Child Abuse & Neglect 46 (2015) 182–189

Contents lists available at ScienceDirect

Child Abuse & Neglect

Research article

Predictive analytics and child protection: Constraints and opportunities

Jesse Russell

National Council on Crime and Delinquency, 426 South Yellowstone Drive, Suite 250, Madison, WI 53719, USA

ARTICLE INFO

Article history:
Received 4 October 2014
Received in revised form 8 May 2015
Accepted 27 May 2015
Available online 2 July 2015

Keywords:
Predictive analytics
Risk assessment
Child protection
Decision making

ABSTRACT

This paper considers how predictive analytics might inform, assist, and improve decision making in child protection. Predictive analytics represents recent increases in data quantity and data diversity, along with advances in computing technology. While the use of data and statistical modeling is not new to child protection decision making, its use in child protection is experiencing growth, and efforts to leverage predictive analytics for better decision-making in child protection are increasing. Past experiences, constraints and opportunities are reviewed. For predictive analytics to make the most impact on child protection practice and outcomes, it must embrace established criteria of validity, equity, reliability, and usefulness.

Introduction

Increases in data quantity and data diversity, along with advances in computing technologies, have made predictive analytics a powerful tool for helping to guide decision making. IBM has suggested that across the globe, 2.5 quintillion bytes of data are created every day. The volume of new data creation is so great that ninety percent of data in the world today was created in the last two years (IBM). In response to these increases in data availability, predictive analytics applications have become common in business settings and are increasing in social services settings.

As some look to leverage the potential of data and predictive analytics, it is important to consider the opportunities and the constraints for using predictive analytics in child protection decision making. The use of data to inform child protection decision making is not new; neither is the use of data and statistical modeling to estimate likelihoods of particular events and assign predictive scores to child protection agency clients. The question is how the emergence of predictive analytics might inform, assist, and improve the use of data to guide decision making in child protection.

Decision Making in Child Protection

Child protection agencies ask "who" questions, "how" questions, and "why" questions. Who questions are about whom the agency should serve, allocate resources to, and target interventions to. Who questions are about which children and families the agency should be most concerned about, who might be appropriate for a particular intervention, or who might be the best opportunity for prevention. Who questions help child protection systems identify which system-involved families they should focus on.

These questions also help answer "who not?" questions. Given all the families who are referred to the system, which families are unlikely to experience further maltreatment, or which families would not benefit from a particular intervention?

In contrast, "how" questions are about how child welfare agencies should serve children and families, the programs they should develop, or the practices they engage in. How questions can be specific to a particular child or family, a particular

http://dx.doi.org/10.1016/j.chiabu.2015.05.022

program, or general to agency practice. How questions could include things about how to develop an effective service plan for a family, whether a specific program is effective, or whether an agency should change its practice or policies.

"Why" questions, which are most commonly addressed by researchers, are about the causes and consequences of child protection involvement. These questions ask why some families might be more likely than others to experience maltreatment, what drives outcomes, or what are the causes of child welfare outcomes?

Child protection agencies have a mandate to respond to and prevent future child abuse and neglect. Though agencies receive many reports of child maltreatment, no agency has the resources to investigate or serve every family mentioned in every report. Moreover, the interventions child protection agencies offer are not always welcome or entirely benign. It is central to the mandate of child protection agencies to decide which families and children to serve—a "who" question.

Predictive models directly relate to these types of questions to inform specific decision points in child protection. For example, a question about which families might be more likely to experience future maltreatment (a family's risk of future child abuse or neglect) can inform the decision of whether to open a case for services. Informed by the outcome of a risk assessment that provides an estimate of the likelihood of future maltreatment, the agency may end its involvement with the family or decide to open a case for services. In California, for example, low- and moderate-risk families are generally recommended for closure unless safety threats remain unresolved, while high- and very high-risk cases are recommended for case opening (Wicke Dankert & Johnson, 2014).

The focus of this article is on these types of decisions—responses to who questions—that can be informed by predictive models.

