Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the 'Black Box' of Machine Learning

Philip Gillingham*

School of Social Work and Human Services, University of Queensland, St Lucia Campus, Brisbane, Queensland, Australia

*Correspondence to Dr Philip Gillingham, Senior Research Fellow, School of Social Work and Human Services, University of Queensland, St Lucia Campus, Brisbane, Queensland, Australia. E-mail: p.gillingham@uq.edu.au

Abstract

Recent developments in digital technology have facilitated the recording and retrieval of administrative data from multiple sources about children and their families. Combined with new ways to mine such data using algorithms which can 'learn', it has been claimed that it is possible to develop tools that can predict which individual children within a population are most likely to be maltreated. The proposed benefit is that interventions can then be targeted to the most vulnerable children and their families to prevent maltreatment from occurring. As expertise in predictive modelling increases, the approach may also be applied in other areas of social work to predict and prevent adverse outcomes for vulnerable service users. In this article, a glimpse inside the 'black box' of predictive tools is provided to demonstrate how their development for use in social work may not be straightforward, given the nature of the data recorded about service users and service activity. The development of predictive risk modelling (PRM) in New Zealand is focused on as an example as it may be the first such tool to be applied as part of ongoing reforms to child protection services.

Keywords: Predictive risk modelling, risk assessment, preventative intervention

Accepted: March 2015

Introduction

Preventing child maltreatment, rather than responding to provide protection to children who may have already been maltreated, has become a major concern of governments around the world as notifications to child protection services have risen year on year (Kojan and Lonne, 2012; Munro, 2011). One response has been to provide universal services to families deemed to be in need of support but whose children do not meet the threshold for tertiary involvement, conceptualised as a public health approach (O'Donnell *et al.*, 2008). Risk-assessment tools have been implemented in many jurisdictions to assist with identifying children at the highest risk of maltreatment in order that attention and resources be directed to them, with actuarial risk assessment deemed as more efficacious than consensus based approaches (Coohey *et al.*, 2013; Shlonsky and Wagner, 2005). While the debate about the most efficacious form and approach to risk assessment in child protection services continues and there are calls to progress its development (Le Blanc *et al.*, 2012), a criticism has been that even the best risk-assessment tools are 'operator-driven' as they need to be applied by humans. Research about how practitioners actually use risk-assessment tools has demonstrated that there is little certainty that they use them as intended by their designers (Gillingham, 2009b; Lyle and Graham, 2000; English and Pecora, 1994; Fluke, 1993). Practitioners may consider risk-assessment tools as 'just another form to fill in' (Gillingham, 2009a), complete them only at some time after decisions have been made and change their recommendations (Gillingham and Humphreys, 2010) and regard them as undermining the exercise and development of practitioner expertise (Gillingham, 2011).

Recent developments in digital technology such as the linking-up of databases and the ability to analyse, or mine, vast amounts of data have led to the application of the principles of actuarial risk assessment without some of the uncertainties that requiring practitioners to manually input information into a tool bring. Known as 'predictive modelling', this approach has been used in health care for some years and has been applied, for example, to predict which patients might be readmitted to hospital (Billings *et al.*, 2006), suffer cardiovascular disease (Hippisley-Cox *et al.*, 2010) and to target interventions for chronic disease management and end-of-life care (Macchione *et al.*, 2013). The idea of applying similar approaches in child protection is not new. Schoech *et al.* (1985) proposed that 'expert systems' could be developed to support the decision making of professionals in child welfare agencies, which they describe as 'computer programs which use inference schemes to apply generalized human expertise to the facts of a specific case' (Abstract). More recently, Schwartz, Kaufman and Schwartz (2004) used a 'backpropagation' algorithm with 1,767 cases from the USA's Third National Incidence Study of Child Abuse and Neglect to develop an artificial neural network that could predict, with 90 per cent accuracy, which children would meet the

criteria set for a substantiation of abuse. Schoech (2010) describes how technological advances which connect databases from different agencies, allowing the easy exchange and collation of information about people, can ‘accumulate intelligence with use; for example, those using data mining, decision modelling, organizational intelligence strategies, wiki knowledge repositories, etc.’ (p. 8). In England, in response to media reports about the failure of a child protection service, it has been claimed that ‘understanding the patterns of what constitutes a child at risk and the many contexts and circumstances is where big data analytics comes in to its own’ (Solutionpath, 2014).

The focus in this article is on an initiative from New Zealand that uses big data analytics, known as predictive risk modelling (PRM), developed by a team of economists at the Centre for Applied Research in Economics at the University of Auckland in New Zealand (CARE, 2012; Vaithianathan *et al.*, 2013). PRM is part of wide-ranging reform in child protection services in New Zealand, which includes new legislation, the formation of specialist teams and the linking-up of databases across public service systems (Ministry of Social Development, 2012). Specifically, the team were set the task of answering the question: ‘Can administrative data be used to identify children at risk of adverse outcomes?’ (CARE, 2012). The answer appears to be in the affirmative, as it was estimated that the approach is accurate in 76 per cent of cases—similar to the predictive strength of mammograms for detecting breast cancer in the general population (CARE, 2012). PRM is designed to be applied to individual children as they enter the public welfare benefit system, with the aim of identifying children most at risk of maltreatment, in order that supportive services can be targeted and maltreatment prevented.

