

SECTION 2

Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County

by Tim Dare and Eileen Gambrill

INTRODUCTION

This report comments on two linked papers produced by Rhema Vaithianathan, Nan Jiang, Tim Maloney and Emily Putnam-Hornstein as part of the development of a predictive risk modeling tool to improve child protection decisions being made by the Allegheny County Department of Human Services (DHS) (Vaithianathan, et al., 6 Feb, 2016, and Vaithianathan, et al., 23 March, 2016). The details of the predictive risk model are presented in those papers and we do not here attempt to repeat that presentation. We assume those reading this ethical assessment will be familiar with the papers.

Since our assessment depends on the accuracy of our understanding of the tool, however, we begin with a brief summary so that it will be clear what we are taking them to have proposed.

SUMMARY OF THE PROPOSED ALLEGHENY FAMILY SCREENING TOOL

In short, in 2014, DHS sought partners to work with them on using their integrated data systems to make better child protection decisions. The consortium of researchers led by Vaithianathan was awarded the contract and commenced work on building a predictive risk modelling tool. Following discussion and preliminary work, it was decided to develop a tool that would provide a risk assessment when a call about an allegation of maltreatment was received by the DHS call center, rather than at the birth of a child.

The Allegheny Family Screening Tool (AFST) will produce a risk score which will help call screeners decide whether a call warrants a visit and whether there is a justification for screening the child in and carrying out an investigation.

Once the call is established as a referral, call screening staff will be able to search KIDS, the child welfare electronic information system, to determine whether any of the people named in the referral are already in the system. If so, there will be an ID number for those people, which will allow immediate linking of data held about them from various sources including health or court records and previous welfare contacts. (Temporary IDs will be created where none is held or where there is insufficient information to identify a person. Permanent or corrected IDs will be added retrospectively once all the information is established). Once identity and basic relationships are established — typically a few hours after the call arrives — a risk score and data visualization will be generated.

Calls typically refer to multiple people and the risk score will relate to the *call as a whole*. The risk score will present the *maximum* risk score for all children in the referral. While calls will identify a child who is named as a victim and other children living in the house as “other children,” the AFST will score every child in the referral regardless of whether they were identified as the victim.

PARTICULAR ETHICAL ISSUES

a. Consent

Predictive risk modeling often generates significant difficulties around obtaining meaningful consent from those whose information is used and for whom risk profiles are generated. Typically, data will be aggregated in ways that make it difficult to trace clear relationships between data-providers and end-users, and data collected for one purpose will typically be used for another. Under those circumstances it is difficult — perhaps impossible — to design effective informed consent procedures. (These difficulties are exacerbated where individuals really have no choice about whether to provide the information at the outset. That will be the case *de jure* with criminal justice and birth data and may be the case *de facto* if individuals cannot, for instance, access essential services or support without providing the data.)

This is one of a number of points at which we think that it is ethically significant that the AFST will provide risk assessment in response to a call to the call center, rather than at the birth of every child. In the latter case there is no independent reason to think there are grounds to override default assumptions around consent. The fact there has been a call, however, provides at least some grounds to think that further inquiry is warranted in a particular case.

In addition, accessing data in response to a call will reduce the numbers of families or individuals whose data is being accessed by the tool and so reduce the overall incidence of access to family or individual information.

Finally, if DHS were *already* entitled to access the data gathered by the tool in response to a call, then it seems legitimate to regard the use of the tool at that point as a new and more effective way of doing something already permitted. The force of this point depends, we think, on the extent to which the AFST delivers information that would have been available, *in principle*, to a diligent call screener.

b. Information about other family members

As noted, calls typically involve multiple people: the victim, other children in the home, the mother, father and other adults. The risk score will be based on information held about all of these people. It may seem that there are significant issues around access to information about those individuals who are not the primary concern of the call. They might wonder about the justification for using information about them as part of an assessment to which they are, perhaps, only peripherally related.

We think that there should be protocols around the use of this information about individuals who are not the primary concern of the call.

Notwithstanding the need for such protocols, we believe the fact that it is at the point of call that risk assessment is carried out again has ethical significance. As above, the fact information about ‘other’ individuals is accessed in response to a call raising concerns about the welfare of a child provides grounds for access; accessing information only where there has been a call will reduce the numbers of families or individuals whose data is being accessed by the tool; and, while access to such information may have been more haphazard prior to the introduction of the AFST, we assume that the model does not create new rights of access to that information — that a diligent child welfare call screener would already have been entitled to gather the information now to be accessed by the tool.

c. False Positives/False Negatives

All predictive risk models will make some errors at any threshold for referral, and so, in the child protection context, identify as low risk some children who go on to experience abuse or neglect and identify as high risk some children who do not.