Predictive Analytics

An emerging approach with the potential to inform decision making in child protection, predictive analytics looks at the past experiences of an organization to estimate the likelihood of future events. Predictive analytics looks at that past by using computer algorithms to sort through an organization's data to produce and "train" (or shape) a model that can then estimate likelihoods of particular events and assign predictive scores to the organization's clients.

Predictive analytics can include a broad set of statistical and analytical tools that identify trends, relationships, and patterns within data that can be used to predict a future event or behavior. Predictive analytics as a broad category can include the concepts and methods associated with "big data," data mining, machine learning, classification and regression trees (CARTs), and random forest modeling, among others.

Standards for Judging Predictive Models

The standards by which predictive models in child protection should be judged are well-established. They should be valid, reliable, equitable, and useful (D'Andrade, Austin, & Benton, 2008). Even if the methodology by which they are developed and used in practice varies and changes, these are the standards by which they must be evaluated.

Validity, in general, refers to how well a test or task matches the attribute or the domain of the knowledge we wish to assess. Validity is about whether the test measures what it is meant to measure. While the concept of validity is clear, actual measurement of validity can be challenging. Most commonly, validity is statistically assessed through the receiver operating characteristic (ROC) or the area under the ROC curve (AUC) (see, inter alia, Fogarty, Baker, & Hudson, 2005).

The ROC is a graphical plot of the true positive rate against the false positive rate (Fogarty et al., 2005). In a binary test (with one "positive" outcome and one "negative" outcome) the true positive rate represents how many correct positive results are achieved among all positives in the sample—or how often the test says the result is positive when the outcome was actually positive. The false positive rate, on the other hand, represents how many incorrect positive results are achieved among all negatives in the sample—or how often the test says the result is positive when the outcome was actually negative. This concept allows analysts to weigh the costs and benefits of a particular test along with the trade-offs between having fewer false positives and more true positives.

Calculating accuracy from the ROC is done simply by adding the number of true positives and the number of true negatives as a fraction of the total sample. In this way, ROC accuracy represents how often the test produces a correct result.

A single measure that can be derived from the ROC is the AUC, which represents the probability that for a randomly drawn pair (one from the positive group and one from the negative group), the test will rank or score the positive case higher than the negative case (assuming that positives are higher on the scoring scale). In other words, the AUC measures the percentage of randomly drawn pairs for which the test correctly classifies both cases.

Interpretation of the AUC in one sense is easy: High AUC values represent more accurate classification. The AUC is often seen as the standard measure for comparing validity across classification models because it is a single metric that can reduce the appearance of subjectivity (Hand, 2009; Lobo, Jiménez-Valverde, & Real, 2008). The potential of the AUC is that two models can be directly compared in a sort of validity competition—models with higher AUC scores might be understood as better classifiers or as more valid (see Rice & Harris, 2005). The literature on AUC and potential alternatives is robust (see, for example, Hand, 2009, or Hanczar et al., 2010).

While the greatest endorsement of the AUC may be that it performs better than other single measures of model validity (Bradley, 1997), trouble in interpreting the AUC can stem from reducing the ROC to a *single* measure rather than using the ROC framework to examine multiple trade-off options (Powers, 2012). As far as single measures of model validity go, the

AUC is the most commonly used and appears to have some advantages over alternative measures. However, reliance on any single measure, AUC or otherwise, may be problematic. The performance of a discriminant model consists of trade-offs and nuances; any assessment of that performance should take into account those trade-offs.

The core sense of the accuracy of a particular predictive model rests in how distinct differences in outcome rates are across classification levels. Well-functioning models ought to produce a distribution across categories (such as high, moderate, low, or any other scale of categories) that corresponds to actual outcome rates. For example, if families classified as high risk do not actually have substantially higher rates of future maltreatment than do families classified as moderate risk, then the risk assessment is simply not producing accurate results. It is also critical that accuracy of the predictive model not vary by type of maltreatment (e.g., abuse or neglect allegations) or multiple-type situations, it must be robust across different situations. Most critically, accuracy is demonstrated when outcome rates clearly correspond to classification categories.