The reforms to the child protection system have stimulated debate in the media in New Zealand, with senior professionals articulating different perspectives about the creation of a national database for vulnerable children and the application of PRM as being one means to select children for inclusion in it. Particular concerns have been raised about the stigmatisation of children and families and what services to provide to prevent maltreatment (New Zealand Herald, 2012a). Conversely, the predictive power of PRM has been promoted as a solution to growing numbers of vulnerable children (New Zealand Herald, 2012b). Sue Mackwell, Social Development Ministry National Children’s Director, has confirmed that a trial of PRM is planned (New Zealand Herald, 2014; see also AEG, 2013). PRM has also attracted academic attention, which suggests that the approach may become increasingly important in the provision of welfare services more broadly:

In the near future, the type of analytics presented by Vaithianathan and colleagues as a research study will become a part of the ‘routine’ approach to delivering health and human services, making it possible to achieve the ‘Triple Aim’: improving the health of the population, providing better service to individual clients, and reducing per capita costs (Macchione *et al.*, 2013, p. 374).

The application of PRM as part of a newly reformed child protection system in New Zealand raises a number of moral and ethical concerns and the CARE team propose that a full ethical review be conducted before PRM is used. A thorough interrogation of these concerns is provided by [Keddell \(2014a\)](#) and the aim in this article is not to add to this side of the debate. Rather it is to explore the challenges of using administrative data to develop an algorithm which, when applied to families in a public welfare benefit database, can accurately predict which children are at the highest risk of maltreatment, using the example of PRM in New Zealand. As [Keddell \(2014a\)](#) points out, scrutiny of how the algorithm was developed has been hampered by a lack of transparency about the process; for example, the complete list of the variables that were finally included in the algorithm has yet to be disclosed. There is, though, sufficient information available publicly about the development of PRM, which, when analysed alongside research about child protection practice and the data it generates, leads to the conclusion that the predictive ability of PRM may not be as accurate as claimed and consequently that its use for targeting services is undermined. The consequences of this analysis go beyond PRM in New Zealand to affect how PRM more generally may be developed and applied in the provision of social services.

The application and operation of algorithms in machine learning have been described as a 'black box' in that it is considered impenetrable to those not intimately familiar with such an approach ([Gillespie, 2014](#)). An additional aim in this article is therefore to provide social workers with a glimpse inside the 'black box' in order that they might engage in debates about the efficacy of PRM, which is both timely and important if [Macchione *et al.*'s \(2013\)](#) predictions about its emerging role in the provision of social services are correct. Consequently, non-technical language is used to describe and analyse the development and proposed application of PRM.

PRM: developing the algorithm

Full accounts of how the algorithm within PRM was developed are provided in the report prepared by the CARE team ([CARE, 2012](#)) and [Vaithianathan *et al.* \(2013\)](#). The following brief description draws from these accounts, focusing on the most salient points for this article.

A data set was created drawing from the New Zealand public welfare benefit system and child protection services. In total, this included 103,397 public benefit spells (or distinct episodes during which a particular welfare benefit was claimed), reflecting 57,986 unique children. Criteria for inclusion were that the child had to be born between 1 January 2003 and 1 June 2006, and have had a spell in the benefit system between the start of the mother's pregnancy and age two years. This data set was then divided into two sets, one being used to train the algorithm (70 per cent), the other to test it

(30 per cent). To train the algorithm, probit stepwise regression was applied using the training data set, with 224 predictor variables being used.

In the training stage, the algorithm ‘learns’ by calculating the correlation between each predictor, or independent, variable (a piece of information about the child, parent or parent’s partner) and the outcome, or dependent, variable (a substantiation or not of maltreatment by age five) across all the individual cases in the training data set. The ‘stepwise’ design of this process refers to the ability of the algorithm to disregard predictor variables that are not sufficiently correlated to the outcome variable, with the result that only 132 of the 224 variables were retained in the final model. Each predictor variable is given a numerical weighting and, when it is applied to new cases in the test data set (without the outcome variable), the algorithm assesses the predictor variables that are present and calculates a score which represents the level of risk that each individual child is likely to be substantiated as maltreated. To assess the accuracy of the algorithm, the predictions made by the algorithm are then compared to what actually happened to the children in the test data set. To quote from CARE:

Performance of Predictive Risk Models is usually summarised by the percentage area under the Receiver Operator Characteristic (ROC) curve. A model with 100% area under the ROC curve is said to have perfect fit. The core algorithm applied to children under age 2 has fair, approaching good, strength in predicting maltreatment by age 5 with an area under the ROC curve of 76% (CARE, 2012, p. 3).

