

Administrative Data and Predictive Risk Modeling in  
Public Child Welfare:  
Ethical Issues Relating to California

Brett Drake, Ph.D. and Melissa Jonson-Reid, Ph.D.  
Brown School of Social Work, Washington University in St. Louis

2018

## Table of Contents

Introduction and Scope of Work.....	3
Structure of the Report.....	3
Executive Summary.....	4
Key Review Conclusions.....	4
Recommendations.....	6
The Child Welfare System: Assumptions and Evidence.....	8
CWS: Definitions and Purpose.....	8
Is CWS more akin to a Public Health or a Criminal Justice System?.....	8
False Positives and False Negatives in the CWS Context.....	10
Bias in CWS.....	13
Risk Assessment: What Works and What Doesn't.....	13
Administrative Child Welfare Data and PRM.....	14
Sources and Types of Administrative Data.....	14
Testing PRM Systems before Deployment: A Methodological Note.....	18
Ethical Issues Associated With the Use of PRM in California.....	19
Over- and Under-Identification: Using the Right Tools and Hitting the Sweet Spot.....	19
Stigmatization.....	23
Data Access and Data Use: Ethical Considerations.....	25
Could PRM Cause Unfair Treatment of Particular Subpopulations?.....	29
Is it ethical to identify situations for which no effective remedy is available or feasible?.....	30
Is it permissible to employ variables in PRM without the consent of the involved persons?.....	30
Is it ethically permissible to use data from "peripheral" persons?.....	31
What ethical limits exist relative to the use of PRM for other likely "use cases?".....	31
Is the relative "lack of understandability" of PRM ethically problematic?.....	31
Is it ethically permissible to include "opt-in" data in addition to universal data in PRM?.....	31
Is it ethical to use data for a reason other than that for which they were originally collected?.....	32
Are there ethical concerns regarding the person or system the risk score is attached to?.....	32
Can "Bright Line" applications of data be ethical?.....	33
Conclusion.....	35
References.....	36

# **Administrative Data and Predictive Risk Modeling in Public Child Welfare: Ethical Issues Relating to California**

## **Introduction and Scope of Work**

The central focus of this report is to describe ethical issues pertaining to the use of administrative data and predictive risk modeling (PRM). California has not made a decision about whether to utilize PRM; however, this paper explores potential use cases in California. California is considering the following use cases: (1) to support Red Team reviews, (2) identification of candidates for providing prevention services under FFPSA, (3) explorative research on allocating limited intensive resources post Family Reunification, and (4) to assist with administrative CQI on referral screening practices.

Because the proposal under consideration in California uses data already “in house” (i.e., data already held by the California Department of Social Services through its Child Welfare Services Case Management System (CWS/CMS) and fully accessible to caseworkers), the use of child welfare data for prediction purposes are a particular focus of this review.

Our primary approach in this review is to use empirical evidence to evaluate the degree to which ethical and social justice objectives are met by the use of PRM. Specific ethical concerns (e.g., unnecessary intrusion due to false positive screening decisions) can be evaluated based on available evidence (e.g., comparison of false positive rates between current practice and PRM modeling). Said another way, the report focuses on ethical conclusions made possible by a review of all relevant facts, rather than philosophical conjecture.

### **Structure of the Report**

This report begins with an executive summary encapsulating the paper’s main findings and recommendations. Three subsequent sections provide greater depth and specificity: (1) **“The Child Welfare System: Assumptions and Evidence”** describes the functioning of the child welfare system and how ethical issues can be considered in this context; (2) **“Administrative Child Welfare Data in the Context of PRM”** explains the kind of data available and discusses prior uses of administrative data, including PRM; and (3) **“Ethical Issues Associated with the Use of PRM in California”** provides a detailed analysis of ethical issues associated with the proposed use of PRM in California. We do not purport to provide a legal analysis. Note: Throughout the report we use CWS to refer to the “child welfare system” as a whole.

## **Executive Summary**

California is considering utilizing PRM to generate risk scores which can be used as a tool for the following: (1) to support Red Team reviews, (2) identification of candidates for providing prevention services under FFPSA, (3) explorative research on allocating limited intensive resources post Family Reunification, and (4) to assist with administrative CQI on referral screening practices. While some work in PRM has used multiple sources of cross-program data (e.g., criminal justice, public benefit), California plans to use only information already internally available through their administrative child welfare system (CWS) records. This report draws upon existing empirical evidence and prior ethical frameworks and analyses regarding the use of PRM to provide an analysis and set of recommendations specific to California's proposed use of PRM.

### **Summary of Findings**

California's proposed use of PRM generates no unique or new ethical concerns. This does not mean, however, that the implementation of a PRM program in California is free of ethical challenges. As with any technology, tool, or procedure, PRM carries the potential for misuse. Were PRM to be misused, significant ethical concerns could arise. We describe these potential problem areas and suggest steps for avoiding them. PRM, however, also may confer ethical advantages. To the degree that PRM provides a more accurate risk assessment than is currently available, it carries the potential for improved and more ethical CWS practice. To the degree that PRM improves standardization, transparency, and equity of practice, that is also ethically desirable. Transparency can be supported through making algorithms open to academic and public / community review.

The above conclusions are largely based on the observation that PRM, as envisioned in California, is very limited, both in terms of the data to be used and its proposed use of the risk scores. The proposed use is also limited compared to a range of applications beyond the CWS suggested in other cases (Siegel, 2013). Should California decide to expand either the types of data or the types of outcomes modelled, the ethics of PRM use in California should be re-examined.

## **Key Review Conclusions**

- Universal-level "case finding" PRM-based interventions, such as those contemplated by New Zealand and (more recently) Allegheny County, and others already operating in US programs, such as "Birth Match," include far broader and greater ethical concerns than what is being considered for California. Yet, even those broader uses of PRM outside of California, have been evaluated as ethically sound (Dare, 2013; Dare & Gambrill, 2016) provided that necessary safeguards (such as appropriate training and evaluation) are employed.
- The ethics of PRM must be considered in the context of current CWS policy, practice, and outcomes. This requires an understanding of the goals, functioning, and paradigmatic assumptions underlying the CWS.
- PRM is a form of risk assessment and stratification. The literature is clear in finding that well-designed risk assessment tools are a useful supplement to, but not a replacement for, worker judgment. The literature clearly shows that actuarial risk assessment methods, when used properly and in conjunction with worker judgment, are superior to "worker alone" or "consensus-based" risk assessment. For clarity, "actuarial" risk assessments are those based on statistical relationships between observable characteristics and maltreatment, rather than opinion. The emerging literature suggests that PRM is markedly superior to prior actuarial tools.

- Given that PRM can be tested in advance of implementation in the field, such preliminary work, including proof of concept studies, should occur to demonstrate: (1) superior predictive capacity over current practice; and (2) fairness with regard to subpopulations. Should PRM be found lacking in either regard, its application is ethically suspect. In simple terms, if PRM doesn't work better than current practice, or if it worsens systematic biases against particular subpopulations which cannot be resolved, it should not be used.

#### Summary of Key Ethical Issues

- Any decision-making system related to prediction of behavior in the future will result in false positive and false negative determinations. More efficient systems make fewer false determinations (either positive or negative) and are ethically preferable. Changing thresholds can shift the balance of false positives and false negatives, one against the other, but cannot reduce both at the same time. Determining an optimal ethical balance of false positives to false negatives is an important CWS function, and may vary by case factors, such as having lower thresholds for very young children because of their increased vulnerability to harm.
- Relative ethical tolerance for false negatives and false positives is dependent upon many factors, including the consequences of either false determination. This varies according to the point in the process a decision is made (e.g., screening-in a report vs. terminating parental rights). In general, there is more tolerance for false positives and less tolerance for false negatives at the screen-in and case opening levels than at the case disposition levels, and rightfully so. The consequences for the caregiver and family of false positives are much greater later in the process when decisions about court involvement, placement into foster care, or possible termination of parental rights are made. This report focuses on the early stage of the process (i.e., at the screening of calls and supportive service decision-making stage) as this is the currently proposed application of PRM in California.
- PRM can provide an ethically desirable means of addressing issues of possible racial and class bias proactively, as rates of predicted outcomes (e.g., re-referral) can be checked against actual output from PRM models and compared among different subpopulations. Comparisons of PRM based on modifiable factors as compared to past events and demographic characteristics can also be made to see if such concerns persist. This can, and should, be done prior to implementation of PRM.
- With regard to other potential uses, such as using PRM risk scores to assign particularly complex cases to more experienced workers, we find no new ethical issues or problems. Such decisions are already being made on the basis of potentially much less accurate information. To the degree that PRMs are better at predicting risk than current practice, the ability to more accurately match a case to a worker's level of experience is enhanced. Rather than an ethical concern, we see this as an ethical improvement to practice.
- Concerns about stigmatization secondary to implementation of PRM (e.g., due to confirmation bias or unwarranted trust in PRM outputs) do not differ markedly from existing stigmatization concerns under the current process of assessing risk using the Structured Decision Making tools, which also score risk levels and guide prioritization. Protocols for training of staff and procedures for the use of decision aids are already in place in California and can be easily extended to the use of PRM.
- Many other concerns relative to PRM (e.g., use of data without informed consent, use of data regarding peripheral actors named in referrals, degree of transparency in PRM calculations) are not unique to PRM, but apply to current screening and assessment practices. To the degree that these concerns remain unchanged, PRM is ethically neutral. To the degree that those concerns are minimized or made more transparent, or system functioning is improved

(e.g., through enhanced predictive data), PRM is ethically preferable to current practice. Ideally, any PRM algorithm (not the confidential data) would be open for public review and discussion.

#### Key Conclusions Relating to the CWS Use of PRM in California

- In the currently proposed use case of PRM in California, three factors fundamentally condition all ethical considerations:
  - PRM, as envisioned, will only utilize CWS data
  - PRM will be only applied to referrals that have already been made to the CWS hotline and where CWS must make a decision (i.e., the tool will not be used for “case finding”)
  - PRM may be used in RED team reviews, identification of candidates for providing prevention services under Families First Prevention Services Act (FFPSA), explorative research on allocating limited intensive resources post family reunification, or to assist with administrative CQI on referral screening practices
- These factors set the proposed use of PRM apart from other more universal case finding approaches.
- California’s choice not to use a private vendor to develop a state algorithm is ethically sound. Unless such vendors are willing to make their algorithms (“proprietary intellectual property”) completely public, transparency – and potentially ongoing evaluation and adjustment – is compromised.
- California’s capacity to test and continually evaluate the model can assure that PRM improves over current practice and does not disadvantage any specific subpopulation.

#### Recommendations

- **Prior to implementation, any PRM system should be tested and validated.** The validation should include prior cases and should, at a minimum, demonstrate:
  - *Better predictive capacity than current practice.* PRM scores should be better predictors of subsequent outcomes (e.g., re-referrals) than current classification systems in use (i.e., as reflected by current screen-in decisions).
  - *Equity in accuracy across subgroups.* Validation efforts must determine how well PRM works for different groups of children and families. We would suggest that there be a statistical evaluation of the performance of PRM for population subgroups, particularly at risk or structurally disadvantaged groups, including racial and ethnic groups, immigrant groups, and those who are low income. At a minimum, two questions should be explored: (1) what is the predictive validity of PRM across groups? and (2) what is the balance of false positives to false negatives within and between groups? PRM approaches can be evaluated that focus on modifiable factors and behaviors as compared to those including unchangeable caregiver demographic factors to further assess how the models can balance accuracy and utility to case planning processes. It should be noted that we endorse an “outcome-based” evaluation here rather than a “process-based” evaluation. In our view, the specific variables and kinds of data used by the PRM (so long as they are ethically obtained) are of less importance than the degree to which the model, as a whole, functions in an egalitarian and socially-just manner.

- **If PRM improves assessment over current practice and passes tests of equity across groups, workers using PRM should be thoroughly trained in its proper use prior to implementation.** At a minimum this would include the following:
  - *Workers must understand what PRM is and its intended use.*
  - *Workers must regard PRM outputs as information that will assist professional decision making, not as information that requires a decision.* Workers must understand clearly that their professional judgment, including their assessment of the child, family, and associated systems, is informed by PRM risk scores, not replaced by them. Procedures should exist that allow workers (perhaps with supervisory approval) to make case decisions (such as screening in a referral) that are not consistent with the PRM scores. For example, because PRM is necessarily limited by information already possessed, it is possible that an assessment or investigation may reveal a key factor in screening or deciding to open a case to services that was simply not available to the computer algorithm. This requires that a process be in place to allow for flexible use of the risk score given the present circumstances.
  - *Workers must understand that PRM is fallible.* Workers must understand that PRM is not a magic crystal ball, and must be trained to consciously avoid “confirmation bias,” the tendency to find information and make judgments consistent with preconceptions.
- **Following implementation, the PRM system should be continually monitored.** In the same way that it is ethically necessary to model the accuracy and potential bias of a PRM system in advance of deployment, the system should be monitored on an ongoing basis. For example, cut-off points for a screen-in decision should be re-evaluated based on evidence, such as the degree to which referrals screened out at a specific risk level are re-reported, or referrals screened in turn out to have been unnecessarily included. Similarly, PRM accuracy should be continually reevaluated by racial or other group characteristic (e.g., age, gender) as shifts in demographics or resources available could result in needed changes to the algorithm over time. As with any scientific process, this ongoing evaluation should be public and transparent, serving as an ongoing accountability feature.
- **The use of predictive analytics should not be under control of a proprietary organization, if possible.** Unless the proprietary organization is willing to: (1) be *completely* transparent with regard to the algorithms used; (2) show clearly how non-relevant concerns (e.g., profit motives) are excluded from system decisions; and (3) allow for periodic reviews and any needed updates to the algorithm over time without undue cost burden to the state, the use of a public or nonprofit entity is ethically preferable to the use of a private proprietary agency.