Equity is the second standard by which predictive models are judged. While validity considers whether a test measures what it is meant to measure overall, predictive models must also demonstrate equity. Equity is determined by how similarly a predictive model functions across sub-populations. Equity is the degree to which a model classifies outcomes the same way across subgroups (i.e., "risk" is the same for boys and girls, across major race and ethnicity groups, and between major geographic areas) and is an essential measure of instrument validity. When a predictive model is not performing equitably across subgroups, "crossover" is observed. For example, crossover occurs when higher-risk clients from one group have similar (or lower) outcome rates compared to lower-risk clients from another subgroup. In other words, higher-risk clients from one subgroup have crossed over and, in reality, experience the outcome more similarly to lower-risk clients from a different subgroup.

Reliability represents how often different users of a predictive model come to the same conclusions from the same information. This is important because child protection agencies want to ensure that child protection decisions are made according to policy and that policy is applied consistently between workers and across local offices. Inter-rater reliability measures variations in judgments across appraiser, and the goal is to minimize these variations in order to evaluate the extent to which a test or measure yields the same results on repeated trials (Thorndike, 1985).

Research has shown that case decisions based on clinical judgment alone are not highly reliable. One study (Rossi, Schuerman, & Budde, 1996), which compared the inter-rater reliability of case recommendations among 27 nationally recognized child welfare experts and 103 workers based on a number of case vignettes, found that child welfare experts and workers had rates of agreement of 65% when deciding whether or not to place a child in foster care and workers had rates of agreement of 64%, with corresponding kappas of .45 and .35. A study by Regehr, Bogo, Shlonsky, and LeBlanc (2010) investigated reliability and workers' confidence in their assessments. The study found "considerable variability" in appraisals (p. 621), even among highly skilled and trained workers, though there were high level of worker confidence in their appraisals. Using a case scenario, they found that two thirds of workers indicated a safety threat was present, and the other third did not. With the same scenario, on a risk scale of 0–18, the range of scores given by workers ranged from 1 to 15. They concluded that even when presented with the same information and the same assessments, workers were "highly variable" in their conclusions (p. 626). When a model is not reliable, child protection decision making is based on worker discretion rather than policy, making it difficult to achieve validity and equity (Johnson & O'Connor, 2009).

Usefulness is the final standard by which predictive models are judged. Utility is less about methodology and more about implications for practice. A useful predictive model has to provide useful information and practicable guidance for workers making decisions in the field. It also must be easily understood and not overly burdensome for workers to use. For example, a predictive model that classifies families according to a tertiary outcome, or simply an administrative output, might not have the potential to impact outcomes for children and families in a meaningful way. When no potential exists for a particular predictive model to impact practice, improve systems, or advance the well-being of children and families, then that model is inadequate.

Applications of Predictive Analytics in Child Protection

Efforts to apply predictive analytics approaches to public agency data have only recently begun. Predictive analytics has had two main applications in child protection: (1) in traditional risk assessment, to estimate the likelihood of a future report of maltreatment or substantiated report of maltreatment; and (2) to estimate the likelihood of other child protection outcomes, such as child deaths, case failures, and other outcomes.

Predictive Models of Future Maltreatment Report/Substantiation

Risk assessment has been at the fore of approaches to estimating the likelihood of future maltreatment, future reports of abuse or neglect, or substantiation. Many child protection agencies currently use risk assessment for the purpose of targeting resources and informing service decisions. A variety of such risk assessments are used by child protection agencies across the United States and much of the world. An early review of risk assessments in child protection (English & Pecora, 1994) found that the use of risk assessment was a widespread phenomenon as early as the 1990s.

