Child Abuse & Neglect 154 (2024) 106943

Contents lists available at ScienceDirect

Child Abuse & Neglect

journal homepage: www.elsevier.com/locate/chiabuneg

# Child welfare predictive risk models and legal decision making

Matthew Trail [*,1]

*University of Bonn, Germany*

ARTICLE INFO

*Keywords:*
Predictive risk models
Legal decision making
Child welfare/foster care

ABSTRACT

*Background:* Child welfare agencies around the world have experimented with algorithmic predictive modeling as a method to assist in decision making regarding foster child risk, removal and placement.
*Objective:* Thus far, all of the predictive risk models have been confined to the employees of the various child welfare agencies at the early removal stages and none have been used by attorneys in legal arguments or by judges in making child welfare legal decisions. This study will show the effects of a predictive model on legal decision making within a child welfare context.
*Participants and setting:* Lawyers, judges and law students with experience in child welfare or juvenile law were recruited to take an online randomized vignette survey.
*Methods:* The survey consisted of two vignettes describing complex foster child removal and placement legal decisions where participants were exposed to one of three randomized predictive risk model scores. They were then asked follow up questions regarding their decisions to see if the risk models changed their answers.
*Results:* Using structural equation modeling, high predictive model risk scores showed consistent ability to change legal decisions about removal and placement across both vignettes. Medium and low scores, though less consistent, also significantly influenced legal decision making.
*Conclusions:* Child welfare legal decision making can be affected by the use of a predictive risk model, which has implications for the development and use of these models as well as legal education for attorneys and judges in the field.

## 1. Introduction

Predictive models have long had supporters in the child welfare world and advocates for their use argue that represent a way for state Child Protective Service (CPS) agencies to make better decisions and identify more children in need of care (Chouldechova et al., 2018; Field et al., 2023; Vaithianathan et al., 2017). However, the models have also been the subject of intense criticism of their methods and their built-in biases (Gerchick et al., 2023; Keddell, 2019a; Samant et al., 2021).

Despite their widespread use and history, none of the predictive models have ever been used by attorneys representing parties in a CPS case or by the courts that are hearing the cases. However, courts in other areas of the law, such as criminal recidivism have experience using predictive risk models (Hamilton, 2019), so the likelihood remains that eventually these models will make it to the courtroom even if only as discoverable evidence presented by a party.

How these models might affect legal decisions amongst actual practitioners in child welfare is unknown. The present study is the

* Corresponding author.
*E-mail address:* trail@coll.mpg.de.
[1] **Permanent Address:** Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, 53113 Bonn, Germany.

https://doi.org/10.1016/j.chiabu.2024.106943
Received 15 November 2023; Received in revised form 28 April 2024; Accepted 4 July 2024
Available online 16 July 2024
0145-2134/

first of its kind to test legal decision making with the use of predictive risk models in the U.S. child welfare context. In an experimental survey of attorneys, law students and judges with child welfare experience, the results show that predictive risk models can change legal decisions about the placement and removal of foster children. This has important ethical and policy issues that will be discussed later in this paper.

Research has supported the idea that machine advice can affect human decision making (Bogert et al., 2021; Grgić-Hlača et al., 2022; Logg et al., 2019) and that humans are ambivalent about the machine advice given (Kawaguchi, 2021; Mahmud et al., 2022; Burton et al., 2020). People also tend to view machine advice either through the lens of automation bias, where machine advice is overestimated or through algorithmic aversion where users prefer human advice even when the machine is superior (Horowitz & Kahn, 2023; Jones-Jang & Park, 2022). Relatedly, some scholars have suggested that humans using machine advice can also lead to moral confusion as it is unclear if the ultimate decision is under the control of the human or the machine (Elish, 2019; Hohenstein & Jung, 2020).

There are also serious critiques of the use of predictive models across fields especially in the light of their tendency to amplify racial bias and discrimination such as within hiring decisions (Dastin, 2018; Raghavan et al., 2020); medical decisions (Obermeyer et al., 2019), facial recognition (Raji & Fried, 2021), and policing (Collins, 2018) to name just a few examples.

Though this paper could not hope to trace the use of predictive models across fields, nor even to identify each and every predictive model used in a child welfare context, a brief highlighted history of their use in the child welfare system to the present day will hopefully help put this current research in the proper context.

## 2. Literature review

Samant and colleagues, with the American Civil Liberties Union (ACLU) reported that as many as 26 states had experimented with child welfare predictive models, and at least 11 were currently using them, though this number might be higher as it is not always clear when and where the models are being used. The ACLU noted that most of the models are used for some sort of child abuse risk modeling, though those models could come in different forms such as individual risk models for individual children or even neighborhood risk models for all children in a certain location (Samant et al., 2021). Researchers who looked at one model out of Washington state found that the model being used did not reflect many subjective factors that caseworkers were considering and that the factors it was using were not significantly correlated with abuse (Saxena et al., 2022).

Saxena et al. (2020) conducted a meta-analysis of the literature regarding predictive models used in child welfare and found that most of the models being used are focused on removal decisions and not with decisions related to stability or placements. The researchers also wrote that they do not take into account factors like sibling placement, proximity to family or agency characteristics, which leads to incomplete models (Saxena et al., 2020). Saxena et al. (2023) also came to similar conclusions when they studied the use of caseworker notes and predictive data, finding that most models do not include a holistic view of the case and the child and focus solely on easy to quantify data. They also noted that since even within the child welfare world, sever child abuse is actually a very rare occurrence, so models that predict severe abuse are often flawed (Saxena et al., 2020).

Though as described above, most models are focused on risk and removal, other uses of predictive models in child welfare include their use to predict permanency outcomes of youth already in care (Ahn et al., 2021; Stepura et al., 2020; Elgin, 2018), to predict the stability of reunification (Purdy & Glass, 2023) to predict the most stable placement (Moore et al., 2016), and which cases could best be served by a team of child welfare experts (Willcott & Stewart, 2021).

Leslie et al. (2020) wrote that the use of predictive models in child welfare was predicated on the possible benefits of the models such as more consistent and rational decision making, better CPS accountability and better resource management. Drake et al. (2020) noted that the use of actuarial tools for prediction in child welfare is very common and even those tools require flawed human input.

One of the common complaints levied against the use of predictive models is that the models are biased. Gillingham (2019) noted that the biases can be introduced both through the absence of the correct data within the case files from which it is based and also the biases inherent in human decisions that are reflected within the data. Keddell (2019a) also discussed that the child welfare system includes both parents and children and that sometimes those parties have conflicting interests and that even a model that was built to be fair to one side, might still be biased against the other. Keddell (2019b) went on to write that different child welfare trends in different locations will also impact the model so that potentially biased decisions from one location become data in the new model.

Perhaps the most discussed of the current models in the Allegheny Family Screening Tool (AFST) used in Allegheny County, Pennsylvania, the county that encompasses Pittsburgh. Since the AFST is the basis for the hypothetical model used in my experimental survey, I will discuss its history, purpose, controversy and current use at more length.

In the report chronicling the origin and use of the AFST, its creators described how predictive modeling had been in use already in medical settings and that the AFST was created following a 2014 request from Allegheny County CPS. The county is special because it maintains an integrated database of all of the county services (Vaithianathan et al., 2017).

The researchers built the model analyzing data from all of the CPS referrals from 2008 until 2016 and then included other data from other systems including juvenile justice, criminal justice, public benefit programs, census data on neighborhood poverty and behavioral and mental health programs (Vaithianathan et al., 2017). The result was a model based on 800 different variables from the merged datasets that produced an individual prediction that a child would either be re-referred to CPS within two years or placed into care by CPS within two years (Vaithianathan et al., 2017). The researchers claimed that knowing the possible outcome of a child would be at risk in the future would be the best way to support initial case decision making (Vaithianathan et al., 2017).