When considering the significance of these ineliminable errors for the AFST it is essential to keep in mind that decisions informed by predictive risk modeling tools will in almost every case have been made by some other means prior to the use of the tool and will continue to be made if such tools are not adopted. Consequently, ethical questions about predictive risk modeling tools are essentially and unavoidably *comparative*: they are questions not simply about the costs and benefits of a particular predictive risk modeling tool, but also about how those costs and benefits compare from an ethical perspective with the costs and benefits of plausible alternatives. They must be considered in light of alternatives that carry costs of their own.

And, while it is true that all predictive risk modeling tools will make errors at any threshold, at is also true that they are both more accurate than any alternative — they make fewer errors than manually driven actuarial risk assessment tools and even very good child protection professionals relying on professional judgement and experience — and they are more *transparent* than alternatives, allowing those assessing a tool's performance to accurately identify likely error rates and to accommodate them in responses to the predictions of a particular modeling tool. The greater accuracy and transparency of predictive risk modeling tools also allows them to serve as (inevitably imperfect) checks against well-understood flaws in alternative approaches to risk assessment.

So, while one should of course reduce the false-positive/negative rate as far as possible (by, for example, choosing higher thresholds for intervention, though that will carry its own costs), one can also reduce the ethical significance of false-positives and negatives by, *for instance*:

1. Providing opportunity for experienced child welfare professionals to exercise judgment about appropriate responses to a family's identification as at-risk. (*We note that one possible response to high risk scores under the AFST are mandated home visits, which would provide just this sort of opportunity*)
2. Ensuring that professionals who are using information provided by predictive modeling tools understand the potential of those tools to mis-categorize families
3. Providing training to guard, in so far as possible, against confirmation bias in the professional engagement with families identified as low- or high-risk
4. Ensuring that intervention triggered by identification as at-risk is positive and supportive rather than punitive
5. Ensuring that intervention triggered by identification as at-risk is as non-intrusive as possible consistent with the overall aims of reducing child maltreatment risk
6. Identifying and minimizing the adverse effects of identification as at-risk, such as, for instance, possible stigmatization

d. Stigmatization

There are obvious burdens associated with identification as an at-risk child or family. Those burdens may range from those that are fairly straightforward and transparent, and to some extent at least under the control of social services, to the more complex and diverse burdens of social stigmatization. We should not underestimate the significance of stigmatization:

- The associated burdens may be borne in *anticipation* of conduct that might never come to pass.
- In many cases, the burdens that follow from being identified as a member of a group arise from false beliefs about what that identification means. The burdens associated with identification as an at-risk individual or group may actually increase risk of the adverse outcome.

- The burdens of stigmatization often fall upon those who are already the subject of social disapproval or demarcation, ‘appropriating and reinforcing pre-existing stigma’

These are matters for significant ethical concern. Again, however, it must be remembered that that they are not distinctive of predictive risk models. It would be naive to suppose, for instance, that negative conclusions were not already drawn from correlations between child maltreatment and socio-economic position, that existing approaches to child protection did not carry risks of confirmation bias, of unwarranted intrusion on families who were not at risk, of appropriating and reinforcing existing stigma. The point is not to suggest that these costs can be disregarded, but to emphasize the importance of weighing the costs and benefits of implementing the AFST against the costs and benefits of alternatives. Plausibly, for instance, the AFST may reduce some of these potential burdens, allowing child protection professionals to avoid confirmation bias more effectively, and allowing more effective targeting of services that, while not eliminating unwarranted intrusion, may reduce it.

In addition, we believe that there are responses to stigmatization that can at least reduce its impact and which tip the balance in favor of predictive risk modeling. Those responses include:

- i. Maintaining careful control over the dissemination of the ‘product’ of the AFST. Access to risk scores and visualization should be distributed only to those who a) have appropriate training and b) need the information in order to further child protection goals.
- ii. Provide appropriate training targeted at reducing stigmatization and its negative effects. Such training might be expected to:
 - a. Emphasize the possibility of false positives/negatives.
 - b. Emphasize that even given high confidence in risk scores, they are *only* risk scores and predictions. Individuals identified as at high risk must not be treated as though they have already been victims or perpetrators.
 - c. Include training against confirmation bias, one of the most obvious dangers of stigmatization.

In addition, many of the responses to false positives/negatives set out above will also be directly relevant to concerns about stigmatization.

e. Racial Disparity

Many of the issues around false positives/ negatives and stigmatization are manifest in problems associated with racial disparities in the data upon which the AFST would rely. The researchers have established that current decisions around referring and placing children who are the subject of calls are affected by race. Overall, black children are almost three times more likely to have some interaction with the child welfare system than white children. Having been referred, black children are also more likely than white children to be screened in and placed. If they are screened out, black children are more likely than white children to be re-referred and placed.

