Chapin Hall & Chadwick Center

COLLABORATING AT THE INTERSECTION OF RESEARCH AND POLICY

POLICY BRIEF

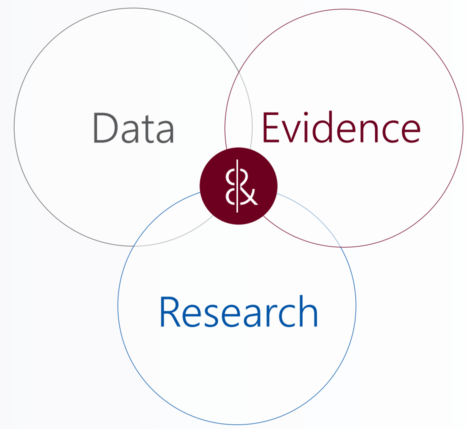
SEPTEMBER 2018

Making the Most of Predictive Analytics: Responsible and Innovative Uses in Child Welfare Policy and Practice

Predictive analytic approaches have received national attention as child welfare leaders seek rigorous empirical strategies to improve the reliability of decisions. The gravity and impact of decision making in child welfare and the complexity of factors that must be understood warrant the application of rigorous analysis. This brief presents the promise and pitfalls of predictive analytics and examines principles for responsible use.

Predictive analytics can be applied with transparency, integrity, and responsibility to improve outcomes for children and families. It can be a powerful tool to target resources and attention to families who may require more intensive interventions, while also identifying children and families who are succeeding, so we can learn from their experiences. Intentional efforts can reduce the risks of misapplication so that the full potential of predictive analytics can be achieved.

Chadwick Center and Chapin Hall are pleased to collaborate to bring together research, implementation science, and evidence-based practices to guide child welfare systems in thoughtful and cost-effective practice and policymaking. Policy briefs created under the collaboration will show decision makers how to leverage data, rigorous research, and evidence to ensure that each child receives services that are proven to effectively meet individual needs and are delivered with fidelity.



What is Predictive Analytics?

Predictive analytics refers to the practice of extracting information from existing data sets and identifying patterns that may help to predict future outcomes. Predictive risk modeling (PRM) applies the outputs of these analyses by using models to generate algorithms, or sets of "if-then" statements, that can be used to calculate a level of risk for each new case based on similarity to previous cases. Predictive analytics uses routinely collected data (called “administrative data”) to identify individuals at risk of an adverse event. For example, a predictive risk model might indicate that a child under three with fewer than two siblings and a mother with substance abuse problems may be more likely to experience future harm than other children.

WHAT DOES PREDICTIVE ANALYTICS DO?

1. Use data we already have to help us understand client characteristics and needs.
2. Provide new information about how risk and protective factors interact to influence risk for an outcome.
3. Detect patterns in big data to help us more quickly recognize where families are on a continuum of risk.
4. Provide support for our clinical judgment.

The application of powerful statistical tools to “big” data has the potential to promote understanding of precursors of both positive and negative outcomes. These benefits are weighed against legitimate concerns among lawmakers, policymakers, community advocates, caseworkers, and the public for potential human risks associated with relying on automated approaches that may heighten the effects of racial disproportionality and perpetuate systemic bias.

In order to harness the tremendous potential of predictive analytics for assisting with some of the most challenging problems faced by children and families, policymakers and system leaders must understand the method and its appropriate applications. In this way, they can leverage extensive data holdings and powerful analytic tools responsibly to maximize precision and sensitivity of decision making in child welfare.

> "We can't control the city's economy, or poverty, but we can **figure out how to apply the limited resources** of our system in the most effective ways possible."
>
> Andrew White, Deputy Commissioner for Policy, Planning & Measurement, NYC Administration for Children's Services

As with any application of sophisticated data analyses, multiple decisions inform the selection of an approach, the definition of a target outcome, the inclusion of appropriate data, and the application of findings. This brief reviews key decisions in these areas and provides guidance for policymakers and system leaders seeking to employ predictive analytics to enhance decision making in child welfare. In addition to providing an overview of the approach, we weigh the application of predictive analytics with an examination of ethical and methodological considerations.

