Journal of Family Violence
https://doi.org/10.1007/s10896-019-00074-y

ORIGINAL ARTICLE

Preventing Infant Maltreatment with Predictive Analytics: Applying Ethical Principles to Evidence-Based Child Welfare Policy

Paul Lanier[1,2,3] · Maria Rodriguez[4] · Sarah Verbiest[1,2,5] · Katherine Bryant[2,5] · Ting Guan[1] · Adam Zolotor[3,6]

Abstract

Infant maltreatment is a devastating social and public health problem. Birth Match is an innovative policy solution to prevent infant maltreatment that leverages existing data systems to rapidly predict future risk through linkage of birth certificate and child welfare data then initiate a child protection response. Birth Match is one example of child welfare policy that capitalizes on recent advances in computing technology, predictive analytics, and algorithmic decision making. We apply frameworks from business and computer science as a case study in ethical decision-making in child welfare policy. Current Birth Match policy applications appear to lack key aspects of transparency and accountability identified in the frameworks. Although technology holds promise to help solve intractable social problems such as fatal infant maltreatment, the decision to deploy such policy innovations must consider ethical questions and tradeoffs. Technological advances hold great promise for prevention of fatal infant maltreatment, but numerous ethical considerations are lacking in current implementation and should be considered in future applications.

Keywords Family violence · Child welfare · Infants, decision making

Algorithms and the data that drive them are designed and created by people – There is always a human ultimately responsible for decisions made or informed by an algorithm. "The algorithm did it" is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes.

-Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) ethical premise

Introduction

Similar to most sectors of government in the United States, the public child welfare system (CWS) is investing in the use of new technology and shifting toward data-driven, computer-powered policy and practice. Over the past decade, greater availability of "big data" and high-speed machine learning and computing has fostered the perception that technology can help solve some of the *wicked* child welfare problems (i.e., persistent problems that defy ordinary solutions; Chouldechova et al. 2018; de Haan and Connolly 2014; Kulkarni et al. 2016; Russell 2015). Increased use of data and computing technology signals at least two major shifts or innovations in child welfare policy and practice. First, the value of clinical prediction based on individual experience and training is replaced by an increasing availability of mathematical "actuarial" prediction and judgement (Brauneis and Goodman 2018, p. 111). In other words, the numerous human decision points involved in child welfare services are increasingly informed by, and in some cases determined by, computer

✉ Paul Lanier
planier@unc.edu

1 School of Social Work, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

2 Jordan Institute for Families, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

3 Injury Prevention Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

4 Silberman School of Social Work, Hunter College, New York City, NY, USA

5 Center for Maternal and Infant Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

6 School of Medicine, Department of Family Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Published online: 07 June 2019

output. Second, taking a cue from business, as new prediction tools and technologies have advanced, leaders have seen an opportunity to shift from "reactive" to "proactive decision making" (Banerjee et al. 2013 p. 6). For CWS, proactive planning means a shift toward prevention. These technological advances present a potentially transformative opportunity to modernize a notoriously antiquated system, and thereby, create a higher-quality and more effective public service.

Recently, a growing number of experts have begun sounding alarms at the unintended consequences of poorly implemented technology-based innovations in social services broadly, and child welfare specifically (Eubanks 2018; Harcourt 2007). Few academic papers have adequately described predictive analytics or machine learning processes in a way that is understandable to child welfare researchers, policy makers, and practitioners. This collective lack of knowledge prohibits thoughtful discussion of the various ethical considerations involved in the use of this technology. Given the considerable pressure on CWS policy makers, it is easy to be seduced by the promise of technology without considering whether policy shifts uphold professional and societal values.

This article is intended to introduce readers to predictive analytics, machine learning, and their current application to CWS decision-making processes. The enthusiasm with which many CWS leaders have embraced predictive analytics in general, and algorithmic decision making in particular, appears to be taking place in the context of a dearth of information regarding its key assumptions, requirements, and suitability for child welfare related tasks. We introduce several frameworks from outside of the field of family violence that are crucial to scrutinizing the application of business models to government services. Last, we apply these frameworks to a specific child maltreatment prevention policy known as Birth Match. The Birth Match policy is intended to prevent severe and fatal infant maltreatment by using CWS data to identify newborns who are likely at risk for maltreatment based on one or both parents' prior CWS involvement. Birth Match exemplifies the two features of innovation in the shifting policy landscape of CWS previously introduced: (a) computer output augments (or replaces) clinical judgment, and (b) the policy goal is proactive by design. Our goal is to help the field develop a more critical appraisal of policies founded in algorithmic decision making and to better understand which questions to ask when considering practice and policy change.

The State of Predictive Analytics

What Is Predictive Analytics? For centuries, both scholars and lay people have been occupied with the desire to predict the future and thereby gain greater certainty in life (Christian and Griffiths 2016). Accurate prediction of future events hinges on the ability to gather and process large amounts of information; whether the goal is to predict the weather tomorrow, who will win a baseball game tonight, or changes in the stock market next year. The current predictive analytics moment has been propelled not only by significant advances in computing, artificial intelligence, data archiving, and storage and retrieval capabilities, but also by the drive toward evidence-based interventions and policies. Further, the global proliferation of smart phones, social media, and the Internet of things (Gershenfeld et al. 2004) have converged to produce 98% of the world's data in the last two years (Marr 2018). The emergence of big data science has made it possible to collect millions of data points about any one person in the world, and to use those data in a number of mathematical models to predict behavioral outcomes of interest. For example, an online social media site can predict which advertisements will be most successful (i.e., effective in producing a sale) for each individual user based on the person's prior purchases and other online activity. In human services in general, and CWS services in particular, this type of analysis is referred to as *predictive analytics* (Russell 2015).

Predictive analytics can be generally understood as a sophisticated form of risk modeling, in which historical data are leveraged to understand relationships between myriad factors to estimate a probability score for the behavior or outcome of interest. Equally important, predictive analytics often use methods hitherto not a part of typical applied social science research: artificial intelligence, data mining, and perhaps most importantly, machine learning.

Numerous social science and public health studies have examined the etiology of child maltreatment using large data sets. However, in most studies, researchers have selected the estimation model (e.g., ordinary least squares regression) and a variable list based on a given theoretical framework. In contrast, machine learning uses iterative model building in which computers use data to identify patterns and make updates to the underlying model without human input. In Table 1, we cite a set of definitions from the business analytics literature of the various types of analytics, ranging from descriptive to prescriptive as defined by Banerjee and colleagues (Banerjee et al. 2013). As new analytic methods are applied to family violence questions, it is helpful to situate a given methodology within the broader taxonomy of analytic approaches from which a method was derived.