In addition to these recommendations, we would highlight the continued use of good clinical practices as steps necessary to safeguard the implementation of PRM (Dare, 2013; Dare & Gambrill, 2016). These include working with families in a cooperative and in most cases, voluntary, manner and limiting any intrusive practices that might promote stigma, insofar as is consistent with child safety.

This last point requires some elaboration. The natural tendency is to think of any risk assessment system as a means of identifying very high risk referrals. This is only half the story, however. Risk assessment systems also identify low risk referrals. PRM systems have the capacity to identify referrals with *very* low

statistical future risk (Allegheny County DHS, 2016). Current risk assessment tools do not have this feature. Many families will obtain the lowest possible score on traditional risk assessment tools, and many of those families do, in fact, have negative future events. PRMs produce a risk score which is more sensitive at both ends of the spectrum. Those families with very low PRM risk scores have a known (and extremely low) statistical risk of future negative outcomes. This new information, when used in screening decisions, can both avert unnecessary agency effort and limit intrusive practices.

## **The Child Welfare System: Assumptions and Evidence**

### **CWS: Definitions and Purpose**

Ethics cannot exist absent context. This section will address a number of key contextual issues and explore underlying, often competing, assumptions. For purposes of this report, we use the “Child Welfare System” (CWS) to refer to *public child protective services agencies existing at the state or county level*. Broadly, such a term includes the continuum of child protection functions from initial screening through possible placement into foster care. We use a narrow definition, referring to the functioning of the state/county agency in particular, as well as agencies specifically contracted to explicitly fulfill state/county agency functions (e.g., a private agency contracted to provide foster care oversight under state guidance). We do not include the vast array of external agencies which interact with these entities (e.g., law enforcement, social service providers, or hospitals).

Readers desiring a detailed description of the purpose, process, and critical stages of the CWS might wish to consult *Appendix A: The Child Welfare System: An Overview*. For our purposes, however, the following points bear consideration:

- CWS agencies exist at the county or state level but are broadly constrained by federal legislation, which sets a series of particular requirements (e.g., timeframes) and establishes key system goals of safety, permanency, and reunification. Among these, child safety is the preeminent goal.
- CWS agencies have no punitive role. They exist only to protect children. For example, CWS interest in domestic violence is secondary to the impact it might have on a child. At the extreme, if a family has only one child who dies due to maltreatment, there is no ongoing CPS case, as there is no child to protect. While protecting children may result in CWS recommendations to a court that may take action to constrain parental rights, this is viewed as a last resort and the decision is not made by CWS agencies.
- The term “Client” can be confusing with regard to CWS. CWS works *on behalf of* children, but interventions are often *focused on* adults. This is because changing adult behavior is often the means through which child safety can be enhanced.
- CWS involvement generally begins with a hotline call. About seven million children per year are included in about four million referrals. Of these, about 60% are screened-in (investigated or assessed) and about 700,000 are substantiated. Only about 200,000 enter foster care. Non-foster care services are provided to about one million other children, many of whom had unsubstantiated maltreatment allegations.

### **Is CWS more akin to a Public Health or a Criminal Justice System?**

There is perhaps no issue more central to this analysis than our understanding of the nature of the CWS. The ethical concerns which arise and the manner in which they are analyzed are entirely dependent on the context in which they exist. We argue that any ethical evaluation of CWS must be based in a clear recognition of the primary fact the CWS is a service agency tasked with enhancing public safety, not an agency designed to assess guilt and mete out punishment.

This is an area of fundamental and widespread confusion, especially as it relates to the use of data and the consideration of false positives and false negatives (see below). Public health systems are designed to improve the health and well-being of the public. Criminal Justice systems also have a key role in directly safeguarding the public (for example, in responding to a 911 call), but also have a clearly punitive role, embodied in the criminal courts through adjudication and sentencing. This has a number of important implications.

- *Punitive vs. protective functions*: Public health systems are not concerned with justice or punishment – their job is to enhance public safety and well-being. Criminal justice systems are largely concerned with justice and punishment.
- *Past/Future orientation*: In public health processes, the past provides context for changing the present and future while criminal courts often adjudicate what has happened in the past as a basis for future punishment. For example, if unsanitary food handling at a distributor causes illnesses, a public health response will stop after threat is contained, but a criminal response may include fines and punishment long after threat has ended.

In general, the CWS is much more like a public health system, albeit with caveats about prevention as compared to an intervention emphasis. CWS responds to threats to child well-being and attempts to mitigate those threats. If a threat no longer exists (for example, if the threat to the child is no longer present due to some radical change in circumstance) then CWS involvement ends. CWS has no authority at all to determine criminal guilt or to sentence.

The idea that CWS has a punitive role is, however, widespread. This is understandable as many of the means through which the CWS keeps children safe, such as removing children into protective custody, represent tremendous intrusions (warranted or not) into families and will commonly be seen in a punitive light by the parents involved. The substantiation decision that follows an initial investigation is another aspect of CWS functioning which seems “criminal justice-like.” Although the substantiation decision may appear to be like a “guilty” verdict in a criminal court, there are a number of critical differences, most specifically that it is not a precursor to sentencing or any other form of punishment. States vary in regard to the legal standard used and in some cases this disposition is a necessary precursor to offering a family services through CWS.

Another way in which CWS seems to be like a criminal justice agency relates to the fact that, in some cases, a family court is involved. As illustrated earlier, progression to court involvement is relatively rare. The goal of a family court proceeding is to consider evidence of past parental behavior as it may relate to ongoing risk to the child. The judge (no jury exists in family court) then arrives at a course of action deemed to meet the best interests of the child, which usually includes remaining with their family or returning home after intervention. Only about 14% of children in foster care have their parental rights terminated in a given year.<sup>1</sup>

These differences between how the criminal justice and CWS responds to events are important to consider with regard to the use of information. *While CWS does use past information to inform present decisions, this is related to trying to assess the future risk to a child – not to determine guilt in order to assign punishment.*

---

<sup>1</sup> <https://www.acf.hhs.gov/sites/default/files/cb/afcarsreport24.pdf>

## False Positives and False Negatives in the CWS Context

In a perfect world, we could achieve absolute predictive certainty of future events we want to avoid, and then make the correct decisions to avoid them. All determinations would therefore be either “true positives” (saying something is true when it is actually true) or “true negatives” (saying something is false when it is actually false). The world, however, is imperfect, and both false positives (mistakenly saying something is true, when it in fact is false) and false negatives (mistakenly saying something is false when it is actually true) occur. In CWS practice, a true positive is screening in a referral (or opening or taking other appropriate action) where risk is actually present, while a false positive involves screening in a referral where no risk is present and no action is required. Similarly, a true negative would be screening out a referral when no action is needed. The most potentially tragic error involves false negatives, or failing to identify and respond to a serious risk. These “failures” to respond or to respond at a sufficient level to prevent harm can result in child fatalities, which often trigger significant public and political outcry (CECANF, 2016).

It is theoretically possible to completely avoid false positives or (not “and”) false negatives, even in an imperfect world. You can avoid false positives by acting like everything is false or avoid false negatives by acting as if everything is true. This of course, isn’t really decision making at all, and would be a preposterous basis for policy. The point, however, is that your chances of error can be adjusted depending on whether you prefer to avoid false positives or false negatives. Establishing a higher threshold for responding will decrease false negatives but increase false positives. A lower response threshold will increase false negatives and decrease false positives. Systems, therefore, can be “slanted” or “tuned” towards either avoiding false positives or avoiding false negatives.

Let’s take the simple example of a 911 call where nothing but indistinct sounds are heard. While it is probable that the call has been made in error, a police unit will be dispatched. Why? Because in responding to 911 calls there is a much higher need to reduce false negatives (cases where you should respond because someone is being hurt or there is a medical emergency but don’t) than to reduce false positives (cases where you don’t need to respond but do and then just apologize for the interruption and leave). As an obvious counterexample, let’s consider a capital murder case. Now the parameters are reversed. We are willing to put in place protections and appeals that may cause error on the side of false negatives (guilty individuals going free) to avoid false positives (innocent individuals being put to death or jailed for life). Our point is that any decision making system must include judgments about how important it is to avoid false positives versus how important it is to avoid false negatives.

For our purposes, this requires that we consider the element of time, the maximum possible risk, and the degree of information available.

### False Positives and False Negatives Depend on Stage and Available Information

The consideration of false positives and false negatives must be informed by several factors. The first two factors are interrelated: Uncertainty and the stage of investigation. As CWS involvement progresses, more information is gained and uncertainty is reduced. Certainly, exceptions exist – a new element unknown to the worker may emerge anytime during a case, for example – but generally, uncertainty decreases with time. Uncertainty is generally highest at the referral level, and so the screening decision is the decision made with the least available information.

Uncertainty due to limited information is a key factor, as the accumulation of information should make the decision making process as a whole more reliable. Specifically, the risk of *both* false positives and false negatives should decrease as more information becomes available. As an example of a case which begins

with very high uncertainty, hotline workers sometimes don't even have a name, only an address, and a concern that a child has a suspicious mark that could be a bruise. The Emergency Response worker will be dispatched, find out who the family is, may determine that the mark is a bruise, and will then set about gaining more information to reduce uncertainty. What is the child's story? What is the parent's story? Does the physical evidence support a given conclusion? By the time a case might reach the point of court involvement, uncertainty is further reduced simply because more is known. For example, for the relatively few cases that reach the level of court involvement, medical testimony will be available, a more complete history will be in place, and the information the court has to work with will be substantially better than that held during the initial investigation and vastly better than the information available when the screening decision was made.

Another critical factor to consider is risk. Take a simple example of a person who has left home for a two-week vacation and worries that he may have left the hallway light on. That person is unlikely to feel compelled to return home, because the risk of not acting is trivial: Merely a few extra dollars on his or her electric bill. On the other hand, let's assume that the person is concerned that they may have left their *gas stove* on. In this situation, the person is very likely to return home, as the potential consequences are much more serious. CWS decision making must be considered in a similar manner. When choosing how to calibrate false positives and false negatives, we must consider not only uncertainty, but the risk associated with making a mistake in either direction. Importantly, from an ethical standpoint, this risk should be related to the child and the family, not the potential liability of the agency.

Avoiding false negatives (determining the child is not at risk when the child is at risk) is unquestionably the most basic concern of CWS. A false negative determination is a potential, perhaps literally "fatal" error throughout the CWS process, as a child can be endangered at any point from the initial call forward. Mitigating this harm is the core charge of the CWS.

False positives are another very serious concern (determining the child is at risk when the child is not at risk), but moves from the child to the caregiver level. Parents have a constitutionally guaranteed right to determine how to raise their own children, and most parents, even those referred to CWS, care deeply about their children. A false positive determination can prevent the parent from exercising this right and cause emotional discomfort, as well as unnecessary intrusion by CWS. Yet, *the degree to which that can happen is dependent upon the stage of the CWS case.*

At the *hotline level*, when the determination is to screen-in or screen-out a referral, there is potential harm done to families through false positives. The downsides to false positives are lower at this stage than at later stages of CWS involvement, such as at removal decisions. CWS agencies must balance the tradeoffs between preventing false positives and preventing false negatives, which could result in harm done to children.

*During the investigation phase*, a number of determinations are made. In egregious cases, where evidence of sufficient harm to the child is immediately apparent, a removal may be made, sometimes with the assistance of law enforcement at the first contact. This is the most serious possible consequence of a false positive determination at this stage, but is very uncommon, since child removals require judicial confirmation of the need to remove within 72 hours. Far more likely is that the case will be investigated and closed with no further services, as mentioned earlier. An intermediate case is one in which the investigation or assessment recommends services when they are not needed, but does not invoke court involvement. Here again, a false determination of risk may have unnecessary impacts, including asking the family to engage in unneeded services and longer discomfort or more embarrassment at having been

engaged by CWS. So the downside of a false positive assessment and case opening is higher than the costs of such a mistake at the hotline level, but far lower than the costs possible at the family court level. It should be noted, however, that the worker who meets with the family and assesses the situation gains more information than was available to the hotline worker, so the chances of error should be reduced. This is an important point. The decision to screen-in a case merely gets a worker to the door. As we have described above, other decisions (e.g., substantiation) do not automatically follow from a screen-in. In fact, substantiation and foster care follow a screen-in decision only in a small minority of cases.

*After a case is opened or at the same time a case is opened, it is possible that a recommendation is made to involve Dependency Court.* At this point, uncertainty is further reduced by more complete information, but the potential negative effects of a false positive determination are extremely severe. It is possible, for example, that if the court is wrongly convinced that a parent represents a serious ongoing threat to their child's safety, that parent's rights may be unjustly and unnecessarily terminated altogether, with severely negative consequences for the entire family.

We have detailed the potential costs of both false negatives and false positives in order to provide context to the ethical evaluation of decision-making. The key findings of our analysis can be summarized as follows:

- 1) False negatives always, by definition, constitute a potential danger to the child. This danger can be severe, up to and including fatality.
- 2) False positives constitute a low risk to families at the screening level, a higher risk to families during the investigation phase, and the highest risk to families if the dependency court is invoked following an investigation.
- 3) Uncertainty should generally decrease from hotline through investigation to case service decisions. Assuming competent continual assessment, false positive and false negative errors should become progressively less likely due to greater access to more information about the family and associated context.

