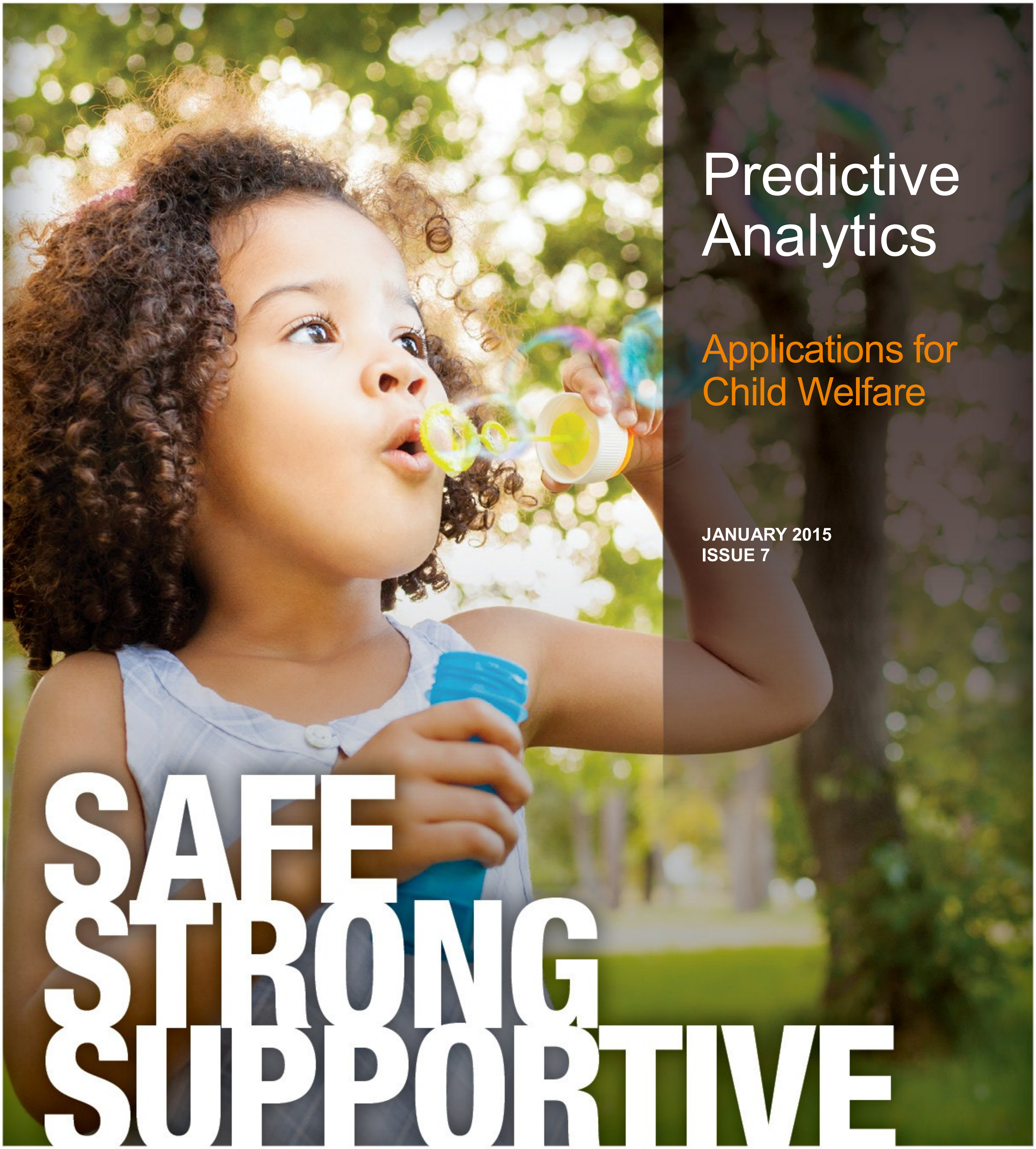


Predictive Analytics

Applications for Child Welfare

JANUARY 2015
ISSUE 7

SAFE
STRONG
SUPPORTIVE

safe children | strong families | supportive communities

FOR INTERNAL USE ONLY

casey family programs | casey.org

Table of Contents

The three main ideas covered in this digest that relate to safe reduction:

- **Predictive analytics is an evolving field and its application and potential for child welfare are just starting to be understood.**
- **Findings from predictive analytic studies can help target preventive services for vulnerable young children.**
- **Using predictive analytics is not an end goal or a solution. It is a tool that needs to be used in combination with clinical social work and a service array that is available to families before, during, and after child welfare involvement.**

INTRODUCTION

By Katharina Zulliger, Co-editor

Predictive Analytics has become a popular term. Yet, while generally agreed to be a method of statistical analysis for the purpose of assessing risk or predicting the likelihood of future outcomes, it has not yet been uniformly defined, particularly as it pertains to use in human services. This issue of the **Practice Digest** offers a number of working definitions informed by perspectives from research, practice and data, and discusses several uses relevant to child welfare.

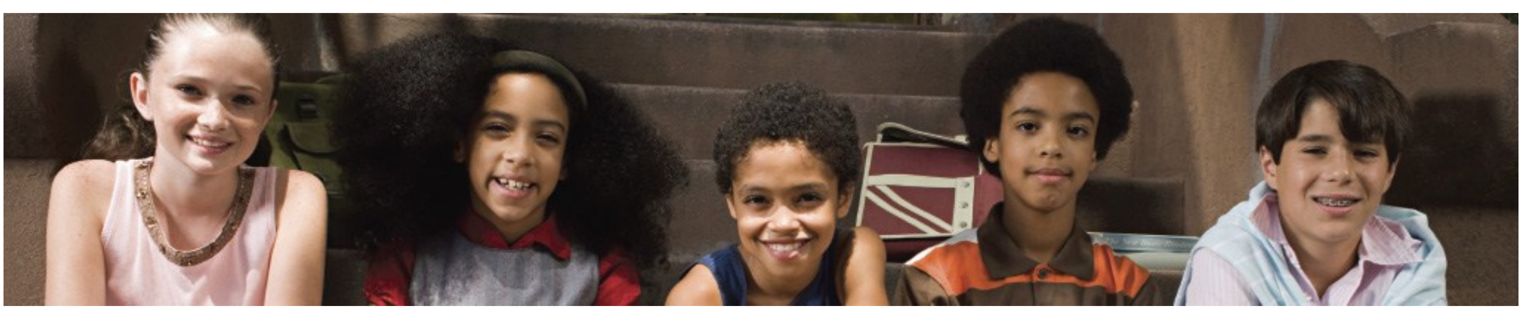
One common use for jurisdictions is to explore ways in which agency data can be employed to estimate risk for children and families they serve. Based on such a model, a referral may warrant intervention or an open case may be identified as needing additional supervision. Population-based efforts, such as using existing data from vital records, education or health care systems to keep children safe are an even more exciting proposition. Combined with advances in computer modeling techniques, linking large and complex administrative data sets in innovative ways has created the opportunity to reform the provision of services to families at risk. However, to quote one of the interviewees featured in this issue: "There always will be 'good' or 'bad' uses of data." So along with a promise of reform comes a obligation for responsible and ethical use of this information.