New predictive analytics methods do not rely solely on the user to identify variables and specify models. Instead, machine learning techniques were developed to automate the analytic process and to optimize model building. Machine learning occurs in three forms: supervised, unsupervised, and semi-supervised. Supervised machine learning uses data in which the outcome of interest has already been observed (along with numerous covariates). An algorithm is given the existing data, the algorithm "learns" the relationships between covariates and the outcome of interest by estimating their functional relationship, and then makes predictions about

Table 1 Taxonomy of business analytics and potential application to prevention of fatal child maltreatment

Type of Analytics	As Applied to Business Analytics[a]	As Applied to Infant Maltreatment Prevention
Descriptive	Describes a phenomenon through different measures that could capture its relevant dimensions. The purpose is to simply unravel "what happened" or alerting on what is going to happen.	National Child Abuse and Neglect Data System (NCANDS) National Survey of Child and Adolescent Well-Being (NSCAW)
Diagnostic	Evaluates "why" something happened. To discover the root causes of a problem, diagnostic analytics needs exploratory data analysis of the existing data or additional data to be collected using tools such as visualization techniques.	Child fatality review committees
Predictive	Seeks options for future business imperatives, predicts potential future outcomes, and explains drivers of the observed phenomena using statistical or data mining techniques. Examples include forecasting sales of a product for the next month or predicting the behavior of a target segment of consumers.	Multi-sector administrative data linkage Eckerd Rapid Safety Feedback Allegheny Family Screening Tool
Prescriptive	Suggests what courses of action may be taken in the future to optimize business processes and achieve business objectives. In other words, this category of analytics associates decision alternatives with prediction of outcomes. Prescriptive analytics use decision analysis, including tools such as optimization and simulation.	Broward County (Schwartz et al. 2017)

[a] Adapted from Banerjee et al. (2013)

outcomes for new cases based on the function it has learned. An example of supervised machine learning in daily life is Google Search, also known as Google Web Search. Google Search first gathers data from the user's prior searches (along with other data Google has gathered on similar searches around the relevant geographic area) using an algorithm known as PageRank and applies its analysis of those data to the user's newly entered search terms to make suggestions about the most relevant information related to the user query (Google 2018; Noble 2018). For example, if you google 'coffee shop near me', PageRank will produce a set of results it estimates will be most likely to make you "click" a link, with the most likely at the very top of the search results. PageRank uses certain parameters to make these estimates. Some examples include: the number of coffee shops in your area, shops that sell a brand of coffee you have purchased online before, shops that other people in your area have "clicked" in previous searches, or shops that have paid to be included at the top of local search results (Noble 2018). The algorithm learns whether the prediction it produced was good (and thereby whether the parameters it used were successful) based on your clicks, then makes adjustments to improve future predictions.

Unsupervised machine learning looks for patterns in a data set in which no outcome of interest has been observed and the variable categories are unknown. Unsupervised learning is typically used to find patterns or structures in data where there is not a defined or measurable outcome of interest. Unsupervised algorithms often look for associations, clusters, or latent structures in data. Extending the prior search example, suppose a dataset contains a list of all the coffee shops in an area and a set of variables describing the coffee shops, but no information about the type of coffee shop it is (i.e. local, chain, café, etc.). An unsupervised algorithm may be used to classify coffee shops by types using the descriptor variables given in the data, searching for patterns in the variables that it can use to group similar coffee shops, thus allowing classification by type. Semi-supervised machine learning, as its name suggests, is a hybrid of the two aforementioned approaches. Semi-supervised learning is commonly used when some of the cases have values for both covariates and outcomes, but the majority of cases have values only for covariates and are missing data on the outcome of interest.

Recent CWS applications of predictive analytics fall into the category of supervised machine learning (e.g., Chouldechova et al. 2018; County of Los Angeles Office of Child Protection 2017; Schwartz et al. 2017; Vaithianathan et al. 2018). Such predictive risk models use data outcomes for a certain period of time to "teach" a model, following which predictions are made on incoming child welfare cases based on the formulated model. In turn, the model's predictions can be used to make decisions about human service interventions, giving rise to the term *algorithmic decision making* (Newell and Marabelli 2015).

An algorithm is simply a set of rules that follows a logical progression to solve a given problem or calculation. An algorithm typically includes a procedure, an input, and an output. In our example of the Birth Match policy, which is described in detail later in this article, the decision-making algorithm can be a relatively simple procedure. Although state policies vary, generally the input information includes whether a newborn child was born to a parent who has experienced a prior termination of parental rights (TPR), which takes a binary *yes/no* value. The output is the decision to initiate a child protective services (CPS) assessment, which also takes a binary *yes/no*

value. The algorithmic procedure specifies "if prior TPR = *yes*, then CPS = *yes*; if prior TPR = *no* then CPS = *no*." Although this process does not use machine learning (the algorithm is entirely defined by models developed by humans), it is an example of algorithmic decision making because the process is automated and, once implemented, requires minimal human input.

Algorithmic Decision Making: Assumptions and Requirements When coupled with machine learning, algorithmic decision making implicitly assumes the data used to teach the algorithm is a sample of observed outcomes and covariates such that the causal relationship between both is exemplary of the real world. The data used to develop algorithms is typically referred to as the "training" set. The relationships identified in the training set are then used to classify and make predictions on new data (the "test" set). This means that the same relationships found in the training data are assumed to exist in the test set. This key assumption follows Bayesian logic: With enough good prior information about the past, you can make accurate predictions about the future (Christian and Griffiths 2016; Pearl and Mackenzie 2018). This article does not disparage Bayesian statistics. Indeed, without Bayesian methods we could not engage in this conversation. However, this article strives to call out the obvious: Assuming the future will look like the past necessitates agreeing that what occurred in the past will continue to occur in the future.

Given this assumption of continuity, algorithmic decision making has two broad data requirements to ensure the ground truth is accurately represented. The first data requirement is clean data; a training and test set used for algorithmic decision making should be tabular, such that each row represents one observation and each column represents one variable of interest (Wickham and Grolemund 2016). Further, both data sets should be as complete as possible, with few to no missing values. Although this requirement might seem to be common sense, typically the majority of time spent on any given machine-learning project is devoted to the data cleaning process (e.g., imputing missing values, ensuring variables are named appropriately, minimizing measurement error, etc.), particularly when projects use administrative data. When an analysis includes vast numbers of records, improper or insufficient data cleaning can lead to errors and biases, enough to inspire algorithmic approaches to automate the data cleaning process itself (Chu et al. 2016). However, it remains to be seen whether automating data cleaning is a useful or proper solution given the complexity of the task.