Predictive analytics has been applied directly to developing risk assessments. For instance, a predictive analytics approach (using neural network models) was taken to develop a risk assessment by modeling the likelihood of recurrent maltreatment in child protection cases (Jolley, 2012) and distinguish between static and dynamic risk factors. Another study that explored the use of predictive analytics for developing a risk assessment used a classification and regression tree (CART) approach (Sledjeski, Dierker, Brigham, & Breslin, 2008). This study explicitly compared the ability of the predictive analytics approach (the CART) versus a regression model approach and concluded that the CART was better at prediction for high-risk groups. One similar comparison using neural networks (an approach to pattern recognition inspired by biological neural structures; for a full description of the method, see West, Brockett, & Golden, 1997) arrived at a similar conclusion, based on an analysis of the true positive rate (Schwartz, Kaufman, & Schwartz, 2004). Another comparison study found that a neural network approach did *not* outperform alternative methods (Flaherty & Patterson, 2003).

In contrast, some studies have touted the benefits of particular modeling techniques, like neural networks (Marshall & English, 2000) or a mixed-methods approach (Silver & Chow-Martin, 2002). However, any study with the conclusion that one particular technique outperformed another should be interpreted with caution. No technique can develop a risk assessment without a human to carry out the analysis and re-interpret the data. Some techniques might rely upon more or less involvement by an expert analyst, but the human involvement is essential. To suggest that a neural network, CART model, or any other particular technique performed better in one study or did not perform better in another study misses the important question of how those techniques are being incorporated into procedures for developing risk assessments for the agency as a whole. These approaches need to be tested in the real world of the actual operational environment of the agency.

A substantial limitation to studies on the relative merits of any particular modeling technique is that different techniques are more easily applied in some practice situations than in others and are more appropriate for some decision points than others. In particular, risk assessment in child protection practice normally occurs at a point when few data elements are available and the need for expert judgment on the results of the risk assessment is important. Risk assessment is most commonly used at an early stage of decision-making in a case, often at the point of deciding whether to open a case for services. At this point, the agency may have relatively little existing information about the family and have few data points from which to draw. Because of this, risk assessment most commonly requires the worker to gather new information. Predictive analytics can provide insights into what information ought to be collected by the worker, but the assessment is still predicated on a worker collecting the necessary information.

Further, for it to be useful, the future event being modeled with predictive analytics must be a meaningful one for practice. Estimating the likelihood of substantiation based on intake data, for example, might be of questionable utility because it is unclear how it would directly inform a specific decision point. The likelihood of substantiation is not itself an indicator of safety, danger, or future maltreatment. While such a predictive model might be valid, reliable, and equitable, its utility is unclear. For predictive analytics to be impactful for child protection, it must abide by all four standards.

Predictive Models of Other Outcomes

Beyond traditional risk assessment, predictive analytics has been used in a number of jurisdictions to explore the possibility of identifying which children and families might be at higher likelihoods of a number of different negative outcomes. For example, the Florida Department of Children and Families recently worked with consulting and technology partners to analyze child fatality data. This work explored the potential to apply predictive analytics to identify which children had a higher likelihood than others of a premature death (Florida Department of Children and Families, 2014). They used data from reports to a child abuse and neglect reporting hotline about a child fatality that contained allegations of abuse or neglect as factors leading to the child's death. The analysis focused on all reports, whether or not the family had prior agency involvement.

The analysis was based on statistical models of the odds of death based on the data collected from the hotline information. The analysis identified 14 consistent risk factors. Researchers noted limitations due to problems in data quality and ambiguity, and emphasized that a more inclusive cross-system data would strengthen the analysis and make a predictive model more useful for practice.

Another study examined how a predictive analytics technique can highlight unhypothesized relationships among factors involved with worker decisions to investigate child maltreatment reports (Johnson, Hendricks Brown, & Wells, 2002). A predictive analytics approach also has been used to examine risk of re-reports and re-abuse in children who remain in the home after an abuse report (Dakil, Sakai, Lin, & Flores, 2011). The aim of work in this vein has been to use predictive analytics to construct models that can be applied to specific points in the life of a child protection case, such as investigations or case management.

