Predictive analytics in child protection

Prof Eileen Munro, Durham University and London School of Economics

CHESS Working Paper No. 2019-03
[Produced as part of the Knowledge for Use (K4U) Research Project]
Durham University
April 2019

The K4U project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 667526 K4U) The above content reflects only the author's view and that the ERC is not responsible for any use that may be made of the information it contains

CHESS working paper (Online) ISSN 2053-2660

Predictive analytics in child protection

Prof Eileen Munro

K4U Project
Philosophy Department
Durham University
50 Old Elvet
Durham DH1 3HN

Emeritus Professor of Social Policy
Department of Social Policy
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
E.Munro@lse.ac.uk

In child protection work, difficult risk assessments need to be made when deciding what actions to take when helping a child who has been abused or is likely to suffer abuse. Predicting the future is difficult and errors can be false negatives (leaving a child in danger) or false positives (removing a child who would not have been harmed). Both types of errors in child protection work have a high cost, firstly to the child and family and secondly to the professionals involved. Society is harshly critical when a child is left with parents or carers and is subsequently killed but there is also, though less frequently, severe criticism when children are thought to be removed without sufficient grounds.

Prediction has also become more important in recent years because of the political interest in 'social investment' – providing early help to those seen as likely to be problematic in some way in later life. Prevention services tackle a range of problems, including being maltreated, behaving badly and having poor health and educational outcomes. One political option is to provide universal services so there is no need to work out which children, families or communities to target. This however is less widely used nowadays and there is a shift towards targeted services. It therefore becomes necessary to find some means of determining who should receive additional help to flourish. One solution is to make services available and leave parents to seek help but this raises concerns that some families may not be willing to come forward.

Improving risk predictions has therefore attracted a lot of research attention. In the past, instruments to help professionals have either been a form of guided professional judgement or an actuarial instrument. In recent years, however, there has been a surge of interest in using predictive analytics to address important decision points with the hope that they will increase accuracy. This is facilitated by scale of datasets now available for data mining as more and more agencies develop computerized records that can be linked. Data mining in child welfare and protection has linked family members' datasets from health, education, police, income and housing, building a detailed set that enables profiling.

Predictive analytics are seen as having the ability or potential ability to identify the children (or even foetuses) who should be targeted for additional help.

I am using the term predictive analytics to refer to decision making systems that use data mining to identify patterns in large datasets and use algorithmic processes, including machine learning, to automate or support human decision making. Machine learning is the process by which a computer system trains itself to spot patterns and correlations in (usually large) datasets and to infer information and make predictions based on those patterns and correlations without being specifically programmed to do so. Typically, these systems involve 'profiling', the processing of personal data about an individual in order to evaluate personal characteristics relating to their behavior, preferences, economic situation, health etc.

Data mining has been considered valuable in other sectors. In healthcare, considerable work is going into using predictive analytics to improve decision making. But the possibility of using it in child protection causes considerable discussion and disagreement. There is general recognition that it raises serious technical, legal and ethical challenges but there are differing views on whether these can be overcome for particular decision tasks. Some are more optimistic (Cuccaro-Alamin, Foust, Vaithianathan, & Putnam-Hornstein, 2017; Schwartz, York, Nowakowski-Sims, & Ramos-Hernandez, 2017). Others are concerned that it will be used in ways that have a negative impact on children and their families ((Church & Fairchild, 2017; Eubanks, 2017; Keddell, 2015; Oak, 2015).

There has also been a mixed pattern of development in usage and the decision they are designed to support. Some places have developed decision tools based on predictive analytics and then dropped them because of concerns about accuracy or ethics (New Zealand's Predictive Risk Model for early identification of future harm; Illinois Department of Children and Family Services Rapid Safety Feedback for rating referrals to the agency hotline). Others however are in use (for example Allegheny Family Screening Tool (2017); London Councils Children's Predictive Safeguarding Model).

There are many decision points at which predictive analytics might be used in child protection so there is no single answer to this debate. My aim here is to identify the range of factors that need to be considered in deciding whether they are useful, legal and ethical in a specific decision context and to make some comments on the distinctive features of the child protection context.