The second data requirement of algorithmic decision making is a clear understanding of the key variables the algorithm is set to learn or predict and their cause-and-effect relationships (Brynjolfsson and Mitchell 2017). To date, the vast majority of research in human services and child welfare has been limited to identifying associations. However, a strong causal relationship between dependent and independent variables is necessary for algorithmic decision making. Causal relationships are typically best inferred using randomized controlled trials, which are not always feasible or desirable in human services and child welfare research. Because analytic techniques that use tools of causal inference can help identify these relationships, quasi-experimental techniques are garnering increasing attention. Although casual inference has received some attention in human services research (e.g., Cook et al. 2014; Rose and Stone 2011; Rose 2018), few researchers have implemented casual inference analysis in child welfare (e.g., Doyle 2013; Foster and McCombs-Thornton 2013).

The following example illustrates key issues surrounding the use of predicative analytics in child welfare. Suppose a child welfare jurisdiction is trying to understand the likelihood of a CWS-involved family experiencing foster care placement. The team uses all available administrative data to create an algorithm and finds that CWS-involved parents whose homes are visited by case workers at least twice a week are **more likely** to have their children taken into foster care: that is, the frequency of case worker visits appears to be a predictive factor for foster care placement. Without appropriate causal mapping, a logical next step would be to consider whether changing the visitation policy so case workers visit homes no more than once per week would have a demonstrable effect on decreasing foster care placements. To decide whether changing the visitation policy would actually be beneficial, it would be of paramount importance to examine whether the finding for twice-weekly visits is a mere association or indicative of a causal relationship. For example, it could be that parents with a greater severity of problems not only need more visits from case workers but also are more likely to have their children taken into care. That is, the relationship between home visits and entering foster care might be confounded by the extent of family need (which the jurisdiction in this example has not operationalized, nor reliably measured). If the jurisdiction decides to tell case workers they should not visit families more than once per week, then this policy decision might not have any impact on the likelihood of foster care placement, and could lead to unintended negative consequences. Worse yet, if the change in visitation policy is found to be associated with a desired effect, but the effect was caused by something other than the policy change, then the desirable outcome might be misattributed to the policy change: For example, the desired effect was caused by an intervention only delivered at the one case worker visit. Said another way, predictive analytics in general, and machine learning in particular, are not immune to confounding variables or erroneous conclusions unless the proper analytic steps have been taken (i.e., clean data and review of causal mechanisms; Guyon and Elisseeff 2003). The proliferation of algorithmic decision making without clear knowledge of the conceptual and analytic steps required to assure actionable

results has led to a marked vulnerability to what is termed *algorithmic bias,* meaning the amplification of human bias embedded in data (see Garcia 2016).

Example of Algorithmic-Decision Making and Predictive Analytics in Child Welfare With this foundational understanding of predictive analytics and algorithmic decision making, we now turn to a specific policy application underway in several CWS jurisdictions. The problem that policy makers and researchers are attempting to solve could not be more pressing. Nearly 2000 child fatalities from maltreatment occur each year, with infants experiencing the highest rate of CPS investigations and child maltreatment fatalities across all child age groups (U.S. Department of Health and Human Services 2018). Without question, any tool available to protect the most vulnerable in our society should be explored. Although overall rates of official maltreatment victims have declined over the past decade, from 2012 to 2016 child maltreatment *fatalities* increased by 7%, leading to an ongoing policy goal to prevent severe and fatal child maltreatment.

Following overwhelming bipartisan support in the U.S. Congress, the Protect Our Kids Act of 2012 established the Commission to Eliminate Child Abuse and Neglect Fatalities. A viable policy solution that emerged from this work involves data sharing and multidisciplinary support toward the goal of identifying high-risk children at birth (Commission to Eliminate Child Abuse 2016). This policy, known as Birth Match, was first reviewed as a strategy to protect newborns by Shaw et al. (2013). In addition to their review, the Commission report cited a population-based study that identified infants at risk for maltreatment by using linked birth certificate and child welfare records (Putnam-Hornstein 2011). Shaw et al. (2013) provided detailed descriptions of Birth Match in three jurisdictions: New York City, Maryland, and Michigan. A more recent report identified Minnesota and Texas as having implemented Birth Match as well (Barth et al. 2016).

Although important nuances exist across jurisdictions, similarities also exist in jurisdiction's Birth Match policies. Under Birth Match, all newborns are automatically assessed for maltreatment risk using linked data, including CWS records. A positive match triggers a CPS response (e.g., assessment, investigation). Linkage is typically focused on identifying two risk factors readily measured in available data: (a) whether a sibling has been removed from the home in the past, or (b) whether parental rights have been terminated in the past.

The Birth Match policy directly confronts a seminal CWS question: how confident does CWS need to be that a child is at risk of future harm to step in and protect the child? Or, as one scholar framed the question, "How aggressive should child protective services be?" (Doyle 2013, p. 1143). The unknown potential outcomes of the two alternatives (investigate vs. not investigate) must be weighed by the decision-maker. Because CWS decisions cannot be randomly assigned, the causal impacts of interventions on child well-being, as compared with true counterfactual conditions, are not well understood. On one hand, research has indicated that depending on the quality and stability of services, young children placed in foster care are at higher risk for negative outcomes in behavioral, emotional, and physical health (Lawrence et al. 2006; Rubin et al. 2007). One of the few rigorous causal analyses of foster care in one state found evidence for increased likelihood of juvenile delinquency and elevated use of emergency healthcare (Doyle 2013). Rigorous studies have indicated that foster care confers much better outcomes than placements in institutional settings (Humphreys et al. 2015), but it remains unclear whether foster care is better than in-home services. On the other hand, children who experience maltreatment, particularly chronic maltreatment, are at greater risk for a host of negative outcomes (Gilbert et al. 2009; Jonson-Reid et al. 2012). Further, one study found that a report to CPS was a strong independent risk factor for injury mortality in children younger than 5 years (Putnam-Hornstein 2011).

Application of Ethical Principles in Machine Learning to the Birth Match Policy

Methodological decisions generally, and specifically those regarding research and the application of predictive analytics, are also ethical decisions (Cuccaro-Alamin et al. 2017; Sobočan et al. 2018). The decision to engage in an ethical decision-making process is often guided by the use of specific frameworks, such as the DuBois (2008) 4-point SFNO model (stakeholders, facts, norms, options). The assumptions regarding each of the four domains would be explored by a multidisciplinary case review group, to help ensure all ethical aspects are considered (Sobočan et al. 2018).

In addition to generalized ethical frameworks, scholars are beginning to develop ethical reviews standards that reflect the new questions arising as technology presents new ethical challenges. In considering emerging ethical questions related to artificial intelligence and predictive analytics, we identified two frameworks to apply to the Birth Match policy generally. Although Birth Match involves ethical questions beyond those presented here, given the policy's reliance on algorithmic decision making, the current analysis focuses only on this aspect. For example, although there is ongoing legal debate regarding TPR as a social practice (see Sankaran 2017), the ethics of TPR is not considered in the domain of the present discussion.