Note that these disparities are to be found in the existing data. They exist independently of predictive risk modeling. The difficulty for the AFST is that such disparities in the data are potentially reinforcing. If the AFST relies upon existing data it will see evidence that black children are at higher risk than white children. If the disparities in the data reflect genuine underlying differences in the need for protection – perhaps because ethnicity tracks socio-economic disadvantage – they may not be of cause for concern: they might reflect underlying need rather than bias. If the disparities do reflect race-based bias, however, they may be ethically problematic.¹

¹ The researchers seem to show that poverty is not sufficient to explain the different referral and placement rates.

A well-known and ethically problematic example of racial disparity and its effects on predictive risk modeling occurs in the criminal justice context. In the U.S., young black men are more likely to be stopped and searched by police than their white counterparts, and having been stopped and searched are more likely to be arrested both because the stop and search provides opportunity to find evidence of offending such as drug possession, and because police are more likely to arrest young black men for offences for which their white counterparts are more likely to receive a warning. It is clear that these contacts and arrests arise to a significant extent because of racial bias. The contacts and arrests appear in the data used by predictive risk modeling tools to predict offending. Since those tools find greater evidence of contact and arrest for young black men, they are likely to place young black men in a higher risk category than their white counterparts, and since the contact and arrests reflect bias and not underlying criminality, that risk classification is unwarranted. The use of predictive risk modeling in such contexts requires at least great care lest it reinforce stigmatization, bias and disadvantage.

Examples such as the stop and search case might lead one to think that predictive risk modeling is inappropriate in contexts where one cannot be sure that data is not affected by racial bias, or at least that one should ensure that race is not taken into account by tools used in those contexts. However, there are important differences between the stop and search case and the modeling proposed in the AFST. A predictive policing tool may well recommend stopping and searching young black men *because* they have been stopped and searched in the past. That intervention is not designed to prevent future stops and searches. We think it matters in the AFST case that while a history of engagement with child protection services may lead the AFST to overstate the actual risk status of a child or family, the intervention which flows from that classification is designed and intended precisely a) to identify that family or individual's actual risk status through home visits and professional judgement, and b) to address in so far as possible any risk factors which are found to exist. It matters, ethically, this is to say, that a high risk score will trigger further investigation and positive intervention rather than merely more intervention and greater vulnerability to punitive response. We believe, that is, that the fact that the AFST will prompt further detailed inquiry into a family's situation and that any intervention is designed to assist gives grounds to think the model is not vulnerable to the legitimate concerns generated by the existence of disparities in data used in punitive contexts.

We note that the research — although not intended to show the effectiveness of field screening — suggests that such screening *reduces* the effects of disparities in the child protection data. Under the current system as we understand it, all children under seven who are the subject to a call must be field screened. Field screens appear to correct for the bias that sees a disproportionate number of black children referred and placed. The researchers write that:

We find that when call screeners were forced to field screen, they were more inclined to screen out black children, whereas when they did not have to conduct field screens (age seven and older), they were more inclined to screen in Black children compared to White children. This suggests that the requirement for more information (i.e. via a field screen) reduced the disparities in screening (Vaithianathan et al, 23 March, 2016, 8)

Note, as an aside, that this appears to be an example of the additional transparency of predictive risk models over alternatives, suggesting that it is possible to track and correct for disparities that may have remained hidden under alternative approaches. More generally, it is important not to understate the burden that engagement with child protection services may place on families, but it is also important not to respond to the disparity issue in ways that worsen or leave unaddressed the position of children who might be helped.

f. Professional Competence/Training

As we have mentioned at a number of points, it is essential — if predictive risk modeling tools are to operate ethically — that staff using and relying upon them are competent with their use and interpretation. The use of such tools must be accompanied by appropriate training to ensure that competence. We set out some specific elements of such training under the stigmatization discussion above where we mentioned training to recognize the possibility of false positives/negatives; to see that even given high confidence in risk scores, they are *only* risk scores and predictions; and to recognize and guard so far as possible against common reasoning flaws and biases.

g. Provision and identification of effective interventions

Predictive risk modeling is a form of screening. So regarded, it is natural to suppose that it is subject to ethical constraints taken to apply to screening programs. One of the current reviewers has discussed the relevance of the standard statement of these constraints, the WHO Screening Principles, for predictive risk modeling in the child maltreatment context. We will not repeat that analysis here, but simply indicate that accurate predictive risk models appear to perform well under the principles (see Dare, 2013, pp. 36-47).