PREDICTIVE ANALYTICS FOR CHILD WELFARE

In the child welfare context, predictive analytics is most commonly associated with identifying maltreatment risk levels. While the approach works by quantifying risk, it can be used to target services to children, families, and communities that are most likely to benefit, thus ensuring that limited resources are used efficiently and effectively. In the context of new incentives to utilize both prevention and treatment resources strategically, predictive analytics can provide empirical guidance for planning and practice decisions. In addition, predictive analytics can be used to examine the characteristics of children and families who have positive outcomes in spite of their risk, in order to determine how best to nurture resilience and fortify protective factors.

EXAMPLE: NEW YORK CITY

The Administration for Children's Services in New York City sought to build predictive models that could identify risk factors among families for "frequent involvement" in the child welfare system. Frequent involvement was defined as multiple investigations within a given period of time resulting in referrals for preventive services or foster care. Models were translated into algorithms that could be applied to:

1. Prioritize "exit conferences" among families completing services who may be at high risk for returning to the attention of child welfare, directing additional attention to families to ensure that needs are met.

2. Develop "risk cohorts" among provider agencies, or groups of providers with populations of similar risk levels, so that provider quality could be measured more fairly.

Both implementations involved the engagement of diverse stakeholders to provide input on the variables used in the models and to oversee the ethical application of the models. New York has been a leader in the development of principles to guide the ethical use and application of predictive analytics, as well as the engagement of stakeholders in formal oversight groups to inform and advise their use.

Considerations for the Use of Predictive Risk Modeling

WHAT QUESTIONS/PROBLEMS DO WE SEEK TO ADDRESS?

While the most important decisions made by child welfare workers have to do with assessing safety and risk, local examples illustrate other opportunities for the application of predictive models. Given the potentially serious consequences of imprecise or inaccurate decisions in the child protection system, leaders will likely continue to seek to enhance the decision making of hotline staff and investigators with rigorous empirical approaches. As the field's understanding of predictive models has grown, other opportunities to use them have emerged (Russell, 2015). These include the use of predictive models to:

- Prioritize cases by risk for the receipt of services, supervision, or consultation
- Adjust performance ratings of providers serving populations that differ in their collective "risk" of negative outcomes
- Improve the utility of dashboards and other technologies that can guide the day-to-day work of case workers

While our systems aspire to serve every youth and family in need of care, a recognition of the fiscal and programmatic realities has expanded the use of predictive risk models from answering the question "Who is at greatest risk?" to questions like "Who first?" and "Who more?"

Predictive analytics risk rankings should be strategically deployed to inform the level of attention, service, and care that families need to achieve both safety and well-being. Extreme caution should be taken to ensure that the output of analytic models does not drive legal decisions, such as the termination of parental rights. This can be challenging in the context of a system that seeks guidance to avoid deadly outcomes. Careful planning is required to:

> The application of predictive risk models have expanded from answering the question "Who is at greatest risk?" to questions like **"Who first?" and "Who more?"**

- Specify who may receive risk rankings and at what point in the life of a child welfare case they are shared
- Safeguard confidentiality of risk level assignments, thoughtfully allocating access to staff to limit confirmatory bias
- Implement training for staff on the meaning/use of risk levels or enhancements to address elevated risk
- Offer voluntary services based on risk levels

WHAT OUTCOMES DO WE HOPE TO PREDICT?

Predictive analytics quantifies the risk of clearly defined outcomes based on existing data sets. Most outcomes can be defined in multiple ways. For example, while most child welfare system leaders and policy makers would agree that "future harm" is an outcome of interest, there are countless ways to define "harm": frequent reports of harm, reports of harm within a particular time frame, or reports of specific types of child abuse or neglect. Consensus around the definition of outcomes can improve methods, promote understanding of analytic goals and approaches, and facilitate buy-in by system stakeholders whose support is often necessary for successful implementation. This consensus can be difficult to reach, but is best approached using broad engagement of system actors and key stakeholders over a period of time that allows for debate and refinement.

To arrive at reliable predictive models, researchers should seek to predict outcomes that occur frequently enough or with enough severity to warrant serious concern, and not so infrequently that the available data may not contain a sufficient number of cases. Outcomes such as fatalities occur so infrequently within a single system that it is difficult generate reliable models to predict them. Additional considerations may include the window of time in which the outcome may occur, the criteria for establishing that the outcome has occurred (from hotline report to substantiated investigation), and any exclusionary criteria (youth at home vs. youth in foster care) for the sample whose risk will be estimated.