Since its implementation, the AFST has evolved and developed through multiple versions, though its basic design remains the same (Gerchick et al., 2023). During one early version of the AFST, a problem with the coding resulted in mistaken risk scores being given to a small portion of the intake cases (De-Arteaga et al., 2022). Through this unintended natural experiment, researchers found that intake workers were more likely to override mistaken risk scores than correct ones and were especially likely to screen in cases that the tool erroneously scored low (De-Arteaga et al., 2022).

In Goldhaber-Fiebert and Prince's (2019) impact evaluation of the AFST, they reported that it had a small effect on increasing the number of cases that were screened in for investigation and increased the level of accuracy for both Black and white children, but actually decreased slightly the level of accuracy for cases that were screened out. As an estimate as to how many children might be affected by the model, the researchers guessed that on average each month 24 more children would be accurately screened in by use of the AFST and that 11 would be inaccurately screened out. They also found that the AFST did not improve the consistency of intake decisions, though they acknowledge the sample size might have been too small to detect changes (Goldhaber-Fiebert & Prince, 2019).

A similar model began use in 2017 in Douglas County, Colorado known as the Douglas County Decision Aid (DCDA) and was built by the same research team as the AFST (Vaithianathan et al., 2019). In a review of the DCDA, Fitzpatrick and Wildman (2021) found that high risk scores were the most likely to result in investigation and increased the likelihood that scores above the median would be investigated and decreased the chance of those below the median.

How the AFST deals with bias built into the dataset and its effects on minority populations is an ongoing and heated dispute. While acknowledging the inherent racial and human bias in child welfare data, Field et al. (2023) looked at recent data from the model and conducted a difference within difference analysis to determine if the AFST reduced racial differences in intake outcomes and in case outcomes. They found that the AFST reduced racial differences in screening outcomes by 2.5 % and that this effect was most profound for older youth (Field et al., 2023).

Vaithianathan et al. (2020) tested the AFST on children admitted into the hospital for injuries and found that the model was accurate at predicting injuries associated with abuse. They wrote that the 5 % of children identified by the model as the highest risk had significantly higher chances of reported injury than those the model identified in the lower 50 % of risk (Vaithianathan et al., 2020).

However, other researchers have disputed the claim that the AFST reduced bias. When Cheng et al. (2022) looked at the AFST data and interviewed the CPS intake staff workers, they found that the workers themselves did not particularly understand how the AFST model worked, but that they were making conscious choices to screen out more referrals involving black children that the model identified as mandatory. They noted that the AFST model on its own would have screened in 37 % more black children than what was actually screened in (Cheng et al., 2022).

De-Arteaga et al. (2022) also found that the AFST on its own did not reduce racial bias as its screen in scores would have been similar to those prior to the implementation of the tool, nor did the tool on its own have a significant effect on intakes from poorer neighborhoods. Humans looking at factors beyond the risk score seemed to be primary for decisions surrounding race and socio-economics (De-Arteaga et al., 2022).

The ACLU, which has been investigating the AFST, examined the various updated versions of the AFST model and conducted an independent analysis (Gerchick et al., 2023). Their principal argument against the tool is that the design choices made by its creators perpetuate biases in the system and that those choices are arbitrary and not objective. Gerchick et al. (2023) and the ACLU noted that the AFST has three primary problems. It includes some data such as criminal history that is always included no matter the time frame between the child welfare allegation and the past crime (Gerchick et al., 2023). It also combines scores for an entire family, so risk is calculated for all the children together even if they are different ages. Finally, they wrote that by using as much data as possible for the model, the AFST includes racially biased information such as criminal and juvenile justice data (Gerchick et al., 2023).

What all this research and background reflects is a state of uncertainty and debate over the use of predictive models in the U.S. child welfare community and how decisions will be affected in different legal contexts. How these models might affect child welfare legal decisions is an open question that this paper hopes to help answer by showing that legal removal and placement decisions in the child welfare context can be influenced by the use of predictive models.

## 3. Methodology

This is a preregistered quantitative randomized vignette survey study to test how legal removal decisions could be changed by a predictive model with ethical approval by my Max Planck Institute. The survey was sent to law students, attorneys and judges with child welfare and/or juvenile justice experience. Recruitment was completed using multiple avenues. Recruitment began in August 2022 with survey request emails that were sent to child welfare law firms, attorneys working for the state child protective agencies and to attorneys for Court Appointed Special Advocate Guardians Ad Litem (CASA/GAL). The survey was also posted on specialized child welfare and juvenile attorney Facebook groups that had controlled member access restricted to child welfare and juvenile attorneys. Finally, the survey was distributed during two presentations, a webinar for the National Association of Children's Council on February 15, 2023 and during an in-person presentation for the National Council of Juvenile and Family Court Judges in Dallas Texas on March 20, 2023. Recruitment concluded in July 2023 with a total of 255 completed responses.

The survey consisted of seven demographic questions followed by instructions to read the following vignettes and then answer the questions as if they were the judge deciding the case. The broad structure of the survey was that participants read and answered questions about two child removal decisions. They were then given randomized high, medium or low predictive model scores and asked to answer the same questions given the new scores. Lastly, participants were asked 16 short questions taken from the Dalgleish Scale to measure their beliefs about family preservation and child safety (Dalgleish, 2010; Dettlaff et al., 2020; Hollinshead et al., 2021).

Prior to distribution, the vignettes and questions were given to five experienced child welfare attorneys with practices in Virginia, Florida and Texas to make certain that the vignette cases were clear and appropriate and that the language of the questions could be understood by attorneys in different state jurisdictions that might use slightly different terms when discussing the same legal issue. Following qualitative interviews with each attorney, minor changes were made to make the vignettes and questions more universal. Vignette use is common in social science and specifically in child welfare research as they allow researchers to test real life scenarios with practitioner decision making under controlled circumstances (Keddell & Hyslop, 2018; Middel et al., 2022; Stokes & Schmidt, 2012; Taylor, 2005).

This study followed a within subjects design, so all participants saw the same vignettes in the same order. Randomization was done at the treatment level. Participants were first presented with a vignette about a 10-year-old child named Jonas and a removal decision due to issues of neglect and possible medical neglect. The scenario of this vignette was taken from an actual court case in Texas where there was a serious dispute amongst the parties as to removal and placement.

Participants were then asked to answer on a six-point Likert question if they believed Jonas should be removed and taken into state custody with a range from "definitely not" to "definitely". This question was followed by a five-point Likert question asking participants to rate their confidence in the previous decision with a range of "not at all confident" to "extremely confident".

Participants then answered two other similar six-point Likert questions regarding if they would place Jonas in a foster home or leave Jonas in his mother's home. Each of these questions was followed by a five-point Likert question rating their confidence in each decision.

Following their answers, participants were randomly exposed to one of three treatments, a low, medium or high predictive risk model score. The words "low", "medium" and "high" were included in the vignette to eliminate any ambiguity about what the numbers themselves meant. The vignette also includes a statement about the accuracy of the algorithm being 90 % to increase participant confidence.

After their exposure to the treatment, participants were then given the same Likert questions and their corresponding Likert confidence question with the inclusion of the phrase "Given the new predictive risk modeling score…". Questions were given in the same order as before.

Following those questions, participants were given another vignette, this one about a 5-year-old boy named Carlos and possible neglect issues. This vignette was also inspired by a true Texas foster care case where parties disputed the issue of removal and placement.

Participants then followed the same pattern as the first vignette, answering questions and relaying their confidence, then answering the questions again after being exposed to a random, high, medium or low predictive risk model prompt. All told, participants answered 12 questions per vignette, six pretreatment and six post treatment.

Finally, participants were given the Dalgleish Scale, developed by the late Len Dalgliesh and used in multiple studies regarding foster care decision making (Dettlaff et al., 2020; Hollinshead et al., 2021). The Dalgleish Scale presents participants with eight forced choice decisions in which they must choose either a statement that leans strongly towards family preservation or child safety. After each choice, participants are then asked to rate their preference on a 1 to 5 Likert question for their prior response. The final result is that each participant has a score between −40 and 40 representing their overall preferences towards family preservation versus child safety (Dettlaff et al., 2020; Hollinshead et al., 2021).