Given this level of performance, particularly the ability to stratify risk based on the risk scores assigned to each child, the CARE team conclude that PRM can be a useful tool for predicting and thereby providing a service response to children identified as the most vulnerable. They concede the limitations of their data set and suggest that including data from police and health databases would assist with improving the accuracy of PRM. However, developing and improving the accuracy of PRM rely not only on the predictor variables, but also on the validity and reliability of the outcome variable. As Billings *et al.* (2006) explain, with reference to hospital discharge data, a predictive model can be undermined by not only ‘missing’ data and inaccurate coding, but also ambiguity in the outcome variable. With PRM, the outcome variable in the data set was, as stated, a substantiation of maltreatment by the age of five years, or not. The CARE team explain their definition of a substantiation of maltreatment in a footnote:

The term ‘substantiate’ means ‘support with proof or evidence’. In the local context, it is the social worker’s responsibility to substantiate abuse (i.e., *gather clear and sufficient evidence to determine that abuse has actually occurred*). Substantiated maltreatment refers to maltreatment where there has been a finding of physical abuse, sexual abuse, emotional/psychological abuse or neglect. If substantiated, these are entered into the record system under these categories as ‘findings’ (CARE, 2012, p. 8, emphasis added).

However, as [Keddell \(2014a\)](#) notes and which deserves far more consideration, the literal meaning of ‘substantiation’ used by the CARE team may be at odds with how the term is used in child protection services as an outcome of an investigation of an allegation of maltreatment. Before considering the consequences of this misunderstanding, research about child protection data and the day-to-day meaning of the term ‘substantiation’ is reviewed.

Problems with ‘substantiation’

As the following summary demonstrates, there has been considerable debate about how the term ‘substantiation’ is used in child protection practice, to the extent that some researchers have concluded that caution must be exercised when using data about substantiation decisions ([Bromfield and Higgins, 2004](#)), with some even suggesting that the term should be disregarded for research purposes ([Kohl *et al.*, 2009](#)). The problem is neatly summarised by [Kohl *et al.* \(2009\)](#) who comment that ‘lay persons and policy makers often assume that “substantiated” cases represent “true” reports’ (p. 17). The reasons why substantiation rates are a flawed measurement for rates of maltreatment ([Cross and Casanueva, 2009](#)), even within a sample of child protection cases, are explained with reference to how substantiation decisions are made (reliability) and how the term is defined and applied in day-to-day practice (validity).

Research about decision making in child protection services has demonstrated that it is inconsistent and that it is not always clear how and why decisions have been made ([Gillingham, 2009b](#)). There are differences both between and within jurisdictions about how maltreatment is defined ([Bromfield and Higgins, 2004](#)) and subsequently interpreted by practitioners ([Gillingham, 2009b](#); [D’Cruz, 2004](#); [Jent *et al.*, 2011](#)). A range of factors have been identified which may introduce bias into the decision-making process of substantiation, such as the identity of the notifier ([Hussey *et al.*, 2005](#)), the personal characteristics of the decision maker ([Jent *et al.*, 2011](#)), site- or agency-specific norms ([Manion and Renwick, 2008](#)), characteristics of the child or their family, such as gender ([Wynd, 2013](#)), age ([Cross and Casanueva, 2009](#)) and ethnicity ([King *et al.*, 2003](#)). In one study, the ability to be able to attribute responsibility for harm to the child, or ‘blame ideology’, was found to be a factor (among many others) in whether the case was substantiated ([Gillingham and Bromfield, 2008](#)). In cases where it was not certain who had caused the harm, but there was clear evidence of maltreatment, it was less likely that the case would be substantiated. Conversely, in cases where the evidence of harm was weak, but it was determined that a parent or carer had ‘failed to protect’, substantiation was more likely.

The term ‘substantiation’ may be applied to cases in more than one way, as stipulated by legislation and departmental procedures ([Trocmé *et al.*, 2009](#)).

It might be applied in cases not only where there is evidence of maltreatment, but also where children are assessed as being 'in need of protection' (Bromfield and Higgins, 2004) or 'at risk' (Trocmé *et al.*, 2009; Skivenes and Stenberg, 2013). Substantiation in some jurisdictions may be an important factor in the determination of eligibility for services (Trocmé *et al.*, 2009) and so concerns about a child or family's need for support may underpin a decision to substantiate rather than evidence of maltreatment. Practitioners may also be unclear about what they are required to substantiate, either the risk of maltreatment or actual maltreatment, or perhaps both (Gillingham, 2009b).