Moreover, while promising, predictive analytics does not tell us how to respond. For example, it cannot tell us how best to address risk, provide guidance on how to provide services or address the underlying factors that create risk, such as lack of affordable housing, history of maltreatment, or violence in the home and community. There is widespread consensus that prevention services are the key to effectively addressing child maltreatment as a social problem. Having the technology to identify families who need those prevention services most and have the greatest potential to benefit from them holds too much promise for the child welfare field to ignore. Yet, without a responsive, effective and accessible service array, communities and agencies concerned with child welfare will not be able to deliver on that promise.

The *Practice Digest* begins with a general introduction and definition of predictive analytics. Our feature article is a joint interview with **Dr. Barbara Needell** from the University of California Berkeley School of Social Welfare, and **Dr. Emily Putnam-Hornstein**, at the University of Southern California School of Social Work, who discuss the challenges and promises of linking existing administrative data sets for child welfare research and practice. Dr.'s Needell and Putnam-Hornstein explain findings of their research in this area and suggest ways it can inform policy, practice, future research and evaluation. The discussion is followed by a summary of research mentioned in the interview as relevant to this topic.

The practice section discusses considerations for jurisdictions planning to use predictive analytics. The policy contribution explores ethical considerations raised by the potential of targeted outreach to families identified as high-risk.

The **Strategic Perspective** is provided by Linda Jewell-Morgan, Senior Director and Strategic Consultant for Florida. Using Eckerd's model of Rapid Safety Feed-back, the article discusses a practice example deployed within the child protection system and discusses outcomes achieved with the model to date.



What is Predictive Analytics?

By Kirk O'Brien

There is no one answer to the question: "*What is predictive analytics?*" The understanding of predictive analytics and its use are evolving. Searching the internet provides a multitude of answers, many of which are complicated, confusing, and not relevant for child welfare. Before suggesting a working definition, it is instructive to examine some of the characteristics of predictive analytics, which include:

- ***Various statistical methods.*** Many methods fall under the predictive analytics umbrella, including those typically taught in graduate statistics classes (regression, hierarchical linear modeling) and some that require more advanced programming (machine learning). Some predictive analytics employ algorithms, which is following a set of procedures to solve a problem. For example, we each follow a set of procedures to read a pdf. The algorithms differ based on if we read the document on our computer, print it out to read the hard copy, or send to our Kindle to read on our mobile device. The problem is the same: how to read a pdf. The algorithm differs because the procedures differ. Understandably, there is significant complexity involved when a computer algorithm is created to solve a problem related to children in care, such as: "What are the characteristics of youth who age out of foster care?"
- ***Multiple predictors of an outcome(s).*** Predictive analytics looks at how a combination of predictors impacts an outcome. The key here is that multiple predictors have to be involved—otherwise it's a correlation, which is the association between two variables. A correlation does not provide any information on the relative strengths of predictors when examining their combined influence on an outcome. We want to know which variables are more important for predicting a specific outcome.
- ***Opportunities to examine data in deeper ways.*** Child welfare has made great strides in the last decade in understanding the youth who are served, what services they receive, and the outcomes they achieve. Less progress has been made in examining the effect of multiple predictors on outcomes and how information from multiple data sources can be used to examine outcomes (see discussion below).
- ***Re-branding of existing methods.*** Many of the methods used are not new. Re-branding has helped make predictive analytics a buzz word and has provided an opportunity for statistical software companies to take advantage of the excitement around this buzz. Child welfare should exercise caution before entering into relationships with companies making promises around predictive analytics.

So then what exactly is predictive analytics? Although the methods and understanding continue to evolve, predictive analytics can usefully be defined as ***"the practice of extracting information from data sets to determine patterns and predict outcomes and trends."*** [1]

Predictive Analytics has Been Used by Child Welfare Researchers for Decades and its Use is Growing

The purpose of predictive analytics in child welfare is to use information to help understand the youth and families served better, *and* to improve services and outcomes achieved. The analytic methods that contribute to this goal and are part of predictive analytics are not new (although they are being used in new and innovative ways). Child welfare researchers, for example, have used predictive analytics to examine child welfare outcomes such as legal permanency or high school completion. **The question being examined is what variables predict these outcomes?** Researchers collect information including demographics, risk factors, placement information, familial information, etc. and seek to determine which of these variables, when examined as a collective, tell us the most about achieving legal permanency.

There are many other questions about child welfare outcomes that have been examined in the past. Some examples include:

- What youth are most likely to exit foster care without permanent families?[3]
- Which cases are most likely to experience: Failed

Predictive Analytics is Used in Many Fields of Business

We are all contributors, beneficiaries, or targets of predictive analytics: For example, a man walks into a Target department store, irate that his teenage daughter is being sent coupons for baby items. The man's fear? Target was encouraging his daughter to get pregnant, at least that was his accusation. Target's manager apologized which was followed by another apology…from the father, who had recently had some difficult conversations with his daughter, who was due several months later. So how does Target, or any of the many companies and organizations using predictive analytics, know about their consumer's behavior? Simple—they analyze the information that we provide them through purchases, demographic information, and other information (registries we sign up for). In the example above, Target was able to analyze buying behaviors of women on their baby registries and identified spending patterns, which are so strong that Target is almost able to pinpoint delivery dates. The benefit to Target? By sending coupons for items needed during pregnancy, they are encouraging spending behaviors for those and other items.[8]

For companies like Target, the goal of using predictive analytics is to increase profitability. Other companies encourage other kinds of spending behaviors: Amazon suggests what you might also like to buy; Netflix suggests what movies you might enjoy; dating sites suggest who you might prefer as a companion. These companies' bottom line benefits from analyzing the information consumers provide them.

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

reunifications, aging out, crossing over to juvenile justice, or exposure to or becoming a victim of violence?[4]

- Which re-unified children are most likely to re-enter care?

Additional questions of interest to child welfare that can be analyzed using predictive analytics are:[6]

- What services are associated with positive youth outcomes?
- What foster parent characteristics are associated with positive youth outcomes?
- What social/case worker characteristics are associated with positive youth outcomes?

As more data becomes available and more systems can be linked, increasingly sophisticated questions can be asked and answered.

Considerations for Child Welfare Applications

Jurisdictions planning to use Predictive Analytics should consider the following:

Clearly communicate complicated analyses. Not all of the analyses that fall under the predictive analytics umbrella are complicated, but some are. Regardless, when findings from analyses are presented (in reports, journals, presentations), they must be made accessible to disparate audiences. That's not to say audiences must understand the complicated mathematics involved in algorithms or advances statistics, but a general overview of what is being done must be provided in an understandable format.

Implement adaptive analyses. It is important that as more information becomes available (e.g., variables and/or new participants are added), that analyses are adaptive—in other words, they learn from new information. Basically, analyses need to be re-run as new information becomes available.

A good example comes from the Georgia Cold Case Project where a set of variables was used to predict whether youth would exit care without a permanent *family. The cold case project put together teams, which included specially trained lawyers that helped overcome barriers preventing youth from having a permanent family. As the project experienced success, the population changed. As analyses were re-run, a different set of variables became important for predicting whether a youth exited care without a permanent family. These new variables were addressed by the cold case team. Had they relied on the old information, the cold case teams would have been addressing the wrong issues for these youth. As the population changed, the analyses changed, which dictated a change in strategy for serving youth.*[7]

Having a Plan. Jurisdictions must be prepared to act on findings. When engaging in predictive analyses, jurisdictions should have the end in mind. In other words, jurisdictions must be prepared to take action, otherwise they are engaging in a strictly academic exercise. The idea is not to conduct predictive analytics, rather the idea is to use predictive analytics to serve youth more effectively. For example, Georgia's Cold Case Project didn't just run analyses, managers created teams to act on findings. Predictive analytics can be used to tailor services, support decision making, and for a variety of other uses.

