



KIM JOHNSON
DIRECTOR

STATE OF CALIFORNIA—HEALTH AND HUMAN SERVICES AGENCY
DEPARTMENT OF SOCIAL SERVICES
744 P Street • Sacramento, CA 95814 • www.cdss.ca.gov



GAVIN NEWSOM
GOVERNOR

July 25, 2019

**SUBJECT: CALIFORNIA DEPARTMENT OF SOCIAL SERVICES' RESPONSE TO
THE ETHICAL REVIEW OF PREDICTIVE RISK MODELING IN THE
CALIFORNIA CHILD WELFARE SYSTEM**

“Administrative Data and Predictive Risk Modeling in Public Child Welfare: Ethical Issues Relating to California” provides an ethical analysis for the potential use of Predictive Risk Modeling (PRM) in the California Child Welfare System (CWS). This independent ethics review was a required deliverable for a research grant that supported exploration of whether a PRM tool could successfully assess risk of future CWS involvement.

This document summarizes CDSS's perspective on some of the issues raised in the ethical review.

- Best place to utilize PRM: The ethics paper asserts that the hotline is the best place to use PRM because this is the point at which there is most uncertainty and, thus, where additional information is most valuable. This is a compelling argument. However, there are many places in child welfare where more information could be available to support result-oriented practice, such as identifying which children and families would most benefit from supportive services or improving management of staff assignments.
- PRM and decision-making: The paper asserts that PRM should be used to provide additional information but should not be a substitute for social worker judgment or clinical skills. CDSS concurs with this assessment. CDSS does not support “bright line” cutoffs where decisions are made based solely on a risk score. It is worth noting that all models are not 100 percent accurate. CDSS also agrees that training should be provided on the dangers of confirmation bias if PRM is implemented.
- Data sources and ethical concerns: If California moves forward with PRM, it will only utilize CWS data. CDSS concurs that there are no ethical issues with allowing CWS to use its own data, especially since previous reports can be highly predictive of future child safety. More study would be needed to determine whether use of external data sources in the algorithm is ethical. Additional data sources would need to improve the algorithm's accuracy without exacerbating bias or inequities.

- “Universal-level risk stratification”: CDSS disagrees with the ethics paper about this practice; the Department believes that “universal-level risk stratification” is unethical and has no intention to use it now or in the future. Identifying and proactively targeting services to families with no CWS involvement is a violation of families’ privacy and their rights to parent as they see fit. This would be an overreach in the roles and responsibilities of a government agency.
- Access to PRM by staff: CDSS believes that using PRM with RED Teams or for administrative CQI is preferable to having hotline staff use PRM. Since concerns exist about PRM exacerbating bias or overly influencing decision-making, CDSS supports its use for CQI rather than being used more directly in decision-making that immediately affects children and families.
- Transparency and accountability: CDSS concurs with the paper that any algorithm used by CWS must be available to the public and should not be proprietary. CDSS also believes that the algorithm and its application must be explained in an accessible way that does not require advanced mathematics courses to understand.
- PRM and racial equity: CDSS feels that this paper does not sufficiently address concerns about bias and racial equity. The most worrisome critique of PRM is that it could continue racially biased decision-making in child welfare practice. This criticism is made because PRM relies on previous CWS actions to assess current risk; and previous actions may have been influenced by racial bias. This concern is most often raised about Black and Native American families, who have historically been profiled as higher risk. CDSS believes that use of PRM is ethical if, and only if, the algorithm is both more accurate at predicting risk and is implemented in a way that reduces racial inequities relative to current practices. Research on implicit bias suggests that conscious or unconscious racial bias has influenced past decision-making and continues to influence present decision-making in every facet of human life. Even if PRM is not used, concerns remain about how implicit bias affects current CWS practice. PRM may or may not reduce harm to vulnerable communities in comparison to the use of actuarial tools and social worker judgment. PRM’s reliance upon historical data likely introduces bias that cannot be fully eliminated, even if some variables are removed from the algorithm.

However, if properly considered, PRM’s use of CWS data may reduce bias relative to current practice by allowing risk assessments to be driven by factors that are truly predictive of future maltreatment, rather than racial bias. CDSS

believes that the metrics suggested in this paper do not adequately address these concerns. Additional information must be provided to demonstrate that PRM reduces harm to communities of color relative to current practice.

Concerns exist about racial bias influencing both which families are referred to CWS and the decisions that are made by CWS staff. CWS cannot control which families are referred to the system; therefore, even if CWS workers are free of bias, disproportionality in child welfare will continue until society is free of racial bias. In order to compare the PRM algorithm to current practice, researchers must examine a large, random sample of child welfare referrals and compare decisions that would be made incorporating PRM risk scores to decisions that are made using current practice alone. For PRM to be ethical in racial terms, it must move CWS toward reducing bias and racial inequities.

- Continuous assessment and quality improvement: CDSS agrees with the ethics paper that, if PRM is deployed, it must be tested before its release and monitored on an ongoing basis. This is especially important regarding racial equity and disproportionality. These concerns must be continuously addressed for as long as PRM is used. Ongoing monitoring should examine whether families' representation in CWS by racial group is moving toward reflecting the distribution of families by racial group in the California population. If disproportionality is worsening, an immediate assessment must be done to determine whether disproportionality is being driven by PRM's implementation, other tools and practices of social workers, or by disproportionality in which families are referred to CWS. These practices reflect the department's goal of pushing the system to become more racially equitable. CDSS also concurs that ongoing work is needed to ensure that PRM is implemented as intended and is monitored through an active continuous quality improvement process.
- Using race in the algorithm: CDSS does not agree that it is ethical to include race in the algorithm. Removing race from the algorithm does not eliminate bias, since race is correlated with many variables. However, explicitly including race as a variable simply provides an opportunity for more bias to seep into the algorithm's predictions.
- PRM scores and stigmatization: The paper asserts that a high score should not increase stigmatization, particularly because it is more accurate. CDSS does not agree with this assessment. Labeling a family as "high risk" will create stigma, whether that is intended or not. Social worker training must be provided to

counteract the impact of stigmatization, particularly since any algorithm should be used to further inform, but not make, decisions.

- Feedback loops: CDSS concurs that feedback loops are a risk in CWS research with or without PRM. Measures used to assess a tool's performance should be carefully selected so that the outcome measures are not driven by the risk score produced by the tool itself.
- Population estimates vs. individual families: CDSS agrees with the paper's remarks that approximating probabilities in a population does not mean that PRM is able to approximate probabilities for an individual case. Risk scores should be framed as an estimate of average outcomes for similar families and not as destiny. This is a leading reason that PRM should NOT be used to *make* decisions about individual situations, but instead to *inform* decision-making.