child will be safe or not in future), and for some

important aspects of good judgement – including openness to new information, a well calibrated sense of confidence, and the ability to consider a variety of different perspectives. These attributes are associated with more accurate forecasting and are commonly identified as important for good social work decision-making (in the UK Professional Capabilities Framework; British Association of Social Workers, 2018); they are also skills that can be practiced and developed (Mellers et al., 2015a; 2015b). We hypothesise that if (or when) you have these skills, you can make more informed judgements, measured by the accuracy of your forecasts about different actions, events, and outcomes, and this leads to better, more ethical, decision-making.

This hypothesis forms the basis for an ongoing programme of work (Wilkins and Forrester, 2019). More broadly, the proposed relationship between social worker's judgement and decision-making is based on Dalgleish's (1988) General Assessment and Decision-Making (GADM) model, whereby judgements are made prior to the decisions that follow, and in which judgements inform but do not determine the course of action (Figure 2).

[Figure 2 to be inserted here]

Method

This paper presents the results of a study in which social workers were asked on two separate occasions, six to eight months apart, to read a series of case vignettes as part of a survey and make forecasts about the likelihood of different actions, events, and outcomes. The accuracy of these forecasts was scored with the aim of determining 1) the aggregate level and the range of forecasting accuracy, 2) whether accuracy is stable over time, 3) whether different

vignettes are harder or easier to forecast and 4) whether personal or professional factors are predictors of forecasting accuracy.

The study consisted of two online surveys hosted in Qualtrics (https://www.qualtrics.com). In survey one (T1), social workers were asked to read four vignettes, based on real-life referrals to one local authority in England. This survey formed part of a randomised controlled trial of decision-making interventions (the results of which have been reported elsewhere; Wilkins et al., 2020). In survey two (T2), the same social workers were asked to read six vignettes, some of which they had seen before at T1 and some of which were new. After each vignette, respondents were asked four questions, to assess the likelihood of different actions, events, and outcomes (see Table 2 for some examples). Respondents were given clear instructions about the task and completed a practice question prior to viewing the vignettes. This ensured that respondents understood the task and emphasised to them that they were being asked to make forecasts, rather than to give a view on what should happen, or to make a (simulated) decision.

[Table 2 to be inserted here]

Data collection

Respondents were asked to respond to the questions using a scale from 0 to 100, where 0 indicated the action, event, or outcome would definitely not happen, and 100 indicated it definitely would happen (Figure 3). Respondents were able to choose any number between 0 and 100. As the vignettes were based on real-life referrals, the subsequent actions, events, and outcomes were known from the case files and this information was used to generate the questions for each vignette. This allowed the accuracy of each respondent's forecasts to be

calculated using Brier scores (see below). In doing this we sought to evaluate the extent to which respondents made accurate forecasts about what happened following the referral. Participants were not asked to make their own decision or recommendation about what should have happened, nor to pass judgement about the real-life decisions that had been made.

[Figure 3 to be inserted here]

In addition to the vignette questions, respondents were also asked about their personal and professional characteristics (including age, gender, current role, and length of experience). T1 data were collected between August and October 2019. T2 data were collected between March and June 2020.

Sample

In total, there were 283 respondents for the T1 survey, all of whom were qualified social workers or student social workers in England. Respondents were recruited via social media announcements and by contacting social work employers. Of the T1 respondents, 103 consented to be contacted again in the future. Of these, 72 took part in the T2 survey, and 63 completed it ($\geq$ 98%). The findings presented here relate to the 63 respondents who fully completed both surveys.

Case vignettes

A total of twelve case vignettes were used in the T1 survey, all of which were based on real-life referrals obtained from one local authority in England. The referrals were purposefully selected so that a range of actions, events, and outcomes could be included. At T1, all respondents were allocated the same two vignettes and were then each randomly allocated

another two vignettes from the remaining pool of ten. At T2, the respondents were allocated the same two vignettes from the start of T1, followed by a random selection of four vignettes from the same sample. A mixture of random and pre-selected vignettes was used to explore whether different vignettes might be easier or harder to forecast, and to assess forecasting accuracy over time.