It is easy to see how the CWS system was built upon at least an implicit recognition of these principles. In most states, the screen-in process is designed to tolerate false positives and avoid false negatives, with most cases (about 60%: DHHS 2017) being screened in. The tolerance of false positives may be especially high in cases of great perceived risk. For example, the final report from the federal Commission to Eliminate Child Abuse and Neglect Fatalities (CECANF, 2016), suggests that *all* hotline reports for children under age three be investigated (CECANF, 2016, Recommendation 2.1e). Universal screen-in is already policy in some jurisdictions nationally, with some counties investigating all child maltreatment reports and some investigating all reports in a given category (e.g., very young children). This represents an intentional and maximal "skewing" of the system in favor of minimizing false negatives (if all young children are screened in, then false negative screening decisions for this age group *cannot* exist). On the other hand, dependency court proceedings include a large number of checks intended to avoid action based on false assessments of risk, including longer time periods to assess and provide services before additional determinations are made, legal representation for the state and often an advocate representative for the child in the form of a Guardian ad Litem or Court Appointed Special Advocate volunteer, and, in some cases, legal representation for the parents. CWS advises the court, but does not make the final determinations at this point. At the extreme end of the spectrum, the number of parents who have their parental rights terminated is a tiny fraction of all reported cases and a small percentage even of children in foster care - about 15% of children in foster care have a TPR designation (USDHHS, 2015).

## **Bias in CWS**

There is a longstanding belief, shared by the laity and professionals alike, that the CWS includes a number of biases that impact how it functions. By “bias” we mean beliefs or actions on the part of CWS personnel based on case-irrelevant factors such as race or income. The two most commonly suggested kinds of potential bias in CWS are race and class, although surveillance bias due to exposure to service providers is also often mentioned.

In the attached *Appendix B: Bias in the Child Welfare System*, we overview the existing evidence for the three kinds of bias mentioned above. Research has evolved substantially in the past five years. Before that time, most researchers (and others) assumed that race, class and surveillance bias might have large effects. Recent well-controlled studies are casting substantial doubt on these conclusions. We will briefly cover what is known on these fronts, emphasizing not raw disproportionality (differences in overall numbers of reports), but disproportionality after other key factors, mainly income, are controlled for. This allows an “apples to apples” comparison.

Virtually all recent studies employing controls for risk factors show disproportionality to be much smaller than was previously expected. In some cases, the identified disproportionality is the reverse of what was anticipated. Of course, while it is appearing increasingly unlikely that CWS can be empirically shown to be making decisions that are influenced by widespread or practically large levels of bias, this in no way mitigates valid ethical concerns that bias may exist in individual cases.

## **Risk Assessment: What Works and What Doesn't**

It would, of course, be tremendously useful for CWS to have a “crystal ball” which would help workers to see the future. Specifically, the most critical information needed by CWS has to do with the future risks to child safety, as mitigating this is the central focus of the system. The pursuit of this crystal ball has a long history.

Up until the 1990s the nascent CWS relied primarily on clinical judgment in assessments of risk. With the increasing awareness that clinical judgment alone was not a good means of predicting risk (Rycus & Hughes, 2003), agencies looked to the use of more formal instruments. Attempts to create such instruments, usually in the form of checklists or sets of risk factors that the worker can rate, have been going on for at least three decades (Ammerman & Hersen, 1992). There have been two general types of risk assessment: “consensus-based” instruments, which include items that experts have determined should theoretically predict future risk of harm; and “actuarial” instruments, which are based on items empirically demonstrated to be predictive of future risk of harm. A good amount of research has been completed on risk assessments, especially in the 1990s and early 2000s. It has long been established that actuarially-based tools outperform instruments based on consensus and also outperform clinical judgment alone (Baird & Wagner, 2000; Coohy et al, 2012; Gambrill & Schlonsky, 2000; Van Der Put, Assink, & Stams, 2016).

### *Actuarial Instruments*

In particular, actuarial variables referencing prior maltreatment and prior maltreatment reports have consistently been found to be strong predictors of future risk, commonly being the most predictive factor in actuarial instruments (Dorsey, et al., 2008; Marshall & English, 1999; Sledjeski, Dierker, Brigham, & Breslin, 2008). Research is clear, however, that once screened in, unsubstantiated and substantiated cases are at similar risk of recidivism. In other words, substantiation status makes little or no difference in predicting future risk. This has been found using state-level samples (e.g., Drake et al., 2003) as well as national data (Hussey et al., 2005; Kohl et al., 2008). One of the things we don't know enough about is the

distinction between future risk among screened-in and screened-out cases. Work from Allegheny County (Allegheny County DHS, 2016) suggests that screened out cases are somewhat *more* likely to be re-reported than screened-in cases, although they are less likely to have a subsequent report resulting in a foster care placement. Additionally, other elements of risk are key to strong performance of actuarial measures like parental criminality, mental health, or substance abuse (Coohey et al, 2012).

### *Predictive Risk Modeling*

“Predictive Risk Modeling” (PRM) has recently received attention as another way to generate probabilities of future risk to children in the child welfare system. PRM is similar to actuarial methods in that it aggregates available information into a risk score. What is different is that a computer is used to sort through the information to try to improve upon what can be generated through checklists based on a single worker’s assessment. PRM’s validity is largely judged by its “Area Under Curve” (AUC), a statistical measure of what proportion of predictions are made accurately. AUC is based on the “Receiver Operating Characteristic” (ROC) curve, which plots rates of true positives against rates of false positives at various thresholds. An AUC of 1.0 is perfect, and an AUC of .5 represents no improvement over random chance. Large scale trials modeling the use of PRM have been done in New Zealand (Vaithianathan, Maloney, Putnam-Hornstein & Jiang, 2013), resulting in an “Area Under Curve” (AUC) assessment of 76%.

In the United States, Allegheny County tested a PRM at the screening level drawing on multiple data sources. Preliminary results suggest that the PRM model considerably outperforms current screening decision making. For example, among those cases classified as being at highest risk by the PRM, 27% of cases were screened out during usual service. Among those cases, one in three were re-referred *and placed* within two years of the initial call. Using the same model, about half of the cases identified by the PRM as being lowest risk were screened in, and virtually none of those cases (1.4%) were re-referred and placed within two years.

Preliminary findings from the PRM algorithm developed in California (Putnam-Hornstein, Vaithianathan, Prindle, Cuccaro-Alamin, Nghiem, & Gupta, 2018) have shown very high AUC’s ranging from 0.80 to 0.94 when CWS data is used to predict future levels of system involvement (e.g., 3+ additional referrals, foster care placement). Algorithmic classification performance of future system involvement was also shown to be superior to current practice as defined by the use of an existing actuarial risk assessment tool (i.e., the SDM® Family Risk Assessment). From a statistical perspective, this confirms and extends prior work done in Allegheny County showing that the model was more accurate than traditional methods for screening child maltreatment referrals.

## **Administrative Child Welfare Data and PRM**

This section will provide an overview of the nature of administrative child welfare both in general, and relative to its use in PRM specifically. **Potential applications of PRM that are being explored in California would only use CWS data.** While we discuss the ethical issues attendant to the use of other forms of data, this is being done to provide the reader with a more complete and general sense of the issues involved. **Many of the themes we explore below go beyond the specific business use case in California.**

### **Sources and Types of Administrative Data**

There are many different kinds of administrative data that exist and could possibly be used in PRM. These have radically different ethical issues, ranging from minor or non-existent (i.e., accessing public domain

data) to extremely serious (i.e., use of private healthcare data). It is therefore useful to consider the various kinds of data available.

- **Universal:** Some data exist for everyone (or effectively everyone). Among these are birth and death records, marriage records, divorces, and other publicly registered milestones. In California, birth records are detailed and provide information predictive of maltreatment reports (Putnam-Hornstein & Needell, 2011). Some of these data are freely available to the public, but most are not.
- **CWS data:** The CWS holds its own internal data on individuals who have been part of a child protection referral. This is not a universal dataset but is “gated” by someone having made a referral alleging the abuse or neglect of a child. It is important to note that the decision to make a referral is the only mechanism at play. There is no intermediary decision-maker, such as might occur with medical data if a person decides to go to a public clinic as opposed to seeing a private doctor who may handle reporting responsibilities differently. CWS data are never publicly available, but may be aggregated for public reporting purposes (e.g., the California Child Welfare Indicators Project at UC Berkeley).
- **Other State-Held Data:** These data are created when an individual comes into contact with a state administered system, either involuntarily (e.g., criminal justice) or voluntarily (e.g., when seeking income assistance). A wide range of state-held data exist in different states, and some (e.g., South Carolina) have integrated these data sources into single systems. These data vary enormously in accessibility. For example, arrest records may be publicly available, but income maintenance data generally are not. It is again important to remember that each of these datasets have different mechanisms through which people enter them. Most of these differ markedly from the CWS.
- **Public Domain Data:** Various forms of data are available to the general public and could be accessed by CWS. Examples include sex offender registries, conviction records, Census, or aggregated arrest data describing neighborhood characteristics. Each of these could be of value in PRM or to the worker in the field.
- **Other Data:** Much data exists in the hands of non-public institutions which could be relevant to CWS functions. For example, private hospitals maintain data on emergency department usage which could be of value in modeling. Each of these data sources have a different set of ethical concerns attending them, which we describe below in the “*Ethics of Access...*” “*Ethics of Analysis...*” and “*Ethics of Use...*” sections.

## The Nature of Administrative Data

It is important to not simply think about “Administrative Data” in a monolithic fashion. Administrative data are as complex as any other type of information.

- **Format of information recorded:** Data can be text (narrative), ordinal (ranked counts, such as first or second), numeric (“Interval” or “Ratio”), or categorical (race, action taken). These each pose different challenges to the researcher or user, and as yet, textual information has been little used, as it requires a set of complex and evolving procedures to be processed into usable information.
- **Validity (“correctness”) of the Data:** Different kinds of data have different propensity for error. For example, mortality data (dead or not) has little chance of error, but a ruling of the cause of death has a greater chance of error as it is based on the information available and the human judgment of a coroner. For example, sometimes a child’s death may be from maltreatment, but cannot be conclusively judged as such based on available data (CECANF, 2016). Similarly, we might put less faith in the validity of arrest data than we would in conviction data because the further one’s involvement in the criminal justice system, the more information that is available and

required to move forward. Data can be both valid and suspect at the same time. For example, report data to CWS is quite valid in the sense that it accurately reflects that a report was made and what was said by the reporter. Nevertheless, the specific concerns expressed in the report may be in error.

- **Subject level (“unit of analysis”):** Perhaps one of the greatest areas of flexibility and confusion in the use of administrative data has to do with the unit of analysis. Administrative data can be aggregated at the geographic level, can be managed to reflect events within a given family, or they can focus on parents, perpetrators or children. The same base data can be managed to create datasets at all these levels and more. It is therefore important to understand that in a mathematical sense, the data aren’t “about” any given actor or system per se, they can be managed to pertain to any or all actors or levels of analysis, and they can be identified or de-identified. Some levels of analysis are intrinsically de-identified relative to individuals (e.g., geographic analyses). This issue becomes critically important from an ethical perspective. The questions “Who is this data about” or “whose interests and privacy concerns are important in these data” have not been addressed systematically in CWS data or analysis. We believe that some tentative principles can be advanced, however.

The subject of the data, from a CWS mission perspective is, first and foremost, the child. CWS exists to support child safety as its primary function. Other actors are critically important (e.g., caregivers and alleged perpetrators), as they relate to child safety. From an ethical perspective relative to the child, using the information available to support child safety is the primary CWS concern, but maintaining confidentiality interests of the child is also a key concern.

Secondarily, key CWS functions (permanency, well-being) require data about other key actors, including caregivers and alleged perpetrators. These data are necessary for CWS to pursue key functions. Again, safeguarding the confidentiality of all involved persons is a key CWS concern.

### Current Predictive uses of CWS Administrative Data

It is standard practice for CWS data systems to provide internal (CWS) data to CWS workers. This is not only ethically permissible, but is ethically obligatory given that prior CWS data on reports are among the most predictive data which can be used in assessing child safety (see above). How this information can be accessed (i.e., easily summarized compared to requiring review of case files) and therefore how frequently such information is used likely varies a great deal. Different workers may access different information, and may use that information differently. Many local and state systems allow CWS workers to manually access other state data sources, such as arrest data, although this may be onerous to do and may be impractical for instant use in screening or emergency response circumstances. Sometimes there is additional cost involved to the CWS. These are financial, political and technical rather than ethical barriers. There is no current disagreement that CWS may access such databases. There is no reason, given current technology that such record linkages could not be established in real-time. Having workers fully informed seems advantageous, from both a practical and ethical perspective. This is particularly true given the history of bad child outcomes later explained as occurring because “nobody noticed” or “the child slipped through the cracks.”

### In Context: California, New Zealand, Allegheny County and Birth Match

The PRM system currently proposed in California is remarkably conservative compared to the systems envisioned in New Zealand, Allegheny County, and in states operating Birth Match Systems. California is the only jurisdiction proposing to use only data already held by the CWS system. New Zealand and the Birth Match states also propose (and Allegheny County discusses) using administrative data prior to any

report being made, which California does not. Possible ethical concerns attendant to the use of non-CWS data, which are present in the other cases, are therefore absent in the case of California. Substantial work has already been done evaluating the ethics of PRM as studied in New Zealand (Dare, 2013) and Allegheny County (Dare & Gambrill, 2017), and various articles have been published providing ethical commentary on the matter, primarily regarding the New Zealand case (de Haan & Connolly, 2014; Gillingham, 2015; Keddell, 2015)

### A Clarification: “Universal-Level Risk Stratification” vs. “Referral Level Risk Stratification”

One of the most important distinctions in the use of PRM in child welfare is between what we will term “universal-level risk stratification” and “referral-level risk stratification.” For purposes of conducting an ethical evaluation of a PRM in California, we would note that California is not considering population level risk stratification. As such, ethical concerns regarding universal monitoring are not germane to the California case. That said, we have offered a short description of the two use cases. We include a discussion of this issue simply to provide context. Ethical reviews have already been performed on the use of PRM in applications which go well beyond those planned in California. Those evaluations have been generally supportive of the use of PRM. In this regard, California’s proposed use case could be seen as “conservative” and less ethically risky than broader approaches with a greater number of ethical concerns.