EXAMPLE: ILLINOIS DEPARTMENT OF CHILDREN AND FAMILY SERVICES

As part of an effort to reduce reliance on congregate care placements, the Illinois Department of Children and Family Services sought to understand who was at greatest risk of placement in congregate care settings. When models identified a set of predictors that included specific behavioral and emotional problems, the Department was able to design and implement a pilot of a therapeutic foster care intervention that could serve as an alternative to congregate care. The predictive models yielded algorithms that were translated into eligibility criteria for the programs, so that a target population could be more precisely identified and served. The models also served to guide customized programming that could meet the complex needs identified analytically as contributing to risk.

WHAT DATA ARE AVAILABLE AND APPROPRIATE TO INCORPORATE INTO PREDICTIVE ANALYTICS?

Effective predictive analytics depend on data. To engage in predictive modeling, policymakers should consider the breadth, depth, and quality of available data.

1. **Breadth** refers to the availability of data on many of the variables that may be related to the outcome of interest for a sufficient portion of the population.

It also refers to the availability of data at the time end users will apply the models. For example, child welfare systems require that certain data elements be entered into the jurisdiction’s electronic record keeping system, resulting in a pool of items that are available for the majority of children in the system. Breadth ensures that sufficient data exist to ensure that the analyses are generalizable to the system population as a whole, and not driven by limited data from a subset of children and families. Optimal breadth standards should be locally determined, and data quality issues addressed, prior to the start of any analyses.

2. **Depth** refers to the availability of data of a duration sufficient to observe outcomes of interest in a cohort over time.

While many jurisdictions are interested in longer-term outcome indicators—such as re-reports of maltreatment and re-entry to out-of-home care—these require maintenance and collection of child-level data over a period of months or years to track outcomes. Ensuring appropriate depth allows policy makers to examine outcomes over the time periods that are of most interest to child welfare systems; appropriate depth will vary by the timeframes necessary to examine the selected outcomes.

3 **Quality** refers to the reliability, validity, and comprehensiveness of available data.

The quality of child welfare data depends on the consistent, informed use of data collection systems by frontline staff. Thus, assessment of data quality should incorporate input and feedback from these staff. Data quality may be jeopardized by missing data, data entry errors, and the pervasive or systematic misuse of data fields, such as when frontline staff commonly omit or repurpose data fields. Agencies need to provide training for staff on data entry to ensure consistency and quality across workers and also review data quality on a regular basis through examination of missing or incomplete data. In addition, data should be collected in ways that are both efficient and effective. For example, outcomes of interest should not be entered into text fields, which are time consuming to enter and require recoding to be of use for analytics.

WHAT ANALYTIC APPROACH SHOULD WE USE?

Approaches to the technical work of building and running predictive models fall into two broad categories.

Machine learning approaches, such as random forest and neural networks, are atheoretical; they incorporate all available variables to arrive at a set of predictors regardless of hypothesized relationships or prior research findings.

Regression model approaches may incorporate theoretical hypotheses about the mechanisms of risk; model builders select variables that prior research suggests affect outcomes for inclusion in models that are then applied to another sample to test their accuracy (Russell, 2015; Cuccaro-Alamin et al., 2017).

The choice between these approaches depends on the background and training of researchers, and their philosophical and personal preferences. Machine learning approaches make use of a larger array of potential predictors, but may identify predictors that are not "actionable." Regression modeling approaches may be subject to a researcher's own hypotheses, and may omit key variables that might be salient predictors if included. Despite their differences, each yields an algorithm that can produce risk scores, or a list of individuals or families ranked according to risk of a particular outcome.

> "We have **so much information** that could be better used to identify which kids need which services—and predictive analytics can help us get to that point."
>
> Wendy Henderson, Director, Wisconsin Bureau of Youth Services

How Can Predictive Analytics Models be Applied to Improve Child Welfare Systems, Processes, and Outcomes?

PROVIDING SERVICES TO THOSE IN GREATEST NEED

While predictive analytics is most commonly associated with identifying maltreatment risk levels, it can also be used to more appropriately target services to an individual child, family, or community. Particularly when deploying evidence-based practices in risk reduction efforts, use of predictive analytics can ensure that limited resources are used where they can have the most benefit. For example, Nurse-Family Partnership (NFP) is a well-established evidence-based home visiting program for the prevention of child abuse. It is targeted at first-time mothers. While many families may benefit from receiving NFP, start-up and ongoing costs may prohibit universal delivery of the practice to all first-time mothers. Predictive analytics can be utilized to identify which families meeting criteria for NFP are most at risk for poor outcomes, and recruitment and engagement can then be focused on those in greatest need of the resource.