A prior framework attempted to identify standards for evaluating predictive models when applied to CWS practice. Russell (2015) suggested predictive models should be valid, reliable, equitable, and useful. These standards were a restatement of a formulation proposed by D'andrade et al. (2008) that instruments used in child welfare to assess risk and safety should be evaluated for instrument reliability,

validity, outcomes, and appropriateness for use with children and families of color. D'andrade et al.'s discussion was presented in the context of the CWS dilemma of whether actuarial risk-assessment instruments should replace consensus-based instruments when applied to allegations of maltreatment. Actuarial risk assessment tools use objective measures to assign a current risk score and a determination of risk for future maltreatment. In contrast, consensus-based assessment are theory-based and help the caseworker organize information and document subjective decision-making (see Mendoza et al. 2016). However, the current discussion moves beyond whether algorithms conducting risk assessment can outperform humans in predicting future risk; the question of Birth Match is whether algorithms can (a) make allegations of potential future child harm (a form of maltreatment) and (b) automate a process within the CWS to begin an assessment or investigation. Our understanding of Birth Match policy is based in part on Shaw et al.'s (2013) work and our ongoing review of Birth Match policies available in public documents.

The first ethical framework, provided by Brauneis and Goodman (2018), describes "desirable documentation" that supports transparency in applying an algorithm to the decision-making process (see Table 2). The second framework comes from the community of researchers known as Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) that has developed a set of principles and guiding questions for considering algorithmic decision making. We attempt to ascertain whether these principles and categories have been publicly applied and documented in any implementation of the Birth Match policy (see Table 3).

Desirable Documentation for Algorithmic Transparency

Brauneis and Goodman (2018) identified eight categories of transparency and accountability for algorithmic transparency. Table 2 presents an outline of the key questions to consider for each category in reviewing policies with an ethical review lens. This section provides a review of these categories and briefly applies them to the Birth Match policy. The first category seeks to understand the transparency of the general predictive goal and how the policy will be applied. The goal of the Birth Match policy is clear: Birth Match is intended to identify infants who might be at risk of maltreatment, and thereby prevent fatal infant maltreatment from occurring.

The second category asks if data is relevant, available, or collectable. Generally, jurisdictions use a limited number of variables for prediction with Birth Match, even though ample data are available that could contribute to this prediction. For example, data might be available on maternal substance use, a factor that can influence the level of future risk, and the presence of which would trigger the decision to allege maltreatment. Next is the

Table 2 Application of "desirable documentation" for algorithmic transparency in predictive analytics to birth match policies

Category of Transparency and Accountability[a]	Key Question	Applied to Birth Match Policy
1. General predictive goal and application	Has government clearly articulated the general goals in using a predictive algorithm?	*Yes*. Predicting which infants are at highest risk for severe and fatal maltreatment.
2. Data: Relevant, available, collectable	Is there documentation of all the possible available data that could be conceivably relevant to making the prediction?	*No*. In most cases, governments are using only one or two variables in the prediction algorithm.
3. Data exclusion	Is there documentation of what available data were excluded because of data quality concerns, susceptibility to manipulation, time and place limitations, lack of relevance, and other policy considerations?	*No.* Not clear why the main criteria variables (i.e., *sibling in foster care, prior termination of parental rights*) were selected. Not clear whether risk factors or causal mechanisms have been fully identified and validated in the literature to encompass these two variables.
4. Specific predictive criteria	Are the criteria used for predictions clearly documented?	*Unclear*.
5. Analytic and development techniques used	Are the analytic techniques used to discover correlations between characteristics of the subjects of predictions clearly documented?	*Yes*.
6. Principal policy choices	Are policy choices and tradeoffs in the predictive algorithm (e.g., relative weighting of false positive to false negatives) documented clearly?	*No*.
7. Validation studies, audits, logging, and nontransparent accountability	Has post-implementation validation analysis determined the predictive strength of the algorithm?	*No*.
8. Algorithm and output explanations	Is a plain-language description of the predictive algorithm and output available to the public?	*Yes*.

[a] Categories from Brauneis & Goodman (2018, p. 167–175)

Table 3 FAT/ML principles for accountable algorithms and social impact questions applied to birth match policy

Principle[a]	Guiding Questions	Applied to Birth Match Policy
1. Responsibility	• Who is responsible if users are harmed by this product? • What are the reporting process and process for recourse? • Who has the power to decide on necessary changes to the algorithmic system during design stage, pre-launch, and post-launch?	Child welfare jurisdictions? Termination of parental rights (TPR) appeal? If bought from a company, no one.
2. Explainability	• Who are your end-users and stakeholders? • How much of your system / algorithm can you explain to your users and stakeholders? • What extent of information about the data sources can you disclose?	Case workers? Supervisors? Child welfare-involved parents?
3. Accuracy	• What sources of error do you have and how will you mitigate their effect? • How confident are the decisions output by your algorithmic system? • What are realistic worst-case scenarios in terms of how errors might affect society, individuals, and stakeholders? • Have you evaluated the provenance and veracity of data as well as considered alternative data sources?	Potential concerns regarding human data entry error and human bias in reporting.
4. Auditability	• Can you provide for public auditing (i.e., probing, understanding, reviewing of system behavior) or is there sensitive information that would necessitate auditing by a designated third party? • How will you facilitate public or third-party auditing without opening the system to unwarranted manipulation?	No process for auditability identified.
5. Fairness	• Are there particular groups that might be advantaged or disadvantaged in the context in which you are deploying the algorithm / system you are building? • What is the potential damaging effect of uncertainty / errors to different groups?	What do we know about families with TPR? See: Meyer and Moore (2015).

FAT/ML = Fairness, Accountability, and Transparency in Machine Learning

[a] Principles and Social Impact Statement from FAT/ML retrieved from http://www.fatml.org/resources/principles-for-accountable-algorithms

category of data exclusion. It is critically important to understand why certain variables were selected and why others were excluded in the development of the algorithm. Were the data chosen because of quality, ease of access, and/or its ability to be analyzed? For example, what historical time parameters should be set for Birth Match to look for a prior TPR? In Maryland, records for TPR are reviewed for only the past 5 years. In Michigan, the Birth Match review extends back as far as digital records exist (Shaw et al. 2013). Further, Birth Match's reliance on "birth identification" might fail to identify children who are born at home without a recorded birth certificate. Did the excluded variables have limitations that would present a challenge to integrating them into the algorithm? Or perhaps the variables were not relevant to the prediction? Although the data applied to Birth Match are generally clear (i.e., having a sibling in foster care or a prior TPR), it is not clear why these variables were selected and why other data that might also be risk factors for child maltreatment were excluded. Although the effect sizes for prior CPS involvement might be the highest for predicting infant maltreatment, why limit the algorithm to one or two variables? For example, are variables related to the number of interactions with CPS an indicator of risk for maltreatment?