Another example comes from the Lead Community Care Agency in Tampa, Florida. In partnership with a technology consultant, they used predictive analytics to develop a case escalation tool. The predictive analytics model was used to identify which cases "have a high probability of failure" (Mindshare Technology, 2014). The analytics were based on existing data from the Florida State Automated Child Welfare Information System (SACWIS). Case failures were defined as child protection cases that included outcomes related to failed reunifications, aging out, juvenile justice crossover, not graduating from high school, runaways, and exposure to violence.

In Hillsborough County, Florida, the Department of Children and Families and Eckerd Community Alternatives analyzed data on child deaths (Department of Children and Families, 2013). They examined 1,500 child maltreatment cases, including eight child deaths. The analysis suggested nine factors related to the risk of child death. These factors were used to produce a checklist to be administered four times a year in the homes of children under the age of three. The results of the checklist can then initiate a response for the case manager to take actions to increase the safety of the child.

In addition to these examples, a number of jurisdictions have used predictive analytics as part of Title IV-E waiver programs. These have included identifying children at risk of re-entry into substitute care within 12 months of reunification, youth at risk of aging out of foster care, and families at risk of homelessness, among other. Across all the examples presented here, there is a common theme of using existing data to identify children and families at higher risk of potential negative outcomes. If the families and children most at risk of these outcomes can be accurately, equitably, and reliably identified, then child protection agencies will have established a newfound opportunity to effectively prevent these negative outcomes.

Future Challenges and Opportunities

Data Availability Remains a Challenge

The potential impact of predictive analytics depends on the amount of available quality data. Four aspects of data quality matter: the number of records available, which contributes to statistical power; the number of variables available; the amount of missing data, such as information omitted or not measured; and the quality and consistency of data entry. These data quality and diversity factors are important for any predictive model that is based on data; they are especially crucial for models grounded in computer algorithms. First, a sufficient number of cases or records must be available in order for the underlying algorithm(s) to converge on a stable model. If the number of available cases is not large or selected randomly, then the ability to generalize findings beyond the current sample is substantially limited. Second, a large number of variables or data fields must be available. Most predictive analytics methods explore interactive and combinatorial effects among variables. If a particular dataset does not have a large number of variables available, then the predictive analytics algorithm(s) will have insufficient fodder for establishing these interactions. Third, large amounts of data should not be missing. Some child protection agency management information systems (MIS) often contain data fields that are empty for many cases. Most predictive analytics algorithms have clever solutions to missing data (such as the process of identifying "surrogate" variables that have similar predictive qualities), but large amounts of missing data will reduce the efficacy of any predictive analytics effort. Fourth, the quality and consistency of data entry is hugely critical. The adage "garbage in, garbage out" applies here, meaning that the results of the analysis can only be as good as the data that are used. Data in child protection systems can lack quality in numerous ways, mainly due to how, when, and by whom data are entered.

An important consideration to this discussion is that the above four aspects of data availability and quality apply not just overall, but must be considered specifically at the time in a case when a predictive model might be applied. For any predictive model to be conducted when an agency is deciding whether to end its involvement with a family at the close of an investigation or initial assessment or to open a case for services, the information available (and the corresponding amount of data in the MIS) could still be quite limited. Though some research has demonstrated how a relatively small number of child and family risk factors can be used to predict maltreatment reports, injuries, and even death (see, for example, Putnam-Hornstein & Needell, 2011), the information needed to identify those factors must be proactively collected and recorded. More information could be collected as the case moves forward, but not much is yet known at the time of case opening, for example.

Further, data in other systems (e.g., court data, juvenile probation data, or public health) may be useful for the child protection agency to have at the time of applying a predictive model. In most jurisdictions, however, case-level data from other agencies are not available to the child protection agency. The potential for interoperability—the ability of two or more systems to exchange information and use the information that has been exchanged—is not new but has recently gained considerable traction (Smith, 2008). Interoperability, or any form of data sharing, integration, or coordination, can be complementary to predictive analytics. The value of achieving greater access to more sources of data can be substantial for a child protection agency, whether or not that agency has any use for predictive analytics. Data diversity, or having access to more information from more sources, can help improve decision making of all sorts, strengthen applied research, and help improve ongoing agency operations management. Data diversity is important beyond any particular method for using those data.