Researchers have also drawn attention to which children may be included in rates of substantiation (Bromfield and Higgins, 2004; Trocmé *et al.*, 2009). Many jurisdictions require that the siblings of the child who is alleged to have been maltreated be recorded as separate notifications. If the allegation is substantiated, the siblings' cases may also be substantiated, as they might be considered to have suffered 'emotional abuse' or to be and have been 'at risk' of maltreatment. Bromfield and Higgins (2004) explain how other children who have not suffered maltreatment may also be included in substantiation rates in situations where state authorities are required to intervene, such as where parents may have become incapacitated, died, been imprisoned or children are unaccompanied refugees. They also point out that, because legislation may frame maltreatment in terms of acts of omission or commission by parents and carers, maltreatment of children by anyone outside the immediate family may not be substantiated.

Data about the substantiation of child maltreatment may therefore be unreliable and misleading in representing rates of maltreatment for populations known to child protection services but also in determining whether individual children have been maltreated. As Bromfield and Higgins (2004) suggest, researchers intending to use such data need to seek clarification from child protection agencies about how it has been produced. However, further caution may be warranted for two reasons. First, official guidelines within a child protection service may not reflect what happens in practice (Buckley, 2003) and, second, there may not have been the level of scrutiny applied to the data, as in the research cited in this article, to provide an accurate account of exactly what and who substantiation decisions include.

The research cited above has been conducted in the USA, Canada and Australia and so a key question in relation to the example of PRM is whether the inferences drawn from it are applicable to data about child maltreatment substantiations in New Zealand. The following studies about child protection practice in New Zealand provide some answers to this question.

A study by Stanley (2005), in which he interviewed seventy child protection practitioners about their decision making, focused on their 'understanding of risk and their active construction of risk discourses' (Abstract). He found that they gave 'risk' an ontological status, describing it as having physical properties and to be locatable and manageable. Accordingly, he found that an important activity for them was finding facts to substantiate risk. Wynd

(2013) used data from child protection services to explore the relationship between child maltreatment and socio-economic status. Citing the guidelines provided by the government website, she explains that

a substantiation is where the allegation of abuse has been investigated and there has been a finding of one or more of a number of possible outcomes, including neglect, sexual, physical and emotional abuse, risk of self-harm and behavioural/relationship difficulties (Wynd, 2013, p. 4).

She also notes the variability in the proportion of substantiated cases against notifications between different Child, Youth and Family offices, ranging from 5.9 per cent (Wellington) to 48.2 per cent (Whakatane). She states that:

There is no obvious reason why some site offices have higher rates of substantiated abuse and neglect than others but possible reasons include: some residents and neighbourhoods may be less tolerant of suspected abuse than others; *there may be variations in practice and administrative procedures between site offices*; or, all else being equal, there may be real differences in abuse rates between site offices. It is likely that some or all of these factors explain the variability (Wynd, 2013, p. 8, emphasis added).

Manion and Renwick (2008) analysed 988 case files from 2003 to 2004 to investigate why high numbers of cases that progressed to an investigation were closed after completion of that investigation with no further statutory intervention. They note that siblings are required to be included as separate notifications in any report to child protection services. In their sample, 30 per cent of cases had a formal substantiation of maltreatment and, significantly, the most common reason for this finding was behaviour/relationship difficulties (12 per cent), followed by physical abuse (7 per cent), emotional (5 per cent), neglect (5 per cent), sexual abuse (3 per cent) and suicide/self-harm (less than 1 per cent). Identifying children who are experiencing behaviour/relationship difficulties may, in practice, be important to providing an intervention that promotes their welfare, but including them in statistics used for the purpose of identifying children who have suffered maltreatment is misleading. Behaviour and relationship difficulties may arise from maltreatment, but they may also arise in response to other circumstances, such as loss and bereavement and other forms of trauma. Additionally, it is also worth noting that Manion and Renwick (2008) also estimated, based on the information contained in the case files, that 60 per cent of the sample had experienced 'harm, neglect and behaviour/relationship difficulties' (p. 73), which is twice the rate at which they were substantiated.

Manion and Renwick (2008) also highlight the tensions between operational and official definitions of substantiation. They explain that the legislation

specifies that any social worker who 'believes, after inquiry, that any child or young person is in need of care or protection . . . shall forthwith report the matter to a Care and Protection Co-ordinator' (section 18(1)). The implication of believing there is a need for care and protection assumes a complicated analysis of both the current and future risk of harm. Conversely, recording in

CYRAS [the electronic database] asks whether abuse, neglect and/or behaviour/relationship difficulties were found or not found, indicating a past occurrence (Manion and Renwick, 2008, p. 90).

The inference is that practitioners, in making decisions about substantiation, are concerned not only with making a decision about whether maltreatment has occurred, but also with assessing whether there is a need for intervention to protect a child from future harm.