Of the 251 completed responses, 208 or almost 83 % were female, 40 or almost 16 % were male and two participants chose non binary and one chose not to answer. The overall national percentage of female attorneys in 2022 was 38 % (American Bar Association, 2022). The uneven distribution in this sample population most likely comes from the specific type of attorney that practices child welfare and juvenile law, which leans heavily on nonprofit, public service and legal aid law firms. The Legal Services Corporation (2021), which is the largest funding organization for legal aid law firms across the United States does track the gender of the lawyers in the legal aid firms it funds and women were the majority of attorneys across their 899 legal aid offices.

The participant's reported ethnicity is listed below. The majority of individuals within this sample are White or Caucasian, constituting 77.68 % of the total population, followed by Black or African American (7.30 %), Hispanic or Latino (6.01 %), and Asian or Pacific Islander (3.43 %). Additionally, there are smaller percentages of Native American or Alaskan Native (2.58 %) and Multiracial or Biracial (2.15 %). These figures are similar to the overall national demographics reported by the ABA with white lawyers at 81 %, Black lawyers 4.7 % and Hispanic at 5.8 % (American Bar Association, 2022).

Out of the sample nearly 43 % had worked as an attorney for 18 or more years, with nearly 30 % having at least a decade of experience. Private attorneys represent the largest percentage of individuals at 74 responses, followed by 53 attorneys representing CASA or GAL programs and 45 nonprofit legal firm attorneys, and 19 CPS attorneys. In addition, 10 judges, 13 law students and 2 law professors completed the study. Finally, 37 attorneys selected the category *Other Attorney*.

The majority of participants were private practitioners. Though it is unclear exactly what role the 37 attorneys that chose *Other Attorney* have in the child welfare system, the survey was also distributed to groups that include attorneys with a primary juvenile justice orientation, so it is likely those responses come from those groups. In any follow up study, juvenile attorneys should be included as their own group. Given that the survey was distributed only to attorneys, judges and law students and there was no financial incentive to take or complete it, it seems unlikely that non-legal field participants would have completed the survey. Additionally, time recording data within the survey indicates that these responses match times when other attorney responses were recorded.

Though the survey was sent to groups with nationwide representation and responses were recorded for all but 13 states, a majority of the responses came from Florida with 41, Texas with 40, Arizona with 15, California 15, Colorado 15, Pennsylvania 12 and Georgia 11. Because of my former work as a foster care attorney in both Florida and Texas, I was able to draw from a more robust network and thus recruit more attorneys in both of those states. Since Texas and Florida are the 2nd and 3rd most populous states, the weight of responses from those two states does not seem problematic.

## 4. Results

The analysis of the data was done with structural equation modeling (SEM) using Stata to answer the underlying question of if the inclusion of a predictive model would affect the legal decision of lawyers and judges regarding initial child removal and placement. (See Appendix 1 for regression tables).

The three primary questions following the Jonas vignette were;

1. How likely are you to order that Jonas be taken into the custody of the state protection agency as a ward of the state?
2. How likely are you to order that Jonas be removed and placed in a therapeutic foster home?
3. How likely are you to order that Jonas remain in his mother's home?

The three primary questions following the high, medium or low treatment risk scores were;

1. Given the new predictive risk modeling score, how likely are you to order that Jonas be taken into the custody of the state protection agency as a ward of the state?
2. Given the new predictive risk modeling score, how likely are you to order that Jonas be removed and placed in a therapeutic foster home?
3. Given the new predictive risk modeling score, how likely are you to order that Jonas remain in his mother's home?

### 4.1. Jonas vignette

For each set of Jonas questions, the best predictor of the second response is what participants selected on their initial response. The coefficient between the two questions are 0.801 for ward of state, 0.76 for foster placement and 0.88 for placement with the mother. This indicates that though the risk score treatments do affect attorney decision making, their initial decision still is the primary predictor. Beginning with the Jonas questions using the low treatment as the baseline reference category, the results of the model are shown in Table 1.

For every Jonas question, high risk score treatments had an effect on legal decision making, meaning lawyers were more likely to support Jonas coming into foster care, being placed in a foster home and not being placed with his mother. Medium risk score treatments also made lawyers more likely to support Jonas being placed into care and less likely to approve placement with the mother. Lawyers who saw the low risk scores were less likely to approve of Jonas coming into care and were more likely to approve of placement with his mother.

### 4.2. Carlos vignette

Similar to the Jonas results, the initial selection by attorneys was the best predictor of what their second response would be. The coefficients for each pair of questions are 0.895 for Carlos being a ward of the state, 0.896 for him being placed into foster care, and 0.848 for being placed with his mother.

The Carlos vignette had essentially the same questions and structure as the Jonas questions, so the analysis is the same. The Carlos question using the low treatment as the baseline reference category results of the model are shown in Table 2.

The high-risk score treatments for Carlos also had an effect on legal decision making, meaning lawyers were more likely to support

**Table 1**
Jonas treatment effects.

| Scenario | Risk score level | Coefficient (effect size) | Statistical significance | Interpretation | Effect size interpretation |
|---|---|---|---|---|---|
| Jonas Ward (being taken into state custody) | Medium | 0.294 | $p < 0.010$ | Slight increase in willingness to place Jonas under state agency | Small |
| Jonas Ward | High | 0.611 | $p < 0.001$ | Increase in willingness to place Jonas under state agency | Large |
| Jonas Ward | Low (vs High) | −0.6108 | $p < 0.001$ | Less likely to order Jonas into care by state agency compared to high treatment | Large |
| Jonas Foster Home | Medium | 0.17 | Not Significant | No significant change in decision to place Jonas in foster home | Small |
| Jonas Foster Home | High | 0.7235 | $p < 0.001$ | More likely to place Jonas into a foster home | Large |
| Jonas Foster Home | Low (vs Medium) | 0.183 | Not Significant | Not Significant | Small |
| Jonas Mom | Medium | −0.309 | $p < 0.005$ | Less likely to place Jonas back with his mother | Small to Moderate |
| Jonas Mom | High | −0.881 | $p < 0.001$ | Significantly less likely to allow Jonas to stay with his mother | Large |
| Jonas Mom | Low (vs High) | −0.316 | $p < 0.002$ | More likely to leave Jonas in his mother's home compared to high treatment | Small to Moderate |

Carlos coming into foster care, being placed in a foster home and not being placed with his mother. Medium risk scores also made lawyers more likely to support Jonas being placed into care, being placed into a foster home and less likely to approve placement with the mother. Lawyers who saw the low treatments were less likely to approve of Carlos coming into care, being placed in a foster home and were more likely to approve of placement with his mother.

The effects of the treatments were not uniform across the questions and varied between Jonas and Carlos scenarios. The Carlos vignette shows more consistent significant effects across all questions and treatments. The Jonas vignette shows mixed results, with some non-significant effects in the foster placement question, indicating potentially less consistent variation. The differences in the vignettes might be the result of the different fact patterns of each vignette including that Carlos is younger than Jonas.

In addition to the primary variables, participants were asked demographic questions about their age, race, gender and experience in years. Overall, these demographic factors had mixed significant effects on the questions. Attorneys with the most experience were less likely to place Jonas into care as a ward ($p < 0.05$). Older attorneys were more likely to leave Jonas with his mom ($p < 0.05$). Similarly, for the Carlos ward question, the oldest attorneys were less likely to favor placing him into care as a ward ($p < 0.05$). Taken together, there is a slight indication to suggest that older attorneys might be more skeptical of taking children into care in the context of the vignettes, but the findings are not significant across all the questions, so this interpretation should be taken with caution.

The other somewhat consistent demographic finding is that Black or African American attorneys were more likely to favor both Jonas staying with his mother ($p < 0.05$). They were also more likely to disapprove of Carlos being taken into care as a ward ($p < 0.05$), and being placed into a foster home ($p < 0.05$). Though again, there are not significant finding for all questions and given the small number of Black or African American participants, these should also be interpreted with caution.