- **Universal-level Risk Stratification:** Examples of universal-level risk stratification include Birth Match (implemented in several states), New Zealand’s proposed preventative intervention program, and the future preventative programs outlined (and currently being explored) by Allegheny County (Allegheny County DHS, 2016). All of these programs use existing data (e.g., prior histories of termination of parental rights, birth records, income maintenance records) to identify at risk children *among the entire population of children*. This represents a system for “case finding” (at least in a preventative sense) separate from the traditional CWS hotline in which reports of abuse or neglect are fielded. Some of these datasets (e.g., birth records) are truly universal, but with some exceptions (e.g., children born out of state). Other datasets (e.g., income maintenance) are universal in the sense that everyone in the population can be identified as either using or not using the service, but details within the dataset are only available to service users. Universal-level risk stratification can be used to generate CWS contacts without a referral being made. This is a fundamentally different way to conduct CWS services, and is currently in place on a very small scale in states with Birth Match programs. Adoption of a program along the New Zealand lines would supplement the “referral-gated” CWS process with an additional “data-gated” case generation system. Universal-level risk stratification raises a host of ethical issues attendant to the identification of new CWS cases. These ethical issues do not arise in referral-level risk stratification, as no new cases are identified as the “gate” – a hotline referral, remains the same.
- **Referral-Level Risk stratification:** PRM can be used after a referral is received to better predict the future risk and determine the level of CWS response required. This is the situation being considered in California. No new referrals or reports are generated, but depending on the parameters of the system, some cases that are not currently investigated or opened might be investigated or opened, and some cases which are currently investigated or opened might not be under the proposed system. The overall goal would be to make screening, opening, and service provision determinations more reflective of the actual risk to the child.

## The Ethics of Electronic vs. Physical Data

We would like to close this section with a brief discussion of the differences between electronic scoring and human-centered scoring of data. Paper copy (or even computer-entered) risk assessment tools or structured systems with aspects akin to risk assessment (e.g., SDM®) have been used for years. Indeed, the CWS data available for PRM has human origins. Someone had to call and provide information. If a prior report was investigated, a worker generated data about that family.

The use of a more complex form of computerized risk assessment (PRM) seems to be causing a quantum increase in ethical concerns about data use, despite the fact that the new and old processes and instruments may use the same data.

We feel this is ironic, as we can identify no specific ethical concerns attendant to the use of computerized systems, but we can identify many ethical advantages. The proposed PRM in California does not include any data which the worker does not already have the right to access – and which it is expected that they review, time allowing. PRM is therefore different only in the increased efficiency, transparency, and standardized use of data it provides. It is more efficient and standardized because it “sees” all the data, not just what the worker might notice, or the fewer items which might be included in a traditional risk assessment tool. Indeed, research on decision-making indicates that even given available information, there is great variability as to the attendance to and prioritization of types of information when assessing a child protection case based on training and education of workers (BenBenishty, Segev & Surkis, 2002; Ruscio, 1998). Additionally, there is error associated with the human decision-making process itself (Fluke, 2011). To the degree PRM has superior predictive ability compared to reliance on an individual’s assessment alone (guided by a structured tool or not), it is more efficient. PRM is also more amenable to corrective action in another sense. Being fully computerized, the use of the PRM can be simulated in advance of implementation and potential ethical issues (such as systematic over- or under-assessment of risk in a given racial group) can be addressed in advance. Indeed, that is what California has been exploring with its proof-of-concept built from historical records. Finally, computerized processes are fully mathematically “knowable,” in a way that worker decision making never can be. In this way, PRM adds a tool to the decision making process which has the net effect of increasing transparency. PRM confers other ethical advantages secondary to the improved resistance to error inherent in computer systems. A PRM algorithm does not suffer burn-out, does not tire, and does not tally risks inaccurately. We all would like our CWS workers to be all-seeing, not forgetful, and 100% consistent in how they use information. Within the narrow bounds of what it is asked to do, a PRM system meets all these standards.

## Testing PRM Systems before Deployment: A Methodological Note

PRM systems can be run on historical referrals which have already received services and have known outcomes. For example, in Allegheny County (Allegheny County, 2016), risk categories assigned by PRM models were compared to screening and service decisions made in known cases. An observable subsequent outcome measure (e.g., subsequent re-report and/or foster care placement) was used to compare the accuracy of PRM projections against decisions made using current procedures. This kind of analysis can provide very useful data. Should those referrals highlighted as high-risk by the PRM recidivate at a higher rate than those for which current risk is identified (e.g., screened-in vs. screened out cases), then this stands as evidence that PRM is a superior risk detection instrument than current processes. Such evidence is useful and can serve as a basis for making decisions about the employment of PRM.

Evidence of this type is not perfect, however, there is no intention to use PRM as a “stand alone” decision making tool. Should PRM be implemented it would be as a tool to assist workers in making decisions. While the PRM’s risk score can be identified using research designs of the above nature, the precise

human decisions which would be based on the PRM score cannot. Such “validation” PRM research can be a powerful indicator of the accuracy of PRM assessments compared to the accuracy of current decision-making, but cannot fully model PRM as it would be used in real time in the field. For this reason, while testing PRM is important for scientific, policy, and ethical reasons, it should be followed up with monitoring of the impact of PRM after it goes “live,” should it do so.

## **Ethical Issues Associated With the Use of PRM in California**

Questions regarding the degree to which the use of PRM in California poses new or unique ethical issues relative to the general principles above must begin with a simple question: “Are there any ways in which PRM is different than current practice with regard to access or use of data?” We believe that the answer to this question is “no.” California does not plan to use any data to which they do not currently have ethical and legal access. Further, the uses to which the data are to be put (making decisions regarding screening and case opening) are not new – these decisions are currently being made using the same data, simply in a less structured manner. The PRM data will simply provide a new tool for making more accurate decisions and raises no new general issues from an ethical perspective. The proposed change will “use the same old data for the same old purposes.” It will merely use those data more efficiently and produce more accurate predictions.

Given that ethical analyses of PRM in child welfare are already extant, we will first summarize the main issues, points and conclusions from those analyses and relevant commentaries. This section draws heavily from three prior documents, “Predictive Risk Modeling and Child Maltreatment, An Ethical Review,” written by Tim Dare, a philosophy professor at the University of Auckland. This report was published in response to a recommendation for an ethical review made in New Zealand’s Vulnerable Children Report (Vaithianathan et al., 2012). Similarly, a paper focusing on ethical aspects of PRM was developed for Allegheny County (Dare & Gambrill, 2017), who also provided their own feedback on Dare & Gambrill’s analysis (Allegheny County DHS, 2017). We omit content specific to the New Zealand or Allegheny population-based modeling, since this type of modeling is not being contemplated in California. While we use these reviews as background, this report is not meant to be merely an application or extension of the prior work. The ethical analyses and conclusions herein are ours, and are specific to California.

### **Over- and Under-Identification: Using the Right Tools and Hitting the Sweet Spot**

#### Using the Right Tools

As has been previously discussed, in an ideal world, all referrals would be correctly classified as needing a response or not, and all decisions to open, or not open, a case would be correct. We do not live in an ideal world. Therefore, we understand the following:

- 1) Incorrect classifications will be made.
- 2) Less data makes classification more difficult, meaning early decisions (such as screening) can be harder to make.
- 3) Thresholds can be set to reduce false positive classifications or false negative classifications, but not both simultaneously.
- 4) Some means of classifying cases (e.g., actuarial methods) work better than others.

Given these observations, it is clearly ethically incumbent upon the CWS to use the best possible decision aids for classifying referrals available (“the right tool”) and to find the threshold which finds the best balance (“the sweet spot”) between avoiding false negatives and allowing children to remain unprotected on one hand, and avoiding false positives, and burdening families with unnecessary system contact on the

other hand. It is also worth noting that there are a large number of purposes to which any classification scheme could be applied. You could have a process designed to make a correct diagnosis of a currently existing medical condition, for example. The purpose here is somewhat different, and involves prediction of human behavior in the future, among the more difficult tasks which can be attempted.

As previously discussed, unsupplemented worker decisions and consensus-based tools have been shown to be inferior to actuarial tools in their predictive utility. From an ethical perspective, the use of empirically validated actuarial tools therefore constitutes a moral imperative. Any other approach involves knowingly using a less accurate means of understanding the risks to the child in the case, which is ethically indefensible. The question is more complex than it may seem on the surface, for a reason familiar to all research methodologists. The ability of a tool to predict probabilities in a population does not necessarily mean that the tool will apply to every individual case. This is commonly recognized in CWS practice. Many CWS agencies have a policy that states that given a particular risk score, a particular action must be taken. These policies, however, often include an "override" clause, whereby a worker may take a different action (often with a required supervisory sign-off) should there be valid reasons why the instrument may not fit the family or situation well. The use of actuarial tools in CWS is therefore supplementary to worker decision-making, rather than rigidly directive of the worker's actions. This idea is increasingly forwarded as a means to accommodate the ethical requirement to use the most effective tool, coupled with the ethical requirement to be sensitive to specific situations (Coohy et al, 2012; Schlonsky & Wagner, 2005).

With regard to PRM, we encounter a simple extension of the above issues. PRM is simply a more sophisticated actuarial method. The sophistication of PRM compared to "pencil-and-paper" or more likely "screen entry" actuarial tools is no reason to prefer it. Useful evidence pertaining to which is preferable can only be found in comparing the predictive capability of one to the other. The ethical question with regard to over and under identification therefore boils down to a purely empirical question: "Does PRM more correctly classify cases than the best available standard actuarial tools?" If the answer to this question is "Yes" then the use of PRM is ethically preferable to current practice. If the answer is "no" then it is ethically inferior.

This is, as stated, an empirical question. We have a rare case where an ethical issue can be settled mathematically. This is very unusual. Many ethical issues (e.g., end of life decisions) involve weighing abstract values, and that inevitably requires subjective evaluation. In the case of child safety, we have a range of available empirical predictors which are relatively straightforward (re-reporting, later placement, other outcomes) which provides a strong set of empirical indicators for ethically preferable outcomes. Any responsible use of PRM requires that the entity adopting PRM demonstrate that it is a more accurate classification system (or at least an equally accurate system) to current practice. Fortunately, this can be done fairly easily. Any agency desiring to adopt PRM in a decision-making capacity should build PRM models for particular tasks and compare their effectiveness scientifically to existing decisions.

In the case at hand, we feel that California should develop a PRM model or PRM models relevant to supporting worker decisions for specific tasks (e.g., screening in vs. screening out or cases being opened vs. cases not being opened). Dependent variables (outcomes) of interest could include re-reports or subsequent placements or other variables that signal future system involvement. Some key variables of that are pure measures of child safety (e.g., fatalities) may not be possible to analyze due to methodological and statistical concerns such as small cell sizes, but could serve as a method for validating PRMs trained to target other outcomes. All analyses should determine if a predictive risk model has better predictive capability (e.g., in predicting re-reports) than existing classification systems or tools.

Such an analysis can be examined using existing data without changing current practice simply by comparing the predictive capacity of existing risk assessment tools to a PRM, for example. Some methodological limitations would exist. It would not be possible, for example, to monitor the effects of things like the above described “overrides” without taking the PRM procedure “to the field.” Still, it should be possible to get an informative estimate of the relative accuracy of PRM compared to existing practice.

Dare (2013) makes the point that to the degree that a PRM approach is more accurate, it is ethically improved. Given the nature of PRM, as the data it draws upon increase, the predictive utility of the model may also increase. This means that the use of cross-sector data (e.g., arrests, medical, birth data) may provide a more accurate PRM tool, and thus (from an efficacy perspective) a more ethically sound tool.

### Conclusion and Recommendation

From an ethical perspective, and considering only the issue of “the right tool,” should PRM prove to be more accurate than existing classification tools and other potential alternatives, its use becomes not only ethically permissible but ethically necessary. It would be fair for any consumer of CWS services to ask “Did CWS use the best possible means available in responding to my family?” It is ethically desirable that CWS be able to respond to that question and to do so on the basis of sound evidence. We would further endorse the use of PRM as a tool to assist worker decision making in current practice, and not a system (Coohey et al, 2012; Dare, 2013; Schlonsky & Wagner, 2005) to replace worker decisions. We would further endorse the utilization of cross-sector data as ethically valuable in increasing the accuracy of the tool, subject to other ethical concerns attendant to those data.

### Hitting the Sweet Spot

All decisions require thresholds. When should a given “risk score” cause a referral to be screened-in or opened for service? We would make the following observations from an ethical perspective.

Dare (2013) cautions us to recall the consequences of errors in classification. In this sense, decisions made early in the process are ethically different from decisions made late in the process. We discussed this extensively above (“How False Positives and False Negatives Depend on Stage and Available Information”). The essential points are as follows: Child safety is always a key goal at every stage of intervention. This means the potential harm caused by a false negative is always high. A child who dies as a result of being screened-out and a child who dies as a result of reunification are equally dead. The risk to the child and family borne by a false positive, however, changes radically over time. This suggests that risk or PRM models “early” in the CWS process, like screening-in, should have a lower threshold than those for later decisions, like case opening.

In addition, there are situations in which different thresholds might be applied due to specific empirically determined risks. As previously stated, the Federal Commission to Eliminate Child Abuse and Neglect fatalities suggests a fully inclusive screen-in threshold of 100% when referrals involve very young children. This is due to the fact that fatalities are so strongly clustered among the very young.