In summary, the studies cited about how substantiation is both used and defined in child protection practice in New Zealand lead to the same concerns as other jurisdictions about the accuracy of statistics drawn from the child protection database in representing children who have been maltreated. Some of the inclusions in the definition of substantiated cases, such as 'behaviour/relationship difficulties' and 'suicide/self-harm', may be negligible in the sample of infants used to develop PRM, but the inclusion of siblings and children assessed as 'at risk' or requiring intervention remains problematic. While there may be good reasons why substantiation, in practice, includes more than children who have been maltreated, this has serious implications for the development of PRM, for the specific case in New Zealand and more generally, as discussed below.

The implications for PRM

PRM in New Zealand is an example of a 'supervised' learning algorithm, where 'supervised' refers to the fact that it learns according to a clearly defined and reliably measured (or 'labelled') outcome variable (Murphy, 2012, section 1.2). The outcome variable acts as a teacher, providing a point of reference for the algorithm (Alpaydin, 2010). Its reliability is therefore crucial to the eventual predictive accuracy of the algorithm. In the case of PRM, substantiation was used as the outcome variable to train the algorithm. However, as demonstrated above, the label of substantiation also includes children who have not been maltreated, such as siblings and others deemed to be 'at risk', and it is likely these children, within the sample used, outnumber those who were maltreated. Therefore, substantiation, as a label to signify maltreatment, is highly unreliable and a poor teacher. During the learning phase, the algorithm correlated characteristics of children and their parents (and any other predictor variables) with outcomes that were not always actual maltreatment. How inaccurate the algorithm will be in its subsequent predictions cannot be estimated unless it is known how many children within the data set of substantiated cases used to train the algorithm were actually maltreated. Errors in prediction will also not be detected during the test phase, as the data used are from the same data set as used for the training phase, and are subject to similar inaccuracy. The main consequence is that PRM, when applied to new data, will overestimate the likelihood that a child will be maltreated and include

many more children in this category, compromising its ability to target children most in need of protection. A clue as to why the development of PRM was flawed lies in the working definition of substantiation used by the team who developed it, as mentioned above. It appears that they were not aware that the data set provided to them was inaccurate and, additionally, those that supplied it did not understand the importance of accurately labelled data to the process of machine learning. Before it is trialled, PRM must therefore be redeveloped using more accurately labelled data.

More generally, this conclusion exemplifies a particular challenge in applying predictive machine learning techniques in social care, namely finding valid and reliable outcome variables within data about service activity. The outcome variables used in the health sector may be subject to some criticism, as Billings *et al.* (2006) point out, but generally they are actions or events that can be empirically observed and (relatively) objectively diagnosed. This is in stark contrast to the uncertainty that is intrinsic to much social work practice (Parton, 1998) and particularly to the socially contingent practices of maltreatment substantiation. Research about child protection practice has repeatedly shown how using 'operator-driven' models of assessment, the outcomes of investigations into maltreatment are reliant on and constituted of situated, temporal and cultural understandings of socially constructed phenomena, such as abuse, neglect, identity and responsibility (e.g. D'Cruz, 2004; Stanley, 2005; Keddell, 2011; Gillingham, 2009b).

In order to create data within child protection services that may be more reliable and valid, one way forward may be to specify in advance what information is required to develop a PRM, and then design information systems that require practitioners to enter it in a precise and definitive manner. This could be part of a broader strategy within information system design which aims to reduce the burden of data entry on practitioners by requiring them to record what is defined as essential information about service users and service activity, rather than current designs that aim to capture 'everything' (Gillingham, 2014). The challenge of deciding what can be quantified in order to generate useful predictions, though, should not be underestimated (Fluke, 2009). Further complicating factors are that researchers have drawn attention to problems with defining the term 'maltreatment' and its sub-types (Herrenkohl, 2005) and its lack of specificity: '... there is an emerging consensus that different types of maltreatment need to be examined separately, as each appears to have distinct antecedents and consequences' (English *et al.*, 2005, p. 442).

With existing data in child protection information systems, further research is required to investigate what information they currently contain that may be suitable for developing a PRM, akin to the detailed approach to case file analysis taken by Manion and Renwick (2008). Clearly, due to differences in procedures and legislation and what is recorded on information systems, each jurisdiction would need to do this individually, though completed studies may offer some general guidance about where, within case files and processes, appropriate information may be found. Kohl *et al.*

(2009) suggest that child protection agencies record the levels of need for support of families or whether or not they meet criteria for referral to the family court, but their concern is with measuring services rather than predicting maltreatment. However, their second suggestion, combined with the author's own research (Gillingham, 2009b), part of which involved an audit of child protection case files, perhaps provides one avenue for exploration. It might be productive to examine, as potential outcome variables, points within a case where a decision is made to remove children from the care of their parents and/or where courts grant orders for children to be removed (Care Orders, Custody Orders, Guardianship Orders and so on) or for other forms of statutory involvement by child protection services to ensue (Supervision Orders). Though this might still include children 'at risk' or 'in need of protection' as well as those who have been maltreated, using one of these points as an outcome variable might facilitate the targeting of services more accurately to children deemed to be most vulnerable.