---

[1]Edited from http://www.webopedia.com/TERM/P/predictive_analytics.html

[2] http://www.casey.org/Resources/Initiatives/PermanencyRoundtables/

3http://www.nacac.org/adoptalk/coldcase.html

4 http://www.businesswire.com/news/home/20140311006273/en/Mindshare-Technology-Applies-Predictive-Analytics-Child-Welfare#.U8Qbmk2Ybcs

[5] http://www.dcf.wisconsin.gov/children/TitleIV-E/demonstration/PDF/rpm_summary.pdf

Ethical Issues Raised by One Type of Predictive Analytics: Risk Modeling *by Steve Christian*

Depending on its intended use, predictive risk modeling (PRM) may raise ethical questions. At present, no work is being done that identifies individual families for targeted outreach before a child welfare case has been opened. In order to take that next step, a strong ethical framework will need to be in place that includes (at a minimum) protocols for the following: confidentiality of information, careful training of agency staff, development of appropriate policy and thoughtful engagement of families. For example, the research in New Zealand, which linked benefits data with child welfare records, has engendered a discussion concerning some of the following topics:

Does PRM violate people's right to privacy? PRM involves the use of personal information without the consent of the individual in order to generate an individually identifiable risk score. That information may be shared within the child welfare agency, private service providers and others for the purpose of intervening with the family to prevent child maltreatment that may or may not occur, given that PRM also returns false positives.

When risk is identified, what are the ethical obligations of the child welfare agency? What are the obligations of the agency to provide services or to the family if services are refused? From a practical standpoint, will families be likely to refuse services if and when they find out how they were identified and does PRM thus interfere with family engagement?

Does PRM raise issues of due process and fairness to families? Should families have the right to contest a risk score in the same way that they have the right to appeal a substantiation of maltreatment?

How do agencies balance the need for transparency in the PRM process against the likelihood that data may be misinterpreted? The use of certain demographic and economic predictor variables may reinforce stereotypes and prejudices regarding race, poverty and ethnicity.

Is PRM likely to lead to a more risk-averse, coercive and deficit-based child welfare system? Would reliance on PRM make risk the central organizing principle of child welfare, undermining the best practice paradigm that emphasizes family strengths and resiliency?

See, e.g., Kiddell, E. (2014). The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool? Critical Social Policy, published online at http://csp.sagepub.com/content/early/2014/07/23/0261018314543224; Centre for Applied Research in Economics, University of Auckland (2012). Vulnerable Children: Can Administrative Data be Used to Identify Children at Risk of Adverse Outcomes? https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/vulnerable-children/auckland-university-can-administrative-data-be-used-to-identify-children-at-risk-of-adverse-outcome.pdf

Interview with Emily Putnam-Hornstein, PhD, and Barbara Needell, PhD: Research, Predictive Analytics and Data Linkage

By Dee Wilson

Will you please describe how you have been able to use birth data in California to predict CPS reports and study injury related deaths?

EPH: We were interested in using data that would allow us to better understand children reported to child protective services in the context of the larger populations of children in our state – and birth records and death records were a natural place to start. The current work we are doing actually builds on the work that Barbara had done almost a decade prior, which was to use information from birth records to explore the characteristics of children reported to child protective services. Starting with birth records allowed us to think about children's CPS involvement *longitudinally* and from the perspective of the child over time. So rather than an annual snapshot of how many children were reported or placed in foster care, it allowed us to follow an entire birth cohort of children who were born in California and answer questions around how many of those children were reported for alleged maltreatment, substantiated as victims, and/ or entered foster care *at some point* between birth and age five. And what emerged was that a much larger share of children have contact with CPS than annual rates would suggest – and, not surprisingly, there were some very notable differences in the rates of CPS involvement based on information that could be harvested from universally collected data at birth.

We then became interested in whether, analogous to work that was taking place in New Zealand, birth record data could be used to screen an entire population of children and then risk-stratify children based on the likelihood of later involvement with the child protection system. At this point, it is still just a thought exercise. However, the modeling work we have done suggests that with a reasonable degree of accuracy, it is possible to differentiate risk of being reported to CPS among the roughly 500,000 children born in California each year. The next step is to consider whether these data can help us to think more strategically about how we allocate resources at a community level – and whether within communities - it might be possible to use data to ensure we are prioritizing the highest risk families for the limited number of voluntary service slots we currently have for home visiting, early intervention and other programs. Data suggest that there is no reason we cannot move strategically upstream in our prevention efforts, creating services and supports that are tailored and targeted to those families at greatest risk of child maltreatment. And we can provide these services during the peak period of a child's developmental and physical vulnerability, when child maltreatment fatalities are highest.

In the area of injury related deaths, linkages between birth, CPS, and death records allowed us to look at death as an outcome for children under age 5 – both those with and without prior referrals to CPS. And these population-based data indicate that a child's report to CPS for maltreatment is not random, nor is it simply a function of poverty. After adjusting for other risk factors, children reported for maltreatment sustained inflicted fatal injuries at 5.9 times the rate of children who had not been reported. A prior allegation of maltreatment was the single strongest predictor of an intentional injury death – much stronger than poverty, maternal age, child health, or other risk factors. In health and public health research, death is frequently employed as a marker of population-level differences in health and well-being. These data demonstrate that a report of maltreatment to CPS is more than just a marker of poverty—it is an important signal of child risk.

As a technical matter, are unintentional injury related deaths and intentional/maltreatment related deaths the same?

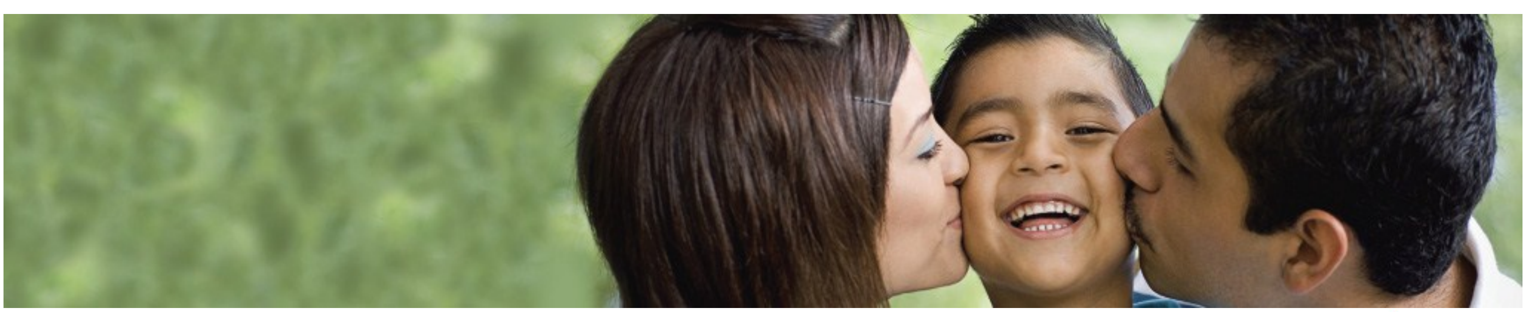
EPH: Our research on child deaths has largely focused on the broad category of injury–related deaths, which includes both those deaths that are deemed accidental or unintentional, and those that are intentional or maltreatment. And these deaths are tragically the same in that a child has died and at least among infants and young children, research suggests these injury deaths are largely preventable. A focus on all injury fatalities is also driven by what we felt we could confidently glean from available data. The coding and recording of deaths in death certificates continues to make it challenging to figure out exactly which deaths were or were not maltreatment related. I should add that when I say most injury deaths were preventable, I am not saying that injury deaths are always the fault of an individual. As an example, we can consider environmentally preventable deaths, where--if we had different safety protocols in place, or had we thought differently about traffic patterns--deaths might have been prevented. But I still come back to the fact that a child died and at a population-level, studying patterns and disparities in injury deaths provides information about the health of families and communities. Linked data from California indicate that children known to CPS are at much greater risk of not only intentional injury death, but also unintentional injury deaths and SIDS and other sleep-related deaths.