INFORMING SERVICE ARRAY PLANNING, TRAINING, AND CONTRACTING

A primary purpose of predictive analytics in child welfare systems is to identify children and families at risk for future maltreatment who would benefit from additional supports or intervention. It is crucial that appropriate evidence-based services are available to meet the needs of the population and ameliorate the risk. Predictive analytics can be a key source of information when selecting and implementing an array of evidence-based practices:

- By providing valuable information on the characteristics and distribution of risk in the community to be served, predictive analytics can inform the needs assessment process that should guide program selection.
- The distribution of risk throughout the population and across geography can inform contracting decisions about needed capacity to train and deliver evidence-based practices.
- Location information incorporated into predictive analytics can be used to determine where to deliver services to be most effective, such as the most appropriate community in which to locate a family resource center to address unmet need.

Systems should ensure that services are appropriate, evidence-based, and delivered with fidelity; routine examination of the outcomes of the service array can determine whether the services are having the desired impact.

ILLUMINATING RESILIENCE AND SOURCES OF STRENGTH

Predictive analytics also gives us an opportunity to observe the unique qualities of children and families that succeed in spite of elevated risk, employing the principles of the "Positive Deviance" approach (Pascale, Sternin, & Sternin, 2010). Predictive models are developed using a "training" sample, and then tested retrospectively using a "test" sample in which the accuracy of predictions can be judged, prior to employing the model in a prospective manner. Careful observation of cases that do not experience predicted negative outcomes may yield crucial insights into the characteristics that can shield families and protect them in the face of risk. However, our ability to observe strengths and protective factors depends on measurement strategies that capture the full array of child and family functioning, incorporating information about needs and strengths.

In addition to hoped-for improvements in our ability to effectively serve children and families, predictive analytics exercises may also offer improvements to child welfare agencies in the following areas:

- Capacity to work with administrative data
- Data integrity
- Engagement of internal and external stakeholders in planning and development of an empirical approach to understanding risk
- Understanding of the constellation of factors that heighten risk among children and families
- Empirical answers for policy decisions regarding the allocation of resources (time & attention)

Integrity: Address existing systemic bias

Concerns about bias are pervasive in child welfare systems, which may respond differentially to different groups of people. Consequently, some fear that the use of child welfare data can reinforce this bias when administrative data used in models (e.g., dates, ages, and characteristics of clients) are supplemented by system responses to clients (e.g., services and placements). Therein lies the potential for institutionalizing bias: if biased system responses are used to indicate high levels of risk, the use of these decisions as "predictors" may magnify the effect of biased decision making. However, using data about the needs, strengths, and functioning of children and families can help us to avoid the potential of predictive analytics to amplify disproportionality. This strategy requires a concomitant shift in our assessment and measurement strategies to include well-being, strengths, and protective factors. Child welfare jurisdictions that use tools to measure these components can then structure models that rely on family need, and not on possibly biased system responses.

Responsibility: Ensure that automated approaches enhance human judgment

Critics of predictive analytics sometimes express concern that the approach removes human judgment from the child welfare decision making process. The optimal use of predictive analytics should blend guidance derived from predictive algorithms with the judgment of those involved in the case. Predictive analytics provides powerful empirical guidance based on the analysis of thousands of cases that have come into contact with the child welfare system; when used effectively and responsibly, predictive analytics supplement, but should not replace, expertise, clinical judgment, and critical thinking about strengths, needs, and contextual factors. Accordingly, systems should be put in place to allow for overriding a predictive analytics-identified response, with appropriate justification and supervisor approval.

Transparency: Make information about model developent and operation clear and available

Given the complicated mathematical techniques that comprise predictive analytics, it is not surprising that a lack of transparency in the models themselves is a primary ethical concern. Models and analyses are often conducted by external contractors, with systems leaders and stakeholders unable to access details on the model due to the proprietary nature of some model purveyors. Given the potential negative consequences of being identified as a family at high risk for child maltreatment, efforts should be made to ensure that the predictive analytics process is as transparent as possible, including the use of documents describing the process in plain language, oversight committees that review the entire process, and regular monitoring of the reliability and validity of the models developed.