Echoing our recommendation under “The Right Tool,” we would stress that PRM models allow different outcomes to be modeled in advance of their implementation. A PRM model can easily be set to output a binary decision (e.g., “screened-in” vs. “screened-out”) at any desired threshold, and the true/false percentages (based on one or several carefully chosen outcome variables, perhaps “re-report,” “future placement,” or “fatality” can be determined for both decisions. Again, a fully virtual trial can be very useful in comparing PRM models to existing models, but would need to be reassessed once field use were to begin.

## Conclusion and Recommendation

Any agency adopting PRM should determine if the same or different models should be used for different decision points, and should also determine what thresholds produce the best mix of true positives, true negatives, false positives, and false negatives. Stage of process (screen-in vs. case opening, vs. other decisions) should be considered relative to setting thresholds, with earlier decision points having less tolerance for false negatives. Determining the “best” mix of true and false positives and negatives will require human judgment, and is not a purely mathematical exercise in the same way that comparing actuarial and consensus-based instruments can be. For example, is it ethical to accept that, on the average, one child who should be screened in won’t be to assure that, say, ten children who should not be screened in are accurately classified? Should the number be lower, perhaps five? Should it be higher, perhaps a hundred? These are Solomonesque decisions to be sure, and there is an ethical requirement for them to be made on the basis of the best available evidence.

### *False Positives: Mitigation*

Dare (2013) suggests a series of practical steps which can be used to mitigate false positives. Essentially, the list involves ways in which systems and workers can be alert to the chance that they are acting on the basis of a false positive, and can act to mitigate any attendant harm. Dare suggests the following practices:

#### **Suggested Means of Reducing the Consequences of False Positives (Dare, 2013)**

- Providing opportunity for experienced social service professionals to exercise judgement about appropriate responses to a family’s identification as at risk;
- Ensuring that such professionals understand the potential of the PRM to miscategorize families;
- Providing training to guard, in so far as possible, against confirmation bias in the professional engagement with families identified as high-risk;
- Offering rather than requiring engagement as a consequence of identification as at risk;
- 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;
- Ensuring that intervention triggered by identification as at risk is positive and supportive rather than punitive; and
- Identifying and minimizing the adverse effects of identification as at risk, such as, for instance, possible stigmatization.

We would characterize these recommendations as sound basic social work practice and endorse their usage. Many of these principles are extensions or restatements of already existing good practice guidelines (e.g., “ensuring that intervention... is positive and supportive”).

### *False Negatives: Mitigation*

We may also be able to take some small steps in mitigating false negatives. While such mitigation may appear impossible (how can a worker help when a case is never screened in or has been closed?), there may be steps which can be taken. First, we suggest keeping data regarding all calls, screened in or not, opened or not, indefinitely and incorporating them into any PRM to the extent that they improve the model. Second, in cases where a case is not opened, we suggest that the worker still make the maximum effort to provide help to the family, perhaps in the form of informational packets supporting family

material, parenting, or other needs. In this way, at least some minimal service is provided in cases which are inappropriately closed.

### *Perspectives*

As a final note, we would suggest that one's interpretation of the relative cost of false positives and false negatives will be fundamentally colored by how one views CWS. If CWS is viewed from a public health perspective, then it would be natural to view false negatives as more adverse than false positives, because the goal of a public health system is to minimize risk. If the CWS is viewed from a criminal justice perspective, then there will be greater discomfort with false positives, as incorrectly determining "guilt" is something we try very hard to avoid within this perspective. The next section discusses one of the key potential negative outcomes attendant to false positives: Stigmatization.

Another related perspective which can inform this debate has to do with timing. Imagine a family that will develop child maltreatment issues in the future. Low threshold systems will have a higher chance of providing earlier interventions at a time when truly preventative work is possible. This will come at the cost of unnecessarily contacting a number of families that will never be at risk. Again, from a public health perspective, this may not be an unfamiliar or unacceptable cost. A program testing for a given problem (such as lead exposure) with a very low positive rate may still be deemed successful. High threshold systems will reduce the number of false positives and true positives, as described, but may also tend to move intervention "forward" to a later time in the development of the maltreating behavior. From this perspective, threshold setting may influence the degree to which CWS services can be truly preventative (e.g., can be the basis for secondary as opposed to tertiary prevention). To the degree that earlier prevention is morally preferable, lower thresholds become more ethically attractive.

### **Stigmatization**

Dare's (2013) second area of focus is "Stigmatization and the Costs Associated with Identification as At Risk." He presents a number of key points:

- Stigma can exist both in cases where child abuse and neglect may be correctly or incorrectly assessed.
- Stigma may arise from false beliefs about the nature of the stigmatized action or group. As an example, we would use the case of head lice. Head lice are often incorrectly thought to be a sign of poor child care or slovenliness. In fact, head lice are spread by contact with other exposed children and such infestations occur indiscriminately. Despite this, few families are willing to talk openly about their child's head lice, for fear that they will be harshly judged. We concur with Dare (2013) that to some degree this situation may manifest in child maltreatment. Child maltreatment referrals happen to more than one in three children (Kim & Drake, 2017) and are a fairly common experience in our society. This does not mean that a visit from CWS will ever be interpreted as anything like a normal experience. If the CECANF recommendation on screening in all young children were to be put into effect in the laudable cause of saving children's lives, then thousands of families would be contacted based on relatively less complete or serious reports. Despite the lower threshold for investigating, it is unlikely that this would reduce the stress or potential stigmatization faced by the family. Dare (2013) cautions us to respect this form of stigma as particularly ethically concerning, as the stigma is based on unfounded assumptions. In particular, Dare suggests that in mitigating such stigma, "Policy makers who trigger such costs must take them into account even though they may think them completely without foundation and even though, in some sense, the policy makers are not responsible for their imposition on the bearers of illegitimate stigma" (2013, p.26).

- Stigma may increase the risk of adverse outcomes. Fear of being placed in a stigmatized group may inhibit disclosure or involvement in treatment. It may be possible that normalization or other approaches by CWS staff meant to reduce stigma may have positive effects in this regard.
- Stigma may be cumulative. While the CWS system does not cause this dynamic, it is worth considering. Poverty causes stress for families and can be a driver into the child welfare system. It is necessary that CWS responses within this population work to reduce existing and cumulative stigma effects to the degree possible.

Dare argues that avoidance or reduction of stigma is meritorious, but does not excuse the government from its primary responsibility to protect its own citizens. In particular, he quotes John Stuart Mill as stating that, "the only purpose for which power can be rightfully exercised over any member of a civilized society against his will is to prevent harm to others" and that "it is one of the undisputed functions of government to take precautions against crime before it is committed" (2013, p.29).

We are thus again at a balancing point. Stigma must be reduced through whatever means CWS may employ, but some stigma may be unavoidable if CWS is to achieve its primary function of keeping children safe.

#### *Implications for PRM*

Beyond broad and obvious efforts to decrease stigmatization such as using less intrusive means while still adequately ensuring child safety, Dare makes several suggestions with regard to PRM.

***Must PRM data be more closely held than other data?*** Can stigmatization be lessened if PRM data are closely held within the CWS agency? We quote Dare's recommendation in full here: "The most obvious response to at least some of these concerns about stigmatization is to maintain careful control over the dissemination of the 'product' of the predictive risk model. I suggest that such information should be disseminated as narrowly as possible, consistently with achieving the benefits of the program. This may mean, for instance, that only senior social service professionals should have access to such information, that they be carefully trained as to how to manage the information they possess, so as to limit opportunity for stigmatization of individuals or groups, and that consideration be given to what level of detail is required to make effective use of the model's predictions." (Dare, 2013, p.30-31).

We do not concur with this recommendation. We see no reason to treat PRM data as more ethically "radioactive" than currently existing CWS data. To our knowledge, no such restrictions have ever been in place with regard to risk scores from actuarial instruments. To the degree that PRM scores are more accurate than existing actuarial scores (and they shouldn't be used if they aren't) their use is less ethically troubling than the use of existing data, not more so. We see no ethical reason to change current policy simply because PRM is novel in how it works. There is an affirmative reason not to do as Dare suggests. To the degree that PRM scores provide useful evidence relevant to improved decision making, their availability to well-trained workers should be maximized, not minimized. A later report (Dare & Gambrill, 2016) seems to roll back Dare's (2013) recommendations, suggesting that PRM scores "should be distributed only to those who a) have appropriate training and b) need the information in order to further

#### **Dare and Gambrill's (2016) Training Recommendations:**

- 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

child protection goals...” They further make specific recommendations regarding training for workers using PRM scores (see text box) with which we agree. We see this as a simple reaffirmation of general principles of sound practice when any tool is used, not a new issue raised uniquely by PRM.

**Stigmatization due to confirmation bias.** Arguably not a stigmatization issue per se, Dare (2013) makes two points about confirmation bias – the human tendency to see what we expect to see based on a range of preconceptions, including the PRM score. First, high and misleading PRM scores could encourage workers to wrongly perceive the family as more at risk than they otherwise would. On the other hand, PRM scores may provide a more objective “yardstick” (Dare, 2013, p. 31) than is currently available to work against preconceptions or biases held by the worker. We would add another dimension to this issue, asking how use of PRM compares to existing practice. We see no new issues raised by PRM. Any concerns in this sense already exist relative to current risk assessment measures.

**Stigmatization as a “before the fact” issue.** One area in which we very much agree with Dare (2013) is in recognizing that CWS referrals can be preventative rather than “after the fact.” The risk of stigma is qualitatively different; individuals can feel stigmatized for interacting with CWS around something that has not yet happened. Dare suggests mitigation through consistent emphasis on supportive and preventative worker conduct, rather than punitive conduct. Again, we trace this issue squarely back to the innate tension between a public health paradigm (which this issue clearly represents) and a criminal justice paradigm. We feel that Dare makes a critical point here. PRM (like any risk assessment) does not require prior wrongdoing to have occurred. We strongly suggest that any workers using PRM scores be thoroughly trained on this and are able to work with families in a way that emphasizes this point. We would again note that this represents no new ethical issue – these conditions already exist when any risk assessment tool is used. Workers must always be trained to be aware of internal biases, be sensitive to additional information and be prepared to use clinical judgment to augment tools (Coohey et al, 2012; Detlaff et al, 2011). Again, to the degree that the ethical tool is accurate, the use of that tool becomes less ethically troublesome.

### *Conclusion and Recommendations*

We see very few ways in which PRM provides new ethical challenges with regard to stigma. We make this assertion based on a simple question: “How is PRM data more likely to cause stigma than existing data and procedures?” The only plausible reason we see that PRM could worsen ethical issues with stigma is if workers are badly trained in its use and wrongly hold high PRM scores as dispositive. It is absolutely essential that workers be trained that PRM is a tool which supplements their professional work, not some kind of irrefutable oracle. If workers use PRM scores in the same way they use existing risk assessment scores and other data, and if the PRM scores are more accurate (which they should be if they are to be used at all), then ethical concerns around stigmatization will necessarily be reduced by PRM, not made worse. This is because a more accurate risk assessment system will assure that interventions are delivered more precisely to those who need the services and unnecessary stigmatization will be avoided. This may not reduce the overall experience of stigma *per se*, but it will reduce the proportion of cases in which stigma is unnecessarily brought about, which is of ethical value.

### **Data Access and Data Use: Ethical Considerations**

This section will discuss data access and use using an “inside to outside” approach, beginning with internal CWS data, progressing to data that CWS currently has access to and then considering other data sources. The first two sections are the most comprehensive, as these are most relevant to the California case. We would like to specifically note that we divide the ethics associated with CWS Data into two categories: Access and use. Many ethical concerns about PRM involve the ethics of accessing various data

sources and the degree to which CWS may have a “right” to access these data, or persons may have the “right” for the CWS not to have access to those data. We do not address these questions for one simple reason: For all data sources which CWS currently has access to, an ethical decision has already been made regarding their right to access those data. That decision is unrelated to PRM. What is entirely relevant to PRM is the question of data use. While data access may be ethical, misuse of such data can never be ethical. We therefore focus primarily on the ethical use of data below in relation to PRM.

#### *CWS Internal Data: Access and Use*

Data access and use are different issues. *There are no grounds whatsoever for denying CWS systems access to their own data in an a priori sense.* In the case of California, all the data to be used in the PRM are data which CWS already has an established right to access and is already accessing now. We know of no ethical, legal or practical reason that an agency can’t look at its own data, and many ethical, legal and practical reasons why they must. CWS agencies are expected to use their past experiences with families to inform practice. It would clearly be negligent, for example, for CWS to take action without consideration of all the facts it has available, including past maltreatment reports and information gained in prior field work.

*“if the county 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” (Dare & Gambrill, 2016, p.3).*

In our view, the ethics of the use of CWS data comes down to a single question: Does the use of the data result in improved service? We believe that it is ethically incumbent upon CWS to make the best possible use of their data in maximizing the quality and effectiveness of service they provide. This is obviously challenging, as data can be predictive, it can be simple noise, or it can be misleading. Good ethical practice requires that the data be evaluated in the most accurate way possible, informed by the best available evidence and research. Arbitrary (not evidence-based) use of data is unethical, but not using data can also be unethical. For example, if CWS consistently ignored or failed to act on credible reports of maltreatment and a child later died (something too common nationally), they could be seen as guilty of unethical practice. We would therefore make the following assertion: *The ethical use of internal CWS data is that which promotes the best and most effective practice.* We would also suggest a corollary: *Failing to use internal CWS data in a way which promotes the best and most effective practice is unethical.*

By “best and most effective practice” we mean effective processes and interventions in all phases of CWS practice, from correctly making screening and case opening decisions to providing appropriate preventative services, to making appropriate decisions about removal. These practices should be empirically demonstrable in their superiority through means such as tracking recidivism, placement stability, reunification, etc.