Technical adequacy

On one aspect, the technical argument for predictive analytics is persuasive. Computers can analyse much larger datasets than humans can manage. This analysis also comes with a degree of accuracy and speed that outstrips human capabilities. It can uncover trends and insights a human might discount or not even consider. Machine learning systems are trained using large datasets provided by the system designer. Once trained, it can infer information or make predictions based on additional data inputted to the system and processed according to the algorithm.

However computers work with the data created by humans in specific social, historical and political conditions and consequently do not avoid biases and prejudices that may be buried in that data. Caliskan et al's study shows that '*standard machine learning can acquire stereotyped biases from textual data that reflect everyday human culture'* (Caliskan, Bryson, & Narayanan, 2017 p.183). Artificial intelligence and machine learning may perpetuate cultural stereotypes and they conclude: '*caution must be used in incorporating modules constructed via unsupervised machine learning into decision-making systems*' (Caliskan et al., 2017 p.185).

The capacity of predictive modeling to contain hidden biases is a major concern in child protection because of the nature of the datasets used. Their reliability and completeness are open to serious challenge. As discussed in Chapter 6, the core concept of child maltreatment is problematic and there is no universal, fixed and detailed definition of what it means. Professional judgments about whether a child is experiencing or is likely to experience maltreatment have low reliability, i.e. low inter-rater agreement. Several studies show this not just in making judgments about what counts as maltreatment but also which cases reach the threshold for initial investigation or for removal from the family and also whether or not practitioners are using decision support tools (e.g. Arad-Davidzon & Benbenishty, 2008; Britner & Mossler, 2002; Jergeby & Soydan, 2002; Regehr, Bogo, Shlonsky, & LeBlanc, 2010; Schuerman, Rossi, & Budde, 1999; Spratt, 2000). Therefore data is influenced by the particular practitioner who entered it.

Gillingham (2015) raises further concerns about the way the data is constructed:

> '*As information is entered into an information system, it has to be categorized according to the fields built into the information system, which may or may not fit the circumstances the practitioner has observed. This can happen in many ways (see Author's own) but an obvious example is the level of detail required by the information system. For example, a common question in risk assessment tools concerns illicit drug use by caregivers. Ticking a yes/no box in response to such a question is not only overly simplistic but confounding. In terms of data, caregivers who occasionally smoke marijuana after the children have gone to sleep are put in the same category as caregivers who inject heroin two or three times a day and spend much of their time finding the means to do so. Dick (2017) calls this the 'flattening effect' of categorizing data. Clearly there are different levels of risk of harm or neglect posed by each scenario'.*

The degree of unreliability in the dataset has significance for the overall accuracy of any predictions: 'When associations are probed between perfectly measured data (e.g. a genomic sequence) and poorly measured data (e.g. administrative claims health data), research accuracy is dictated by the weakest link' (Khoury & Ioannidis, 2014 p.1054).

The incompleteness of many of the datasets that are being used for predictive analytics is also a concern since child protection datasets, in particular, are known to be incomplete in a non-random way. They include children and families who have been referred to the service and these are known to cover only a percentage of children who suffer maltreatment. Studies of people's self reporting of maltreatment in childhood reveal a much higher number than official statistics of cases known to child protection services (Stoltenborgh, Bakermans-Kranenburg, Alink, & van Ijzendoorn, 2015). However, these studies also have wide variation among their results (Radford, 2011 Appendix D). Getting a reliable measure of a phenomenon such as child maltreatment is very difficult. For example, in England, Gilbert (2009) estimates that only 10% of cases are reported. Jud (2018) summarises a number of studies of the incidence of maltreatment which show not only that a large majority of cases are not known to services but also that the incidence varies depending on whether minor and moderate maltreatment are included as well as serious. Moreover, there is evidence that the dataset has persistent biases in the over-representation of low income families and ethnic minorities (Cawson, Wattam, Brooker, & Kelly, 2000).

The dataset of referrals to child protection also includes large numbers who, on subsequent consideration, are deemed not to need a child protective service. In the US, 57.6 % referrals were screened in during 2017 and 42.4% were screened out (Children's Bureau, 2018). In England, 37.9% of referrals were deemed not to need a service during 2017-18 (Department for Education, 2018).