Apply Established Standards

A different area of challenges for predictive analytics in child protection is keeping the focus on the established standards of validity, equity, reliability, and usefulness. While many predictive analytics applications in the business world are oriented around prediction and building predictive models, child protection agencies have a greater need for classification and resource allocation efficiency. The goal of a child protection agency is not to only predict that a particular family will face new reports or new substantiations in the future. Rather, the goal of a child protection agency is to strive to prevent further maltreatment. It is not enough that a child protection agency identify families most at risk of a particular negative event; the agency must be prepared to offer an appropriate intervention in those cases. For example, if an agency can estimate the likelihood of a youth becoming homeless, for that estimate to be useful, the agency would need to be able to offer an appropriate service intervention to prevent homelessness. Often, appropriate service interventions cross traditional agency and department boundaries. A child protection agency might be able to estimate the likelihood of future delinquency, for example (see Bogie, Johnson, Ereth, & Scharenbroch, 2011), but if delinquency prevention is not the purview of the child protection agency itself, intervention responsibility can become unclear. In these cases, political will and creative funding strategies may be necessary.

Additional Opportunities for Predictive Analytics

While this paper has focused on the question of how predictive analytics might be applied to "who" questions in child protection, other opportunities for child protection agencies to make use of predictive analytics may exist. In particular, predictive analytics could be used by child protection agencies as a tool for operations management. The potential for predictive analytics to inform performance measurement and the tracking of key performance indicators is an emerging idea with much potential for exploration. As more social services agencies increase their data holdings, establish more data sharing agreements, and build more data warehouses holding more types of data, the potential for predictive analytics to inform day-to-day practice decisions is huge.

Predictive analytics might also be used by child protection agencies as a method for strengthening applied research. For child protection agencies, state social services divisions, federal agencies, foundations, and nonprofit organizations, predictive analytics could be an effective and efficient mechanism to support system reform efforts. Predictive analytics applied to child protection can help uncover unseen patterns. For example, predictive analytics could be used to explore the type of client characteristics that relate to program effectiveness, to analyze how race or ethnicity interacts with other case information to contribute to disparities, or to demonstrate how different program elements combine to be most effective. Those data findings can be translated into actionable practice and policy insights for protecting children and making families more successful.

Predictive analytics has also been applied to foster care data to identify family characteristics and casework practices that relate to reunification outcomes (Cordero, 2004). In addition, a predictive analytics approach has been used for evaluating specific child protection practices and programs. For example, Lalayants, Epstein, and Adamy (2011) used a predictive analytics approach to evaluate a multidisciplinary consultation program using agency records along with data from other sources. The evaluators were able to use the approach to identify patterns in outcomes by consultation type.

Good Information Analysis and Decision Making Depend on More Than Just New Tools

Of course, the human analytic element remains. The computer cannot do it all—even a supercomputer with a lot of data and a sophisticated algorithm. This is true for both assessing risk and for what to do with it. The interaction between data scientist and practitioner (from agency directors to managers, supervisors, and caseworkers) is critical. The tools alone lack the contextual knowledge and practice insights to provide truly valuable analysis.

Computer algorithms cannot understand decision-making processes, practice implications, or the relative costs and benefits of particular approaches. For example, an algorithm must first be told the best unit of analysis. In child protection, that could be a case, an episode, a child, a family, or a household. Further, an algorithm alone cannot determine the utility of how many gradations of risk might be useful for a child protection agency. If an agency does not have five different levels of service intervention, then an assessment with five levels might not be appropriate. The interaction between data analyst and agency management is necessary.