Data analysis

Data were analysed using SPSS (version 25) and Microsoft Excel. For the forecasts made in response to the vignettes, Brier scores were calculated for each possible outcome. Brier scores are used to measure the accuracy of probabilistic predictions and range from 0 to 2, where 0 indicates complete accuracy and 2 indicates complete inaccuracy (see Table 3 for a worked example). 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 (Tetlock & Gardner, 2016, p. 64). Brier scores are calculated using the following formula, where x = the forecast for the outcome that occurs and y = the forecast for the outcome that does not occur:

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

Brier scores for each vignette were calculated by averaging the scores for the associated questions. An overall mean Brier score for each respondent was calculated by averaging their Brier scores for all of the vignettes they were allocated.

[Table 3 to be inserted here]

Findings

This section presents findings in relation to the following:

1) Overall levels and the range of forecasting accuracy in the sample

2) The reliability of forecasting accuracy between T1 and T2

3) Forecasting accuracy in relation to the different vignettes

4) Associations between forecasting accuracy and personal and professional characteristics

Forecasting accuracy and range

The overall mean Brier score for the combined T1 and T2 forecasts was 0.47 (SD = .11) which is slightly more accurate than you would expect by chance (Table 4). The range of accuracy within the sample was considerable, from 56% more accurate than you would expect by chance (a Brier score of 0.22) to 76% less accurate than you would expect by chance (a Brier score of 0.88).

[Table 4 to be inserted here]

Forecasting reliability

To determine whether respondents made similar forecasts at T1 and T2, their Brier scores for the two vignettes repeated in both surveys were explored. A paired-samples t-test determined that there was no statistically significant difference in Brier scores between T1 (M = 0.41, SD = 0.16) and T2 (M = 0.41, SD = 0.15), t (62) = -0.17, p > .05 (two-tailed). A Pearson's correlation coefficient test showed there was a medium, positive correlation between Brier scores for the two time points, r = .48, n = 63, p < .001. This indicates that on average respondents made reasonably similar forecasts each time when presented with the same vignettes.

The reliability of forecasting accuracy over time was analysed by comparing overall mean Brier scores for T1 (calculated using all four vignettes, none of which respondents had seen before) with their Brier scores for the randomly allocated vignettes in T2. A paired-samples t-test determined there was no statistically significant difference in Brier scores between T1 (M = 0.47, SD = 0.11) and T2 (M = 0.50, SD = 0.11), t (62) = -1.53, p > .05 (two-tailed). A Pearson's correlation coefficient test showed there was also a medium, positive correlation between overall mean Brier scores and Brier scores for the randomly allocated vignettes, r = .37, n = 63, p = .003. This could indicate that respondents who made more accurate forecasts in T1 also made more accurate forecasts in T2 (and vice versa). We also examined how many respondents had overall Brier scores better than chance at T1 and whether the same respondents also scored better than chance at T2. At T1, there were 39 respondents with Brier scores better than chance. Of those, 28 (72%) had a Brier score better than chance at T2. There were 24 respondents at T1 whose overall Brier score was equal to or worse than chance. At T2, 11 of those respondents still had Brier scores worse than chance, while 13 respondents improved and had Brier scores better than chance.

Variability in case vignette forecasting difficulty

At T1, respondents made more accurate forecasts for the two baseline vignettes (seen by every respondent) compared with those they were randomly assigned. For the two seen by everyone, 73% of respondents were more accurate than chance, while only 41% were more accurate than chance for those randomly assigned. At T2, respondents also made more accurate forecasts in relation to the two baseline vignettes, compared with those randomly assigned (76% compared with 57.1%). A paired-samples t-test confirmed that Brier scores for previously seen vignettes (M = 0.41, SD = 0.15) were significantly lower than Brier scores for

randomly allocated vignettes (M = 0.50, SD = 0.11), t (62) = -4.74, p < .01 (two-tailed). The mean difference in Brier scores was -0.09 with a 95% confidence interval ranging from -0.12 to -0.05.