Testing the accuracy of predictive tools is limited by the imperfect feedback available. If a prediction that a child is in too much danger to remain with their family leads to the child's removal into alternative care, the prediction is never tested. If a child gets a low rating and stays at home, testing is limited to the feedback from repeat referrals to child protection. Therefore, the ability to learn and rectify any errors in the predictive algorithm is weak and limits its technical adequacy.

Legal factors

There is considerable discussion in the literature on two key legal matters: problems of anonymising data and the transparency and accountability of decision making.

Developers and users of predictive analytics need to pay attention to the local laws on privacy and sharing of confidential and sensitive information without consent. In many jurisdictions, confidentiality restrictions can be lifted in child protection cases. In the English law, the threshold is if there is concern that a child is suffering or likely to suffer significant harm. This restriction has different impact on the various decision tasks that predictive analytics are designed to support. If the decision relates to assessing the risk of maltreatment of a child referred to child protection then the tools developed have typically been using the set of data that is already available to the professional decision maker. However, the growing interest in broadening the range of risk assessment to preventive services and of combining data from a wider range of datasets from other public and private services raises new legal questions.

One solution that is offered is of anonymising the data so that it cannot be linked to an identified or identifiable individual; the individual's privacy is still preserved. However, the research need is for *linked* data – enabling development of a rich profile of an individual – and this creates a problem.

Paul Ohm (2009) conducted a major review of the literature and reached the daunting conclusion: *'Data can be either useful or perfectly anonymous but never both' (2009 p.1704) .* A similar point is made in the Royal Society Report on 'Science as an Open Enterprise' (2012):

> *'It had been assumed in the past that the privacy of data subjects could be protected by processes of anonymisation such as the removal of names and precise addresses of data subjects. However, a substantial body of work in computer science has now demonstrated that*

the security of personal records in databases cannot be guaranteed through anonymisation procedures where identities are actively sought'.

Korff and Georges (2015) clarify why this is so:

'The main problem is that effective anonymisation does not just depend on stripping away direct identifiers (name, address, national identification number, date of birth) from a data set. Instead, the relevant measure is the size of the "anonymity set" – that is, the set of individuals to whom data might relate. If you're described as "a man" the anonymity set size is three and a half billion, but if you're described as "a middle-aged Dutchman with a beard" it is maybe half a million and if you're described as "a middle-aged Dutchman with a beard who lives near Cambridge" it might be three or four ' (Korff & Georges, 2015).

Pseudonymisation is offered as a partial solution. It is defined in the EU General Data Protection Regulation (GDPR) as 'the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information'. The problem lies in the phrase 'without the use of additional information'. As databases increase, additional information is becoming increasingly available.

An added danger for children comes from the potential linkages between welfare-related datasets and others so that the profiles of children and their parents can be more detailed and hence more readily de-anonymised. In the UK, for example, there are companies that pull together numerous datasets and offer a service to help you understand the profiles of households and postcodes and have also developed classifications covering health, retail and leisure activities. One such company says it offers:

'a geodemographic segmentation of the UK's population. It segments households, postcodes and neighbourhoods into 6 categories, 18 groups and 62 types. By analysing significant social factors and population behaviour, it provides precise information and an in-depth understanding of the different types of people' (Acorn, 2019)

With such detailed additional information, identifying individuals becomes more probable.

The second major legal concern is the lack of transparency in decisions made according to an algorithm and the consequent difficulties this causes if anyone wishes to challenge a judgment made about them. One US state requires details of any algorithm to be made public but many are being developed by private companies who refuse to publish on the grounds of it being their intellectual property, and commercial concerns. Even if the details are available, few people would be able to scrutinize it or understand the computation. The increasing number of 'expert' systems that create feedback loops to continuously improve the underlying algorithm create another barrier to transparency. The problem comes in several forms:

This problem is often termed 'algorithmic opacity', of which three distinct forms have been identified. The first is intentional opacity, where the system's workings concealed to protect intellectual property. The second is illiterate opacity, where a system is only understandable to those with the technical ability to read and write code. And the third is intrinsic opacity, where a system's complex decision-making process itself is difficult for any human to understand. More than one of these may combine – for example, a system can be intentionally opaque and it be the case that even if it wasn't then it would still be illiterately or intrinsically opaque. The result of algorithmic opacity is that an automated system's decision-making process may be difficult to understand or impossible to evaluate even for experienced systems designers and engineers' (Cobbe, 2018 p.5).