### *4.3. Dalgleish scores*

The last explanatory variable tested with the participant's Dalgleish number, representing their internal beliefs about child safety and family preservation. The results of this were mixed across the two vignettes.

For the Jonas ward question, the positive coefficient with a $p$-value of ($p < 0.01$), the Dalgleish number, implies that for each unit increase in the score, there's a slight yet significant increase in the likelihood of recommending Jonas to become a ward of the state. For the Jonas mom placement question, a negative coefficient and p-value ($p < 0.01$) suggests that attorneys with high Dalgleish scores are less likely to favor a placement with the mother.

Overall, for the Jonas questions, this suggests that attorneys who are inherently more cautious and risk-averse when it comes to child safety are slightly more likely to recommend making Jonas a ward of the state, a tiny bit more likely to suggest foster care placement (though not at a significant level) and slightly more likely to disapprove of placement with the mom.

As for the Carlos vignette, the Carlos ward of the state question had a positive coefficient but was not significant. The Carlos foster care question has a positive coefficient and a $p$ value of ($p < 0.05$), suggesting that as Dalgleish score increases, attorney decisions to place in a foster home is also expected to slightly increase. Finally for the Carlos Mom question with a negative coefficient of $-0.008$ and p-value ($p < 0.05$) the Dalgleish score has a significant negative effect on decision to place Carlos back with his mom, however the coefficient indicates a only a very small decrease.

All totaled, the Dalgleish score seemed to have less of an impact on the Carlos questions than on the Jonas questions, but the direction of the coefficients were similar for both sets of questions, indicating that core beliefs about the purpose of foster care do have some effects on decisions in the context of these vignettes.

**Table 2**
Carlos treatment effects.

| Scenario | Risk score level | Coefficient (effect size) | Statistical significance | Interpretation | Effect size interpretation |
|---|---|---|---|---|---|
| Carlos Ward (being taken into state custody) | Medium | 0.373 | $p < 0.001$ | More likely to support taking Carlos into care of the state agency | Moderate |
| Carlos Ward | High | 0.794 | $p < 0.001$ | More likely to support Carlos being taken into care | Substantial |
| Carlos Ward | Low (vs High) | −0.794 | $p < 0.001$ | Less likely to approve of putting Carlos into CPS custody | Substantial |
| Carlos Foster | Medium | 0.351 | $p < 0.001$ | More likely to recommend Carlos be placed in a foster home | Moderate |
| Carlos Foster | High | 0.781 | $p < 0.001$ | More likely to place Carlos into a foster home | Substantial |
| Carlos Foster | Low (vs High) | −0.781 | $p < 0.001$ | Less likely to place Carlos in the foster home | Substantial |
| Carlos Mom | Medium | −0.308 | $p < 0.006$ | Less likely to place Carlos with his mother | Small to Moderate |
| Carlos Mom | High | −0.775 | $p < 0.001$ | Even less likely to place Carlos with his mother | Substantial |
| Carlos Mom | Low (vs High) | 0.776 | $p < 0.001$ | More likely to place Carlos with his mother compared to high treatment | Substantial |

### 4.4. Confidence questions

Turning now to the confidence questions, Table 3 shows the means of the combined questions for both vignettes.

As can be seen by both the central tendency numbers, there is very little change in the confidence of attorneys and judges following the risk score treatment. When the confidence questions were run through the same SEM models as the primary questions, none of them produced any significant results. Neither the treatment effects, the demographic questions, nor the Dalgliesh scores were significant for any of the Jonas confidence questions. The Carlos vignette did show some significant effects. For both the ward of the state and the foster placement questions, the medium treatment showed significant effects and for the mom placement question a higher Dalgleish score was associated with decreased confidence. However, given the lack of significance in any of the Jonas questions and that somewhat strangely, only the medium treatment showed any significance, these results should be interpreted with caution.

## 5. Discussion

From the models above, it seems that at least in this context, attorney's legal decisions about child welfare removal and foster care or home placement can be influenced by the inclusion of a predictive model risk score. This is consistent with Engel and Grgic-Hlaca's (2021) findings that predictive model scores in a legal context could change people's opinions. However, this is the first time that lawyers and judges have ever been directly tested in a child welfare context.

Though it should not be surprising that a predictive model score can influence legal decision making, the results should be sobering. The decisions regarding removal and placement are often the first ones to be reviewed by lawyers and judges at the beginning of a foster care case and represent essentially the core issue of the legal dispute. The role of judges and attorneys is to make legal decisions and legal recommendations in line with the law, applying the facts of a case to the applicable statutory and case law. The predictive child risk models add new facts to a case that can influence legal conclusions.

The results from the vignette study show small but significant differences between pre and post treatment decisions. This implies that for difficult cases, like the ones presented in the vignettes, a predictive model can swing removal and placement decisions. While this result might be expected when attorneys saw high and low risk scores, the ability of even medium scores (10) to push legal decisions towards removal and away from placement with the mother, implies that even the quantification of a moderate risk tends to raise concerns about child safety and lead to more conservative decisions. This is however in keeping with the results from Fitzpatrick and Wildman (2021) that found that predictive scores to the DCDA above the median did increase the likelihood that Colorado CPS would begin an investigation. In their study, the mean DCDA score for a child was just above 8 (Fitzpatrick & Wildman, 2021).

In a related study of an Ontario CPS actuarial risk instrument, King et al. (2021) found that cases with moderate risk scores represented 45 % of cases transferred to open investigations and even the cases that the risk assessment showed only one risk factor were more likely to be transferred. Bosk and Feely (2020) also wrote that many states have risk assessment policies that reinforce risk scores over human discretion so that in borderline cases where the intake worker disagrees with the risk assessment score, the tool is given priority. So medium scores can have an impact on CPS decision making.

Rachlinski et al. (2015) demonstrated in both criminal and civil cases that judges are subject to the anchoring effect both in damages they award and the criminal sentences they give out. They wrote that judges make unreliable decisions and that "anchoring thus undermines the rule of law by introducing an element of arbitrariness into judicial decisions." (Rachlinski et al., 2015, p. 737). Fraidin (2013) wrote that child welfare judges in particular hear the same case potentially dozens of times and might be subject to a bolstering bias where they weigh their original opinions higher than new evidence.

Taken together, the potential for a predictive model to both anchor and then subsequently affect future decisions raises serious issues for their use in the courtroom. A judge or a CPS attorney responsible for prosecuting the agency's case could be influenced even by a moderate score to order or recommend removal and foster placement and then this score could potentially affect reunification decisions at a future date. The long-term effect of predictive models on child welfare decisions is a topic that future studies in the field should consider, but is unknown at the moment.

As discussed in the literature section, these models are often built using historical data that in itself might be inaccurate, biased or not up to date, so a family might get a moderate score with some ease. One of the ACLU criticisms of the AFST is that the variable weights used within the model are not published and understood, so how a score is developed is a mystery to those that see it (Gerchick et al., 2023). In the survey, attorneys were presented as part of the vignette with a 90 % accuracy score for the model, however none of the models discussed so far have an accuracy that high. For instance, the AFST reports an area under the curve (AUC) accuracy between 70 and 80 % (Allegheny County Department of Human Services, 2018). Gerchick et al. (2023) argued that this primary focus on AUC performance is not a reliable measure for how the model actually performs across multiple factors. How attorneys might respond to a predictive model with a lower reported accuracy remains another open question for further research.

One general ethics recommendation for the implementation of child welfare predictive models often is the inclusion of stakeholders

**Table 3**
Combined confidence means.

| Category | Jonas 1 | Jonas 2 | Carlos 1 | Carlos 2 |
|---|---|---|---|---|
| Ward | 3.541 | 3.632 | 3.648 | 3.717 |
| Foster | 3.545 | 3.554 | 3.677 | 3.629 |
| Mom | 3.462 | 3.549 | 3.593 | 3.62 |

in the development (Saxena et al., 2020; Casey Family Programs, 2018; Rahman & Keseru, 2021). In the early creation days of the AFST, stakeholder meetings were held that included court staff and the ACLU, but there is no record of how attorneys responded to the tool (Allegheny County Department of Human Services, 2018). However, as Saxena et al. (2020) argued, true stakeholder engagement involves more than just a simple initial meeting, a human centered approach involves stakeholders at every stage of development through the implementation and interpretation of data. How attorneys perceive these tools and their ethical duties using them remains another open question for further research.