It may be difficult to know how to effectively apply the predictive analytic models in real-life child protective service agency settings. In many case, the standards and data quality requirements that are needed are simply not in place. Often available data sets are messy, limited, and full of errors. There may not be a sufficiently large number of cases to adequately conduct sophisticated analyses, data may be missing, and data quality may be poor.

Because the insights that can be established through predictive analytics are predicated on quality gathering and inputting data in the field, the potential impact of predictive analytics remains limited. These limitations can be mitigated along with bigger shifts in agency culture, to embrace the increasing role and impact of data-driven insights in both case-level decision making and agency-level strategy. Predictive analytics is one emergent opportunity that might be part of a larger shift in how child protection agencies target resources and interventions to the children and families most in need of them. Predictive analytics, however, will ultimately become more useful for research and applied decision making when taken as one aspect of a larger shift in child protection agency culture.

Discussion

Predictive analytics has received considerable ballyhoo in business applications, politics, and in social services settings. However, when talking about predictive analytics it is important to note that the approach is essentially an outgrowth of data analysis and statistical work that has been occurring for decades. The core distinction from a statistical and methodological point of view is how a question is framed, as well as what data are used and how they are analyzed.

The predictive analytics approach does not begin with a hypothesis. Because there is no hypothesis, analysis focuses only on what can be discovered in the data themselves. Thus, the approach is to begin without a particular idea and embrace any patterns that might emerge from the data. Much of child protection practice is based on theory, e.g., ecological models of maltreatment, however. Although predictive analytics and data mining might identify predictors without starting with hypotheses, some models are developed with child protection systems data that actually fit into a number of different theoretical models. Nonetheless, a shift from hypothesis testing to empirically driven insights can imply a shift in perspective on developing insights from data.

Whether data driven or theory driven, child welfare leaders have a responsibility to be diligent, critical consumers of information. For the examples of predictive analytics discussed above, there is limited public information available. The applications of predictive analytics are not academic endeavors; they often involve private, for-profit partners, and trade secrets may apply. Thus, details of model development, performance metrics, and statistical methodologies are not always available. Based on publicly available information, it is difficult to assess how valid the models are, how often they get it right, how trade-offs between false positives and false negatives were made, and how different approaches compare to each other in terms of accuracy.

Similarly, the reliability of the models in practice appears to be underexplored. There is little information on how workers in practice setting are able to use the models with consistency and apply the results to decision making with regularity and reliability. Further, questions of equity are largely left out of the information for most of the models. Equity can be a problem for predictive analytics applications if it is not examined directly. How predictive analytics models interact with efforts to reduce disparities and disproportionality in child welfare must be explored and understood better.

Finally, utility should be further assessed for the application of the models in practice. It is one thing to be able to accurately identify "who" might be the most appropriate person to target interventions to; it is another thing entirely to identify the right intervention, implement that intervention, and effectively apply that intervention in that case. It is also important to consider, in a world where data are still limited, whether the only practices that should be embraced are those that can be guided by data.

Predictive analytics by itself does not represent a paradigm shift in using data and statistical modeling to guide decision making in social services settings. It represents a more incremental advance: the accumulation of more data and computing technology for deriving practice insights from data. Neither increases in data and computing ability nor the emergence of new analytical approaches should indicate a bending of standards to accommodate the changes. For predictive analytics to fulfill its potential, it must embrace the established criteria of validity, equity, reliability, and usefulness. Any particular application of predictive analytics must be judged according to the criteria discussed in this article.

References

Bogie, A., Johnson, K., Ereth, J., & Scharenbroch, C. (2011). *Assessing risk of future delinquency among children receiving child protection services*. Madison, WI: National Council on Crime and Delinquency, Children's Research Center.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159.

Cordero, A. E. (2004). When family reunification works: Data-mining foster care records. *Families in Society: The Journal of Contemporary Social Services*, *85*(4), 571–580.

D'Andrade, A., Austin, M. J., & Benton, A. (2008). Risk and safety assessment in child welfare: Instrument comparisons. *Journal of Evidence-Based Social Work*, *5*(1/2), 31–56.