The use of predictive analytics is creating new challenges for legal systems as they alter the transparency of decision making and the protection of privacy.
Most jurisdictions are now implementing regulations on predictive analytics. Transparency, accountability and a 'positive impact on society' are among the key values. However, Zeuderveen Borgesias offers a word of caution:

'*Several caveats are in order regarding data protection law's possibilities as a tool to fight AI-driven discrimination. First, there is a compliance and enforcement deficit. Data Protection Authorities have limited resources. And many Data Protection Authorities do not have the power to impose serious sanctions (in the EU, such authorities received new powers with the GDPR [the EU General Data Protection Regulation]). Previously, many organisations did not take compliance with data protection law seriously. It appears that compliance improved with the arrival of the GDPR, but it is too early to tell'* (Zuiderveen Borgesius, 2018 p.24) .

How will the predictive tool be used?

A tool cannot be appraised in isolation. It will be used by people with human abilities and limitations in a physical and cultural context. Will the interaction between these be constructive or not?

A key problem will be in people's understanding of how to interpret the results. The 'base rate fallacy' is well evidenced as a common intuitive error. For professionals using predictive instruments, the practical issue is how much confidence they should have in the results. If this instrument predicts that Parent X is likely to harm her child, how likely is this to be true? If positive results are often false positives, then professionals know they need to treat the result with caution.

A famous study, 'The Harvard Medical School Test', illustrates the prevalence of the base rate fallacy in evaluating predictive tests. Staff and students at Harvard Medical School were told of a diagnostic test that had a high sensitivity of 95 per cent (of accurately identifying those with the disease) and a superb specificity of 100 per cent (no one with the disease would test negative).

They were asked the probability of someone who tested positive actually having the disease. The majority of respondents gave the answer of 0.95 – the rate of true positives – overlooking the significance of the base rate in determining the accuracy (Casscells et al., 1978). As the following section will explain, these estimates are far from accurate and, depending on whether the illness being diagnosed was common or rare, this test might or might not be clinically valuable.

Bayes theorem is the formal probability calculation to work out how likely it is that the positive or negative result is accurate but it is not intuitively obvious. When the underlying calculations are presented in terms of probability formulae, people tend to find them hard to follow but Gigerenzer and his colleagues at the Max Planck Institute for Human Development in Germany have found that people are well able to understand the reasoning when it is presented in more familiar ways (Gigerenzer, 2002).

To judge predictive accuracy – i.e. to judge among those who get positive results on the test, how many are cases of abuse, we need to have the values of three variables:

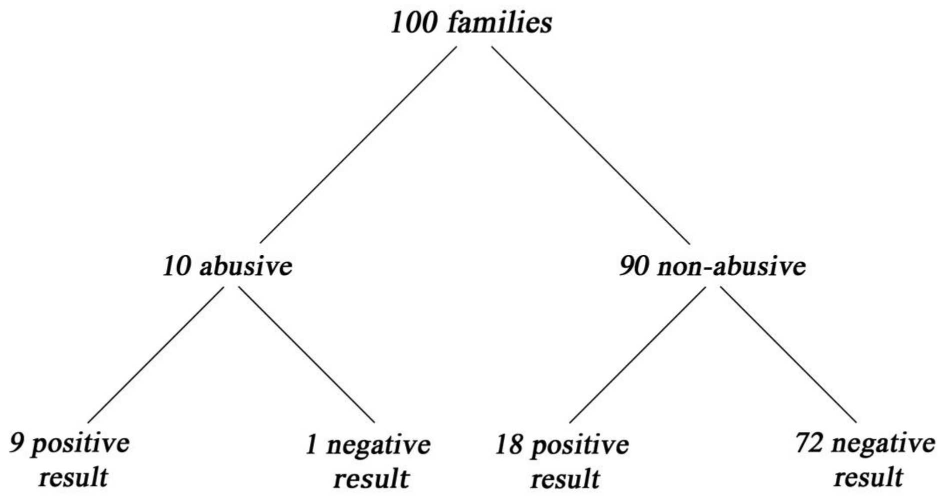
> *Sensitivity:* among many cases of abuse, how many will it predict accurately (true positives)
> *Specificity:* among non-abusive families, how many will it identify correctly (true negatives)
> *Base rate or prevalence* of the phenomenon: how common it is in the population in general.