The dearth of attorneys is especially concerning given that these models have now been shown in this context to have the ability to change the outcomes of legal decision making. The literature focus has thus far been limited to discussions of caseworkers and CPS agencies, but all of their decisions are ultimately litigated by attorneys and approved by judges. In the U.S. system, judges are the final decision makers in child welfare. There are no CPS decisions that cannot be approved or overruled by a judge. Removal, placement, adoption, normalcy, permanency and reunification decisions are all ultimately legal decisions made by a court. Attorneys through participation in case and family conferences, mediation and advocacy are intimately involved in each of these decisions.

In the interview with Pittsburgh child welfare attorney, Robin Frank, she stated that courts and attorneys had been discouraged from inquiring about the AFST number. This is part of the design of the tool itself (Allegheny County Department of Human Services, 2018). Mills (2019) argued that this lack of transparency was actually a core ethical component of the AFST in that it protected judges from being influenced by the score.

In its use for CPS intake workers, the AFST scores are hidden until the workers make an initial investigation decision and this is done to prevent the score from biasing the worker's decision, but scores above 18 are considered a mandatory screen in and intake supervisors are required to override a mandatory score. (Chouldechova et al., 2018). However, if a score prompts an investigation and an open CPS case, then that score is evidence in the case and therefore should be part of the legal record just like a parents' past substance use, criminal conviction or prior CPS history. The score is evidence in the case just like any other.

As discussed above, anchoring and other biases are often built into judicial opinions, so in some ways withholding risk scores might seem like a reasonable approach to prevent bias from creeping into legal decisions. However, legal decisions come with legal and ethical duties imposed by the law and shielding judges and attorneys is contrary to both. One of the interesting aspects of the exclusion of the legal community from these predictive model conversations is that lawyers actually have an ethical duty to be involved. ABA Model Rule 1.1 describes that lawyers have a duty of competence and in their commentary the American Bar Association (2019) wrote that this duty includes understanding technology that affects the practice of law and their clients.

ABA Model Rule 8.4 goes further and prohibits an attorney from discriminating against a protected class of persons, which implies that attorneys need to actively understand how models that their clients are using or are being used against their clients might be biased and discriminatory (American Bar Association, 2019). Judicial codes also prohibit judges from being themselves biased against protected categories of people and allowing bias from the attorneys practicing in their court (Florida Code of Judicial Conduct, 2023; Texas Code of Judicial Conduct, 2019).

Finally, attorneys have a duty under ABA Model Rule 1.3 to diligently represent their clients which included the often-used idea of zealous advocacy (American Bar Association, 2023). Zealous representation would at minimum require basic discovery, which is the formal request to the other parties for information relevant to the litigation, usually in the form of documents, or written interrogatories, but it can also include sworn deposition testimony. Rule 26 of the Federal Rules of Civil Procedure (2000) states that unless the court orders otherwise, all parties are entitled to any nonprivileged material.

A similar rule can be found in every state. Without delving too deeply into the legal weeds, the duty to represent and the obligations to conduct discovery together mean that attorneys have to ask about any predictive models being used and there is no obvious legal reason why CPS agencies can deny providing them.

To be succinct, this experiment has shown that predictive models can influence child welfare legal decisions. Prior research tells us that judges are subject to the same anchors and biases as the rest of humanity. This concern led model designers to exclude the legal community from much of the knowledge related to their regular use. The legal community however has an ethical duty to represent their clients, hear cases impartially and be aware of possible biases that technology brings, which ultimately will entail being exposed to predictive model scores. All the attorneys on a CPS case and the court hearing the case have a legal and ethical duty to evaluate any CPS risk score. This obligation is mandatory and not subject to the ethical preferences of outside researchers. Judicial and attorney ethical rules demand that they consider all relevant evidence and once a risk score is created and used for a CPS case, the legal practitioners on that case do not have a choice to ignore it.

Though this paper has focused greatly on the AFST, this is just the timeliest example as the push to make better decisions on behalf of children will lead to more child welfare predictive model systems being developed around the world. Stakeholder input needs to include the legal community at all stages and it is imperative that the legal community have a greater say in the development and use of predictive models that impact the lives of the clients they represent and the case before the court.

It is also important that the child welfare legal community itself be trained on the uses, potential and biases of predictive models. Arguably the lack of legal attention until now led to the type of circumstances that prompted a Department of Justice investigation of the AFST (Ho & Burke, 2023). A person would be ill advised to write a contract without consulting a lawyer, certainly deploying a predictive model that can bring the weight of CPS authority upon a family should merit a bit of legal advice.

This present study is limited by the sample size and the number of judicial participants, so if judges have a different response than attorneys is still unknown. A larger sample of attorneys would also allow subsequent research to determine if the role an attorney plays in the child welfare system affects their response to a predictive model score. Follow up research that focuses on the legal community's responses where these models have been deployed would also be helpful to learn how the models might be affecting attorney representation.

The use of specific vignettes also means that the results cannot be broadly interpreted to all child welfare legal decisions. Certainly, any changes to the specific factors within each vignette, such as the age, gender and race of the child, could have significant effects on legal decision making and should be the subject of further research.

## 6. Conclusion

Predictive models in many forms are coming for the child welfare field. Many of them are already here. The machine advice they give can change legal conclusions and the legal community has a duty to ethically engage with and debate this advice. How the child welfare legal community thinks about and will respond to this changing technology and how different models will affect their decisions are open research questions that need to be answered hopefully before more of these predictive models are deployed.

## Funding statement

This work was funded by the Max Planck Institute for Research on Collective Goods.

## CRediT authorship contribution statement

**Matthew Trail:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Declaration of competing interest

None.

## Data availability

My data and my Stata results are both available at OCRID.

## Appendix 1

| Variables[a] | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Jonas_2_Ward | / | Jonas_2_Mom | / | Jonas_2_Foster_ | / |
| How likely are you to order that Jonas be taken into the custody of the state? | 0.749*** | | | | | |
| | (0.666–0.832) | | | | | |
| Medium Treatment | 0.365*** | | −0.401*** | | 0.174 | |
| | (0.131–0.599) | | (−0.624 to −0.179) | | (−0.062–0.410) | |
| High Treatment | 0.631*** | | −0.916*** | | 0.745*** | |
| | (0.401–0.861) | | (−1.139 to −0.694) | | (0.512–0.977) | |
| Years Since First Licensed as Attorney = 3–6 years | −0.041 | | −0.025 | | 0.425 | |
| | (−0.770–0.688) | | (−0.731–0.682) | | (−0.316–1.165) | |
| Years Since First Licensed as Attorney = 6–9 years | −0.440 | | 0.516 | | −0.086 | |
| | (−1.144–0.264) | | (−0.164–1.197) | | (−0.802–0.630) | |
| Years Since First Licensed as Attorney = 9–12 years | −0.438 | | 0.047 | | 0.244 | |
| | (−1.156–0.280) | | (−0.647–0.742) | | (−0.485–0.973) | |
| Years Since First Licensed as Attorney = 12–15 years | −0.196 | | 0.080 | | 0.320 | |
| | (−0.931–0.539) | | (−0.629–0.790) | | (−0.427–1.067) | |
| Years Since First Licensed as Attorney = 15–18 years | −0.601 | | 0.299 | | −0.029 | |
| | (−1.328–0.125) | | (−0.403–1.002) | | (−0.767–0.709) | |
| Years Since First Licensed as Attorney = 18+ years | −0.640* | | 0.369 | | −0.075 | |
| | (−1.342–0.063) | | (−0.309–1.048) | | (−0.789–0.639) | |
| Years Since First Licensed as Attorney = Not licensed/Law student | −0.487 | | 0.222 | | 0.355 | |
| | (−1.186–0.213) | | (−0.452–0.897) | | (−0.352–1.062) | |
| How old are you? = 25–34 years old | 0.402 | | 0.261 | | −0.206 | |