Finally, proponents of PRM may argue that the conclusion drawn in this article, that substantiation is too vague a concept to be used to predict maltreatment, is, in practice, of limited consequence. It could be argued that, even if predicting substantiation does not equate accurately with predicting maltreatment, it has the potential to draw attention to individuals who have a high likelihood of raising concern within child protection services. However, in addition to the points already made about the lack of focus this might entail, accuracy is crucial as the consequences of labelling individuals must be considered. As Heffernan (2006) argues, drawing from Pugh (1996) and Bourdieu (1997), the significance of descriptive language in shaping the behaviour and experiences of those to whom it has been applied has been a long-term concern for social work. Attention has been drawn to how labelling people in particular ways has consequences for their construction of identity and the ensuing subject positions offered to them by such constructions (Barn and Harman, 2006), how they are treated by others and the expectations placed on them (Scourfield, 2010). These subject positions and expectations, in turn, impact on the extent to which service users engage constructively in the social work relationship (Munro, 2007; Keddell, 2014b). More broadly, the language used to describe social problems and those who are experiencing them reflects and reinforces the ideology that guides how we understand problems and subsequently respond to them, or not (Vojak, 2009; Pollack, 2008).

Conclusion

Predictive risk modelling has the potential to be a useful tool to assist with the targeting of resources to prevent child maltreatment, particularly when it is combined with early intervention programmes that have demonstrated success, such as, for example, the Early Start programme, also developed in New Zealand (see Fergusson *et al.*, 2006). It may also have potential to

predict and therefore assist with the prevention of adverse outcomes for those considered vulnerable in other fields of social work. The key challenge in developing predictive models, though, is selecting reliable and valid outcome variables, and ensuring that they are recorded consistently within carefully designed information systems. This may involve redesigning information systems in ways that they might capture data that can be used as an outcome variable, or investigating the information already in information systems which may be useful for identifying the most vulnerable service users.

Applying predictive models in practice though involves a range of moral and ethical challenges which have not been discussed in this article (see [Keddell, 2014a](#)). However, providing a glimpse into the ‘black box’ of supervised learning, as a variant of machine learning, in lay terms, will, it is intended, assist social workers to engage in debates about both the practical and the moral and ethical challenges of developing and using predictive models to support the provision of social work services and ultimately those they seek to serve.

Acknowledgements

The author would like to thank Dr Debby Lynch, Dr Brian Rodgers, Tim Graham (all at the University of Queensland) and Dr Emily Kelsall (University of Otago) for their encouragement and support in the preparation of this article. Funding to support this research has been provided by the Australian Research Council through a Discovery Early Career Research Award.

References

- AEG (2013) *Report to the Vulnerable Children's Board—November 2013*, He Korowai Tamariki—Advisory Expert Group on Information Security, Wellington, Ministry of Social Development, available online at www.childrensactionplan.govt.nz/assets/Uploads/CAP-AEGIS-Report.pdf.
- Alpaydin, E. (2010) *Introduction to Machine Learning*, 2nd edn, Cambridge, MA, MIT Press.
- Barn, R. and Harman, V. (2006) ‘A contested identity: An exploration of the competing social and political discourse concerning the identification and positioning of young people of inter-racial parentage’, *British Journal of Social Work*, **36**, pp. 1309–24.
- Billings, J., Dixon, J., Mijanovic, T. and Wennberg, D. (2006) ‘Case finding for patients at risk of readmission to hospital: Development of algorithm to identify high risk patients’, *British Medical Journal*, **12**, pp. 333–27.
- Bourdieu, P. (1997) *Language and Symbolic Power*, Cambridge, Blackwell Publishing Ltd.
- Bromfield, L. M. and Higgins, D. (2004) ‘The limitations of using statutory child protection data for research into child maltreatment’, *Australian Social Work*, **57**(1), pp. 19–30.
- Buckley, H. (2003) *Child Protection Work: Beyond the Rhetoric*, London, Jessica Kingsley.