Each of these three variables plays a distinctive part in working out the overall usefulness of an instrument, but it is the final one – the base rate – that is most often over-looked or misunderstood. Put briefly, the rarer the phenomenon being assessed, the harder it is to develop an instrument with a clinically useful level of accuracy. Conversely, the higher the base rate, the easier it is. Hence, researchers face a harder task trying to develop a risk assessment instrument to screen the general population, where the incidence of abuse is relatively low, than if their target population was specifically families known to child protection agencies, where the base rate will be much higher.

Let us take a practical example as illustration of the impact of the base rate and show how it leads to different results even when the sensitivity and specificity remain the same and are fairly high. Suppose we have an instrument where, the sensitivity is 90 per cent, and the specificity is 80 per cent and the base rate is 10 per cent

> Ten out of every 100 families in this population are abusive (the base rate). Of these 10 families, 9 will get a positive result on using the instrument (the sensitivity of 90 per cent).
>
> Of the other 90 families, around 72 will accurately get a negative result but some 18 will get a (false) positive result (the specificity of 80 per cent).
>
> Imagine the instrument has given a positive result for a group of families. How many of these families with a positive result will actually be abusive? This tree diagram helps to

make the calculation clearer.

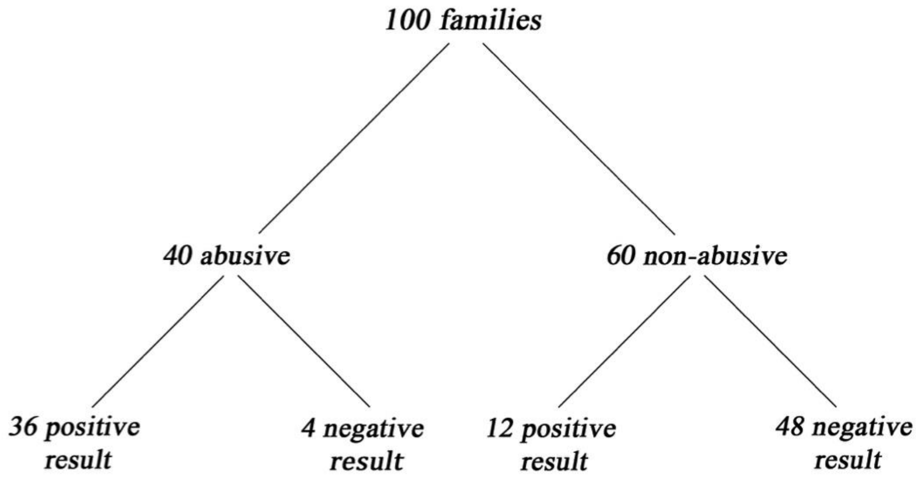


The calculation shows that, in total, 27 families will get a positive result, of which 9 will be true positives and 18 false positives. Thus the probability of a positive result being a true positive is: 9 divided by 27 = 0.33.

In short, about two thirds of the families judged dangerous by the instrument will not be.

In contrast, if the tool is used on a sub-group of the population where we have reason to assign a higher base rate of 40% then the equivalent figures are:

Forty out of every 100 families in this population are abusive (the base rate). Of these 36 families will get a positive result on using the instrument (the sensitivity of 90 per cent). Of the other 60 families, around 48 will accurately get a negative result but some 12 will get a (false) positive result (the specificity of 80 per cent).



In total, 48 families will get a positive result, of which 36 will be true positives and 12 false positives. Thus the probability of a positive result being a true positive is: 36 divided by 48 = 0.75. In short, a quarter of the families judged dangerous by the instrument will not be so, i.e. false positives.