A further concern is that Birth Match might not identify individuals at high risk for maltreatment. Two examples from Maryland exemplify this issue. First, parental rights cannot be terminated for a caregiver culpable in a child's death if the caregiver is not the child's biological or adoptive parent. In a case reviewed and documented by the Baltimore City Child Fatality Review Team (2017, p. 13), "a father drowned his newborn son after serving 5 years in prison for killing the child of his previous partner. As a result of this loophole, he was not matched, and the birth of his son, who was clearly at very high risk of fatality, did not come to the attention of Baltimore City DSS." Second, in a more recent case, a man served almost 3 years in prison following conviction for the child abuse death of his 18-month-old son. However, after his release, he was charged with the death of his partner's son (Prudente and Calvert 2018, July 23). Because he was a caretaker, and not the biological parent, the man was not matched at the birth of the second child, although the man would have been considered at high risk of perpetrating maltreatment (Baltimore City Child Fatality Review Team 2017).

The next transparency and accountability consideration is the algorithm's specific predictive criteria. Although the variables used in the Birth Match algorithm are generally known, the threshold for a prediction is unclear. For example, it is unknown whether a 95% or 99% positive match meets threshold limits to indicate need for a CPS referral. Although the aforementioned category of transparency and accountability is not met by Birth Match, it is clear that jurisdictions use matches of specific criteria on the birth certificate with specific criteria in the CPS system to indicate risk. Although the specifics of the linkage process of these data files can vary, the methods for determining a match are relatively clear.

The sixth category focuses on the principal policy choices made for the predictive algorithm and whether clear documentation of policy tradeoffs is available. In the case of Birth Match, the tradeoffs and choices are not well documented. The policy considerations of the harm of a false positive, which would lead to a CPS referral for a family not at risk, and the effect of such referrals on the family are not clear. For example, do false positives result in overuse of limited CPS resources? Do unnecessary CPS referrals contribute to increased negative perceptions of these organizations and/or the family involved? What impact does anxiety and potential bonding disruption have on the life course of the mother, infant, and family? Additionally, what is the incidence of false negatives and what are the tradeoffs for failing to identify infants at risk? Do other opportunities exist to reach this population? The documentation of such policy considerations is critical to achieving algorithmic transparency and accountability.

Relatedly, is it possible that the Birth Match policy perpetuates known systemic inequities such as the over-surveillance of marginalized groups? This key policy question deserves additional attention when considering the historical context of CWS. Whether certain groups are currently more likely to come to the attention of the CWS due to surveillance bias or discrimination has been debated in the literature (e.g., Kim et al. 2018). Model developers must be aware of these considerations and the specific historical context of institutional racism, classism, heterosexism, and other forms of discrimination that may impact families who would come into contact with CWS. For perhaps an extreme but very real example, consider a jurisdiction with a large Native American population. If records from the 1970s were used to predict the neglect of indigenous children, the algorithm would possibly yield an indication of very high risk for neglect for Native American children based on this demographic characteristic alone. Risk modelers would need to understand the institutional factors that were the true cause of the CWS involvement and examine subsequent policy corrections that were made. This is just one example of how historical data could lead to additional harm and perpetuation of the inequities we hope to eliminate.

Birth Match policy does not fulfill the transparency category regarding validation studies, audits, logging, and nontransparent accountability. Complete analyses providing evidence for the predictive strength of the algorithm have not been identified. For all jurisdictions, the availability of rigorous analyses are critical for policy makers to weigh the continued use or spread of the Birth Match policy.

The final category of transparency and accountability is algorithm and output explanations. Are results of the predictive algorithm clearly documented in plain-language and available to the public? For Birth Match, it is clear that the algorithm is designed to link birth certificate data and CPS data in an effort to identify matches between a parent who has had a prior TPR (or another child in foster care) and infants who might be at risk for maltreatment. Although this information is available, it is important to consider how this information is broadly communicated and if it is easily accessible to stakeholders.

Principles of Algorithmic Decision Making The FAT/ML guidelines for accountable algorithmic and social impact include five key principles: responsibility, explainability, accuracy, auditability, and fairness. Table 3 lists a series of guiding questions that align with each principle and are meant to foster dialogue and careful consideration. The principle of responsibility asks about who is held accountable if users are harmed by the algorithm's product. The responsibility principles also inquires about who holds the power to decide on changes and report on progress. When applied to Birth Match, ultimate responsibility seems to be carried by child welfare agencies that make the decision to investigate an individual flagged as a high-risk match. Policy makers must consider how TPR appeals will be handled; this effort should include collecting information on other outcomes such as maternal mental wellness and maternal/infant attachment. Equally important, if the underlying algorithm generating the Birth Match predictions was purchased from a third-party vendor, making changes to the algorithm during any stage is likely to be difficult or impossible if there are no in-house algorithmic engineers who are involved in the build.

The principle of explainability asks how well end users and stakeholders understand the inputs and outputs of the model(s) being deployed. In the case of Birth Match, child welfare case workers and supervisors are the end users, whereas stakeholders include parents, health care providers, communities, and infants. Given that Birth Match requires that all pregnant women are essentially screened for the risk of child neglect and abuse, communication about the algorithm and related policy mechanism requiring the use of Birth Match needs to be disseminated to a large audience with varying experiences, languages, and educational levels. A family flagged by an algorithm and notified by CWS that they will be the subject of an assessment or investigation might respond to that notification with suspicion or misunderstanding. Further, to date, no evaluations have been conducted regarding the potential side effects of such expansive surveillance on individuals, families, or the communities in which they reside. Given the prevalence of generally negative attitudes toward government services among communities, the principle of explainability is an imperative for practice when considering Birth Match implementation.

The FAT/ML principle of accuracy stresses the importance of reviewing data sources and describing all potential sources of error. A well-established foe of administrative data sets is the likelihood of human error in data entry. Additionally, in the case of child welfare, human bias in reporting is another less

obvious but pervasive risk. For example, some jurisdictions record race based on the perception of the case worker conducting the home visit. That is, a case worker looks at a person during a home visit and decides their race based on how the case worker interprets their skin color: this method of identifying race has been shown to be problematic at best (e.g., see Meissner and Brigham 2001). Further, the data used to make predictions may be grossly out of date. If a parent is flagged as high risk of child welfare involvement, is there current data that might mitigate the predicted risk? For example, has the parent survived an abusive relationship or successfully completed drug treatment and rehabilitation? How might the number of years between child removal and the birth of a new child influence outcomes? The accuracy principle requires contemplation of the realistic worst-case scenarios that might occur if the data are wrong, as well as the impact of reproducing such errors on families and society. In the case of Birth Match, relying on the algorithm might prevent child welfare workers from correctly evaluating the present-day risk and resilience factors in a flagged family.