From your perspective, what do you think is the greatest significance of your findings to child protection? What are some of the limitations?

EPH: I think that the data from California are significant not because of the specific findings in and of themselves, but rather because they offer an empirical foundation for renewed discussions about maltreatment prevention and coordinated services for families. Linking CPS records to other data sources has helped broaden the discussion by not only underscoring just how many children experience conditions that lead to CPS involvement, but also how poor the outcomes of maltreated children typically are relative to other children. When agencies and researchers present and study case management data from the child protection system, there are a small number of stakeholders who pay attention to the discussions. But when we are able to take that population of vulnerable families and children who are involved with the child protection system and present information about them in the context of other systems they have touched or will touch, importantly, that brings a number of other service sectors and interested audiences to the table. In terms of its limitations, the adage that administrative data are a mile wide, but only an inch deep is certainly true here. By drawing upon information collected through a number of different data systems, it is possible to present a more complete picture of children and families involved in the child protection system. But most of the variables we are able to look at are simply surrogate markers or proxies for individual dynamics we wish we could measure. For example, when we look at children with missing paternity at birth, we see that their risk of later maltreatment is much greater than for other children, even after adjusting for other risk factors. But we do not know exactly *why* that is – or whether there are actually multiple dynamics at play that lead to that increased risk. These data are rich in that they are already being collected and they provide large-scale, epidemiological information that can be used for population-level screening. But they are weak if the goal is to establish causal relationships.

You have worked extensively with child welfare administrative data, primarily in regards to measuring outcomes. What are some of the challenges and advantages to using this data?

BN: As a social worker, I was trained to start with the strengths. My career has been based on the use of administrative data to inform child welfare social work, so clearly I think that there are advantages. A few that stand out are: administrative data often relates to an entire population that is being studied, e.g. *all* children reported to child welfare. It is not a sample, which is an advantage since we are not troubled with issues we often have with sampling, such as sample size, sampling bias or non-representative samples. The fact that you have the entire population to analyze is a great plus. In addition, administrative data is often data that we are not inventing or creating, it is data that already exists, originally for other purposes. CWS/CMS (California's current SACWIS system) is a system that California child welfare workers need to manage their cases. But it also includes information on every child at every step of their way through the system, so that it can be used for administrative data analysis. This is true for many other administrative data sets in which many of the same children and families are often involved. Of course, linkage with birth records is particularly exciting as it provides information on the day a child is born, before any child protection involvement has occurred. But we also have the potential to use education, welfare, and health records, information from other systems that also have population-based, administrative data. We are now developing the skills, knowledge, and relationships to access and link to this data. We can now add to what we know about the children and families in the child welfare system or at-risk children by broadening the work to include linked data. Of course, there are also challenges in the use of administrative data. As Emily mentioned, we have to remember that



administrative data tells us a whole lot about "who, what, when, and where," but little, if anything, about "why." Answering the question of *why* we are seeing what we are seeing in terms of performance and outcomes often has to be gleaned through other kinds of research, or additional analysis, and through a link to practice. The accuracy of the data is often dependent on the primary reason it is being collected, as well as on the motivation to have the data be entered correctly. When it comes to child welfare data, it is often entered by child welfare workers who have a lot of other things to do besides data entry. It is important that people are well trained to enter the data, and that everyone understands the importance of this so the data is the most accurate it can be. Data that is associated with money or with cost tends to be more accurate. When using any data we must understand how good it is, and, in cases when it is not as accurate as it can be, how to work to improve it. Lastly, when it comes to challenges, I am a strong believer in making data public. Of course not in making individual, identifiable data public, but making sure performance outcomes at the state and county level in particular are available to everyone. I am proud of the work we have been able to accomplish in California, which makes much of our child welfare data publicly accessible. When you do make data public, however, people need to have training and education in how to use it, what it means, and what it cannot tell us. It is not infrequent that people abuse the data once it is available. Sometimes people will have an agenda and "fish" in the data to make a point they want to make, and sometimes individuals will make simple mistakes. It is an ongoing challenge to determine the best ways to use data by spotting questionable interpretations of data when they occur and educating the public about appropriate and inappropriate data use.

Predictive analytics and predictive risk modeling have become much used terms. Do you have thoughts regarding the proper use and misuse of these terms in child welfare?

EPH: I don't know that there is a proper use or misuse at this time. These are fairly new terms as applied to child protection and other fields of health and social services. That said, I believe that one of the key aspects of "predictive risk modeling" or "predictive analytics" is the use of computers and automated, statistical algorithms as opposed to assessment tools that require an individual who must be trained, motivated, and willing to objectively and accurately enter data. To provide a concrete example: Structured Decision-Making (SDM) is an actuarial risk assessment tool insofar as it was developed based on statistical models and historic data to attempt to figure out which factors should be considered when helping social workers and child welfare workers make decisions about risk and safety for a particular family. But while statistics and computers were used to develop the weights and items included in the tool, ultimately we hand this tools over to a caseworker to complete. The minute you introduce that human element, the statistical models become confounded in that caseworkers quickly learn what the different weighting assignments are, and therefore what boxes to check if they have already concluded the case in front of them is high or low risk. So the assessment process gets muddied. Clinical judgment is incredibly important. There is also significant power and potential in using statistical models. But if you put them together in one tool, I believe you don't get the best of either and maybe the worst of both. Predictive analytics and predictive risk modeling will never replace clinical judgment, but I do believe that computers crunching numbers in a somewhat of a "black box format" can be useful checks and supports to decision-making.

BN: It reminds me in some ways of the struggle when we started to be interested in talking about racial disparity and disproportionality, and the arguments and differences that emerged in simply defining the terms. Finally the field, for the most part, zoomed in on some definitions that were clear and distinctive, so that we could move forward with the work. We might have much the same opportunity now with predictive analytics. We know that the term is used in many different ways and is used to mean different things. Perhaps it is time to begin defining predictive analytics the way Emily has done, where she distinguishes it from clinical judgment tools. Maybe those of us trying to bring people along in this work would want to consider promoting such a definition.

We understand that there has been some "pushback" to the use of predictive analytics to target high risk families for voluntary services, namely potential stigmatization of clients and issues of proper response by child welfare agencies. What is your response to this resistance?

BN: It is not the data's fault. There always will be "good" or "bad" uses of data. As a researcher, I am hard pressed to find a piece of data that I don't like or think I should not be allowed to use or access. But the question is: once you have access to these powerful tools, how do use them? It would be great if everyone who has access to the kind of information we have in these databases worked from a strength-based perspective and tried to figure out how to use this new and enhanced information to build a better future for children and families.