- CARE (2012) *Vulnerable Children: Can Administrative Data Be Used to Identify Children at Risk of Adverse Outcomes?*, Centre for Applied Research in Economics, University of Auckland, available online at www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/vulnerable-children/auckland-university-can-administrative-data-be-used-to-identify-children-at-risk-of-adverse-outcome.pdf.
- Coohey, C., Johnson, K., Renner, L. M. and Easton, S. D. (2013) 'Actuarial risk assessment in child protective services: Construction methodology and performance criteria', *Children and Youth Services Review*, **35**, pp. 151–61.
- Cross, T. P. and Casanueva, C. (2009) 'Caseworker judgments and substantiation', *Child Maltreatment*, **14**(1), pp. 38–52.
- D'Cruz, H. (2004) *Constructing Meanings and Identities in Child Protection Practice*, Melbourne, Tertiary Press.
- English, D. E. and Pecora, P. J. (1994) 'Risk assessment as a practice method in child protective services', *Child Welfare*, **73**(5), pp. 451–74.
- English, D. J., Bangdiwala, S. I. and Runyan, D. K. (2005) 'The dimensions of maltreatment: Introduction', *Child Abuse and Neglect*, **29**(5), pp. 441–60.
- Fergusson, D. M., Grant, H., Horwood, L. J. and Ridder, E. M. (2006) 'Randomized trial of the Early Start program of home visitation: Parent and family outcomes', *Paediatrics*, **117**(3), pp. 781–6.
- Fluke, J. (1993) 'Evaluation of the Pennsylvania approach to risk assessment', in T. Tatura (ed.), *Seventh National Roundtable on Child Protective Services Risk Assessment: Summary of Highlights*, Washington, DC, American Public Welfare Association, pp. 113–72.
- Fluke, J. (2009) 'Allegory of the cave: On the theme of substantiation', *Child Maltreatment*, **14**(1), pp. 69–72.
- Gillespie, T. (2014) 'The relevance of algorithms', in T. Gillespie, P. Boczkowski and K. Foot (eds), *Media Technologies*, Cambridge, MA, MIT Press, pp. 167–93.
- Gillingham, P. (2009a) 'Practitioner perspectives on the family risk evaluation tool: An aide to decision making or "just another form to fill in"?'', *Developing Practice: The Child, Family and Youth Work Journal*, **23**(4), pp. 46–55.
- Gillingham, P. (2009b) 'The use of assessment tools in child protection: An ethnomethodological study', University of Melbourne, available online at <http://repository.unimelb.edu.au/10187/4337>.
- Gillingham, P. (2011) 'Decision making tools and the development of expertise in child protection practitioners: Are we "just breeding workers who are good at ticking boxes"?'', *Child and Family Social Work*, **16**(4), pp. 412–21.
- Gillingham, P. (2014) 'Electronic information systems in human service organisations: The what, who, why and how of information', *British Journal of Social Work*, **45**(5), pp. 1598–613.
- Gillingham, P. and Bromfield, L. M. (2008) 'Child protection, risk assessment and blame ideology', *Children Australia*, **33**(1), pp. 18–24.
- Gillingham, P. and Humphreys, C. (2010) 'Child protection practitioners and decision making tools: Observations and reflections from the frontline', *British Journal of Social Work*, **40**(8), pp. 2598–616.
- Heffernan, K. (2006) 'Social work, new public management and the language of "service user"', *British Journal of Social Work*, **36**, pp. 139–47.
- Herrenkohl, R. C. (2005) 'The definition of child maltreatment: From case study to construct', *Child Abuse and Neglect*, **29**(5), pp. 413–24.

- Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J. and Brindle, P. (2010) 'Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: A validation study', *Heart*, **94**, pp. 34–9.
- Hussey, J. M., Marshall, J. M., English, D. J., Dawes Knight, E., Lau, S., Dubowitz, H. and Kotch, J. B. (2005) 'Defining maltreatment according to substantiation: Distinction without a difference?', *Child Abuse and Neglect*, **29**(5), pp. 479–92.
- Jent, J. F., Eaton, C. K., Knickerbocker, L., Lambert, W. F., Merrick, M. T. and Dandes, S. K. (2011) 'Multidisciplinary child protection decision making about physical abuse: Determining substantiation thresholds and biases', *Children and Youth Services Review*, **33**, pp. 1673–82.
- Keddell, E. (2011) 'Reasoning processes in child protection decision making: Negotiating moral minefields and risky relationships', *British Journal of Social Work*, **41**(7), pp. 1251–70.
- Keddell, E. (2014a) 'The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool?', *Critical Social Policy* Advance Access published July 28, 2014, 10.1177/0261018314543224.
- Keddell, E. (2014b) 'Theorising the signs of safety approach to child protection social work: Positioning, codes and power', *Children and Youth Services Review* Advance Access published March 21, 2014, 10.1016/j.childyouth.2014.03.011.
- King, G., Trocmé, N. and Thatte, N. (2003) 'Substantiation as a multitier process: The results of a NIS-3 analysis', *Child Maltreatment*, **8**, pp. 173–82.
- Kohl, P. L., Jonson-Reid, M. and Drake, B. (2009) 'Time to leave substantiation behind: Findings from a national probability study', *Child Maltreatment*, **14**(1), pp. 17–26.
- Kojan, B. H. and Lonne, R. (2012) 'A comparison of systems and outcomes for safeguarding children in Australia and Norway', *Child and Family Social Work*, **17**, pp. 96–107.
- Le Blanc, V. R., Regehr, C., Shlonsky, A. and Bogo, M. (2012) 'Stress responses and decision making in child protection workers faced with high conflict situations', *Child Abuse & Neglect*, **36**(5), pp. 404–12.
- Lyle, C. G. and Graham, E. (2000) 'Looks can be deceiving: Using a risk assessment instrument to evaluate the outcomes of child protection services', *Children and Youth Services Review*, **22**(11/12), pp. 935–49.
- Macchione, N., Wooten, W., Yphantides, N. and Howell, J. R. (2013) 'Integrated health and human services information systems to enhance population-based and person-centered service', *American Journal of Preventative Medicine*, **45**(3), pp. 373–4.
- Manion, K. and Renwick, J. (2008) 'Equivocating over the care and protection continuum: An exploration of families not meeting the threshold for statutory intervention', *Social Policy Journal of New Zealand*, **33**, pp. 70–94.
- Ministry of Social Development (2012) *The White Paper for Vulnerable Children. Children's Action Plan: Identifying, Supporting and Protecting Vulnerable Children*, Wellington, New Zealand Government, available online at www.childrensactionplan.govt.nz/the-white-paper.
- Munro, E. (2007) 'The dangers of information sharing', *Social Policy Journal of New Zealand*, **31**, pp. 41–55.
- Munro, E. (2011) *The Munro Review of Child Protection: Final Report: A Child Centred System*, Department for Education, available online at www.education.gov.uk/munroreview/.
- Murphy, K. P. (2012) *Machine Learning: A Probabilistic Perspective*, Cambridge, MA, MIT Press.