#### *A Methodological Note About Process vs. Performance: Selection of Variables in PRM*

One issue which we feel is likely to arise is the question of including variables which may be useful in improving the accuracy of a PRM model, but may appear arbitrary or even ethically unacceptable as components of the model. Perhaps the most serious example of such an issue is the inclusion of factors that have nothing to do with maltreatment *per se* but may be predictive. For example, should the child’s race be included? In our view, the first step to answering such a question is to ask, “does the inclusion of the variable improve the predictive capability of the PRM model.” We see no ethical downside to omitting variables which are politically or theoretically concerning unless the omission of those variables markedly degrades the performance of the model overall, or for specific subgroups.

A separate question arises for variables that increase the capability of the model to correctly predict risk but seem arbitrary or even oppressive in use. Such variables can carry predictive utility for many reasons, such as standing as proxies for other variables that are not included in the model. In the most troubling possible case, a variable such as Black race might improve the functioning of a PRM model. The model developers will then be forced to balance two competing ethical concerns: Should they include an apparently arbitrary indicator, such as race, which may have a long history of misuse in decision making (e.g., redlining, unequal sentencing, etc.) and accept degradation of the predictive power of the PRM tool with attendant ethically troubling increases in false positives and/or false negatives? Conversely, should they include whatever variables will produce the most accurate, and therefore, most ethically preferable predictions, while accepting the apparent ethical cost of using variables that have been misused in the past? Our view is that as child safety is the overriding concern of CWS, and limiting false positives (unnecessary intrusion into families) is also a central and ethically valuable CWS goal. For this reason, we feel that the “proof is in the pudding” – a tool which serves people better is an ethically preferable tool, assuming the performance of the tool is improved across all subgroups. Degrading the quality of CWS decision making in the service of avoiding inclusion of apparently troubling variables may be politically wise, but ethically questionable. One exception may be when inclusion of a variable is so potentially troubling (as may be the case with including racial group affiliation) that using said variables is simply politically impossible. We would further note that this is a serious concern only in the (unlikely) circumstance that the predictor in question is a powerful predictor which markedly increases the predictive power of the model as a whole. When no large increase in predictive ability is present, removing the variable becomes less troubling.

#### CWS Internal Data: Specific Issues

There are a number of potential issues regarding how internal CWS data should or should not be used.

**Prior case data as predictive.** While there are not ethical bars to accessing internal agency data, there are substantial ethical reasons supporting such access. It is not ethical for an agency to make decisions not fully informed by the available data. As previously discussed, the best available predictor of child risk is the child’s prior history of CWS involvement, including both substantiated and unsubstantiated cases. A strong case can be made that not using such data is unethical and harmful to children. A recent case in point can be found in Chicago, where, following several fatalities, the Director of the Illinois Department of Child and Family Services is considering making previously unavailable data about unsubstantiated cases available to workers to allow workers to form a more complete picture of the case (Jackson, Marx & Eldeib, 2017).

There are some who believe that allowing a worker to see prior case files may somehow bias the worker, and that it may be more “fair” for the worker to “start fresh” without preconceptions of the family. This contrasts starkly with the established utility of prior reports in predicting future risk, a known value relative to child safety (CECANF, 2016; Coohy et al, 2012; Jackson et al, 2017).

**Prior case data as necessary to good practice.** Child welfare workers, particularly hotline screeners and emergency response workers, necessarily work on the basis of limited information. Frequently, the key limiting factor in the ability of a worker to do his job is the simple lack of data. Many times, a clinical picture cannot be formed on the basis of current data only. Many child welfare investigations involve preverbal or nonverbal children, or include persons, from the victim to family members to the alleged perpetrator, who are not forthcoming. In such circumstances, the ability of the worker to make a valid practice judgment (such as opening a case or not) is severely limited and practice effectiveness is compromised. Lacking access to prior case data will further compromise the worker’s situational

awareness and degrade performance as patterns over time cannot be recognized and considered. Prior case data can be cause for further concern or can be exculpatory. For example, prior case reports showing histories of harassment, or showing conclusive negative findings from a prior report can work to clarify the situation in the favor of the suspected perpetrator. These factors aside, children have a moral right to expect CWS agencies to work in their best interests, especially their safety interests, based on the totality of data available to the CWS agency.

We would here repeat a point made by Dare (2013) with regard to the use of PRM scores. If potential for misuse of those scores is possible, then workers must be trained to use the scores correctly (e.g., not as dispositive). Precisely the same point applies to the use of prior case data. Any case data can be used irresponsibly, but it also has great value. Workers must be trained specifically around avoiding bad practices in use of prior data, such as assuming that any family with a previously substantiated case is currently maltreating their child. Precisely the same principle applies: *Workers must be trained to make the best possible professional use of all information available.*

#### CWS Currently Available External Data: Specific Issues

*We note that these data are not being considered for use in California at the present time. This section is therefore not directly relevant to the currently envisioned PRM tasks.*

All CWS agencies currently have access to data not held directly by the CWS agency. Examples include law enforcement data or some kinds of health data in some areas. Some data sources like law enforcement may have data which is publicly available to anyone (not just CWS), as well as data which they will not make publicly available but will share only with CWS. For example, arrest data are often public domain, but CWS agencies may have access to more data through law enforcement information systems than the general public has. In addition, all CWS agencies have access to public geospatial data, such as socioeconomic characteristics of given neighborhoods (tracts, zip codes, cities, counties) through the Census, although these data are rarely, if ever, used. Some CWS agencies have agreements with agencies holding health data or vital statistics data to share those data to inform case planning.

**Accessing Currently Available External Data.** Again, there is no ethical question with regard to accessing these data in cases in which it has already been established that CWS should have access for the purposes of case and service planning. No new ground is being broken. Of course, that is not to say that all of these data sources which CWS can access are being accessed. We are in a transitional period in our nation's history with regard to data. Legacy systems often involve CWS workers requesting data from a given source (e.g., law enforcement) which may then arrive several hours or days later. Perhaps more commonly, CWS simply does not access data that they have permission to access.

We argue here that the failure of any CWS to use available data can be unethical. For example, if an emergency response worker is not supplied with the information that the alleged perpetrator is listed on the California Sex Offender Registry (many of whom have not had prior CWS contact), this clearly represents bad practice and a threat to the child.

There are reasons that CWS workers are not currently given needed information in a timely fashion. The most important reason is that CWS data systems are generally legacy systems. Current data systems were not designed to take advantage of cross-linkable databases providing live "as needed" data. Such a system requires new programming, may involve hardware purchases and therefore represents a capital investment and necessitates changes in procedures.

From an ethical perspective, it remains undesirable for a child to have his or her safety compromised due to the CWS agency's inability to use available data.

**Use of Currently Available External Data.** Precisely the same points apply as were made in the previous section. We argue that since CWS already has the right to access data pertinent to case decision-making regarding child safety, the only pertinent ethical question is if the use of such data will improve practice. Again, this is not a philosophical question, but a scientific, and more specifically, a statistical one. To the degree that the use of currently available external data can improve the accuracy of PRM systems, the use of those data is ethically sound. Should the data not improve the accuracy of PRM predictions, the data should not be used. Again, this can be estimated prior to implementation, through inclusion of cross-sector data as PRM variables and estimation of improvements in estimates against some future benchmark, such as re-reporting.

#### Data Currently Unavailable to CWS: Specific Issues

We include this brief section merely because not including some recognition of the issue seems an unwarranted omission. It is entirely possible that data exist which are currently not available to CWS, but could dramatically improve service delivery. There may be ethical issues involved in obtaining access to these data. For example, medical records could be extremely valuable in assessing risk to children. An Emergency Response worker who has access to prior emergency room visit data on a preverbal child with a suspicious bruise may be in a much better position to make a correct assessment. Determining if access to these databases is ethical or not requires weighing the benefits of such access (enhanced child safety) against drawbacks inherent in such access (violation of privacy concerns, possibilities that families will fail to access needed but observed services). Such an analysis is beyond the scope of this paper, as the current issue at hand relates to referral-level modeling using internal CWS data only. It should be noted, however, that prior ethical reviews of universal-level risk stratification (e.g., Dare, 2013) must confront this issue. Dare (2013) believes that such population-based risk stratification "is ethically justified," provided that specific recommendations he forwards (many of which are cited above) are followed. In the United States, Birth Match programs constitute another example where the use of data otherwise unavailable to CWS has been judged to be ethical (CECANF, 2016; Shaw et al., 2013). In comparison, the ethical issues confronted in California are of a much smaller magnitude.

We would also touch upon a series of specific outstanding questions that we believe should be addressed in any comprehensive review of the ethics of PRM in CWS.

#### **Could PRM Cause Unfair Treatment of Particular Subpopulations?**

We have consciously avoided this issue until the current time, as it bears on all aspects of our report, and raising it each time it was appropriate would have been onerously repetitive. We consider this one of the key ethical considerations in PRM and it deserves specific attention. The number of subpopulations which could be parsed from the CWS population easily numbers in the hundreds – you could cut by family structure, economic status, area of residence, national origin, or many other factors. The single kind of subpopulation most likely to garner attention however, is race.

One valid ethical concern regarding PRM is that it may not work well for all populations. For example, a given PRM could have a high predictive ability as measured by a high Area Under the Curve (AUC) for the entire population, but it may have a much lower AUC for specific subpopulations. This could be rephrased as "PRM might not have the same predictive accuracy for some groups." While this is a serious concern, a much more serious concern also exists. The ratio of false positives and false negatives may vary by group. These concerns could be rephrased as "PRM might consistently overestimate risk for, and therefore be

biased against, group X” or “PRM might consistently underestimate risk for, and therefore fail to protect, group X.”

#### Conclusion and Recommendations

In our view, any responsible use of PRM will “build in” a test of any possible racial or ethnic bias from the ground up. The good news is that this is entirely feasible. We use the same reasoning as we used above in describing how PRM systems could be tested for predictive accuracy. The only difference is that concerns about racial bias in PRM systems can be tested by repeating those same tests and omitting or including particular variables (such as race or other variables considered to introduce racial bias) or by running the PRM systems on samples from different racial groups and determining if the PRM system is equally accurate across groups.

In any ongoing testing, it would be interesting to compare screening or other decisions which are consistent with the PRM and not consistent (i.e. overrides). Put in simple language, it is not sufficient to worry that the PRM system might perform differently for different racial/ethnic groups. It is necessary to test the PRM system against some outcome benchmark(s) such as re-report or later placement, and compare the characteristics of the Receiver Operating Curve for each group.

Fortunately, this can be done before the PRM system ever goes live. These tests can and should be done in advance of any introduction of a PRM system. In our view, worry about racial bias can be satisfactorily addressed using empirical means. The system either systematically misclassifies individuals of a given race / ethnicity or it does not. The numbers will tell. Should analyses show that the system works better or less well for a given race/ethnicity, the system will need to be adjusted so that such biases are eliminated.

Implicit bias, including racial bias and bias against the poor, is a problem across all cultural and institutional systems. In the Child Welfare System, disparities are primarily driven by differences in socioeconomic status.

#### **Is it ethical to identify situations for which no effective remedy is available or feasible?**

We mention this issue only because it has been commonly debated in the past (Dare, 2013; Dare & Gambrill, 2017). This issue, however, is not relevant to referral-based risk stratification. It is only relevant to population-based risk stratification when new cases might be generated by the CWS. While it is beyond the scope of the present paper, there are similar issues that appear in public health surveillance of disease for which there is no cure at a given time (e.g., past history with HIV, Zika, etc.) that provides data that then informs the ability to serve that population (Dondero, Pappaioanou & Curran, 1988). What is proposed in California does not involve the identification of new situations, just the more effective response to already identified situations.

#### **Is it permissible to employ variables in PRM without the consent of the involved persons?**

This is a question of data access. If the California PRM system were to use data to which CWS currently does not have access, then the ethics of accessing those data, including an analysis of ethical costs and benefits would be warranted. Similar assessments made in New Zealand and Allegheny County have concluded that such access is ethical. In the present case of PRM in California, however, the question does not arise, as CWS already has access to all data intended for use.

### **Is it ethically permissible to use data from “peripheral” persons?**

Sometimes data relates to people other than suspected victims and perpetrators. We would simply reiterate the above paragraph – no new data access is planned, and only data which are currently available to CWS are going to be used. In this case, as in the prior, the data will simply be used in a more structured and statistically predictive way.

### **What ethical limits exist relative to the use of PRM for other likely “use cases?”**

We do not mean to limit our review specifically to screen-in and case opening decisions *per se*. The principles covered in this paper can be applied more broadly. There are other potential uses that PRM data could be used for. For example, a supervisor assigning cases to workers might assign a case with a higher PRM score to a more experienced worker, or, alternately, could use PRM scores to “balance” the difficulty of cases each worker is assigned. *We see no ethical reasons that these practices should be suspect so long as they represent continuations of existing practices using already ethically available and more accurate data.* In our view, any new PRM-related policy or practice which satisfies the above conditions is ethically acceptable, and, due to increased accuracy of PRM data, should be considered ethically preferable in most cases. To the degree that PRM data would be used in new policies, practices or procedures, then the ethical acceptability of those policies, practices, or procedures should be evaluated in the same manner that CWS would (and has in the past) evaluated the ethics of any change in practice, policy or procedure.

### **Is the relative “lack of understandability” of PRM ethically problematic?**

The complex statistical nature of PRM could be argued to be “non-transparent” or “hard to understand.” There could be concerns that such an opaque process would not be open to scrutiny by individuals seeking to challenge a decision in which PRM was used. There are two reasons why we feel that this is not a compelling concern.

First, PRM supplements worker judgment. In order for PRM to be ethically undesirable in this context, it would have to be less “transparent” than worker decision making. This is not the case. Understanding why any person makes a particular decision is conjectural at best and open to interpretation. PRM models are mathematical, and even though they may be complex, they are inherently logical and explainable—something human judgment is not. From this perspective, PRM makes practice and decision making more, not less, understandable.