EPH: We are just beginning to have some very important conversations about the ethics of using data that was collected by one system to potentially target and serve families at high risk of appearing in a downstream system. Likewise, we are just beginning to explore whether we can improve decision-making through predictive analytics. I think some of the “pushback” is not just around stigma, but also the belief that any discussion of targeting somehow undermines efforts to garner public support and enhanced resources for universal services. But if we have services that are not funded at the universal level, I believe we have a moral and fiscal obligation to ensure that available service slots go to the children and families where data would indicate there is the greatest need – or to children and families where data suggest the impact and benefit will be greatest. Frankly, I don’t think the conversation should be framed as one versus the other. But the reality is that we have limited service slots in most programs. And even in countries where universal services exist, there is an appreciation that enhanced, targeted supports are often required. I worry less about stigmatization and debates around universal services – and more about whether the slots we do have go to those families that most need them.

Are there other policy issues that are raised by use of predictive modeling?

BN: We mentioned one before, as an example, using predictive risk modeling for fatality prevention. While that is something computer software allows you to do, the caution that has to be used if you use modeling for individual risk assessment is not always clearly understood. Using a model is different than being able to identify what will happen to a particular child in a particular family.

EPH: From a policy standpoint, we should be exploring how we can better use administrative data to answer all sorts of questions. Although modeling processes behind predictive modeling are complex, the statistics that tell you how a model performs and whether it provides a good means of risk classifying and risk stratifying populations of individuals are fairly basic. The good news is that as long as there is transparency, we should quickly be able to tease out where models are useful. I feel strongly that the usefulness of predictive risk modeling when looking at an extremely rare event, such as a fatality, will be limited. From a prediction standpoint, what we have been exploring in California is far less challenging because we are talking about a large and heterogeneous population of children, all children born in the state. Based on the data we have in California, about 15% of kids are reported to CPS before age five, so it’s not an exceptionally rare outcome. So the question is simply, how do we identify who those 15% are likely to be so we can provide services? Which is a different question than trying to identify from a very high-risk population of children known to child protection, which children are most likely to experience the worst type of outcome, a near fatal or fatal injury. But I could be completely wrong and the good news is that whether we can accurately predict a fatality is an empirically answerable question. That said, I have been a disappointed with some of the early examples of vendor attempts to develop predictive analytics for child fatalities because basic classification and model performance statistics have not been shared in the reports released to the public. Transparency and ensuring that agencies have the information that they need to be informed consumers when different predictive analytic tools are pitched to them is critical.

> *We have a moral and fiscal obligation to ensure that services go to the children and families where data indicates that they are most in need.*

What interventions or services do you believe parents identified as “high-risk” at their child's birth should be offered?

EPH: Thinking of targeting services right at birth, or even better prenatally and for the first few years of a child’s life, is an incredible opportunity. However, I suspect that we will need to rethink and revisit much of what those early intervention programs look like in terms of the intensity, duration, and dosage. If our goal is to target services to the highest risk families who otherwise will end up involved in the child protection system, we will also need to re-think how we effectively engage families in voluntary services and ensure that what is being offered holds real value. I don’t know that right now we have a good menu of interventions and services for this highest risk group. Linked data can help us better understand who is currently being served by existing programs, and how those programs function, and thus help us answer the question of what we need to do to tailor prevention and early intervention services to families who have an even higher concentration of risk factors present.

BN: I agree. We are just at the beginning. We have never had the quality of evaluation research that we need when it comes to CPS interventions. We don't know nearly enough about what works and what doesn't work and why, and that knowledge is something that has to be built on the ability to identify which families are at highest risk.

Dee Wilson: We were surprised recently here in Washington State to find out that many of the CPS families are never going to be offered the services that have been added through federal home visiting funds. Home visiting services are not being provided to most CPS families because of selection criteria that makes many of the families ineligible.

EPH: We have experienced the same thing in Los Angeles County with a program intended to be delivered to young women who are in the foster care system and pregnant or parenting. The collaboration was with Nurse Family Partnership, but it turns out most of the young women in foster care were not eligible for the program. The problem was that we were either not identifying these young women early enough in the pregnancy to refer them to the program during the specified window during which services must begin, or the teen was not eligible to receive services as the program was designed for first-time mothers.

Understanding what predictive analytics can tell us about children and families strengthens social worker practice.

Data has been used for a long time in child welfare and public health, but is becoming more sophisticated, more personal and more linked. How do you see the integration or balance of more traditional social work and mathematical models to identify individuals at risk?

BN: When we started trying to convince child welfare caseworkers that data would be useful to them, we would get pushback along the lines of "I didn't become a social worker to do statistics or data entry." We have seen an enormous change in that attitude over time, and now have an even greater opportunity to use predictive analytics to provide a certain kind of information that can free up social workers to do what they are trained to and what only they can do: clinical assessment and judgment. Years ago, we began to show how understanding basic outcomes and performance goals of their agency was useful and important to any social worker in their day-to-day work. Understanding what predictive analytics can tell us about children and families also strengthens social worker practice.

From my understanding, the predictive accuracy of work completed in New Zealand is very impressive. What data did they have available and what how did they approach their work?

EPH: New Zealand's predictive risk modeling is based on incredibly rich, integrated data. Models were developed based on hundreds of variables spanning data collected over time and across systems. And these models were then used to divide or stratify children into "risk-deciles," (i.e. ten groups of equal frequency) based on the likelihood the child would be substantiated for abuse or neglect during the first few years of life. And the findings are promising. But what is striking is that even through there are different data sources being modeled in different countries, so essentially different populations of children who are reported or substantiated, we have come up with very similar statistics in California. It is important to keep in mind that the similarity in results suggests there may be a ceiling to how well we can predict a complex outcome such as abuse and neglect. New Zealand had significantly more data than we did, but we achieved roughly similar results by simply linking birth records to child protection data. I bring this up because while I certainly hope jurisdictions here in the US soon find themselves with integrated data similar to that of New Zealand, even a single data source may produce results worth pursuing. The other thing to keep in mind is that we don't know how well we may already be informally triaging families into services without any predictive modeling techniques. We don't know for sure if at the local level, through nurses in hospital or other service providers, we may already recognize which families need support and may already be connecting (or trying to connect) them with services. It could be that when we go into a hospital, run a predictive risk model and tell the staff: here are the 20 families you should target for services, these families have already been informally identified as high-risk, but hospital staff just didn't have the right service match for them. So one thing to consider with predictive analysis is not only the classification potential, but how well our statistical approach performs relative to other means of identifying high risk families early in a child's life.

Setting aside prevention, what are some of the practice implications of risk prediction and data linkages for child welfare?

EPH: I am not sure I have the greatest insight because my background in frontline practice is limited relative to so many others in the field. But as mentioned earlier, freeing up child welfare caseworkers to focus on their clinical work with families is a definite advantage. If we can access and link client data from other systems, then less of that information needs to be manually entered by caseworkers. Likewise, if we can use already collected data to make some initial risk assessment and triaging decisions, caseworkers can focus more of their time on engaging the family and assessing strengths and needs that may not be discernible from data. I believe you would be hard pressed to find any caseworker who wouldn't be excited about less data entry and paperwork. From a practice standpoint, we can imagine implementing a predictive risk modeling tool at the hotline to help risk-stratify referrals based on existing data. An agency could then send out more experienced investigators for the cases that are more complex and where the risk is greater. It could be that there are two caseworkers who go out together on certain complex cases based on what the predictive risk modeling indicates. Or, it could be that from a practice standpoint more supervisory oversight for certain cases determined to be high risk through modeling is implemented. For example, one could imagine deploying a model focused on risk-stratifying families with prior referrals based on their likelihood of a second referral within 5 years. In the case of a family identified as high-risk, if a caseworker decides to close a case, or the hotline worker chooses not to open a case for investigation, or really at any chosen decision point, those cases could require a supervisor to sign off on the decision. So the risk score would provide a statistical "check" on clinical decisions. The fundamental practice question is: "How do we use available data to more accurately and efficiently to triage cases, so that we are able to do better work with the cases that need a preventive intervention or need active CPS involvement?"