As respondents were randomly allocated different vignettes at T1 and T2, this suggests some of the variability in forecasting accuracy is related to the nature of the vignettes, rather than whether respondents had seen them before. To explore this, mean Brier scores for each of the vignettes used in both surveys were examined. To allow comparison between the two time points, the original vignette numbering has been retained. Table 5 shows the overall mean Brier scores for the eight referrals used in T1 and T2, as well as the minimum and maximum scores, and the range.

[Table 5 to be inserted here]

Overall, more accurate forecasts than you would expect by chance were made in relation to case vignettes 3, 7, 11 and 12. Worse forecasts than you would expect by chance were made in relation to case vignettes 1, 2, 5, and 8. At both T1 and T2, respondents made the most accurate forecasts in relation to vignette 12 and the least accurate forecasts in relation to vignette 1.

Impact of personal and professional characteristics on forecasting ability

Standard multiple regression was used for exploratory analysis on whether respondents' personal and professional characteristics influenced their forecasting accuracy. The variables included in the regression models were *age, gender, ethnicity, current role, current team, years of experience*, and *highest level of educational qualification*. There was no

violation of the assumptions of normality, linearity, multicollinearity, and homoscedasticity in either model.

At T1, *years of experience* was moderately correlated with Brier scores. The remaining variables had small to no correlation with Brier scores. The model explained 8.1% of the variance in Brier scores (adjusted R^2 = .081, F (7, 53) = 1.76, p = 0.12), and no variable significantly predicted Brier scores. At T2, *age*, *ethnicity*, and *years of experience* were moderately correlated with Brier scores. The remaining variables had small to no correlation with Brier scores. The model explained 12.1% of the variance in Brier scores (adjusted R^2 = .121, F (7, 53) = 2.18, p = 0.05), and no variable significantly predicted Brier scores.

Summary of the study and findings

A sample of social workers and social work students in England were asked to read a series of case vignettes and make forecasts about the likelihood of different actions, events, and outcomes. The case vignettes were based on real-life referrals from one local authority in England. On average, the forecasts made by respondents were 6% better than you would expect by chance, although there was a significant range within the sample – some respondents were much more accurate than chance (up to 56%), while some were much worse than chance (up to 76%). Nearly three-quarters of the respondents who made forecasts more accurately than chance at T1 did so again at T2. Four of the vignettes were consistently easier for respondents to forecast, and four were consistently more difficult.

Years of experience was the only characteristic to show a relationship with forecasting accuracy across T1 and T2 but explained only a small percentage of the variance. None of the characteristics we considered (*age, gender, ethnicity, current role, current team, years of*

experience, and *highest level of educational qualification*) were significant predictors of forecasting accuracy.

Discussion

Social workers in England make decisions about the families they work with, which are informed in part by their forecasts about the likelihood of different future actions, events, and outcomes. The extent to which they make such forecasts explicitly and accurately in their day-to-day practice is yet unclear. This study is a very small, first step in what we hope will be a series of future studies about social work forecasting.

The forecasting accuracy of social workers and students in this study was only slightly better than you would expect by chance, though, it is important to note that the study conditions were highly artificial. Social workers will hardly ever make judgements based solely on such limited information as was provided in the vignettes. Even if they did, they would still have more contextual knowledge of the local authority, thresholds and attitudes to risk, the local community, and the availability of services, all of which are likely to help inform their forecasts. One way of interpreting these results is that *any ability* to make forecasts more accurately than chance is impressive, given the paucity of information available to respondents. It is also worth noting that some of the respondents were able to forecast the likelihood of different actions, events, and outcomes far more accurately than you would expect by chance, while the accuracy of others was far worse.