We think, however, that it is worth specifically mentioning one of the WHO principles. Principle 2 specifies that in order for a screening program to be ethical it must be the case that “[t]here should be a treatment for the condition” for which screening is being carried out. Dare argues that that principle is best seen as resting on the idea that screening programs which might

themselves generate harms must be capable of delivering countervailing benefits (Dare, 2013, pp. 43-44) and argues that there is sufficient evidence that interventions prompted by predictive risk models in the context of child protection meet this demand.

Here we wish to make that point in more general terms. One ethical concern about the AFST springs from the question “why pursue better prediction, if services offered will not be evidence-informed; those most likely to result in hoped for outcomes.” We view this as an ethical problem. And there is another one. Why predict better if staff are not well trained in the conduct of empirically informed assessments? How well trained are they in common factors related to positive outcomes such as empathy and warmth? Yet another is how well trained staff are in gathering valid outcome measures. This raises questions concerning what will happen after risk scores are acted on. What good does it do for example to diagnose more asthma if nothing is done about it that is effective?

Drawing attention to these concerns may be a potential bonus (and an ethical one) of the use of more accurate risk prediction. Professional decision-making is not a one-shot affair. There is a sequence of decisions, each potentially affected by earlier ones, each of which may or may not be acted on as an opportunity to direct decisions in a more positive direction. It is our hope that the use of a more accurate risk estimation will highlight these other issues that affect quality of care for clients.

h. Ongoing monitoring.

The last point leads naturally to another: *Since* professional decision-making in the child protection area is not a one-shot affair, it is essential, we believe, that the County commit to ongoing monitoring of the AFST to ensure that the tool and staff training in its use is maintained, and that the interventions remain as effective as possible. The tool does generate legitimate ethical concerns and those issues must be monitored, and the justification for the burdens the tool imposes requires DHS to identify and implement reasonably effective counter-balancing responses.

i. Resource allocation.

There is an assumption implicit in the discussion in the last few sections that can usefully be made explicit. Whether the AFST is ethical depends to a large extent on its capacity to deliver benefits sufficient to outweigh its costs. We believe that it has the capacity to meet that standard. However, its doing so will require, in addition to training and monitoring and effective intervention, the provision of adequate resourcing. The AFST must not, on ethical grounds, be seen as an opportunity to reduce child protection resourcing or to reallocate child protection professionals in ways that prevent the tool from delivering the benefits upon which its ethical justification relies.

IN SUM

In our assessment, subject to the recommendations in this report, the implementation of the AFST is ethically appropriate. Indeed, we believe that there are significant ethical issues in not using the most accurate risk prediction measure.

Instruments that are more accurate will result in fewer false positives and false negatives, thus reducing stigmatization (false positives) and more lost opportunities to protect children. It is hard to conceive of an ethical argument against use of the most accurate predictive instrument.

As we have emphasized throughout, decisions are being made right now. It is not a matter of making or not making related decisions. The decisions involved are complex ones made in a context of inevitable uncertainty that contributes to inevitable error. Research on decision-making in the helping professions highlights the play of biases and fallacies. Confirmation biases are common in which we seek information that corresponds to our preferred view (e.g., there is no abuse) and fail to seek evidence that contradicts preferred views. Errors of omission (failing to act) are viewed as less harmful than errors of commission (acting - for example, removing a child from the care of her family). The question is, how can we make the fewest errors in our efforts to protect children and families? AFST seems an ethical and potentially important contribution to that effort.

REFERENCES

Dare, T. (2013) *The Dare Report: Predictive Risk Modelling and Child Maltreatment: An Ethical Review*, Ministry of Social Development, Wellington, New Zealand. <http://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/predictive-modelling/00-predictive-risk-modelling-and-child-maltreatment-an-ethical-review.pdf>

Dare, T (2015) 'Predictive Risk Modeling and Child Protection: An Ethical Analysis' in *Challenging Child Protection: Directions in Safeguarding Children* eds. Janice McGhee and Lorraine Waterhouse (Edinburgh; Jessica Kingsley Press, 2015) pp. 64-76

Gambrill, Eileen, and Aron Shlonsky. 'Risk Assessment in Context', *Children and Youth Services Review* 22, no. 11 (2000): 813-37

Gambrill, E. (2012) *Critical thinking in clinical practice: improving the quality of judgements and decisions*. Hoboken, N.J.: John Wiley & Sons

Vaithianathan, Rhema, Nan Jiang, Tim Maloney, Emily Putnam-Hornstein, (6 February 16) 'Implementation of Predictive Risk Model at the Call Centre at Allegheny County'

Vaithianathan, Rhema, Nan Jiang, Tim Maloney, Emily Putnam-Hornstein, 23 March, 2016, 'Developing Predictive Risk Models At Call Screening For Allegheny County: Implications for Racial Disparities'.