BN: A powerful variable in these data sets often is "address." Location information can be used to consider clustering of issues not at the individual level, but at a neighborhood level. This level of analysis can provide information about the need for different services in different neighborhoods within a community.

EPH: Barbara raises a really important point. A colleague and friend recently criticized my continued focus on these data as a means of targeting individuals rather than thinking in terms of targeted place-based or neighborhood initiatives. Certainly, these data can also be used to become even more nuanced and sophisticated in the identification of "hotspots" where there are pronounced service needs. It also goes without saying that Barbara and I think there is tremendous value in using linked data and predictive risk modeling to support program evaluations and the development of cross-system outcome and accountability indicators.

You have done recent work on intergenerational transmission of child maltreatment. What are the major findings from this research?

EPH: Intergenerational findings underscore the importance of thinking intentionally about how services and supports are provided to young mothers.[1] In California, we looked at teen mothers who gave birth in our state. This was a population level examination – we used birth records to identify all teens who gave birth during a two year period – and we then linked information for both the teen mothers and the children of teen mothers to abuse and neglect records. First, we looked backwards to identify which teen mothers had been reported and/or substantiated as victims of child abuse and neglect before becoming pregnant. We looked back to age 10 and defined this as "recent" involvement based on data we had available. Next, we followed all children born to teen mothers from birth to age 5, which is considered a developmentally critical period and the period in which fatality risk is highest. We found a powerful relationship between teen mothers' own history of recent CPS involvement and the likelihood that their child would become involved in the system. The relationship was much stronger than we had expected to see. In fact, the strength of the relationship it is the single finding that has most caught me by surprise to date. We already knew that children of teen mothers were at higher risk of CPS involvement. But children of teen mothers without a child protective history did not look all that different from the "regular" population of children born to non-teen mothers in comparable socio-demographic circumstances. It was the children whose mothers had a history of recent child abuse and neglect who had staggeringly high levels of future child protection involvement. In discussing these findings with others, anecdotally, those teen mothers who seem to be doing the best can point to a relative, mother, aunt, grandfather who have been able to support them, suggesting the importance of an extended family to support the teen. In contrast, recent involvement with CPS may suggest something about the vulnerability of the broader family network. So again, the question remains how do we target support so we can break that cycle of CPS involvement?

Thank you for your time.

1 http://www.hiltonfoundation.org/teenparentsreport

EMILY PUTNAM-HORNSTEIN, PH.D.

Dr. Putnam-Hornstein is a Assistant Professor at the University of Southern California School of Social Work. She joined the faculty in 2011 after completing her doctoral studies at the University of California, Berkeley. She currently directs the Children's Data Network, an agency, university and community collaborative funded by First 5 LA and the Conrad N. Hilton Foundation. The Children's Data Network provides a platform for the linkage and analysis of large-scale, administrative data sources to inform children's policies and programs. Dr. Putnam-Hornstein also maintains an appointment at the UC Berkeley California Child Welfare Indicators Project. She graduated from Yale with a BA in Psychology, received her MSW from Columbia University, and earned her PhD in Social Welfare from UC Berkeley. In 2014, she received the Commissioner's Award from the Administration for Children, Youth, and Families.

BARBARA NEEDELL, PH.D.

Dr. Needell is a researcher at the University of California Berkeley School of Social Welfare. As principal investigator of the California Child Welfare Indicators Project (CCWIP), which is funded by the California Department of Social Services and the Stuart Foundation, she has worked extensively with statewide and county specific administrative data. Dr. Needell and her team at the University of California, Berkeley collaborate with state and county colleagues to produce and publicly disseminate the data used to support the California Child Welfare Outcomes and Accountability System. She is a member of California's Child Welfare Council, and co-chairs the Data Linkage and Information Sharing Committee. Dr. Needell is the recipient of the 2008 Peter Forsythe Award for Leadership in Public Child Welfare from the American Public Human Services Association.

The Research Findings in Brief: Predictive Analytics in Action

Title. Report of Maltreatment as a Risk Factor for Injury Death: A prospective Birth Cohort Study. Putnam-Hornstein, E. (2011)

Question. Is a report to child welfare services a risk factor for dying from intentional or unintentional injury regardless of other risk factors present? Also, does it make a difference if injury death is intentional or unintentional?

Answer. Yes, being reported to child welfare is a risk factor. Children who are reported are at higher risk for both types of injury death.

Why it Matters. The findings of this study suggest that being reported to child welfare services, *regardless of whether or not child abuse or neglect is determined to have occurred,* is a major risk factor for injury death before the age of five. Focusing child fatality prevention efforts on this group of children may be beneficial.

Study Findings. Children, born in California between 1999 and 2006, who had been reported to child welfare services were over five times more likely to die from intentional injuries before the age of five compared to children of similar socio-economic status, who had not been reported. Children with prior reports were also more than twice as likely to die of unintentional injury death.

Title. Predictors of child protective service contact between birth and age five: An examination of California's 2002 birth record. Putnam-Hornstein, E. and Needell, B. (2011)

Question. When matching birth and child welfare records electronically, can information included on a birth certificate, such as paternity, or weight of child, predict future referrals before the age of five? If yes, which of these variables, or which combination of these variables, is likely to provide accurate predictions?

Answer. Yes, data collected from birth records can predict which children are at greatest risk of a report.

Why it Matters. These findings suggests that it is possible to identify a risk pool *using large existing datasets* before abuse and neglect has occurred and to target *prevention* services accordingly.

Study Findings. 11 variables from the birth records of all children born in California in one year emerged as predictive of the likelihood to be referred to CPS during the first five years of life.

Title. Children in the Public Benefit System at Risk of Maltreatment. Vaithianathan, R., Maloney, T., Putnam-Hornstein, E., Jiang, N. (2013)

Question. Is it possible to use public benefit data linked with child welfare data to create a predictive model that will accurately classify children ages 0-5 by risk of experiencing child maltreatment?

Answer. Yes, it is possible to build a model that can predict risk-levels with some accuracy.

Why it Matters. The findings suggest that it is possible to group children into risk pools using automated predictive models. These risk scores can be used in combination with clinical assessments to target *prevention* services to those at highest risk.

Study Findings. A model including 132 variables from public benefit data was able to group children who had any contact with the benefits system into risk categories. Children in the top 10% of risk were 25 times more likely to be substantiated for maltreatment than those in the lowest 10% of risk. 47% of children in the highest risk category had substantiated child abuse and neglect.

Title. A population-based longitudinal examination of intergenerational maltreatment among teen mothers. Putnam-Hornstein, E., Cederbaum, J. A., King, B., Lane, A., Trickett, P. (in press)

Question. Are children of adolescent mothers who have experienced child abuse and neglect more likely to have been reported for maltreatment by the time they are five years old?

Answer. Yes, when linking birth and child welfare records a history of maltreatment after age 10 for the mother increased the child's risk of to be reported to child welfare services by age five.

Why it Matters. The findings suggest that adolescent mothers with prior maltreatment history are distinctly more vulnerable than other teen parents, and that targeting pregnancy prevention as well as parenting support to this subpopulation is advisable.