Two main questions emerge from this study. First, what is a reasonable benchmark for accurate forecasting in social work? Is anything better than chance acceptable, or do we expect social workers to do (even) better than this? Second, how accurately are social workers able to

make forecasts about future actions, events, and outcomes when they have more information? If their forecasts are about families they are working with and are therefore more representative of practice and based on greater availability of information, we might expect them to make more accurate forecasts, thus achieving lower Brier scores than the respondents in this study. Though, it is possible that making forecasts in-context may be more difficult than making them in an artificial scenario. The Theory of Planned Behaviour (Ajzen, 2011) suggests behaviour is influenced by *subjective norms* – what 'relevant others' will think about us if we behave in certain ways. There are many ways in which social workers might be influenced by this. For example, social workers have often been subjected to public opprobrium following tragic cases of child death. The judgement of social workers may be influenced by a fear of being blamed if something goes wrong, far more than a concern with accuracy per se (Kanani et al., 2002). This may lead to workers 'preferring' to make false-positive assessments of risk. If so, workers may end up making inflated forecasts about the likelihood of bad things happening to children, so that if they do, they are vindicated for being right, rather than vilified for being wrong (Warner, 2015).

When considering the results of this study, it is important to keep in mind the nature of forecasting more generally. Consider the following question – why is it easier to predict the weather in Hawaii than in Chicago? The answer relates not to the individual qualities of the forecaster (although these will presumably play some part) but to the nature of the environment. The weather in Hawaii is very stable; no other place on earth has such a narrow range of temperatures (Sanderson, 1993). Once you know this, accurately predicting the weather in Hawaii should be relatively easy. Conversely, the weather in Chicago is notoriously changeable. According to Ralph Kiner (a famous American broadcaster) "if you don't like the weather in Chicago, wait fifteen minutes". Accurately predicting the weather in Chicago is

much more difficult because it also involves judgements about the likelihood of things changing.

Making a contingent forecast – that if things stay the same, X or Y is the more likely outcome – is easier than predicting whether things will change. A child living with domestic violence is very likely, perhaps even certain, to experience serious harm. But whether the child will continue living with domestic violence, or whether their situation will improve and over what timeframe, are much more difficult questions. So, do social workers in general make contingent forecasts, or do they make forecasts about the likelihood of change? We believe they make both. For example, in care proceedings, social workers make forecasts about whether a child continuing to live in their current circumstances will be at risk of significant harm (a contingent forecast), *and* they make forecasts about a parent's ability (with support) to change their child's current circumstances, the likelihood that these changes will take place, in what timeframe, and whether those changes will effectively reduce the risk of harm (forecasts about the likelihood of change).

In practice, social workers consider constellations of questions related to the families they are working with to help determine if a child will be at risk of significant harm in the future. For example, how often is the child attending school? Have the police been called to the house for any reason? Has the child been taken to the hospital with an injury? These types of questions about actions, events, and outcomes can be answered (as we have shown in this study) with measurable rigour and therefore could be useful proxy indicators (and proxy forecasting questions) for the risk of significant harm. In future research studies, we intend to consider how social workers respond to constellations of questions, whether the forecasts from

individual questions might be combined to provide an overall forecast about the likelihood of significant harm, and the accuracy of such forecasts.

Strengths and limitations

The method used in this study (measuring the accuracy of forecasting using Brier scores) is novel to the field of social work decision-making research. We were able to draw on evidence from other fields about the kinds of skills and ways of working that can facilitate more accurate forecasts and thus inform good judgement. While there is still a significant gap in the research on social work decision-making more generally, this study offers a starting point for gaining a deeper understanding of the accuracy of social work forecasting and provides the basis for future studies of forecasting and decision-making in real life.