The crucial message is that it is surprisingly hard to develop a high accuracy rate in predicting a relatively rare event. Even instruments with what seem to be impressively high statistics about how many families they will accurately identify as abusive or safe have a disappointingly low overall accuracy: the majority of the families the instrument identifies as abusive will, in fact, be non-abusive; that is, they will be false positives.

The danger is that the base rate fallacy will adversely influence people's use of the results of predictive analytics. Added to this risk is 'automation bias' - the tendency for people to have undue confidence in the results produced by computers so they are more likely to discount contradictory evidence than people who are making judgments. There is evidence that this is a significant source of error in aviation and medicine (Goddard, Roudsari, & Wyatt, 2011) and so it may be a problem in child protection. A lack of understanding of the importance of base rates is likely to lead to over-confident use and the prevalence of defensive practice in societies that react very punitively when child protection workers fail to protect a child from being killed may increase the automation bias.

Many of the existing tools are, as their name suggests, designed to *support* decisions by professionals. The designers of the tools generally stress that it should be treated as one among several factors that the professional considers. The Allegheney Family Screening Tool, for instance, reports the score for a referral along with text explaining that the system 'is not intended to make investigative or other child welfare decisions'. However, some may be reluctant to use their professional expertise to reach a different decision than the one recommended by the tool. In the event of an adverse outcome, there is a safety in blaming the tool for the decision and a fear that they would struggle to justify going against its recommendation. The more defensive the work culture, the more likely automation bias will occur.

Proponents of predictive analytics point to the increased accuracy of the decisions made with the support of the automated analysis. Critics worry about how it will fit into the whole process of working with families. A balanced assessment of what is the best action to take for a child requires a positive assessment of the rewards in a situation not just the risks. Also, most, if not all, child protection practice approaches involve the worker building a relationship with family members in order to understand their problems and help them provide safer care. How and whether professionals can integrate the predictive tools constructively into this relationship is a concern raised by some (Broadhurst, Hall, Wastell, White, & Pithouse, 2010; Oak, 2015).

Oak also raises the question of whether having the risk assessment performed by a tool will 'lead to the erosion of critical thinking and professional judgment skills, including the ability to define key concepts such as 'risk' or 'abuse' and to recognise that they are socially constructed and contested entities' (2015 p.1215). This seems to overstate the role that predictive analytics are intended to play. Predictive analytics are being developed for the major decision points such as whether to investigate an allegation of harm but workers make many decisions every day. The development of decision support systems does not eliminate the need for professionals to assess risk and make decisions on how best to manage it in their daily work. In some respects, these seem small matters. For instance, workers with heavy workloads (as most are) have to make decisions about how to use their time, which families or other activities need to be prioritized. These decisions will involve risk assessments in deciding which families to prioritise visiting. It is only with hindsight that some of these decisions may be seen to be pivotal in the management of the case – an unplanned home visit seeing evidence of harm or a visit being delayed and swiftly being followed by the child suffering injury.

Ethical factors

The final questions to ask about the use of such predictive analytics relate to whether they are morally acceptable. What benefits will they produce for children and their families? What harm might they do? How do you balance these out?

When used preventively to screen families to identify those children who are likely to develop problems, they raise the standard questions of any screening method. How accurate is the screening tool? Do we have effective services for resolving the predicted problem? Do we have sufficient resources to provide those services?

Judging the accuracy of the screening tool is not just a technical matter but also requires making a judgment about the risk threshold. As discussed earlier, the accuracy of a predictive tool is related on the base rate of the phenomenon you are seeking to predict plus the sensitivity and specificity of the tool. Decision support systems derived from predictive analytics are not 100% accurate and will never be so but this means that decisions need to be made about the risk threshold for action – the balance between the sensitivity and specificity of the predictions. As I discussed earlier in Chapter Four, these are inversely related: if we want to improve the sensitivity (have a low rate of false negatives, of missing children) then automatically we lower the specificity (we increase the number of false positives, inaccurately identifying children).

Do we have effective enough methods to deal with identified needs? A key principle of health screening is that it has benefits for those screened because effective methods area available to mitigate the potential harmful outcome that the screening identifies. In child protection, it is not enough to say that some intervention has been shown to be effective (usually in comparison with another intervention or no treatment in an RCT). We also need to know what percentage of people showed benefit, how great that benefit was, and whether for some there were negative consequences.