The auditability principle not only underscores the need for public oversight in the creation, deployment, and routine use of an algorithm but also encourages public access to data inputs, outputs, and technical documentation. As noted previously, many child welfare jurisdictions have contracted out algorithmic development, making these tools proprietary and limiting public access to the critical information needs noted above (however, notable exceptions include the New York City Administration for Children's Services). One common concern is that machine learning methods (particularly unsupervised) are akin to a "black box", where only the inputs and outputs are discernible, and no insight is given into how the data were analyzed and therefore relate to each other. Indeed, the increasing complexity and autonomy of self-learning processes will severely challenge our ability to understand why and how a certain output is. Therefore, the ability to audit a given process in the future requires proactive attention to design and technical documentation in the beginning planning phases. This will require communication between data scientists and those responsible for future audits and accountability. Given the increasing use of data sharing across sectors of government, identifying details of future auditing processes must be clearly identified. It is unclear whether and how child welfare jurisdictions implementing Birth Match have engaged in any public input regarding the policy, the algorithm(s), its development, or deployment.

Last, the fairness principle is of particular importance in the context of Birth Match. This principle requires entities using algorithmic decision making to question whether and how specific groups are advantaged or disadvantaged in the context in which the algorithm is deployed, and then adjust accordingly. Stated another way, the fairness principle requires an in-depth inquiry into the potential damaging effects of an algorithm on marginalized and vulnerable groups. In terms of Birth Match, a central fairness question asks if known demographic differences exist between families who have experienced a TPR following CWS involvement and those families who did not experience TPR after CWS involvement. Additional key questions include whether those differences are indicative of a causal relationship, or whether differences are confounded by other variables such as socioeconomic status or geography. Does the Birth Match algorithm account for these differences, or does the policy implementation deploy different interventions based on these known differences?

Discussion and Recommendations

Given the complexity of predictive analytics and the high-stakes nature of CPS investigations, policies such as Birth Match should be carefully considered and studied before being implemented. To be clear, our review suggests we can improve the use of algorithmic decision making in child welfare and we should not abandon this powerful tool. However, the increasing concerns around infant safety should serve as a catalyst for thoughtful, interdisciplinary conversations and research, not as a reason to rush the application of a policy with two-generation consequences. Pilot studies and modeling should be conducted to identify the potential risks and benefits to families based on population, method used, and services provided. Equally important, an essential part of this conversation is engaging the voices and experiences of families who have been affected or could be affected by policies such as Birth Match. Likewise, professionals such as social workers and health care providers who would be expected to implement these policies should be invited to share their perspectives on algorithmic approaches, policy development, and evaluation of positive outcomes and unintended consequences.

Another key area for discussion is whether fatal infant maltreatment (and child maltreatment broadly) is an example of what is known in cognitive theory (and applied to business analytics) as a *black swan problem,* which is typically thought of as a random or unexpected event that deviates from the norm and thus is extremely difficult to predict (Kenton 2017). Taleb (2007) coined the contemporary use of this term (see also Hume's Problem of Induction) and defined a black swan by three characteristics: "First, it is an outlier, because nothing in the past can convincingly point to its possibility. Second, it carries an extreme impact. Third, in spite of its outlier state, human nature makes us concoct explanations for its occurrence after the fact, making it explainable and predictable" (Taleb 2007, p. xvii-xviii). Is fatal infant maltreatment a black swan phenomenon?

In the oft-cited study using California data, researchers modeled the risk for 381 intentional injury deaths from a

sample of 4,317,321 births (Putnam-Hornstein 2011). Of those 381 infant deaths, 127 of the parents (33%) had a prior CPS report. Among all children, about 51,000 children were reported to CPS (12%). Thus, those in the family violence field are attempting to prevent an outcome that occurs less than 1% of the time in the highest-risk group (i.e., 127 deaths among those reported to CPS out of 51,000 reported to CPS). Infant death certainly carries an extreme impact. We argue that based on the current literature regarding infant maltreatment, the field should seriously consider whether the black swan label applies, and if so, what does this label mean for maltreatment solutions that center on predictive analytics.

In Shaw et al.'s report that introduced the implementation of Birth Match in three jurisdictions (Maryland, New York City, and Florida), the authors described "the murder of a child or children who had been left in the care of a parent who had, at some point previously, been judged unsafe" (Shaw et al. 2013, p. 220) as the impetus for the policy change. Perhaps anticipating the black swan argument, the authors further stated,

> Although child welfare policy makers often shy away from making policy based on a single low probability incident such as a child death, the difference in these jurisdictions may be that these deaths were not as unpredictable or "random" as is sometimes the case. Given the magnitude of the previous safety concerns with these families and the fact that the courts had taken significant actions to remove children or to involuntarily terminate parental rights, these death were, unfortunately, not unpredictable (Shaw et al. 2013, p. 220-221).

However, Shaw and colleagues made no additional efforts to identify where fatal infant maltreatment falls in the spectrum between completely unpredictable (random) and completely predictable (deterministic). We suggest that more research is needed to determine the causal mechanisms associated with infant maltreatment. The current evidence is not sufficient to determine whether fatal infant maltreatment is an unpredictable black swan event or a predictable human behavior with complex, but understandable, underlying mechanisms.

A series of next steps are critical for continued application of algorithmic approaches to child welfare to be most helpful and least harmful to families and society. First, current applications in child welfare should be rigorously evaluated. The field must answer the question of whether predictive analytics have successfully changed rates of fatal maltreatment in larger jurisdictions where they have been applied. This question will require multiple years of high-quality data and the use of causal inference techniques to understand the potential role of these tools in preventing the most serious consequences of abuse and neglect. Further, a close examination of the families flagged for assessment or investigation by such tools is needed to understand the full impact of predictive analytics policies. How were these families approached by CWS? How did the families perceive the CWS contact? What services were offered? Perhaps most important, what was the outcome of the case? Careful attention should be paid to systemic inequities such as racial bias and their impact on the evaluation of these tools.

Jurisdictions contemplating new applications of algorithmic approaches would be best served by piloting the approach before full-scale implementation. Data can be cleaned and entered into an algorithm, and cases that might be assessed or investigated after implementation can be followed prospectively. If piloting is not possible due to political or funding restraints, the process can be tested retrospectively by inputting data prior to a set point in time (e.g., 4 years ago) and then following births forward for 2 years in conjunction with child welfare and death certificate data for another 2 years. This retrospective approach would allow policy makers the opportunity to understand how many families would be identified, the demographics of such families, and the outcomes for identified children without the algorithmic approach to assessment. It is important to conduct this type of evaluation in multiple (if not all) jurisdictions given the wide variations in reporting, legal definitions, available services, and sources of bias.