The initial sample at T1 was reasonably large, although the number of respondents who participated again at T2 was (inevitably) lower. As a result, the sample included in the final analysis was smaller than desired. A larger sample would have been more beneficial for identifying associations between personal and professional characteristics and forecasting accuracy. The scenario in which social workers were asked to make forecasts in this study was also artificial. As noted already, social workers are not generally required to make forecasts based on such little information or to make them so explicitly in real life. It is important to emphasise that we are not advocating for social workers to use the Brier score methodology as a practice tool.

Conclusions

Whether or not a particular social work intervention has a high or low chance of success, or whether a particular action, event or outcome is thought to be likely or unlikely,

should not alone dictate decision-making. Whatever the judgement of the social worker, parents and children have the right to a private family life, and the right to available support. Parents should also be afforded the opportunity to prove social worker's negative forecasts or predictions wrong - a worker may think it unlikely that a parent will stop misusing alcohol, but still refer the parent to a support service to give them the best chance of success, accepting that they may be wrong about the likelihood of the parent stopping drinking.

This programme of work about social work forecasting is at an early stage. Many elements of our hypothesis about the relationship between more accurate forecasts, good judgement and ethical decision-making need to be further theorised and tested. At present, social workers are regularly asked to make judgements about the future likelihood of significant harm to children, and a belief in their ability to do so may be contributing to the rising numbers of new-borns being taken into care (Critchley, 2020; Hestbæk et al., 2020; Juhasz, 2020). Consequently, it is important to explore and understand what forecasts social workers make, and what helps them to be more accurate. If making reasonably accurate forecasts in social work is not possible, because every situation is unique and the environment is too complex, then we would do well to recognise this. There are few straight-forward decisions in social work and the question of how best to support practitioners as they undertake this critical aspect of their role will continue to be an important focus for research (Helm & Roesch-Marsh, 2017) and for policymaking (Blackwell, 2021).

References

Ajzen, I. (2011). *The theory of planned behaviour: Reactions and reflections.* Taylor & Francis.

Baumann, D. J., Fluke, J., Dalgleish, 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.

Blackwell, A. (2021). *Care review launches with focus on ensuring love, stability and safety for children in families and in care*. Community Care. https://www.communitycare.co.uk/2021/03/02/care-review-launches-focus-ensuring-love-stability-safety-children-families-care/ [Accessed March 2021].

Braye, S. & Preston-Shoot, M. (1998). Social work and the law. In *Social Work* (pp. 56-66). Springer.

British Association of Social Workers. (2018). *Professional Capabilities Framework*. https://www.basw.co.uk/system/files/resources/pcf-social-worker.pdf [Accessed March 2021].

Buturovic, Z. (2010). Putting political experts to the test. *Critical Review*, 22(4), 389-396.

Campbell, J. & Humphreys, J. (1993). *Nursing care of survivors of family violence*. Mosby Incorporated.

Critchley, A. (2020). 'The lion's den': Social workers' understandings of risk to infants. *Child & Family Social Work*, 25(4), 895-903.

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, pp. 317-360). Elsevier.

Ehrendorfer, M. & Murphy, A. H. (1988). Comparative evaluation of weather forecasting systems: Sufficiency, quality, and accuracy. *Monthly Weather Review*, 116(9), 1757-1770.

Evans, T. & Harris, J. (2004). Street-level bureaucracy, social work and the (exaggerated) death of discretion. *The British Journal of Social Work,* 34(6), 871-895.

Forrester, D. & Harwin, J. (2006). Parental substance misuse and child care social work: findings from the first stage of a study of 100 families. *Child & Family Social Work*, 11(4), 325-335.

Fratesi, U. (2009). The regional economist with the crystal ball: a discussion of the possibility of long-term predictions at sub-national level. *The Regional Economist with the Crystal Ball*, 1000-1007.

Helm, D. (2016). Sense-making in a social work office: an ethnographic study of safeguarding judgements. *Child & Family Social Work*, 21(1), 26-35.