Dakil, S. R., Sakai, C., Lin, H., & Flores, G. (2011). Recidivism in the child protection system: Identifying children at greatest risk of reabuse among those remaining in the home. *Archives of Pediatrics & Adolescent Medicine*, *165*(11), 1006–1012.

Department of Children and Families. (2013). *DCF and Eckerd roll out rapid safety feedback to protect at-risk children*. Retrieved from http://www.myflfamilies.com/press-release/dcf-and-eckerd-roll-out-rapid-safety-feedback-protect-risk-children

English, D. J., & Pecora, P. J. (1994). Risk assessment as a practice method in child protective services. *Child Welfare*, *73*(5), 451.

Flaherty, C. W., & Patterson, D. A. (2003). Predicting child physical abuse recurrence: Comparison of a neural network to logistic regression. *Journal of Technology in Human Services*, *21*(4), 93–111.

Florida Department of Children and Families. (2014). *Child Fatality Trend Analysis January 1, 2007 through June 20, 2013*. Retrieved from http://www.dcf.state.fl.us/newsroom/pressreleases/docs/20140106_pressrelease_attachments.pdf

Fogarty, J., Baker, R. S., & Hudson, S. E. (2005). *Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction*. Human Computer Interaction Institute: Carnegie Mellon University.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics*, *26*(6), 822–830.

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*, 103–123.

Johnson, K., & O'Connor, D. (2009). *Implementation and evaluation of Maryland Social Services Administration's Screening and Response Time Assessment*. Madison, WI: Children's Research Center.

Johnson, M. A., Hendricks Brown, C., & Wells, S. J. (2002). Using classification and regression trees (CART) to support worker decision making. *Social Work Research*, *26*(1), 19–29.

Jolley, J. M. (2012). Applying neural network models to predict recurrent maltreatment in child welfare cases with static and dynamic risk factors. *Electronic Theses and Dissertations*.

Lalayants, M., Epstein, I., & Adamy, D. (2011). Multidisciplinary consultation in child protection: A clinical data-mining evaluation. *International Journal of Social Welfare*, *20*(2), 156–166.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*, 145–151.

Marshall, D. B., & English, D. J. (2000). Neural network modeling of risk assessment in child protective services. *Psychological Methods*, *5*(1), 102.

Mindshare Technologies. (2014). *Press release: Mindshare technology applies predictive analytics to child welfare*. Retrieved from http://www.mindshare-technology.com/form_download/announcements/Applied_Predictive_Analytics_Child_Welfare.pdf

Powers, D. M. W. (2012). The problem of area under the curve. In *International Conference on Information Science and Technology*.

Putnam-Hornstein, E., & Needell, B. (2011). Predictors of child protective service contact between birth and age five: An examination of California's 2002 birth cohort. *Children and Youth Services Review*, *33*(8), 1337–1344.

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

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, *29*(5), 615–620.

Rossi, P. H., Schuerman, J. R., & Budde, S. (1996). *Understanding child maltreatment decisions and those who make them*. Chicago, IL: Chapin Hall Center for Children, University of Chicago.

Schwartz, D. R., Kaufman, A. B., & Schwartz, I. M. (2004). Computational intelligence techniques for risk assessment and decision support. *Children and Youth Services Review*, *26*(11), 1081–1095.
Silver, E., & Chow-Martin, L. (2002). A multiple models approach to assessing recidivism risk: Implications for judicial decision making. *Criminal Justice and Behavior*, *29*, 538–568.
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.
Smith, M. (2008). *Building an interoperable human services system*. Smithtown, NY: Stewards of Change.
Thorndike, R. (1985). Reliability. *Journal of Counseling & Development*, *63*(8), 528.
West, P. M., Brockett, P. L., & Golden, L. L. (1997). A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, *16*(4), 370–391.
Wicke Dankert, E., & Johnson, K. (2014). *California risk assessment validation 2013: A prospective study*. Madison, WI: National Council on Crime and Delinquency.