For those considering predictive analytics, a key recommendation is to assess the desirable characteristics and principles (Tables 2 and 3) in currently deployed tools and encourage policy makers and analysts to consider ways of improving the use of these tools. For example, current implementations of Birth Match (and other policies using predictive analytics) could be refined and improved by increasing the availability and quality of the data used to make predictions. Such efforts could include both quantitative data and qualitative data. Additionally, decisions about the exclusion of data should be made rationally and transparently, based on careful examination of empirical evidence. To promote transparency and accountability, the criteria embedded in the algorithm for prediction and technique development should be communicated in lay language to as many stakeholders as possible. Policy analysts and policy makers in a jurisdiction must have a clear understanding of the tradeoffs they are making. What are the risks in trading sensitivity for specificity? Sensitivity refers to a test's ability to correctly identify the true positives. Specificity refers to a tests ability to identify those without the identified outcome. In the case of maltreatment, a risk assessment with high sensitivity would correctly identify all children with future maltreatment as high risk. A test with high specificity would correctly identify all children who do not experience future maltreatment as being low risk. The strength of any test or risk prediction must consider these competing goals.

Although the highest possible sensitivity will identify more children at risk, it will also result in more cases of false positives that flag families who are not in need of assessment, investigation, or resources. This tradeoff can have large system costs, in terms of workforce and resource needs, as well as familial and societal consequences. We do not suggest the tradeoffs will be easy to make, but they should be made consciously, and their benefits and risks communicated as widely and clearly as possible. Algorithm output must not only be clear and useable by child welfare but also understandable to policy makers and families.

Public transparency is critical for the deployment of algorithmic tools. When a jurisdiction is contemplating the use of these tools, engaging in a process that encourages vigorous public comment and oversight is essential. Once implemented, ongoing review of identified cases and outcomes should be the rule, not the exception. A particularly concerted effort should be made to include the affected communities in public comment and oversight. These efforts should include families previously involved in child welfare services as well as vulnerable and/or marginalized communities.

Improved literacy about machine learning among professionals can support a more strategic, comprehensive, and hopefully successful, use of technology in the service of high-risk families. Adding policies and strategies related to machine learning to local social service departments must be done with adequate training, algorithmic transparency, and evaluation of the outcomes. Taking time to lay the groundwork thoughtfully, rather than rushing toward large-scale application, would allow for refinement of methods and approach.

Public education about the role, benefit, and risk of data analytics in all parts of life will be increasingly important to the deployment of these tools. People understand that some data is used to predict rates for insurance (e.g., health, life, auto, and homeowners). But many uses of algorithmic approaches have been focused on voting and selling—and these have been fraught with problems such as personal data breeches, perceived invasion of privacy, and aggressive sales. These techniques have also won elections and helped create corporate giants. If the public is expected to place their trust in the results produced by algorithmic tools, then the public will need assurance, education, and input into government use of these tools in sectors such as health services, human services, and social services.

Concerns regarding systematic implicit and explicit bias and discrimination are not uncommon in the child welfare field or predictive analytics. Families can become at risk due to racial inequities around education, housing, poverty, and/or opportunities. If an algorithm identifies a family as at risk and this leads to unwarranted assessment, services, or child separation, then the CWS can be accused of reinforcing the very inequities it seeks to eradicate. Poverty status in particular would present complicated associations with many other risk variables which may lead to estimation errors and confusion regarding causal directions in prediction algorithms.

We wish to underscore that our intention in writing this paper was not to preach anti-algorithmic decision making. Although the authors' opinions vary, we all believe that algorithmic approaches to child welfare decision making hold great promise for solving wicked problems, that is, previously intractable social problems. Nevertheless, predicting the future based on the past is fraught with assumptions about human behavior that will impede new technology from allowing us to manifest the world we wish to see. Getting this right matters.

References

Baltimore City Child Fatality Review Team, Subcommittee on Child Abuse and Neglect. (2017). *Eliminating child abuse and neglect fatalities in Baltimore City*. Retrieved from http://healthybabiesbaltimore.com/uploads/files/Initiatives/Baltimore%20City%20CFR%20Child%20Abuse%20Report%20January%202017.pdf

Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential? *Vikalpa, 38*(4), 1–12. https://doi.org/10.1177/0256090920130401.

Barth, R. P., Putnam-Hornstein, E., Shaw, T. V., & Dickinson, N. S. (2016). *Safe children: Reducing severe and fatal maltreatment (grand challenges for social work initiative working paper no. 17)*. Cleveland: American Academy of Social Work and Social Welfare Retrieved from http://grandchallengesforsocialwork.org/wp-content/uploads/2015/12/WP17-with-cover.pdf.

Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale Journal of Law & Technology, 20*, 103–176 Retrieved from https://yjolt.org/sites/default/files/20_yale_j._l._tech._103.pdf.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science, 358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062.

Chouldechova, A., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. *Proceedings of Machine Learning Research, 81*, 1–18.

Christian, B., & Griffiths, T. (2016). *Algorithms to live by: The computer science of human decisions*. New York: Henry Holt and Company.

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). Data cleaning: Overview and emerging challenges. In *Proceedings of the 2016 international conference on Management of Data* (pp. 2201–2206). New York: ACM. https://doi.org/10.1145/2882903.2912574.

Commission to Eliminate Child Abuse, & Fatalities, N. (2016). *Within our reach: A national strategy to eliminate child abuse and neglect fatalities*. Washington, DC: Government Printing Office.

Cook, T. D., Tang, Y., & Seidman Diamond, S. (2014). Causally valid relationships that invoke the wrong causal agent: Construct validity of the cause in policy research. *Journal of the Society for Social Work and Research, 5*(4), 379–414. https://doi.org/10.1086/679289.

County of Los Angeles Office of Child Protection. (2017). *Examination of using structured decision making and predictive analytics in assessing safety and risk in child welfare* (item no. 490A, agenda of September 20th, 2016). Retrieved from http://file.lacounty.gov/SDSInter/bos/bc/

1023048_05.04.17OCPReportonRiskAssessmentTools_SDMandPredictiveAnalytics_.pdf

Cuccaro-Alamin, S., Foust, R., Vaithianathan, R., & Putnam-Hornstein, E. (2017). Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review, 79*, 291–298. https://doi.org/10.1016/j.childyouth.2017.06.027.

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. https://doi.org/10.1300/J394v05n01_03.

de Haan, I., & Connolly, M. (2014). Another Pandora's box? Some pros and cons of predictive risk modeling. *Children and Youth Services Review, 47*, 86–91. https://doi.org/10.1016/j.childyouth.2014.07.016.