Study Findings. Teen mothers in California who had experienced substantiated or alleged maltreatment between the age of ten and becoming pregnant were twice as likely to have their children reported to child welfare by age five. Substantiation rates were also higher for these children when compared to children of teen mothers without reports or substantiations. Only teen mothers not in foster care at or after conception were considered.

Considerations for Jurisdictions Planning to Use Predictive Analytics

by Kirk O'Brien and Stephanie Jones Peguero

With the move to increase analytical capacity within human service organizations, it is important to weigh when and how predictive analytics would be useful in understanding how to improve services, practices and outcomes. In child welfare, states and counties are working collaboratively across systems to examine ways to share and access child and family level data in order to delve deeper into outcomes through predictive analytics. There is much work involved in scaling data sharing and analytical ability within and across agencies, and current efforts exist that use predictive models to target specific outcomes. The rise of predictive analytics may seem to place greater burden on existing agencies, but according to the authors of "Analytics at Work", the goal of predictive analytics is for **improved capacity for decision making**, not more reports, more portals, more scorecards or more drill-downs. The American Public Human Services Association (APHSA) developed a capability assessment model that frames analytics in the context of human services. There are three levels of analysis with increasing levels of sophistication: basic, advanced and leading. Eight domains have been identified in the model and predictive modeling is one of the more advanced domains.[1] Their definition provided for leading use of Predictive Analytics may be helpful to understand its potential to enhance decision-making within an agency:

> ***"A data scientist reviews the data available, internally and externally, structured and unstructured, then assesses what is meant by each piece of data. …. Collectively, questions are developed, the precise answers to which are extremely important to the organization. Then, through a series of iterative regression analysis using many variables, an algorithm is developed and validated, and when applied, the best predictor of the question's answer is known. The algorithm is periodically reviewed, retested, and updated.*** "

Although the potential for improved practice is great, challenges associated with data use remain.

The following strategies may be helpful for jurisdictions planning to explore the use of Predictive Analytics further:

- ***Share Data Effectively.*** SACWIS systems are not the only data systems collecting information on youth that may provide valuable insights on what practices and services lead to positive outcomes. Court and education data, for example, may provide critical information for what leads to positive outcomes. Although establishing effective data sharing arrangements can be daunting, they provide the means to a necessary end—critical data for predicting outcomes. The Children's Bureau has produced a video that addresses some of these concerns (http://www.acf.hhs.gov/programs/cb/assistance/program-evaluation/virtual-summit/data-sharing-partnering)

- ***Efficiently link data files.*** Once data sharing agreements are in place, the data have to be connected so that data for youth from one system (SACWIS) can be matched with the same youth's data in another system (Court). Software advancements have made this process easier. Further, as more systems work together, they can often incorporate unique youth identifiers in each other's' systems to simplify connecting data.

- ***Collect data to answer the right question.*** Depending on the question to be addressed, the data best suited to answer particular questions may not have been collected. For example, to answer the question of what foster parent characteristics are associated with positive youth outcomes, data on foster parent characteristics have to be collected. It is critical to have the right data to answer the question—this may require additional data collection.

- ***Collect valid data.*** An essential component to data analyses is having valid data. First, data collected in SACWIS or other systems must be complete. There is often significant information missing because it was not entered. Second, incorrect data are often entered (e.g., incorrect birthdates or case open dates). Ultimately, these issues affect the quality of any analyses. To overcome this barrier, greater care in entering accurate data must be taken.

- ***Lessen worker burden.*** The quality of data entry is often a result of the burden placed on caseworkers responsible for documenting youth information. Large caseloads and inefficient data entry systems put a strain on the ability to enter quality information. In addition to addressing these issues, child welfare jurisdictions must not only emphasize the importance of entering valid information, but they must provide information back to the caseworkers entering it, and demonstrate how it can be used to serve individual youth and families more effectively.

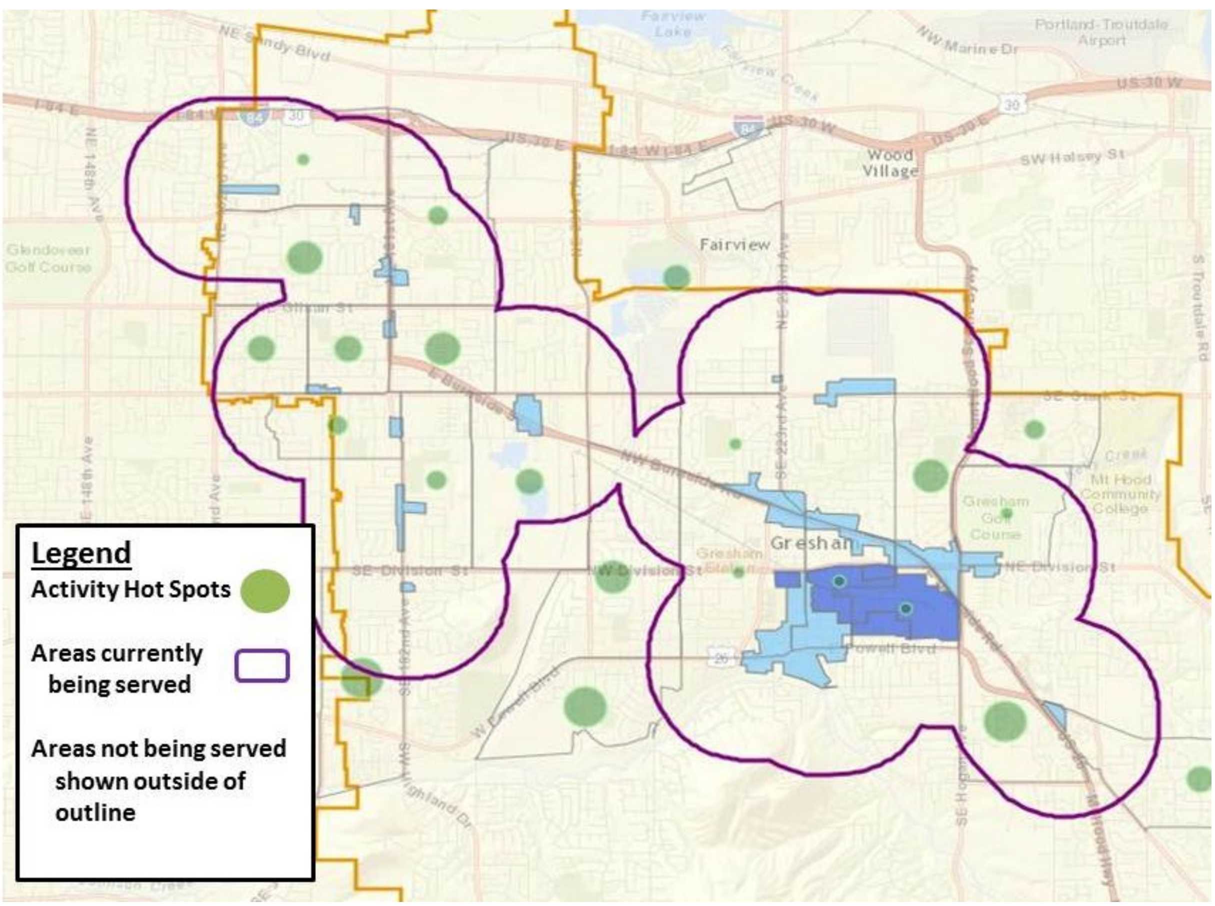
Finally, findings from predictive analyses must be interpreted with caution. For example:

- ***Findings from predictive analytics are only as good as the data that are included.*** For example, if we want to predict whether a youth achieves legal permanency or not, and we don't know their placement history (number of placements, length of time in care), we may be leaving out critical information that can inform legal permanency decisions. As a result, some variables may appear significant because other variables are left out which may be more important.
- ***Predictive analytics is not a search for significance.*** It is a sloppy and unscientific practice to throw every variable into a predictive analytics model. It is a good practice to include variables based on theory/psychological models (e.g., attachment theory). A good theory helps make decisions about what to include/exclude and how to interpret findings.
- ***Predictive analytics is not going to perfectly predict the future and it must be combined with sound clinical judgment.*** For example, findings from predictive analytics cannot determine which individual children will or will not experience abuse or neglect. Findings may say what the likelihood of an event occurring is, which may sometimes be correct and may be wrong at other times.