- New Zealand Herald (2012a) 'Predicting trouble: Child abuse database raises eyebrows', 20 October.
- New Zealand Herald (2012b) 'Tim Dare: Abuse prediction tool too vital to ignore', 25 October.
- New Zealand Herald (2014) 'Children's "vulnerability" like revolving door', 21 July.
- O'Donnell, M., Scott, D. and Stanley, F. (2008) 'Child abuse and neglect: Is it time for a public health approach?', *Australian and New Zealand Journal of Public Health*, **32**(4), pp. 325–30.
- Parton, N. (1998) 'Risk, advanced liberalism and child welfare: The need to rediscover uncertainty and ambiguity', *British Journal of Social Work*, **28**(1), pp. 5–27.
- Pollack, S. (2008) 'Labelling clients "risky": Social work and the neo-liberal welfare state', *British Journal of Social Work*, **40**, pp. 1263–78.
- Pugh, R. (1996) *Effective Language in Health and Social Work*, London, Chapman and Hall.
- Schoech, D. (2010) 'Interoperability and the future of human services', *Journal of Technology in Human Services*, **28**(1–2), pp. 7–22.
- Schoech, D., Jennings, H., Schkade, L. L. and Hooper-Russell, C. (1985) 'Expert systems: Artificial intelligence for professional decisions', *Computers in Human Services*, **1**(1), pp. 81–115.
- Schwartz, D. R., Kaufman, A. B. and Schwartz, I. (2004) 'Computational intelligence techniques for risk assessment and decision support', *Children and Youth Service Review*, **26**, pp. 1081–95.
- Scourfield, P. (2010) 'A critical reflection on the involvement of "experts by experience" in inspections', *British Journal of Social Work*, **40**(6), pp. 1890–907.
- Shlonsky, A. and Wagner, D. (2005) 'The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management', *Children and Youth Services Review*, **27**, pp. 409–27.
- Skivenes, M. and Stenberg, H. (2013) 'Risk assessment and domestic violence: How do child welfare workers in three countries assess and substantiate the risk level of a 5-year-old girl?', *Child and Family Social Work Advance Access published September 10, 2013*, 10.1111/cfs.12092.
- Solutionpath (2014) 'Notes Big Data: Using Big Data to predict child abuse', posted 18 March, available online at <http://solutionpath.co.uk/child-welfare-would-benefit-from-big-data/>.
- Stanley, T. (2005) 'Making decisions: Social work processes and the construction of risk(s) in child protection work', Ph.D. thesis, University of Canterbury, New Zealand.
- Troc, N., Knoke, D., Fallon, B. and MacLaurin, B. (2009) 'Differentiating between substantiated, suspected, and unsubstantiated maltreatment in Canada', *Child Maltreatment*, **14**(1), pp. 4–16.
- Vaithianathan, R., Maloney, T., Putnam-Hornstein, E. and Jiang, N. (2013) 'Children in the public benefit system at risk of maltreatment: Identification via predictive modelling', *American Journal of Preventative Medicine*, **45**(3), pp. 354–9.
- Vojak, C. (2009) 'Choosing language: Social service framing and social justice', *British Journal of Social Work*, **39**(5), pp. 936–49.
- Wynd, D. (2013) *Child Abuse: An Analysis of Child Youth and Family Data*, Auckland, New Zealand, Child Poverty Action Group.