Doyle, J. J., Jr. (2013). Causal effects of foster care: An instrumental-variables approach. *Children and Youth Services Review, 35*, 1143–1151. https://doi.org/10.1016/j.childyouth.2011.03.014.

DuBois, J. M. (2008). *Ethics in mental health research: Principles, guidance, cases*. New York: Oxford University Press.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

Foster, E. M., & McCombs-Thornton, K. (2013). Child welfare and the challenge of causal inference. *Children and Youth Services Review, 35*, 1130–1142. https://doi.org/10.1016/j.childyouth.2011.03.012.

Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal, 33*(4), 111–117. https://doi.org/10.1215/07402775-3813015.

Gershenfeld, N., Krikorian, R., & Cohen, D. (2004, October). The internet of things. *Scientific American, 291*(4), 76–81. https://doi.org/10.1038/scientificamerican1004-76.

Gilbert, R., Widom, C. S., Browne, K., Fergusson, D., Webb, E., & Janson, S. (2009). Burden and consequences of child maltreatment in high-income countries. *Lancet, 373*(9657), 68–81. https://doi.org/10.1016/S0140-6736(08)61706-7.

Google. (2018). How search works. Retrieved from https://www.google.com/search/howsearchworks/

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.

Harcourt, B. E. (2007). *Against prediction: Profiling, policing, and punishing in an actuarial age*. Chicago: University of Chicago Press.

Humphreys, K. L., Gleason, M. M., Drury, S. S., Miron, D., Nelson, C. A., 3rd, Fox, N. A., & Zeanah, C. H. (2015). Effects of institutional rearing and foster care on psychopathology at age 12 years in Romania: Follow-up of an open, randomised controlled trial. *Lancet Psychiatry, 2*(7), 625–634. https://doi.org/10.1016/S2215-0366(15)00095-4.

Jonson-Reid, M., Kohl, P. L., & Drake, B. (2012). Child and adult outcomes of chronic child maltreatment. *Pediatrics, 129*(5), 839–845. https://doi.org/10.1542/peds.2011-2529.

Kenton, W. (2017). Black swan. Retrieved from https://www.investopedia.com/terms/b/blackswan.asp

Kim, H., Drake, B., & Jonson-Reid, M. (2018). An examination of class-based visibility bias in national child maltreatment reporting. *Children and Youth Services Review, 85*, 165–173. https://doi.org/10.1016/j.childyouth.2017.12.019.

Kulkarni, S. J., Barth, R. P., & Messing, J. T. (2016). *Policy recommendations for meeting the Grand Challenge to Stop Family Violence (grand challenges for social work initiative policy brief no. 3)*. Cleveland: American Academy of Social Work & Social Welfare Retrieved from https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1794&context=csd_research.

Lawrence, C. R., Carlson, E. A., & Egeland, B. (2006). The impact of foster care on development. *Development and Psychopathology, 18*(1), 57–76. https://doi.org/10.1017/S0954579406060044.

Marr, B. (2018, May). How much data do we create everyday? The mind-blowing stats everyone should read. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#2667f3d460ba.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3.

Mendoza, N. S., Rose, R. A., Geiger, J. M., & Cash, S. J. (2016). Risk assessment with actuarial and clinical methods: Measurement and evidence-based practice. *Child Abuse & Neglect, 61*, 1–12. https://doi.org/10.1016/j.chiabu.2016.09.004.

Meyer, A. S., & Moore, A. A. (2015). The future of termination of parental rights research and practice: A commentary on ben-David. *Journal of Family Social Work, 18*(4), 253–266. https://doi.org/10.1080/10522158.2015.1079585.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of "datification". *Journal of Strategic Information Systems, 24*(1), 3–14. https://doi.org/10.1016/j.jsis.2015.02.001.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books.

Protect Our Kids Act of 2012. U.S.C. 42 §1305 [Page 126 STAT. 2460; Pub. L.112–275]. (2013).

Prudente, T. & Calvert, S. (2018). For the second time, Baltimore man is charged in the death of a child. *The Baltimore Sun*, retrieved from https://www.baltimoresun.com/news/maryland/crime/bs-md-ci-baby-murder-charges-20180723-story.html

Putnam-Hornstein, E. (2011). Report of maltreatment as a risk factor for injury death: A prospective birth cohort study. *Child Maltreatment, 16*(3), 163–174. https://doi.org/10.1177/1077559511411179.

Rose, R. A. (2018). Frameworks for credible causal inference in observational studies of family violence. *Journal of Family Violence*, advance online publication. https://doi.org/10.1007/s10896-018-0011-3.

Rose, R. A., & Stone, S. I. (2011). Instrumental variable estimation in social work research: A technique for estimating causal effects in nonrandomized settings. *Journal of the Society for Social Work and Research, 2*, 76–88. https://doi.org/10.5243/jsswr.2011.4.

Rubin, D. M., O'Reilly, A. L., Luan, X., & Localio, A. R. (2007). The impact of placement stability on behavioral well-being for children in foster care. *Pediatrics, 119*(2), 336–344. https://doi.org/10.1542/peds.2006-1995.

Russell, J. (2015). Predictive analytics and child protection: Constraints and opportunities. *Child Abuse & Neglect, 46*, 182–189. https://doi.org/10.1016/j.chiabu.2015.05.022.

Sankaran, V. S. (2017). Child welfare's scarlet letter: How a prior termination of parental rights can permanently brand a parent as unfit. *N.Y.U. Review of Law & Social Change, 41*(4), 685–705 Retrieved from https://socialchangenyu.com/wp-content/uploads/2017/11/sankaran.pdf.

Schwartz, I. M., York, P., Nowakowski-Sims, E., & Ramos-Hernandez, A. (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward County experience. *Children and Youth Services Review, 81*, 309–320. https://doi.org/10.1016/j.childyouth.2017.08.020.

Shaw, T. V., Barth, R. P., Mattingly, J., Ayer, D., & Berry, S. (2013). Child welfare birth match: Timely use of child welfare administrative data to protect newborns. *Journal of Public Child Welfare, 7*(2), 217–234. https://doi.org/10.1080/15548732.2013.766822.

Sobočan, A. M., Bertotti, T., & Strom-Gottfried, K. (2018). Ethical considerations in social work research. *European Journal of Social*

Work, advance online publication. https://doi.org/10.1080/13691457.2018.1544117.

Taleb, N. (2007). *The black swan: The impact of the highly improbable*. New York: Random House.

U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. (2018). Child maltreatment 2016. Available from https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/child-maltreatment

Vaithianathan, R., Rouland, B., & Putnam-Hornstein, E. (2018). Injury and mortality among children identified as at high risk of maltreatment. *Pediatrics, 141*(2), e20172882. https://doi.org/10.1542/peds.2017-2882.

Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. Sebastopol: O'Reilly Media.