---

[1] http://www.aphsa.org/content/dam/aphsa/pdfs/NWI/FINAL_NWI%20Analytics%20Capability%20Roadmap_4.17.14.pdf

Using Location Data for Predictive Analytics



As Drs. Needell and Putnam-Hornstein explain in the interview, address can be a powerful variable, and can be used to help identify clusters of need on a neighborhood level and thus help identify target areas for preventive services. GIS tools, such as performing filter queries based on descriptive statistics of prior maltreatment episodes can be useful in combination with predictive analytics. The image to the left shows a hypothetical example of spatial analysis.

Source: Casey Family Programs Data Advocacy

The Strategic Perspective

Predictive Analytics and Rapid Safety Feedback: A Powerful Combination Resulting in Promising Outcomes *By Linda Jewell Morgan, Senior Director Strategic Consulting*

The practice of using patterns found in historical and transactional data to identify risks and opportunities to improve outcomes for children and families is not new to child welfare. However, targeting individual children and their families is only effective when the intervention utilized is appropriate, timely and well-executed.

In Florida, a promising set of outcomes is being attributed to a combination of predictive analytics and a method termed **Rapid Safety Feedback (RSF)**. Eckerd, a community based provider who pioneered this combined approach is attributing the following outcomes to the implementation or RSF:

- Repeat maltreatment of children while placed with their families under child welfare supervision decreased by 22% from 7.09% to 5.58%.
- The number of RSF cases where QA reviews indicated a safety concern that triggered a coaching session has decreased by 21%. This may indicate that case managers and supervisors are applying strategies learned in RSF coaching sessions to other current and new cases.
- No child abuse-related fatalities have occurred since implementation of the model.

In July, 2012, Eckerd became the community based care (CBC) lead agency in Hillsborough County, Florida after the county had experienced nine child abuse-related fatalities in the prior 24 months. As the CBC lead agency, Eckerd is responsible for the provision of all child welfare services following child protection investigation.

Applying learning from a multidisciplinary quality and safety review on 1,500 open in-home cases and reaching out to national experts, Eckerd implemented Rapid Safety Feedback in January, 2013. Rapid Safety Feedback (RSF) utilizes predictive analytics to identify open in-home cases that are at high risk of poor safety outcomes. Case factors include: children under age three, a paramour in the home, history of substance abuse and/or domestic violence and parent(s) previously in the foster care system. Eckerd also contracted with MindShare to provide software to overlay Florida's SACWIS system to allow real time data and dashboards.

Quarterly Quality Assurance (QA) reviews of cases identified through predictive analytics assess 'real time' case manager performance on nine critical safety-related practices.[1] On cases where the QA review identifies safety concerns, QA staff provides supportive coaching within one business day to the case manager and supervisor. Case management and QA staff then jointly develops an action plan to mitigate identified safety concerns. Finally, RSF action plan tasks are tracked electronically to completion by Eckerd QA staff. RSF reviews occur quarterly until the youngest child in the family turns three or the case is closed.

Describing this innovative quality assurance strategy, Bryan Lindert, Eckerd's Director of Quality Management states:

> *"…intervention and joint accountability are best achieved through multiplying the number of coaching opportunities, engaging in dialogue about findings instead of issuing didactic reports, and reviewing the case when critical decisions can still benefit from external input."*[2]

In 2014, the Florida Department of Children's Services (DCF) initiated the implementation of Rapid Safety Feedback statewide on child protective investigation cases. While sufficient time has not passed yet to analyze the impact on federal child welfare outcome measures, the number of RSF cases where QA reviews indicated a safety concern that triggered a coaching session has decreased by one third since the process was first implemented. This may be explained by staff applying the new strategies learned in RSF coaching sessions on all their cases, as well as by training on Florida's new safety model, which is being implemented concurrently.

[1]Quality safety planning, quality supervisory reviews, gathering input from external sources, validating behavior change, and the quality and frequency of home visits, background checks, case planning and follow-up.

[2]Lindert, Bryan, "(October 2014) "Eckerd Rapid Safety Feedback: Bringing Business Intelligence to Child Welfare", Policy & Practice, American Public Human Services Association, Washington, DC.

GLOSSARY OF TERMS

Computer-based Algorithm: A step by step formula for calculations, data processing and more advanced computing functions. Algorithms can be used to find patterns in data. The patterns predict outcomes with varying levels of accuracy, but cannot explain the causal link.

"Big" Data: large data sets, sometimes from multiple sources or a collection of data sets that can be analyzed to reveal patterns, trends, and associations, especially relating to human behavior and interactions. Typically, the term is used in reference to complex data sets that contain so many data elements that it becomes difficult to process them with traditional methods. Instead, patterns are often identified by computers using algorithms.

Predictive Analytics: The practice of extracting information from data sets to determine patterns and predict outcomes and trends. Predictive models typically analyze current and historical data to produce easily understood metrics (quantifiable measure that are used to track and assess such as rates of child protection reports or substantiated cases). For example, these scores rank individuals or families by likely future performance, actions or risk.

Predictive Risk Modeling: A specific type of predictive analytics focused on using data patterns to identify predictors of risk and assign risk categories based on these patterns to individuals or families.

Predictor Variable: a piece of data that is examined to determine its relationship with an outcome.

Regression Analysis: describes one statistical method for determining the strength of relationships among a set of predictor variables and outcome(s). Its focus is on the strength of the relationship between outcome variables (result) and one or more predictor variables. The strength of the relationship between predictor and outcome takes into account other predictors that may also be influencing the outcome.

About the Casey Practice Digest

The Digest provides Casey Family Programs staff with information on research, policy, and practice developments. Each issue is centered on a topical theme, and includes interviews with expert sources, data displaying current trends at a high level, policy and practice considerations, as well as resources for further exploration. The Practice Digest is produced by Knowledge Management and depending on the topic includes editors from Child and Family Services, Data Advocacy, Knowledge Management, Public Policy, and Research Services.

This issue includes contributions by Kirk O'Brien, Research Services; Stephanie Jones-Peguero and Delia Armendariz, Data Advocacy; Linda Jewell-Morgan, Strategic Consulting; and Steve Christian, Public Policy. Edited by Dee Wilson and Katharina Zulliger, Knowledge Management.

Program highlights and examples are descriptive and are intended to provide practice considerations for other jurisdictions. Programs are chosen based on available information and inclusion in the digest does not present an endorsement by Casey Family Programs of these programs over others. For additional information, please contact the local Strategic Consultant or contact identified in the contribution.

For questions or feed-back please contact kzulliger@casey.org or call (206) 352-4230.