In the case of California, PRM is intended as a tool to assist in decision making at the level of referral screening. This decision is representative of the “public health” function of CWS described above. Screening is not a judgment of “guilt,” and carries no formal penalty. There is no intended use for PRM in California for substantiation or custody decisions, which are, in fact, judgments regarding what has happened. These more “criminal justice – like” decisions (e.g., substantiation, removal) should clearly never be made solely on the basis of a risk score which, although useful at a population level, is far from infallible at the individual level.

### **Is it ethically permissible to include “opt-in” data in addition to universal data in PRM?**

One of the more interesting questions regarding PRM, which is raised both in Dare (2013) and Dare and Gambrill (2016) is the issue of universal vs. “opt-in” data sources. There are practical and theoretical concerns.

Universal data (e.g., birth records) are available for everyone, and so, theoretically, everyone shares equally in their exposure and possible violations of privacy. From a practical perspective, the PRM model could include all variables for all persons (absent issues like missing data due to people recently moving to the state).

“Opt-in” data are not available for everyone. Only people who have applied for Temporary Assistance to Needy Families (TANF), for example, are in the TANF database. This generates ethical concerns. Is it ethically sound to use data that is only available on some individuals? Is it problematic when, by definition, only disadvantaged people (the poor) are in such databases? Is it possible that people might forego needed services to remain “off the grid” and thus protect their own privacy relative to that data source? The ethical reviews for both New Zealand and Allegheny County evaluated these concerns and came to the conclusion that the utility of these databases in predicting, and therefore mitigating risk exceeded any potential ethical cost of accessing such databases. In any case, these concerns are just that, concerns. There is no current evidence, to our knowledge, that the use of PRMs in child welfare have changed or will change individual behaviors in the ways hypothesized above.

With regard to the current proposed use of PRM in California, this is again a non-issue because California does not intend to use any data besides CWS data, to which California already has an established ethical right to access.

### **Is it ethical to use data for a reason other than that for which they were originally collected?**

This is a familiar issue in the context of academic research. Principles of informed consent require that research participants understand the nature of their involvement, including the potential uses of their data. This issue is explicitly excluded by human subjects protection frameworks (FERPA, IRB review) regarding an agency’s internal use of their own data.

Besides this, the issue simply does not apply to CWS internal data for two reasons. First, CWS exists to protect children and serve families. By definition, all data they collect, meaning all internal data, were collected for the purpose that PRM serves – to serve children and families at risk of maltreatment. The second reason that the “use” concern does not apply to CWS data is that there is no reason or basis for believing it should, and that the use of held data to improve practice is not only ethically permissible, but is arguably ethically compulsory.

### **Are there ethical concerns regarding the person or system the risk score is attached to?**

Risk scores which include information about an entire family situation are not “about” the child in isolation, nor are they “about” a parent or perpetrator or family in isolation. The idea that a score is “attached” to a given person is misleading. For example, let’s assume that a child receives a given score and is then placed in foster care. Clearly that score is no longer operative or useful. Any PRM score generated is clearly bound to the situation (which includes the whole ecosystem involved) and equally importantly, the time the score is generated. We feel it is a mistake to consider a score as “attached” to any person. *The score is situationally and temporally specific.* Is it ethical to know or consider that an individual of any kind (child, parent, perpetrator, etc...) was involved in a situation that developed a given risk score? We believe that it is, and we believe that it would be ethically questionable for the CWS agency or worker to be intentionally self-blinded to potentially useful information.

## Can “Bright Line” applications of data be ethical?

The previously discussed recommendation of the Federal fatalities commission represents a simple policy example of a “bright line” test based on risk. The commission suggests that all very young children be screened-in because they are far more physically vulnerable than older children. We can conceive of no reason why such a policy would be unethical, unless it is arbitrary and not representative of empirically determined reality. In principle, therefore, bright line tests can be ethical. In principle, we see no reason why “bright line” tests based on a PRM score, which may be far more accurate than simple consideration of a child’s age, cannot be. However, we respect prior thought and discussion on this matter (e.g., Dare and Gambrell, 2016) and would suggest that “work arounds” such as providing the worker the ability to not follow the “bright line” policy, perhaps with supervisory consent, be put in place to meet the needs of atypical situations.

## If CWS does not optimally use what it knows to help families, is that ethically undesirable?

We believe this question answers itself. CWS carries a responsibility to provide the best possible service to families, including both parents and children. Should CWS have information on a child and not use that information to the benefit of the child and family, then that may constitute negligence. PRM systems can help with a part of this problem. PRM scores are automatically generated and will not “overlook” fields they are programmed to access, and will not be subject to human variances in judgment. They do not, of course, provide any help when “what CWS knows” is not available in a format accessible to PRM (e.g., is recorded in case narrative) or when the PRM system is not programmed to access those data. One could easily consider a case where decisions are made after the worker has failed to note a critical piece of prior information from the case record. Such decisions could be ill-informed and catastrophic. PRM can provide a “backup” to such situations, with the worker at least being aware that the PRM score on the family is high, even if the worker has failed to note key prior case characteristics.

## A Note About Transparency

The CWS is a public institution, mandated by, “owned by,” operated on behalf of, and accountable to the people and their representatives. This confers a tremendous potential ethical advantage. Any PRM models used by the state should absolutely be completely open to scrutiny by any member of the public. Private individual data, of course, cannot be, but the mechanism must be. In this way, transparency is not only assured but arguably, vastly improved in comparison to current practice. It is much more simple to explain how a PRM generates a risk score than it is to explain how a human being generates a perception of risk. It is easier to explain how a PRM generates a risk score than to explain exactly how and why a human uses that risk score in the way he or she does.

There are a number of reasons for this. First of all, much of the concern about CPS decision making has to do with matters of *implicit bias*. By definition, it is impossible for a person to report how their decisions have been impacted by implicit bias. Secondly, recall is very problematic. Workers may misremember cases, may confuse similar cases, and are subject to a number of psychological phenomena, such as the tendency to alter perception or recall in pursuit of closure (Dijksterhuis et al., 1996). Third, decision making is often made on the basis of intuition, clinical judgment, accumulated prior experience or “practice wisdom”. Again, these decision making processes are, *a priori*, not subject to precise description by the decision-maker. It is important to note that all of the above concerns represent possibilities for distortion *which are made in good faith*. It is also possible, of course, for recall to be effected by intentional efforts to shade or misreport the truth for personal advantage. Mathematical models have none of these weaknesses.

We believe that having the PRM algorithm open to public scrutiny not only advances the ethical value of transparency to the people being served, but also has the more diffuse and abstract ethical benefit of informing the development of knowledge, allowing replication in other states and service to children and families beyond the borders of California. We would further like to note that we strongly support the decision of the State of California not to employ a private firm to create these models. While there is nothing innately wrong with a private agency doing predictive work of this type for a state, it is likely that any such firm would refuse to allow full and easy public scrutiny of its (proprietary) modeling. This would compromise transparency, which is ethically undesirable. In addition, the presence of a profit motive could be a competing factor in any decisions made by the hired agency with regard to the model. This latter point includes the potential cost burden associated with making changes to a proprietary system if the algorithm needs to be altered to account for population or policy shifts. By opting not to use a private firm to establish the PRMs that may be used, California has clearly sidestepped this entire set of ethical concerns.

### A Note About “Feedback Loops”

There has been recent public concern about the presence of feedback loops in predictive analytics (O’Neil, 2016). Say a person has financial difficulties and then is denied credit and this causes the person to have further financial difficulties. In this way, a decision made by any algorithm (subjective human or objective mechanical) can have a tendency to cause the same decision to be more likely in the future. This could theoretically happen with the use of PRMs in screening. If a screened in prior report is predictive (and prior work suggests it will be) then the tendency of the machine to generate a higher risk score and cause a report to be screened in may “feed back” and make the next report more likely to be screened in. While this is a potential concern, *it is in no way a new concern, nor is it specific to machine algorithms*. This same dynamic could just as easily exist with human decision making. The primary difference between the two algorithms (subjective human and objective machine) goes back to the prior point about transparency. First, you can show why a PRM generated a specific risk score. That is harder with human beings. Second, a machine-generated risk score allows the presence of feedback loops to be mathematically evaluated. For example, if a given parameter falsely inflates the probability of a future screen-in, then that same effect should *not show up* when non-CWS generated outcomes are used (like future hotline calls) and such effects should be minimized when other agencies or actors are involved (such as in the decision to place a child, in which a judge makes the decision). By using non-CWS generated outcome measures such as these as triangulation points, meaningfully large feedback loops within the CWS should be identifiable. Rather than increasing the problems potentially associated with feedback loops, PRM systems increase our ability to see and empirically evaluate such possible looping. This is ethically desirable. Of course, this requires vigilance on the part of the individuals monitoring the PRM and requires intentional and ongoing analyses relative to this specific potential problem.

## Conclusion

We find PRM to be like any new technology. It can be used properly or improperly. We find that there are no *a priori* ethical barriers to utilizing existing CWS data in the PRM context. With regard to the ethics of data access, CWS already has access to its own internal data, and existing empirical studies suggest that those internal data have great value in improving CWS decision making.

In general, an effective PRM-based system which uses data to which CWS already has access and which outperforms existing decision-making is not only ethically permissible but is ethically desirable. CWS has an ethical obligation to use the best tools possible in case decision-making. However, the potential for unethical use of a PRM system does exist. We concur with prior authors (e.g., Dare, 2013) in recommending a series of safeguards to promote the ethical use of PRM in CWS. The specific recommendations can be found above, but fall in to these major categories:

- 1) Workers must be trained how to use PRM outputs effectively and ethically. In particular, workers must be trained to avoid confirmation bias and to realize that predictive models, even when highly predictive at the population level, may or may not be predictive at the individual level.
- 2) PRM does not replace worker judgment. It is a tool for trained workers to use and must be regarded as such; and any outputs from any PRM system should be subject to workers' ability to override those recommendations, perhaps with supervisory sign-off.
- 3) Any PRM system should be validated using historical data and then tested on live data prior to implementation. A PRM system must be shown through appropriate empirical means to outperform existing systems in terms of generation of true positives and true negatives. This can be done by running a PRM on prior cases and comparing the correctness of PRM decisions to field decisions relative to an appropriate outcome variable such as re-report or subsequent placement.
- 4) Concerns about PRM systems being biased relative to specific subpopulations should be addressed using the methods described above and ongoing monitoring. This should happen prior to any implementation of the system.
- 5) Evaluation of the PRM relative to the above concerns (3 &4) must be ongoing. This is not a "one shot" analytic task.
- 6) California's choice *not* to hire a private firm to establish and run the predictive models has potential ethical benefits with regard to transparency.

## References

- Allegheny County Department of Human Services. (2016). *Using predictive modeling to improve outcomes for children in Allegheny County*. Available online at: <https://www.nga.org/files/live/sites/NGA/files/pdf/2016/1607ChildSafety20-Dalton.pdf>
- Allegheny County Department of Human Services. (2017). *Ethical analysis: Predictive risk models in call screening in Allegheny County*. Available online at: <https://www.nga.org/files/live/sites/NGA/files/pdf/2016/1607ChildSafety20-Dalton.pdf>
- Baird, C., & Wagner, D. (2000). The relative validity of actuarial-and consensus-based risk assessment systems. *Children and Youth Services Review, 22*(11), 839-871.
- Ballard, E. D., Van Eck, K., Musci, R. J., Hart, S. R., Storr, C. L., Breslau, N., & Wilcox, H. C. (2015). Latent classes of childhood trauma exposure predict the development of behavioral health outcomes in adolescence and young adulthood. *Psychological Medicine, 45*(15), 3305-3316.
- Bartholet, E. (2015). Thoughts on the liberal dilemma in child welfare reform. *Wm. & Mary Bill Rts. J., 24*, 725.
- Belsky, J. (1980). Child maltreatment: An ecological integration. *American Psychologist, 35*(4), 320.
- Benbenishty, R., Segev, D., & Surkis, T. (2002). Information-search and decision-making by professionals and nonprofessionals in cases of alleged child-abuse and maltreatment. *Journal of Social Service Research, 28*(3), 1-18.
- Broccolo, B., Chen, J., Geetter, J., Mazur, L. S., Parisi, L., & Rutherford, C. (2017, Feb.). HHS finalizes toned-down version of common rule overhaul. *McDermott, Will, & Emery*. Available online at: <https://www.mwe.com/en/thought-leadership/publications/2017/02/hhs-finalizes-common-rule-overhaul>
- Cancian, M., Yang, M. Y., & Slack, K. S. (2013). The effect of additional child support income on the risk of child maltreatment. *Social Service Review, 87*(3), 417-437.
- Center for the Study of Social Policy. (2011). *Disparities and disproportionality in child welfare: Analysis of the research*. Available online at: <https://www.cssp.org/publications/child-welfare/alliance/Disparities-and-Disproportionality-in-Child-Welfare-An-Analysis-of-the-Research-December-2011.pdf>
- Cohen, M. (2015). Birth Match, tool to prevent abuse and neglect, slow to spread in US. *Chronicle of Social Change*, Nov. Available online at: <https://chronicleofsocialchange.org/blogger-co-op/birth-match-tool-prevent-abuse-neglect-slow-spread-u-s>
- Commission to Eliminate Child Abuse and Neglect Fatalities (CECANF). (2016). *Within our reach: A national strategy to eliminate child abuse and neglect fatalities*. Washington, DC: Government Printing Office.
- Coohy, C., Johnson, K., Renner, L. M., & Easton, S. D. (2013). Actuarial risk assessment in child protective services: Construction methodology and performance criteria. *Children and Youth Services Review, 35*(1), 151-161.
- Dare, T., Gambrell, E. (2017). *Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County*. Available online at: <http://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf>
- Dare, T., (2013). Predictive risk modeling and child maltreatment: An ethical review.
- Los Angeles County Department of Children and Family Services (DCFS). (n.d.). Los Angeles County Department of Children and Family Services: 2015-2016 DCFS biennial report. Available online at: [http://dcfs.lacounty.gov/Release/2015-2016BiennialReport\\_r2.pdf](http://dcfs.lacounty.gov/Release/2015-2016BiennialReport_r2.pdf)
- Los Angeles County Department of Children and Family Services (DCFS). (n.d.). Los Angeles County Department of Children and Family Services: About Us. Available online at: <http://dcfs.lacounty.gov/aboutus/index.html>
- de Haan, I., & Connolly, M. (2014). Another Pandora's box? Some pros and cons of predictive risk modeling. *Children and Youth Services Review, 47*, 86-91.