(continued on next page)

(*continued*)

| Variables[a] | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Jonas_2_Ward | / | Jonas_2_Mom | / | Jonas_2_Foster_ | / |
| | (−0.330–1.134) | | (−0.445–0.966) | | (−0.951–0.540) | |
| How old are you? = 35–44 years old | 0.453 | | 0.338 | | −0.226 | |
| | (−0.332–1.238) | | (−0.416–1.093) | | (−1.025–0.572) | |
| How old are you? = 45–54 years old | 0.633 | | 0.183 | | −0.191 | |
| | (−0.143–1.409) | | (−0.565–0.932) | | (−0.980–0.599) | |
| How old are you? = 55–64 years old | 0.685* | | −0.021 | | −0.016 | |
| | (−0.126–1.496) | | (−0.802–0.760) | | (−0.840–0.809) | |
| How old are you? = 65+ years old | 0.526 | | 0.189 | | −0.054 | |
| | (−0.330–1.382) | | (−0.635–1.013) | | (−0.925–0.816) | |
| Race/Ethnicity - Black or African American | −0.076 | | 0.561* | | 0.268 | |
| | (−0.706–0.555) | | (−0.045–1.167) | | (−0.371–0.906) | |
| Race/Ethnicity - Hispanic or Latino | 0.012 | | 0.445 | | 0.246 | |
| | (−0.615–0.639) | | (−0.160–1.051) | | (−0.390–0.883) | |
| Race/Ethnicity - Native American or Alaskan Native | −0.220 | | 0.733* | | −0.480 | |
| | (−1.008–0.568) | | (−0.027–1.493) | | (−1.280–0.321) | |
| Race/Ethnicity - White or Caucasian | 0.219 | | 0.384 | | 0.338 | |
| | (−0.308–0.745) | | (−0.124–0.893) | | (−0.197–0.873) | |
| Race/Ethnicity - Multi-racial or Biracial | −0.055 | | 0.750* | | 0.105 | |
| | (−0.867–0.758) | | (−0.033–1.533) | | (−0.721–0.931) | |
| Race/Ethnicity - A race or ethnicity not listed here. | −0.076 | | 0.865 | | −0.388 | |
| | (−1.224–1.072) | | (−0.242–1.972) | | (−1.554–0.778) | |
| Which of the following best describes you? - Male | 0.230 | | −0.256* | | 0.193 | |
| | (−0.052–0.512) | | (−0.527–0.014) | | (−0.093–0.480) | |
| Which of the following best describes you? - Non-binary | −0.174 | | 0.358 | | 0.096 | |
| | (−1.229–0.881) | | (−0.662–1.378) | | (−0.980–1.171) | |
| Which of the following best describes you? - Prefer not to answer | −0.101 | | 0.241 | | −0.184 | |
| | (−1.515–1.313) | | (−1.123–1.606) | | (−1.620–1.252) | |
| Dalgleish Score | 0.011*** | | −0.011*** | | 0.003 | |
| | (0.004–0.017) | | (−0.017 to −0.005) | | (−0.004–0.009) | |
| var(e.Jonas_2_Ward) | | 0.485*** | | | | |
| | | (0.397–0.574) | | | | |
| How likely are you to order that Jonas remain in his mother's home? | | | 0.765*** | | | |
| | | | (0.690–0.839) | | | |
| var(e.Jonas_2_Mom) | | | | 0.452*** | | |
| | | | | (0.369–0.534) | | |
| How likely are you to order that Jonas be removed and placed in a foster home | | | | | 0.749*** | |
| | | | | | (0.670–0.828) | |
| var(e.Jonas_2_Foster_) | | | | | | 0.501*** |
| | | | | | | (0.409–0.592) |
| Constant | −0.002 | | 0.515 | | 0.070 | |
| | (−1.067–1.064) | | (−0.531–1.561) | | (−1.008–1.149) | |
| Observations | 231 | 231 | 231 | 231 | 231 | 231 |

| Variables | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Carlos_2_Ward | / | Carlos_2_Foster | / | Carlos_2_Mom | / |
| How likely are you to order that Carlos be taken into the custody of the state? | 0.871*** | | | | | |
| | (0.036) | | | | | |
| Medium treatment = 2 | 0.383*** | | 0.422*** | | −0.413*** | |
| | (0.093) | | (0.094) | | (0.119) | |
| High treatment = 3 | 0.808*** | | 0.863*** | | −0.804*** | |
| | (0.099) | | (0.100) | | (0.126) | |
| Years Since First Licensed as Attorney = 3–6 years | −0.222 | | 0.212 | | −0.025 | |
| | (0.302) | | (0.302) | | (0.384) | |
| Years Since First Licensed as Attorney = 6–9 years | −0.134 | | 0.175 | | −0.297 | |
| | (0.291) | | (0.292) | | (0.371) | |

(*continued on next page*)

(*continued*)

| Variables | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Carlos_2_Ward | / | Carlos_2_Foster | / | Carlos_2_Mom | / |
| Years Since First Licensed as Attorney = 9–12 years | 0.138 | | 0.328 | | −0.402 | |
| | (0.298) | | (0.297) | | (0.379) | |
| Years Since First Licensed as Attorney = 12–15 years | −0.125 | | 0.341 | | −0.236 | |
| | (0.304) | | (0.304) | | (0.388) | |
| Years Since First Licensed as Attorney = 15–18 years | −0.230 | | 0.078 | | −0.003 | |
| | (0.299) | | (0.299) | | (0.383) | |
| Years Since First Licensed as Attorney = 18+ years | −0.043 | | 0.133 | | −0.145 | |
| | (0.290) | | (0.291) | | (0.370) | |
| Years Since First Licensed as Attorney = Not licensed/Law student | −0.090 | | 0.307 | | −0.157 | |
| | (0.287) | | (0.288) | | (0.367) | |
| How old are you? =25–34 years old | 0.096 | | 0.069 | | −0.463 | |
| | (0.302) | | (0.302) | | (0.385) | |
| How old are you? = 35–44 years old | 0.158 | | 0.121 | | −0.265 | |
| | (0.323) | | (0.322) | | (0.411) | |
| How old are you? = 45–54 years old | 0.085 | | 0.053 | | −0.199 | |
| | (0.319) | | (0.320) | | (0.408) | |
| How old are you? = 55–64 years old | −0.071 | | −0.018 | | −0.320 | |
| | (0.335) | | (0.334) | | (0.426) | |
| How old are you? = 65+ years old | 0.201 | | 0.005 | | −0.654 | |
| | (0.350) | | (0.351) | | (0.447) | |
| Race/Ethnicity - Black or African American | −0.630** | | −0.499* | | 0.519 | |
| | (0.258) | | (0.259) | | (0.331) | |
| Race/Ethnicity - Hispanic or Latino | −0.451* | | −0.391 | | 0.768** | |
| | (0.257) | | (0.258) | | (0.329) | |
| Race/Ethnicity Native American or Alaskan Native | −0.396 | | −0.496 | | 0.898** | |
| | (0.324) | | (0.326) | | (0.415) | |
| Race/Ethnicity - White or Caucasian | −0.508** | | −0.327 | | 0.800*** | |
| | (0.216) | | (0.216) | | (0.276) | |
| Race/Ethnicity, Multi-racial or Biracial | −0.614* | | −0.133 | | 0.499 | |
| | (0.334) | | (0.334) | | (0.426) | |
| Race/Ethnicity - A race or ethnicity not listed here. | −0.991** | | −0.739 | | 0.365 | |
| | (0.470) | | (0.472) | | (0.601) | |
| Which of the following best describes you? - Male | −0.146 | | 0.119 | | −0.204 | |
| | (0.113) | | (0.113) | | (0.144) | |
| Which of the following best describes you? Non-binary | 0.520 | | 0.046 | | −0.345 | |
| | (0.437) | | (0.439) | | (0.558) | |
| Which of the following best describes you? - Prefer not to answer | 0.018 | | −0.087 | | −0.371 | |
| | (0.583) | | (0.585) | | (0.746) | |
| Dalgleish Scale | 0.004 | | 0.006** | | −0.008** | |
| | (0.003) | | (0.003) | | (0.004) | |
| var(e.Carlos_2_Ward) | | 0.317*** | | | | |
| | | (0.029) | | | | |
| How likely are you to order that Carlos be removed and placed in a foster home? | | | 0.857*** | | | |
| | | | (0.039) | | | |
| var(e.Carlos_2_Foster) | | | | 0.319*** | | |
| | | | | (0.030) | | |
| How likely are you to order that Carlos remain in his mother's home? | | | | | 0.794*** | |
| | | | | | (0.040) | |
| var(e.Carlos_2_Mom) | | | | | | 0.518*** |
| | | | | | | (0.048) |
| Constant | 0.508 | | 0.145 | | 0.978* | |
| | (0.433) | | (0.434) | | (0.580) | |
| Observations | 231 | 231 | 231 | 231 | 231 | 231 |