Do we have sufficient resources to provide the services? Typically, preventive services need to provide help to a large number of families (the false positives) in order to include those who might otherwise have developed serious problems.

In discussing the growing interest in screening for adverse childhood experiences (ACEs), Finkelhor puts the counter argument:

> '*We are going to argue here that it is still premature to start widespread screening for ACES until we have answers to several important questions: 1) what are the effective interventions and responses we need to have in place to offer for positive ACE screening, 2) what are the potential negative outcomes and costs to screening that need to be buffered in any effective screening regime, and 3) what exactly should we be screening for?'* (Finkelhor, 2018 p.175).

What do we know about the actual or potential negative effects of being profiled? Predictions may be carried out by well-motivated professionals who want to help families but that does not necessarily mean that they will have beneficial effects. Problems around parenting and child development are all too easily seen negatively by others. The mere fact of being known to children's services can be stigmatizing and be interpreted by some as a damaging mark against you whether you are an adult or a child. In an English trial of a national database on all children, including all services with which they were in contact, one school Head used the database to screen out all applicants who had a history of being known to Children's Social Care. This was, of course, an illegal use but it still had a harmful effect on the children involved. It would be naïve to assume that criminality would be rare when detailed databases are becoming of increasing practical and commercial value.

Finally, for all decisions in which predictive analytics may be used, there is a significant danger of preserving existing biases and prejudices in professional practice but making them more dangerous because they are hidden from sight in the performance of an apparently neutral scientific mechanism for reaching judgments.

To summarise, predictive analytics may be used for different decision tasks in child protection, the major ones being early identification of families likely to become problematic, decisions on whether to investigate a referral, and decisions on removing or returning a child to their home. Benefits and problems with predictive analytics need to be appraised in relation to the specific decision task they are aiming to support.

Their introduction raises many technical, legal, and ethical concerns.

> '*Machine learning systems are known to have various issues relating to bias, unfairness, and discrimination in outputs and decisions6, as well as to transparency, explainability, and accountability in terms of oversight7, and to data protection, privacy, and other human rights issues, among others'* (Cobbe, 2018 p.5)

A concluding point is that even the most accurate, legal and ethical tools only cover a small part of the task of improving children's safety and well-being. They omit the assessment of the positive aspects of families. Working with a family to provide safe enough care or providing good alternative care will continue to absorb most professional time.

Despite the many counterarguments and concerns about using predictive analytics, many jurisdictions are introducing decisions support systems derived from them to tackle urgent practical problems in targeting limited services. Perhaps they should consider the advice given by Zuiderveen Borgesius on proceeding with caution:

> '*The public sector could adopt a sunset clause when introducing AI systems that take decisions about people. Such a sunset clause could require that a system should be evaluated, say after three years, to assess whether it brought what was hoped for. If the results are disappointing, or if the disadvantages or the risks are too great, consideration should be given to abolishing the system' (Zuiderveen Borgesius, 2018 p.29)*.

When viewed from the narrow point of improving decisions relating to children's safety and well being, predictive analytics look appealing, harnessing the information buried in vast databases to guide professional decision making. However, when this task is placed in the wider context of the technical processes involved and the social situations in which the tools are used, a large number of problems emerge - the hidden bias in the algorithms, the incompleteness and unreliability of the datasets, the lack of transparency, and the impact upon families. Considerable work is going on in artificial intelligence and in improving the law and regulation relating to its use and these may make sufficient progress to reduce some of the difficulties. However, at present, the use of predictive analytics in child protection seems to introduce too many new problems that outweigh their potential benefits

References

Acorn. (2019). *Acorn User Guide*. Retrieved from https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf

Allegheny County. (2017). *Frequently Asked Questions about the Allegheny Family Screening Tool*. Retrieved from Allegheny County, Pennsylvania: https://www.alleghenycountyanalytics.us/index.php/2017/07/20/frequently-asked-questions-allegheny-family-screening-tool/ downloaded 30.01.19

Arad-Davidzon, B., & Benbenishty, R. (2008). The role of workers' attitudes and parent and child wishes in child protection workers' assessments and recommendation regarding removal and reunification. *Children and Youth Services Review, 30*(1), 107-121.