- Dettlaff, A. J., Rivaux, S. L., Baumann, D. J., Fluke, J. D., Rycraft, J. R., & James, J. (2011). Disentangling substantiation: The influence of race, income, and risk on the substantiation decision in child welfare. *Children and Youth Services Review, 33*(9), 1630-1637.
- Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning, 12*(1), 17-26.
- Dijksterhuis, A. P., Van Knippenberg, A. D., Kruglanski, A. W., & Schaper, C. (1996). Motivated social cognition: Need for closure effects on memory and judgment. *Journal of Experimental Social Psychology, 32*(3), 254-270.
- Dondero Jr, T. J., Pappaioanou, M., & Curran, J. W. (1988). Monitoring the levels and trends of HIV infection: The Public Health Service's HIV surveillance program. *Public Health Reports, 103*(3), 213.
- Dorsey, S., Mustillo, S. A., Farmer, E. M., & Elbogen, E. (2008). Caseworker assessments of risk for recurrent maltreatment: Association with case-specific risk factors and re-reports. *Child Abuse & Neglect, 32*(3), 377-391.
- Doyle, J., Kao, G. (2007). Are racial identities of multiracials stable? Changing self-identification among single and multiple-race individuals. *Social Psychology Quarterly, 70*(4), 405-423.
- Drake, B. (2013). Differential Response: What to Make of the Existing Research? A Response to Hughes et al. *Research on Social Work Practice, 23*(5), 539-544.
- Drake, B., & Jonson-Reid, M. (2011). NIS interpretations: Race and the national incidence studies of child abuse and neglect. *Children and Youth Services Review, 33*(1), 16-20.
- Drake, B., Jonson-Reid, M., Way, I., & Chung, S. (2003). Substantiation and recidivism. *Child Maltreatment, 8*(4), 248-260.
- Drake, B., Jolley, J. M., Lanier, P., Fluke, J., Barth, R. P., & Jonson-Reid, M. (2011). Racial bias in child protection? A comparison of competing explanations using national data. *Pediatrics, 127*(3), 471-478.
- Finno-Velasquez, M., Palmer, L., Prindle, J., & Putnam-Hornstein, E. (2017). A birth cohort study of Asian and Pacific Islander children reported for abuse or neglect by maternal nativity and ethnic origin. *Child Abuse & Neglect, 72*, 54-65.
- Fluke, J. (2011). Decision-making in child protection and child welfare: Some considerations for information technology. *Child Welfare 360*, Spring, 8-9.
- Fox, K. (2003). Collecting data on the abuse and neglect of American Indian Children. *Child Welfare, 82*(6), 707-726.
- Gambrill, E., & Shlonsky, A. (2000). Risk assessment in context. *Children and Youth Services Review, 22*(11-12), 813-837.
- 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 & Family Social Work, 16*(4), 412-421.
- Gillingham, P. (2015). Predictive risk modelling to prevent child maltreatment and other adverse outcomes for service users: Inside the 'black box' of machine learning. *The British Journal of Social Work, 46*(4), 1044-1058.
- Harrington, C. (2017). *New ministry set to make same mistakes as CYF*. Newsroom. Available online at: <https://www.newsroom.co.nz/@future-learning/2017/04/11/18926/new-ministrys-risk-based-approach-could-fall-short>.
- Hill, R. B. (2006). *Synthesis of research on disproportionality in child welfare: An update*. Casey-CSSP Alliance for Racial Equity in the Child Welfare System.
- Hussey, J. M., Marshall, J. M., English, D. J., Knight, E. D., Lau, A. S., Dubowitz, H., & Kotch, J. B. (2005). Defining maltreatment according to substantiation: Distinction without a difference? *Child Abuse & Neglect, 29*(5), 479-492.

- Jackson, D., Marx, G., & Eldeib, D. (2017). Illinois child welfare head pledges reform after recent deaths of children. *Chicago Tribune*, July 7, 2017. Available online at: <http://www.chicagotribune.com/news/watchdog/ct-dcfs-sheldon-met-20170516-story.html>
- Jalongo, M. R. (2006). The story of Mary Ellen Wilson: Tracing the origins of child protection in America. *Early Childhood Education Journal*, 34(1), 1-4.
- Johnson-Motoyama, M., Putnam-Hornstein, E., Dettlaff, A. J., Zhao, K., Finno-Velasquez, M., & Needell, B. (2015). Disparities in reported and substantiated infant maltreatment by maternal Hispanic origin and nativity: A birth cohort study. *Maternal and Child Health Journal*, 19(5), 958-968.
- Jonson-Reid, M., Drake, B., Weigensberg, E., Stagner, M., Collier, S., LeBarron, S., & Read, L. (2016, Oct.). *Improving understanding of child maltreatment incidence: A review of data, measures, and methods*. Report for U.S. Department of Health and Human Services Administration for Children and Families (Contract #: HHS P233201500035I).
- Jonson-Reid, M., & Wildeman, E. (2017). Trauma and very young children. *Child and Adolescent Psychiatric Clinics of North America*, 26(3), 477-490.
- Jonson-Reid, M., Drake, B., & Kohl, P. L. (2009). Is the overrepresentation of the poor in child welfare caseloads due to bias or need? *Children and Youth Services Review*, 31(3), 422-427.
- Jonson-Reid, M., Drake, B., & Zhou, P. (2013). Neglect subtypes, race, and poverty: Individual, family, and service characteristics. *Child Maltreatment*, 18(1), 30-41.
- Jonson-Reid, M., & Drake, B. (2018). *After the cradle falls: What child abuse is, how we respond to it and what you can do about it*. Oxford Press.
- 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.
- Kempe, C. H., Silverman, F. N., Steele, B. F., & Droegmueller, W., Silver, H. K. (1962). The battered child syndrome. *Journal of the American Medical Association*, 181, 17-24.
- Kennedy, S., McDonald, J. T., & Biddle, N. (2006). *The healthy immigrant effect and immigrant selection: Evidence from four countries*. (SEDAP Research Paper #164). Available online: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=7599A7E9DA0BA07C273F77C52A5D4346?doi=10.1.1.407.7613&rep=rep1&type=pdf>
- Kim, H., Drake, B., & Jonson-Reid, M. (2018). An examination of class-based visibility bias in national child maltreatment reporting. *Children and Youth Services Review*, 85, 165-173.
- Knott, T., & Giwa, S. (2012). African American disproportionality within CPS and disparate access to support services: Review and critical analysis of the literature. *Residential Treatment for Children & Youth*, 29(3), 219-230.
- Kohl, P. L., Jonson-Reid, M., & Drake, B. (2009). Time to leave substantiation behind: Findings from a national probability study. *Child Maltreatment*, 14(1), 17-26.
- Mallon, G., Hess, P. (2014). *Child welfare for the 21<sup>st</sup> century: A handbook of practices, policies and programs* (2<sup>nd</sup> ed.). Columbia University Press, New York.
- Marshall, D., & English, D. (1999). Survival analysis of risk factors for recidivism in child abuse and neglect. *Child Maltreatment*, 4(4), 287-296.
- Millett, L. S. (2016). The healthy immigrant paradox and child maltreatment: A systematic review. *Journal of Immigrant and Minority Health*, 18(5), 1199-1215.
- Morton, C. M. (2013). The moderating effect of substance abuse service accessibility on the relationship between child maltreatment and neighborhood alcohol availability. *Children and Youth Services Review*, 35(12), 1933-1940.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Books, New York.
- Olson, M. E., Diekema, D., Elliott, B. A., & Renier, C. M. (2010). Impact of income and income inequality on infant health outcomes in the United States. *Pediatrics*, 126(6), 1165-1173.

- Pelton, L. H. (2015). The continuing role of material factors in child maltreatment and placement. *Child Abuse & Neglect*, 41, 30-39.
- Putnam-Hornstein, E., Needell, B., King, B., & Johnson-Motoyama, M. (2013). Racial and ethnic disparities: A population-based examination of risk factors for involvement with child protective services. *Child Abuse & Neglect*, 37(1), 33-46.
- Ruscio, J. (1998). Information integration in child welfare cases: An introduction to statistical decision making. *Child Maltreatment*, 3(2), 143-156.
- Rycus, J., Hughes, R. (2003). *Issues in risk assessment in child protective services: A policy white paper*. North American Resource Center for Child Welfare Policy, Columbus, Ohio.
- Sackett, D. L. (1997, February). Evidence-based medicine. In *Seminars in perinatology* (Vol. 21, No. 1, pp. 3-5). WB Saunders.
- Sedlack, A., & Broadhurst, D. (1996). *Third National Incidence Study of Child Abuse and Neglect: Final Report*. Washington DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Sedlak, A. J., Mettenburg, J., Basena, M., Petta, I., McPherson, K., Greene, A., & Li, S. (2010). *Fourth National Incidence Study of Child Abuse and Neglect (NIS-4): Report to congress*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Sedlak, A., McPherson, K., & Das, B. (2010). *Supplementary analyses of race differences in child maltreatment rates in the NIS-4*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Shaw, T. V., Barth, R. P., Mattingly, J., Ayer, D., & Berry, S. (2013). Child welfare birth match: Timely use of child welfare administrative data to protect newborns. *Journal of Public Child Welfare*, 7, 217-234.
- Shlonsky, A., & 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(4), 409-427.
- Siegel, E. (2013). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die*. Wiley & Sons, Hoboken, New Jersey, USA.
- Sledjeski, E. M., Dierker, L. C., Brigham, R., & Breslin, E. (2008). The use of risk assessment to predict recurrent maltreatment: A classification and regression tree analysis (CART). *Prevention Science*, 9(1), 28-37.
- Social Explorer. (2017). Data downloaded from 5 year (2011-2015) ACS estimate, Field SE:T58, "Median Household Income by Race (in 2015 Inflation Adjusted dollars)". Available online at: [https://www.socialexplorer.com/tables/ACS2015\\_5yr/R11430968](https://www.socialexplorer.com/tables/ACS2015_5yr/R11430968)
- U.S. Department of Health And Human Services (US DHHS), Administration for Children and Families, Children's Bureau. (n.d.). *The Child Abuse Prevention and Treatment Act*. Available online at: <https://www.acf.hhs.gov/sites/default/files/cb/capta2010.pdf>
- U.S. Department of Health And Human Services (US DHHS), Administration for Children and Families, Children's Bureau. (2015, Sept.). *Adoption and Foster Care Analysis and Reporting System (AFCARS)*. Washington DC: Author.
- U.S. Department of Health And Human Services (US DHHS), Administration for Children and Families, Children's Bureau. (2017). *Child Maltreatment 2015*. Administration for Children, Youth and Families, Children's Bureau. Available online at: <https://www.acf.hhs.gov/sites/default/files/cb/cm2015.pdf>
- Van Der Put, C., Assink, M., & Stams, J. (2016). The effectiveness of risk assessment methods: Commentary on "Deciding on child maltreatment: A literature review on methods that improve decision-making." *Child Abuse and Neglect*, 59, 128-129.
- Vaithianahan, R., Jiang, N., Maloney, T., Nand, P., & Putnam-Hornstein, E. (2017). *Developing predictive risk models to support child maltreatment hotline screening decisions*. Available online at:

<http://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf>.

- Vaithianathan, R., Maloney, T., Jiang, N., De Haan, I., Dale, C., Putnam-Hornstein, E., Dare, T., & Thompson, D. (2012). *Vulnerable children: Can administrative data be used to identify children at risk of adverse outcomes?* Report Prepared for the Ministry of Social Development. Auckland: Centre for Applied Research in Economics (CARE), Department of Economics, University of Auckland. Available online at: <http://www.msd.govt.nz/about-msd-and-our-work/publications-resources/research/vulnerable-children/>
- Waldfoegel, J. (2009). Prevention and the child protection system. *The Future of Children*, 19(2), 195-210.
- Wayman, J. C., Cho, V., & Johnston, M. T. (2007). *The data-informed district: A district-wide evaluation of data use in the Natrona County School District*. Available online at: [http://s3.amazonaws.com/academia.edu.documents/30511497/wayman\\_data\\_use\\_evaluation.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1501254555&Signature=0u8TyhYlqy%2Fa2TC1xh4Bmzt2Of8%3D&response-content-disposition=inline%3B%20filename%3DThe\\_data-informed\\_district\\_A\\_district-wi.pdf](http://s3.amazonaws.com/academia.edu.documents/30511497/wayman_data_use_evaluation.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1501254555&Signature=0u8TyhYlqy%2Fa2TC1xh4Bmzt2Of8%3D&response-content-disposition=inline%3B%20filename%3DThe_data-informed_district_A_district-wi.pdf)
- Webster, D. (2011). Using data for child welfare system improvement: Lessons learned from the California Performance Indicators Project. *Child Welfare* 360, Spring, 6 & 13.
- Weinstein, M. C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., & Luce, B. R. (2003). Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health*, 6(1), 9-17.
- Zhai, F., & Gao, Q. (2009). Child maltreatment among Asian Americans: Characteristics and explanatory framework. *Child Maltreatment*, 14(2), 207-224.