Helm, D. & Roesch-Marsh, A. (2017). The ecology of judgement: A model for understanding and improving social work judgements. *British Journal of Social Work*, 47(5), 1361-1376.

Hestbæk, A.-D., Höjer, I., Pösö, T., & Skivenes, M. (2020). Child welfare removal of infants: Exploring policies and principles for decision-making in Nordic countries. *Children and Youth Services Review*, 108, 104572.

Juhasz, I. B. (2020). Child welfare and future assessments–An analysis of discretionary decision-making in newborn removals in Norway. *Children and Youth Services Review*, 116, p. 105137.

Kanani, K., Regehr, C., & Bernstein, M. M. (2002). Liability considerations in child welfare: Lessons from Canada. *Child abuse & neglect*, 26(10), 1029-1043.

King, B., Black, T., Fallon, B., & Lung, Y. (2021). The Role of Risk in Child Welfare Decision-Making: A Prospective Cohort Examination of Families Transferred to Ongoing Child Protection Services after an Investigation. *Children and Youth Services Review*, 125, 105983.

Lancaster, E. & Lumb, J. (2006). The assessment of risk in the National Probation Service of England and Wales. *Journal of Social Work*, 6(3), 275-291.

Landau, R. (1999). Professional socialization, ethical judgment and decision making orientation in social work. *Journal of Social Service Research*, 25(4), 57-75.

Mansell, J., Ota, R., Erasmus, R., & Marks, K. (2011). Reframing child protection: A response to a constant crisis of confidence in child protection. *Children and Youth Services Review*, 33(11), 2076-2086.

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E., & Tetlock, P. (2015a). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of experimental psychology: applied*, 21(1), 1.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., & Ungar, L., (2015b). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science*, 10(3), 267-281.

Mumford, M. D., Steele, L., McIntosh, T., & Mulhearn, T. (2015). Forecasting and leader performance: Objective cognition in a socio-organizational context. *The Leadership Quarterly*, 26(3), 359–369.

Munro, E. (2004). A simpler way to understand the results of risk assessment instruments. *Children and Youth Services Review,* 26(9), 873-883.

Munro, E. (2009). Managing societal and institutional risk in child protection. *Risk Analysis: An International Journal*, 29(7), 1015-1023.

Munro, E. (2010). Conflating risks: Implications for accurate risk prediction in child welfare services. *Health, Risk & Society*, 12(2), 119-130.

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.

Rodrigues, L. B., Calheiros, M., & Pereira, C. (2020). The Psychosocial Process Underlying Residential Care Placement Decisions. *Decision-Making and Judgment in Child Welfare and Protection: Theory, Research, and Practice*, 149.

Sanderson, M. (1993). *Prevailing trade winds: climate and weather in Hawaii*. University of Hawaii Press.

Stenmark, C. K. (2013). Forecasting and Ethical Decision Making: What Matters?, *Ethics & Behavior*, 23(6), 445-462.

Stenmark, C. K., Antes, A. L., Wang, X., Caughron, J. J., Thiel, C. E., & Mumford, M. D. (2010). Strategies in Forecasting Outcomes in Ethical Decision-Making: Identifying and Analyzing the Causes of the Problem. *Ethics and behaviour,* 20(2), 110-127.

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.

Vincent, G. M., Guy, L. S., Perrault, R. T., & Gershenson, B. (2016). Risk assessment matters, but only when implemented well: A multisite study in juvenile probation. *Law and Human Behavior*, 40(6), 683.

Warner, J. (2015). *The emotional politics of social work and child protection*. Policy Press.

Wilkins, D., & Forrester, D. (2019). Predicting the Future in Child and Family Social Work: Theoretical, Ethical and Methodological Issues for a Proposed Research Programme. *Child Care in Practice,* 26(2), 196-209.

Wilkins, D., Foster, C., Sanders, M., & Reid, L. (2020). *Good judgement and social work decision-making: A randomised controlled trial of brief interventions to improve forecasting*. London: What Works for Children's Social Care.