[a] p 0.01 "***" 0.05 "**" 0.1 "*"

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chiabu.2024.106943.

## References

Ahn, E., Gil, Y., & Putnam-Hornstein, E. (2021). Predicting youth at high risk of aging out of foster care using machine learning methods. *Child Abuse & Neglect, 117*, Article 105059. https://doi.org/10.1016/j.chiabu.2021.105059

Allegheny County Department of Human Services. (2018). Frequently asked questions. Retrieved April 27, 2024 https://www.alleghenycountyanalytics.us/wp-content/uploads/2018/10/17-ACDHS-11_AFST_102518.pdf.

American Bar Association. (2019). House of Delegates Resolution. Retrieved April 27, 2024 https://www.americanbar.org/content/dam/aba/directories/policy/annual-2019/112-annual-2019.pdf.

American Bar Association. (2022). ABA profile of the legal profession. Retrieved April 27, 2024 https://www.americanbar.org/content/dam/aba/administrative/news/2022/07/profile-report-2022.pdf.

American Bar Association. (2023). Model rules of professional conduct, rule 1.3 diligence – Comment. Retrieved April 27, 2024 https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_3_diligence/comment_on_rule_1_3/.

Bogert, E., Schecter, A., & Watson, R. T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports, 11*(1), 8028. https://doi.org/10.1038/s41598-021-87480-9

Bosk, E., & Feely, M. (2020). The goldilocks problem: Tensions between actuarially based and clinical judgment in child welfare decision making. *Social Service Review, 94*(4), 659–692. https://doi.org/10.1086/712060

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Casey Family Programs. (2018). Considerations for implementing predictive analytics in child welfare. Retrieved from https://www.casey.org/media/Considerations-for-Applying-Predictive-Analytics-in-Child-Welfare.pdf.

Cheng, H. F., Stapleton, L., Kawakami, A., Sivaraman, V., Cheng, Y., Qing, D., … Zhu, H. (2022). How child welfare workers reduce racial disparities in algorithmic decisions. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3491102.3501831

Chouldechova, A., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In , *Proceedings of Machine Learning Research: 81. Proceedings of the 1st conference on fairness, accountability and transparency* (pp. 134–148). Available from https://proceedings.mlr.press/v81/chouldechova18a.html.

Collins, D. (2018). *'Predictive policing': Big-city departments face lawsuits*. The Associated Press. retrieved from https://apnews.com/article/b11e4bca11e548d3af7a63f24e348c6f.

Dalgleish, L. (2010). Balance of work focus in child welfare. Unpublished manuscript.

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

De-Arteaga, M., Fogliato, R., & Chouldechova, A. (2022). A case for humans-in-the-loop: Decisions in the presence of erroneous algorithmic scores. Retrieved April 23, 2024 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4050125.

Dettlaff, A. J., Hollinshead, D., Baumann, D. J., & Fluke, J. D. (2020). 81Instrumentation to understand the child protective services decision-making processes. In J. D. Fluke, M. L. López, R. Benbenishty, E. J. Knorth, & D. J. Baumann (Eds.), *Decision-making and judgment in child welfare and protection: Theory, research, and practice* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780190059538.003.0004.

Drake, B., Jonson-Reid, M., Ocampo, M. G., Morrison, M., & Dvalishvili, D. D. (2020). A practical framework for considering the use of predictive risk modeling in child welfare. *The Annals of the American Academy of Political and Social Science, 692*(1), 162–181. https://doi.org/10.1177/0002716220978200

Elgin, D. J. (2018). Utilizing predictive modeling to enhance policy and practice through improved identification of at-risk clients: Predicting permanency for foster children. *Children and Youth Services Review, 91*, 156–167. https://doi.org/10.1016/j.childyouth.2018.05.030

Elish, M. C. (2019). *Moral crumple zones: Cautionary tales in human-robot interaction (pre-print) engaging science, technology, and society*. https://doi.org/10.2139/ssrn.2757236

Engel, C., & Grgic-Hlaca, N. (2021). Machine advice with a warning about machine limitations: Experimentally testing the solution mandated by the Wisconsin Supreme Court. *Journal of Legal Analysis.*. https://doi.org/10.1093/jla/laab001

Federal Rules of Civil Procedure. (2000). Rule 26. Duty to disclose; General provisions governing discovery. Retrieved April 27, 2024 https://www.law.cornell.edu/rules/frcp/rule_26.

Field, A., Coston, A., Gandhi, N., Chouldechova, A., Putnam-Hornstein, E., Steier, D., & Tsvetkov, Y. (2023). Examining risks of racial biases in NLP tools for child protective services. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*. https://doi.org/10.1145/3593013.3594094

Fitzpatrick, M., & Wildman, C. (2021). Final report on Douglas County Decision Aid (DCDA) predictive risk modeling randomized control trial experiment. https://www.muckrock.com/foi/colorado-127/child-welfare-predictive-analytics-models-122003/.

Florida Supreme Court. (2023). Code of judicial conduct for the state of Florida. Retrieved April 27, 2024 https://supremecourt.flcourts.gov/Opinions/Judicial-Ethics-Advisory-Committee/Code-of-Judicial-Conduct/Canon-3.

Fraidin, M. I. (2013). Decision-making in dependency court: Heuristics, cognitive biases, and accountability. In *, Vol. 60. Cleveland state law review* (p. 913), 2013 https://ssrn.com/abstract=2358602.

Gerchick, M., Jegede, T., Shah, T., Gutierrez, A., Beiers, S., Shemtov, N., Xu, K., Samant, A., & Horowitz, A. (2023). The devil is in the details: Interrogating values embedded in the allegheny family screening tool. Retrieved April 27, 2024 https://www.aclu.org/the-devil-is-in-the-details-interrogating-values-embedded-in-the-allegheny-family-screening-tool.

Gillingham, P. (2019). The development of algorithmically based decision-making systems in children's protective services: Is administrative data good enough? *The British Journal of Social Work, 50*(2), 565–580. https://doi.org/10.1093/bjsw/bcz157

Goldhaber-Fiebert, J., & Prince, L. (2019). *Impact evaluation of a predictive risk modeling tool for Allegheny county's child welfare office*. Pittsburgh: Allegheny County. Retrieved April 27, 2024 https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-6.pdf.

Grgić-Hlača, N., Castelluccia, C., & Gummadi, K. P. (2022). Taking advice from (dis)similar machines: The impact of human-machine similarity on machine-assisted decision-making. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 10*(1), 74–88. https://doi.org/10.1609/hcomp.v10i1.21989

Hamilton, M. (2019). Debating algorithmic fairness*. *UC Irvine Law Review, 52*, 261. https://doi.org/10.5064/F6JOQXNF

Ho, S., & Burke, G. (2023). *Here's how an AI tool may flag parents with disabilities*. The Associated Press. Retrieved from https://apnews.com/article/child-protective-services-algorithms-artificial-intelligence-disability-02469a9ad3ed3e9a31ddae68838bc76e.

Hohenstein, J., & Jung, M. (2020). AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior, 106*, Article 106190. https://doi.org/10.1016/j.chb.2019.106190

Hollinshead, D., Currie, D., Kroll, K., Feldman, S., Monahan-Price, K., & Fluke, J. (2021). Associations between case, staff, and agency characteristics and the decision to place a child in out-of-home care. *International Journal on Child Maltreatment, 4*, 325–347. https://doi.org/10.1007/s42448-021-00083-9

Horowitz, M., & Kahn, L. *Bending the automation bias curve: A study of human and ai-based decision making in national security contexts*. (2023). Retrieved from https://arxiv.org/abs/2306.16507.

Jones-Jang, S. M., & Park, Y. J. (2022). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication, 28*(1). https://doi.org/10.1093/jcmc/zmac029

Kawaguchi, K. (2021). When will workers follow an algorithm? A field experiment with a retail business. *Management Science, 67*(3), 1670–1695. https://doi.org/10.1287/mnsc.2020.3599

Keddell, E. (2019a). Algorithmic justice in child protection: Statistical fairness, social justice and the implications for practice. *Social Science, 8*, 281. https://doi.org/10.3390/socsci8100281

Keddell, E. (2019b). Risk prediction tools in child welfare contexts: the devil in the detail. retrieved from https://www.husita.org/risk-prediction-tools-in-child-welfare-contexts-the-devil-inthe-detail/.

Keddell, E., & Hyslop, I. (2018). Role type, risk perceptions and judgements in child welfare: A mixed methods vignette study. *Children and Youth Services Review, 87*, 130–139. https://doi.org/10.1016/j.childyouth.2018.02.017

King, B., Black, T., Fallon, B., & Lung, Y. (2021). The role of risk in child welfare decision-making: A prospective cohort examination of families transferred to ongoing child protection services after an investigation. *Children and Youth Services Review, 125*, Article 105983. https://doi.org/10.1016/j.childyouth.2021.105983

Legal Service Corporation. (2021). By the numbers: The data underlying legal aid programs. Retrieved April 27, 2024 https://lsc-live.app.box.com/s/e0ejxcp3lp4msvp77mvrzy9w0z354tf7.

Leslie, D., Holmes, L., Hitrova, C., & Ott, E. (2020). *Ethics review of machine learning in children's social care*. Report. London, UK: What Works for Children's Social Care. Available at SSRN: https://ssrn.com/abstract=3544019.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.2139/ssrn.2941774

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change, 175*, Article 121390. https://doi.org/10.1016/j.techfore.2021.121390

Middel, F., López López, M., Fluke, J., & Grietens, H. (2022). Racial/ethnic and gender disparities in child protection decision-making: What role do stereotypes play? *Child Abuse & Neglect, 127*, Article 105579. https://doi.org/10.1016/j.chiabu.2022.105579

Mills, U. (2019). *'Less bad' bias: An analysis of the Allegheny Family Screening Tool*. MPS Seminar One. Retrieved April 27, 2024 https://medium.com/mps-seminar-one/less-bad-bias-an-analysis-of-the-allegheny-family-screening-tool-ef6ffa8a56fb.

Moore, T. D., McDonald, T. P., & Cronbaugh-Auld, K. (2016). Assessing risk of placement instability to aid foster care placement decision making. *Journal of Public Child Welfare, 10*(2), 117–131. https://doi.org/10.1080/15548732.2016.1140697

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (New York, N. Y.), 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Purdy, J., & Glass, B. (2023). The pursuit of algorithmic fairness: On "correcting" algorithmic unfairness in a child welfare reunification success classifier. *Children and Youth Services Review, 145*, Article 106777. https://doi.org/10.1016/j.childyouth.2022.106777

Rachlinski, J. J., Wistrich, A. J., & Guthrie, C. (2015). Can judges make reliable numeric judgments? Distorted damages and skewed sentences. *Indiana Law Journal, 90*, 6. https://ssrn.com/abstract=2932323.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). *Mitigating bias in algorithmic hiring: Evaluating claims and practices*. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain. https://doi.org/10.1145/3351095.3372828.

Rahman, Z., & Keseru, J. (2021). Predictive analytics for children: An assessment of ethical considerations, risks, and benefits. UNICEF Office of Research - Innocenti Working Paper WP 2021-08 https://www.unicef-irc.org/publications/pdf/WP_2021_08.pdf.

Raji, I. D., & Fried, G. (2021). About face: A survey of facial recognition evaluation. *AAAI 2020 Workshop on AI Evaluation. ArXiv* (abs/2102.00813).

Samant, A., Horowitz, A., Xu, K., & Beiers, S. (2021). *Family surveillance by algorithm: The rapidly spreading tools few have heard of*. ACLU: American Civil Liberties Union. Retrieved April 27, 2024 https://www.aclu.org/sites/default/files/field_document/aclu_family_surveillance_by_algorithm.pdf.

Saxena, D., Badillo-Urquiola, K., Wisniewski, P. J., & Guha, S. (2020). *A human-centered review of algorithms used within the U.S. child welfare system*. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376229.

Saxena, D., Moon, E. S.-Y., Chaurasia, A., Guan, Y., & Guha, S. (2023). *Rethinking "risk" in algorithmic systems through a computational narrative analysis of casenotes in child-welfare*. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3544548.3581308.

Saxena, D., Repaci, C., Sage, M. D., & Guha, S. (2022). *How to train a (bad) algorithmic caseworker: A quantitative deconstruction of risk assessments in child welfare*. Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA. https://doi.org/10.1145/3491101.3519771.

Stepura, K. G., Schwab, J., Baumann, D. J., Sowinski, N., & Thorne, S. (2020). 27Exploration in predictive analysis and machine learning. In J. D. Fluke, M. L. López, R. Benbenishty, E. J. Knorth, & D. J. Baumann (Eds.), *Decision-making and judgment in child welfare and protection: Theory, research, and practice* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780190059538.003.0002.

Stokes, J., & Schmidt, G. (2012). Child protection decision making: A factorial analysis using case vignettes. *Social Work, 57*(1), 83–90. http://www.jstor.org/stable/2371995.

Taylor, B. J. (2005). Factorial surveys: Using vignettes to study professional judgement1. *The British Journal of Social Work, 36*(7), 1187–1207. https://doi.org/10.1093/bjsw/bch345

Texas Supreme Court. (2019). Texas court rules - Code of judicial conduct. https://www.scjc.texas.gov/media/46779/code-of-judicial-conduct.pdf.

Vaithianathan, R., Dinh, H., Kalisher, A., Kithulgoda, C., Kulick, E., Mayur, M., … Prado, D. B. (2019). *Implementing a child welfare decision aide in Douglas County: Methodology report (Tech. Rep.)*. Auckland, New Zealand: Centre for Social Data Analytics. Retrieved April 27, 2024 https://csda.aut.ac.nz/_data/assets/pdf_file/0009/347715/Douglas-County-Methodology_Final_3_02_2020.pdf.

Vaithianathan, R., Jiang, N., Maloney, T., Nand, P., & Putnam-Hornstein, E. (2017). *Developing predictive risk models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation [PDF]*. Auckland: Centre for Social Data Analytics. Retrieved April 27, 2024 https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Methodology-V1-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL.pdf.

Vaithianathan, R., Putnam-Hornstein, E., Chouldechova, A., Benavides-Prado, D., & Berger, R. (2020). Hospital injury encounters of children identified by a predictive risk model for screening child maltreatment referrals: Evidence from the Allegheny family screening tool. *JAMA Pediatrics, 174*(11). https://doi.org/10.1001/jamapediatrics.2020.2770

Willcott, A., & Stewart, L. (2021). Cold case project, annual report of the work of 2019. In *Georgia's child welfare system*. retrieved from http://www.gacip.org/cold-case-project.