RECOMMENDED ACTIONS

Use predictive analytics to understand factors associated with clearly defined outcomes.
Child welfare agencies should identify leads, representatives of key agency functions, and stakeholders to engage in thoughtful exploration of the outcomes to be predicted and the methods for defining these using data. This stakeholder group should have the opportunity to weigh in on all decisions, including the definitions of outcomes to be predicted, the variables to be included, and the ultimate application of predictive models.

Use datasets with appropriate breadth, depth, and quality.
The datasets should be understood with help from frontline staff who can provide clarity and feedback on the use of specific data fields.

Partner with researchers to develop and refine approaches and models.
Predictive analytics is highly technical work. Partnering with researchers to assist with predictive analytics requires clear policies on data sharing and confidentiality and contractual relationships that govern the exchange and use of administrative data.

Use predictive analytics to direct attention to cases requiring more intensive service or supervision, not to impose additional requirements on families.
In addition, plans must be developed to ensure that support continues across placement changes, including after reunification, so that all caregivers have the necessary skills to address each child's needs and re-entry to care is prevented.

Employ assessment strategies that capture elements of well-being, including strengths, protective factors, and functioning.
Incorporating these as variables in administrative databases creates opportunities to maximize the utility of predictive analytics approaches while minimizing reliance on other, potentially biased, sources of information.

Minimize the effects of race and ethnicity on future decision-making.
Racial/ethnic biases may be implicit in the circumstances surrounding child welfare involvement. During the predictive analytics process, systems should explicitly acknowledge the racial/ethnic disparities, avoid predictor variables that signify potentially biased system responses to children and families, and engage ethics review committee with diverse representation. Demographic predictors should be incorporated and interpreted cautiously.

Prioritize human judgment when integrating predictive analytics into service delivery and agency operations.
Since every child and family is unique and may have strengths, protective factors, or environmental protections that may not be incorporated into models but nonetheless can ameliorate risk, predictive analytics should augment, not replace, human judgment. As model outputs are probabilistic, systems should allow workers to override decisions that depend on the output of predictive analytics, with appropriate oversight, to reduce the incidence and consequences of misidentification. This may require modification of existing Medicaid State Plans or other funding stream reimbursement rules.

The ethical use of predictive analytics models depends upon transparency in model development, model application, and model refinement.
All predictor variables should be identified and clearly described, and reports and analytic documentation should be written in language that is accessible and understandable to a wide variety of consumers, including community stakeholders.

Predictive analytics needs to be a collaborative and iterative process.
Models should be periodically revisited and refined based on the availability of additional data, evaluation of their accuracy and impact, and the input of the stakeholder group. Similarly, the validity and reliability of models should be examined on an ongoing basis.

REFERENCES

Cuccaro-Alamin, S., Foust, R., Vaithianathan, R., & Putnam-Hornstein, E. (2017). Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review*, *79*, 291-298.

Lewis, G., Curry, N., & Bardsley, M. (2011). *Choosing a predictive risk model: a guide for commissioners in England*. London: The Nuffield Trust.

Pascale, R., Sternin, J., & Sternin, M. (2010). *The power of positive deviance*. Boston, MA: Harvard Business School Publishing.

Russell, J. (2015). Predictive analytics and child protection: Constraints and opportunities. *Child Abuse and Neglect*, *46*, 182-189.

ACKNOWLEDGMENTS

The Chadwick Center and Chapin Hall would like to recognize Dana Weiner and Jennifer Rolls Reutz for their joint authorship of this Policy Brief. Other important contributions were made by Bryan Samuels and Charles Wilson, as well as Elizabeth Dierksheide and Mallory VanMeeter.

Chapin Hall would like to recognize the financial support provided by Casey Family Programs and their continuing interest in the work.

The Chadwick Center for Children and Families at Rady Children's Hospital-San Diego, would like to acknowledge the California Department of Social Services, Office of Child Abuse Prevention for funding the California Evidence-Based Clearinghouse for Child Welfare (CEBC) as well as recognize CDSS for their leadership in promoting the use of evidence-based practices and data-driven decision making in the state.

SUGGESTED CITATION

Chadwick Center and Chapin Hall. (2018). *Making the Most of Predictive Analytics: Responsive and Innovative Uses in Child Welfare Policy and Practice*. San Diego, CA & Chicago, IL: Collaborating at the Intersection of Research and Policy.

CHAPIN HALL
AT THE UNIVERSITY OF CHICAGO

chapinhall.org

Children's San Diego | Chadwick Center for Children & Families

chadwickcenter.org