Britner, P. A., & Mossler, D. G. (2002). Professionals' decision-making about out-of-home placements following instances of child abuse. *Child Abuse & Neglect, 26*(4), 317-332.

Broadhurst, K., Hall, C., Wastell, D., White, S., & Pithouse, A. (2010). Risk, Instrumentalism and the Humane Project in Social Work: Identifying the Informal Logics of Risk Management in Children's Statutory Services. *British Journal of Social Work, Advance ac*.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183-186.

Cawson, P., Wattam, C., Brooker, S., & Kelly, G. (2000). *Child Maltreatment in the United Kingdom: A Study of the Prevalence of Child Abuse and Neglect*. London: NSPCC.

Children's Bureau. (2018). *Child Maltreatment 2017*. Retrieved from Washington, DC: https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/child-maltreatment

Church, C. E., & Fairchild, A. J. (2017). In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics. *Juvenile and Family Court Journal, 68*(1), 67-81.

Cobbe, J. (2018). Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making.

Cuccaro-Alamin, S., Foust, R., Vaithianathan, R., & Putnam-Hornstein, E. (2017). Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review*.

Department for Education. (2018). *Characteristics of children in need 2017 to 2018*. Retrieved from London:

Eubanks, V. (2017). *Automating Technology, How High-Tech Tools Profile, Police and Punish the Poor*. New York: St Martins Press.

Finkelhor, D. (2018). Screening for adverse childhood experiences (ACEs): Cautions and suggestions. *Child Abuse & Neglect, 85*, 174-179.

Gilbert, R., Widom, C. S., Browne, K., Fergusson, D., Webb, E., & Janson, S. (2009). Burden and consequences of child maltreatment in high-income countries. *The Lancet, 373*, 68-81.

Gillingham, P. (2015). Implementing Electronic Information Systems in Human Service Organisations: The Challenge of Categorisation. *Practice, 27*(3), 163-175.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2011). Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121-127.

Jergeby, U., & Soydan, H. (2002). Assessment Processes in Social Work Practice When Children Are at Risk: A Comparitive, Cross-National Vignette Study. *Journal of Social Work Research and Evaluation, 3*(2), 127-144.

Jud, A. (2018). Current research on child maltreatment epidemiology: BioMed Central.

Keddell, E. (2015). The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool? *Critical Social Policy, 35*(1), 69-88.

Khoury, M. J., & Ioannidis, J. P. (2014). Big data meets public health. *Science, 346*(6213), 1054-1055.

Korff, D., & Georges, M. (2015). Passenger Name Records, data mining & data protection: the need for strong safeguards: T-PD (2015).

Oak, E. (2015). A minority report for social work? The Predictive Risk Model (PRM) and the Tuituia Assessment Framework in addressing the needs of New Zealand's vulnerable children. *The British Journal of Social Work, 46*(5), 1208-1223.

Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. Colorado: Colorado Law School.

Radford, L. (2011). *Child abuse and neglect in the UK today*. Retrieved from London:

Regehr, C., Bogo, M., Shlonsky, A., & LeBlanc, V. (2010). Confidence and professional judgment in assessing children’s risk of abuse. *Research on Social Work Practice, 20*(6), 621-628.

Schuerman, J., Rossi, P. H., & Budde, S. (1999). Decisions on placement and family preservation: Agreement and targeting. *Evaluation Review, 23*(6), 599-618.

Schwartz, I. M., York, P., Nowakowski-Sims, E., & Ramos-Hernandez, A. (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward County experience. *Children and Youth Services Review, 81*, 309-320.

Society Royal. (2012). Science as an Open Enterprise.

Spratt, T. (2000). Decision making by senior social workers at point of first referral. *British Journal of Social Work, 30*(5), 597-618.

Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R., & van Ijzendoorn, M. H. (2015). The prevalence of child maltreatment across the globe: Review of a series of meta-analyses. *Child Abuse Review, 24*(1), 37-50.

Zuiderveen Borgesius, F. (2018). *Discrimination, Artificial Intelligence and Algorithmic Decision Making*. Retrieved from Strasbourg, France: