Mathematica
Progress Together

Arizona Department of Child Safety Next Event Study

Final Report

June 30, 2023

Jonathan Gellar and Allon Kalisher

This page has been left blank for double-sided copying.

Acknowledgements

We thank the people who contributed to and made this report possible.

We are especially grateful to the staff at the Arizona Department of Child Safety. We particularly thank our project officer, Katherine Guffey, for her guidance and support. We also thank Gretchen Camp and Andrew Weller for their work preparing the data used in the study.

At Mathematica, Evelyn Cody provided excellent programming support. John Hotchkiss and Elizabeth Weigensberg provided review of this report. Aden Bhagwat provided project management support.

Allon Kalisher, Project Director
Jonathan Gellar, Deputy Project Director and Technical Lead

This page has been left blank for double-sided copying.

Contents

Tables

Figures

I. Background

As the statutory definitions of child abuse and neglect in Arizona have shifted, along with changes in policy and practice, the Department of Child Safety (DCS) has observed changes in the volume and percentage of child neglect reports to its hotline. Motivated by a desire to reduce unnecessary child protective services involvement and support more Arizona families through less intrusive engagement with community services, DCS contracted with Mathematica to conduct a study that will give policymakers compelling data to inform decisions about the best ways to support and strengthen families.

The primary goal of this research is to deepen understanding of the features of reports that do and do not merit a response from DCS, with a focus on providing evidence to inform Arizona's procedures and policies on when a DCS report and investigation are necessary. A secondary research aim was to determine which children might not need a child protective response if the right community supports were in place, and which prevention services would be most beneficial to each family. Time and data constraints limited this study's focus to the primary research aim. We will explore the feasibility and potential value of analysis for this secondary research aim, along with other analytic priorities, in subsequent analysis.

Ultimately DCS hopes that a better understanding of features associated with impending danger will make it easier to focus services on families most in need, and that fewer families will receive a protective response when they do not need one. This will in turn decrease the number of reports to DCS along with the need for foster care.

In the rest of this section, we provide more background information on the study, including the formulation and purpose of two advisory councils and why we focused on racial bias as we examined the data. We then turn to an overview of our study methods, share our findings, and conclude with a discussion about those findings and planned next steps. Three appendices share the input we received, technical details, additional data tables and graphs, and a copy of DCS's Hotline Report Decision tool.

This page has been left blank for double-sided copying.

II. Study Overview

A. Analytic framework

The primary research objective for this study is to gain a better understanding of when a child protective response is and is not needed—specifically, by using information collected at the first hotline communication to understand the types of calls that are associated with the most risk of impending danger to the child, and conversely, those associated with child safety. This knowledge can, in conjunction with the perspective of people who have lived experience with the child protection system, inform department policy on which cases do or do not merit investigation.

To learn about the cases that are associated with outcomes related to a child's risk of harm, we fit a series of statistical models. Each model relates information available at baseline to an outcome associated with the necessity of a child protection response. Importantly, unlike other recent modeling efforts in child welfare (Goldhaber-Fiebert & Prince 2019, Vaithianathan et al. 2020), the goal of this analysis is not to identify any individual person with a risk score or build a tool to prompt any particular DCS response. Instead, this analysis is being used to identify what case characteristics are most predictive of future child maltreatment. Given this purpose, the primary focus for the statistical model is to gain insights about the model's predictions. The insights we sought to learn were (a) which information available at intake is most associated with the need for a DCS safety response, (b) what the direction and strength of this association is, and (c) which combinations of case characteristics are commonly associated with these responses. Our results emphasize the effects of case characteristics that are most actionable because they can inform Arizona's procedures, policies, and statutes that define when a DCS report and investigation occur.

B. Advisory councils

DCS formed two advisory councils to inform this study: a Data Analytics Advisory Council (DAAC) and a Community Advisory Council (CAC). The DAAC comprised experts in predictive modeling of child welfare data. It was convened to provide a technical review and advice about the study's analytic plan and interpretation of results. The CAC was made up of Arizona community members whose firsthand experience with the state's child welfare system gave them deep insights into the experiences and concerns of communities where DCS intervention is disproportionately common.

1. Data Analytics Advisory Council

The DAAC included four members: Dr. Peter York, a principal and lead of BCT Partners' data analytics team; Dr. John Fluke, a professor and associate director for systems research and evaluation for the Kempe Center for the Prevention and Treatment of Child Abuse and Neglect at the University of Colorado School of Medicine; Dr. Philip Gillingham, associate investigator for the School of Social Science at the University of Queensland; and Brian Clapier, formerly the senior director for Data Advocacy at Casey Family Programs. The DAAC met in April 2022 and June 2022 to discuss the study objectives, orient council members to our plan and timeline, and give feedback on the draft analytic plan. The DAAC met again in May 2023 to discuss findings and obtain additional methodological advice about interpreting

results. Suggestions from the DAAC were incorporated into the study design and in follow-up analysis. Some of their recommendations are also noted as potential next steps in the conclusion of this report.

2. Community Advisory Council

The objectives for the CAC were: (1) to inform community members about the research objectives of the project, and later to share the study findings, and (2) to learn firsthand about the experiences of the community members so they could inform and guide research decisions to improve the study design. In May 2022, DCS consulted with Prevent Child Abuse Arizona to formulate and engage the CAC. Prevent Child Abuse Arizona partnered with four community leaders to conduct focus groups that gave their communities an opportunity to share their insights. The leaders represented African Americans, Native Americans, residents of South Tucson, and residents of rural Arizona. Together, these four groups constituted the CAC. We incorporated the CAC's perspective in interpreting findings and offering relevant suggestions for next steps. The focus groups met again in May 2023 to give their reaction to the preliminary study findings.

The members of the CAC raised several points that affected the study design, primarily through the inclusion of specific covariates in our predictive models. Table II.1 summarizes some of the important points raised during the focus groups and how they affected the study design.

Table II.1. Community Advisory Council feedback and how it affected the study design.

Summary of point raised during focus groups	Effect on study design
Some callers make "bad faith" reports, using DCS as a weapon against neighbors or relatives whom they intend to harm.	We recorded the type of reporter (for example, their profession and whether they are a mandated reporter), and their relationship to the family (such as a relative or neighbor).
There is inherent bias in the DCS system, particularly by race.	We considered that racial bias underlies these data and took steps to mitigate its effect on the results. This is discussed more in Section II.C, in Chapter IV, and in Appendix B.
We should consider geography, such as zip code.	Because Arizona has 368 active zip codes, we did not include flags for all zip codes in the state. However, we did include several geographic characteristics: county (15 counties in the state), flags for rural Arizona and South Tucson (both based on zip code), and zip code–level socioeconomic information.
We should consider family history, such as previous encounters with the child welfare system.	Our plan was always to include previous encounters (for both the child and caretaker) in the models.
Consider school attendance, visits to the emergency room, and signs of children showing fear.	Though this information is relevant to investigations, we did not have the necessary data available at the time of intake, and thus could not include it in our models.
Presenting relative risks as ratios (e.g., "1.25," with RR = 1 indicating "no difference") is confusing.	We present relative risks as a percent increase (positive) or decrease (negative) relative to a reference group.

Notes taken by Prevent Child Abuse America from all focus groups are included in Appendix A.

C. Focus on racial or ethnic bias

Recent literature has raised concerns about the potential for bias in predictive models, particularly those used for child welfare (Cuccaro-Alamin et al. 2017, Drake et al. 2020, Samant et al. 2021). Although these concerns have mostly been aimed at predictive models that are used in conjunction with active case decisions, and that is not true with this study, we recognized the importance of considering how bias may impact our findings. This concern is because the underlying data that is used to build these models are based on subjective decision making that has its roots in systemic bias in the child welfare system, both conscious and unconscious. For example, if reports against African American parents are more likely to be substantiated or result in removal than reports against White parents, even for similar allegations and circumstances, then a model built on these data could conclude that African American children are not as safe with their biological parents as White children are. This phenomenon has been termed "bias in, bias out," and it could potentially lead to more removals of children from African American homes.

For this project, many of the data elements are the result of subjective decision making and could be affected by conscious or unconscious bias. These data elements include all the outcomes we considered—including whether a report is substantiated, whether a child is removed from the home, and whether any additional services are provided to the families. Part of the concern about bias in the predictive models is mitigated by the fact that we are not building a tool that will be directly used to make decisions on a specific child going forward. Nonetheless, the inferences we make based on this analysis could be incorrect if they are based on biased data.

Our analysis addresses these issues in two ways. First, we assess to the best of our ability whether racial or ethnic bias exists in the data by comparing outcomes across different demographic groups. Then, we use a technique called race-blind prediction to mitigate the impact of this bias on our inferences. We describe this technique in detail later in this report.

III. Methods

A. Study period and sample

We based this study on all qualifying contacts (hotline communications[1]) with Arizona DCS that occurred between February 1, 2021, and September 30, 2022. We aligned the start of this study window with the implementation of a new DCS data entry and collection system. The new system, called Guardian, was instituted in February 2021 and tracks much richer data on reports of abuse and neglect than were collected under the previous system. We have access to data through December 31, 2022, but we do not include reports made after September 2022 to allow a three-month period to observe outcomes. We recognize this is a relatively short window, but almost all removals (99.6 percent) took place within 90 days of the hotline communication. Allowing for a longer window to observe outcomes would necessitate shortening an already brief study period. We also recognize that this period overlaps with the tail end of

[1] A telephone call to the DCS child protective services hotline, alleging child maltreatment, is referred to as a hotline communication. If it is screened in to be investigated, the call is referred to as a report. The DCS hotline receives other hotline communications that are not alleging child maltreatment, including those that are screened out (or not investigated).

the COVID-19 public health emergency. This overlap is a limitation of our study, and we cannot say whether or to what degree this may have influenced the data we studied.

We found 141,190 hotline communications during this period, which is our sample for the analysis of screening in. Among those hotline communications, 76,215 (54 percent) were screened-in reports. We then identified all perpetrator/victim pairs within each of the screened-in reports and expanded the data to the report/perpetrator/victim level (152,474 observations). We made the following exclusions:

- We excluded 6,429 observations because DCS was not able to locate the family.
- We excluded 4,368 observations that were not within the jurisdiction of Arizona DCS (either because they occurred outside the state, or because they occurred on a reservation and were therefore under tribal jurisdiction).
- We excluded 194 observations because we could not determine if a case was open for these records due to missing data.

After applying these exclusions, there were 141,483 observations at the report/perpetrator/victim level.

B. Data description

1. Outcomes

A key challenge in this analysis is defining the outcome of interest. Because true maltreatment rates are impossible to determine, we must predict measures of child maltreatment from the child protection system. We look at four decision points which we believe to be correlated with child maltreatment and the need for child protection:

1. Whether the initial hotline communication (typically, a telephone call) was "**screened in**" and investigated. In Arizona, any call making a child maltreatment claim that meets the state definition of a qualifying allegation is automatically screened in and investigated. Once the call is screened in, we refer to it as a *report*, and the time of the call is referred to as an *intake*.
2. Whether, after an investigation, an allegation was **substantiated**, or found to be true. Substantiation decisions are made by the caseworker who investigates a report and the caseworker's supervisor.[2] Our analysis uses the latest information on substantiation, as of the date the data were received on 03/04/2023. We also include proposed substantiations as a substantiation outcome.
3. Whether an investigation resulted in a **case being opened** for services. DCS may provide in-home case management and offer services intended to address families' treatment needs, strengthen family functioning, and avoid removal of the child from the home. Cases are also opened for services (or remain open for services) when children are removed so that safety management services can be offered.

[2] In Arizona, a decision to substantiate may be appealed. In these cases, the final decision is adjudicated by an Administrative Law Judge.

4. Whether the child was **removed** from the home. Our definition only includes removals that last at least 24 hours, which is consistent with prior analyses conducted by DCS. In addition, we only include removals that occurred within 3 months of the hotline communication. This ensures a consistent time window for observing the removal outcome across the entire study period.

We define an additional outcome that we refer to as a **composite measure**. This outcome is defined only for cases that were screened in and is based on the other three outcomes (substantiation, case opening, and removal). The composite measure is a variable with values from 1 to 5 and is ranked in order of least to most intervention: 1 represents no further intervention, and 5 represents a child removal. Table II.2 delineates how each score is defined.

Table II.2. Definition of the composite measure, based on substantiation, case opening, and removal.

Substantiation	Case opening	Removal	Composite measure	Number and percentage of observations	
No	No	No	1	10,2434	72%
Yes	No	No	2	7,770	5%
No	Yes	No	3	10,407	7%
Yes	Yes	No	4	4,347	3%
Either	Either	Yes	5	16,151	11%

For our five outcomes, we use five analytic samples. Each sample consists of observations that correspond with the outcome of interest. The analytic sample for the screening in outcome is the only sample that includes hotline communications that did not advance to an investigation. In this sample, each observation corresponds to a single hotline communication and includes other relevant information, including whether the hotline communication was screened in (the outcome). For substantiation, case opening, or removal of a child to occur, a report must be made. Each report has a perpetrator, the perpetrator's victim, and the allegations made against the perpetrator. Each report can also have multiple perpetrator-victim pairs and each perpetrator-victim pair will have their own associated allegations. Because the relevance of any given allegation could depend on the presence of other allegations on the same report for the same perpetrator-victim pair, we found it most appropriate to keep all allegations that involve the same perpetrator-victim pair as part of the same observation. Thus, for these four outcomes, each observation in their respective analytic samples corresponds with a victim-perpetrator pair. Additional information that is relevant to those pairs is part of those samples, including which allegations were made and what the outcomes for that pair were.

Table II.3. Unit of analysis, definition, and sample for each of the outcomes in this analysis[3]

Outcome	Unit of analysis	Analytic definition	Analytic sample
Screening In	Hotline communication	Whether the hotline communication was screened in	141,109
Substantiation	Report/perpetrator/ victim	Whether at least one allegation was substantiated	140,960
Case opening	Report/perpetrator/ victim	Whether a case was opened for the associated report	141,483
Removal	Report/perpetrator/ victim	Whether the victim was removed from the home within 3 months of intake for at least 24 hours	141,483
Composite measure	Report/perpetrator/ victim	Refer to Table II.2	141,109

Although the substantiation decision is determined separately for each allegation, we examined the substantiation outcome based on whether any allegations associated with the observation (report/perpetrator/victim) were substantiated. This is because we observed remarkable consistency across the allegations that belong to each report/perpetrator/victim observation. We found that 97 percent of all observations have either all allegations substantiated (83 percent) or all allegations unsubstantiated (14 percent). Only 3 percent have a mix of substantiated and unsubstantiated allegations.

Case opening and removal are more naturally defined as the outcomes corresponding to the report and victim, respectively, associated with observation. However, it is noteworthy that a single report or victim can correspond to multiple observations at the report/perpetrator/victim level; in these cases, we simply use the same outcome with all associated observations.

Because some of the outcomes for substantiation were missing at the time of data collection, the final sample sizes for this outcome and for the composite measure (which uses substantiation when no removal took place), are slightly lower than 141,483. The analytic sample sizes for all outcomes are listed in Table II.3.

2. Case characteristics known at intake

Because the primary goal of this project is to identify characteristics of situations in which a DCS report and subsequent investigation might not be necessary, our analysis uses characteristics that were recorded at the time of the initial hotline communication and historical data available in the DCS system. These data include:

- Demographic information (age, gender, and race[4]) of all individuals (perpetrators and victims) associated with an observation

[3]An outcome is defined at a level in the raw data refers to the unit of analysis for which that outcome pertains. For example, the outcome of a substantiation is defined for every allegation, which means the raw data table including the substantiation outcome contains one record for every allegation. The unit of analysis describes the parameters for defining one record in the analytic dataset.

[4] Individuals in our data can have multiple race/ethnicities listed out of 5 options (listed below). In order to limit the number of race categories, we used a hierarchical structure to define a person's race, in this order: 1) American Indian, 2) African American, 3) Asian/Pacific Islander, 4) Hispanic, 5) White. For example, if an individual has Hispanic and African American as their race, we categorize them as African American.

- Geographic information about the family household, including zip code-level socioeconomic status represented using the area deprivation index (ADI)
- History of involvement with DCS, including prior investigations and removals for perpetrators and victims involved in the report
- Information about the source of the report, including whether the caller was a mandated reporter
- Descriptions of the allegations, coded 54 binary flags corresponding to the 54 allegation descriptions that are tracked in the Guardian system

We list the case characteristics used in the models in detail in Appendix B.

The case characteristics used in each model depend on the outcome and unit of analysis. Because screening in is defined for all hotline communications, and some hotline communications do not include qualifying allegations, there are not always clearly defined "perpetrators" and "victims." Thus, we include demographic information summarized across all individuals named during the hotline call, and separate the information by whether the individual is an adult or child. For example, we include the number of adults and children; the mean, maximum, and minimum age of adults and children; the proportion of adults and children who are male or female; and the proportion of adults and children who are classified according to each racial or ethnic category. A history for the screened-in outcome is similarly defined across all individuals named during the hotline communication. For the other four outcomes, which are analyzed at the report/perpetrator/victim level, the demographic characteristics correspond to the one perpetrator and victim associated with the observation.

We are careful not to include any information that is completely deterministic of the outcome in any model. For example, any time there is a single qualifying allegation, the call is screened in. Thus, we cannot include allegation variables in the screening model because the presence of any of these would always mean the call was screened in.

C. Modeling approach

There are many potential modeling approaches that could have been used for this study. When accurate prediction is the primary consideration, a common technique is to fit several types of predictive models and choose the one that best meets some criterion, such as out-of-sample predictive performance. However, because the goal of this analysis is not only to optimize predictions but to draw inferences from the data, we thought there would be minimal benefit to taking such an approach. Instead, to achieve the goal of accurate prediction, we implemented random forests, a method that is both well understood and popular for its ability to produce accurate predictions in a wide variety of contexts. We chose random forests for several reasons:

- Random forests are tree-based algorithms that are not constrained by the assumptions of standard linear regression models. This allows them to capture more complex interactions in the data (for example, allowing the effects of case characteristics to vary based on the type of allegation) which are common in child welfare settings.
- There are well-established methods for interpreting these models, such as variable importance and partial dependence, which we discuss in detail later in this section.

- Models are relatively easier to fit than more complex alternatives such as deep learning and neural networks. Because of the randomness properties of the forest, researchers can be confident that they have avoided overfitting as long as they have fit the forests with enough trees to ensure convergence.
- We have successfully implemented these models in similar child welfare analyses in the past, as have many other researchers (Chouldechova et al. 2018, Weisenberg et al. 2018).

The primary limitation of random forests is that they can be harder to interpret than more traditional statistical models like linear or logistic regression. However, we can still obtain interpretable findings from these models using techniques such as variable importance and partial dependence. We also implement decision tree models to further explore the relationships between our predictors and low rates of outcomes. We describe these calculations and detail our implementation in Appendix B.

IV. Study Findings

We begin this discussion of our findings with descriptive results to reveal the most prevalent allegations and how the distribution of allegations differs among the subgroups we examined. This is followed by an examination of how these allegations and characteristics from our models are associated with specific child welfare outcomes.

A. Descriptive results of allegation types

The Guardian system tracks 53 different types of allegations. Table IV.1 lists the 10 most common allegation types, and the proportion of all report/perpetrator/victim observations that include the allegation. No other allegation type is present in more than 5 percent of report/perpetrator/victim observations. These proportions are listed both overall and for eight key demographic subgroups: six racial or ethnic groups[5] (Table IV.1), and two geographic subgroups of interest (South Tucson[6] and rural Arizona[7]). It is important to acknowledge that the classification by race is based on data entered by DCS staff, and while DCS policy is that staff should document race and Hispanic ethnicity based on client self-identification, this may not always happen. (Some CAC members expressly cautioned their belief that some staff guess the race of clients.) Also, it is important to note that the Native American group does not include reports on Native Americans who live on their reservations since Arizona policy dictates that reports for Native American children who are living on a tribal reservation shall be routed to the corresponding Native American tribe. These tables contain shortened version of each allegation description; the full text of the 53 allegations can be found in Table B.3. Note that the columns are not expected to sum to 100 percent because observations can include more than one type of allegation.

Table IV.1 Proportion of observations (report/perpetrator/victim) that contain each of the 10 most common allegations, overall and by race/ethnicity and geographic region

			Racial or ethnic subgroup						Geography	
Type	Allegation	Overall	White (37%)	Hispanic (28%)	African American (11%)	Native American (6%)	Asian/ PI (1%)	Unknown (17%)	South Tucson	Rural Arizona
Neglect	Exposure to domestic violence	21%	18%	24%	23%	27%	23%	21%	26%	16%
Neglect	Unsafe living environment	17%	22%	16%	15%	17%	10%	13%	19%	23%
Neglect	Caregiver unable to meet child's needs	11%	12%	11%	13%	11%	9%	10%	11%	13%

[5] For screening in, observations are classified into a racial/ethnicity subgroup using the race/ethnicity of the primary caretaker. For the Observations are classified into racial/ethnic subgroups based on the race/ethnicity of the primary caretaker. When the race/ethnicity of the primary caretaker is missing and the race/ethnicity of the victim is known, the race/ethnicity of the victim is used to classify the observation. When the race/ethnicity of the primary caretaker and of the victim is missing, the race/ethnicity of the perpetrator is used. When race/ethnicity of all three individuals is missing, the observation is classified as missing.

[6] South Tucson is defined as any observation that has a zip code in 85705 or 85706

[7] Rural Arizona is defined using Rural-Urban Commuting Area (RUCA) Codes data (Cromartie, 2023). We classify any zip codes not in a metropolitan area as a rural.

Type	Allegation	Overall	Racial or ethnic subgroup						Geography	
			White (37%)	Hispanic (28%)	African American (11%)	Native American (6%)	Asian/ PI (1%)	Unknown (17%)	South Tucson	Rural Arizona
Physical Abuse	Caregiver is imposing/threatening	8%	8%	7%	9%	5%	10%	9%	6%	8%
Neglect	Caregiver unable to perform parental responsibilities	7%	8%	7%	6%	9%	8%	7%	8%	8%
Neglect	Close proximity to domestic violence	7%	5%	8%	8%	8%	6%	7%	9%	4%
Physical Abuse	Bodily injuries (bruises, etc.)	7%	7%	6%	7%	5%	7%	8%	5%	6%
Neglect	Substance exposed newborn/infant	6%	5%	5%	6%	7%	4%	8%	6%	6%
Physical Abuse	Injured during domestic violence	6%	4%	7%	7%	7%	5%	6%	8%	4%
Neglect	Unrealistic/negative view of child	6%	6%	5%	5%	4%	6%	6%	4%	6%

Note: The proportion of observations classified under each racial or ethnic group appears in the column heading.

Although the exact proportions differ, there were relatively consistent patterns across demographic groups in terms of which allegation types are most common. For each group, the two most common allegations were "neglect: exposure to domestic violence" and "neglect: unsafe living environment"; the former is the most common allegation for all subgroups other than White individuals and persons living in rural Arizona. Other common allegations were "neglect: caregiver unwilling/able to meet child needs" (the third most common allegation in seven of the eight subgroups), "physical abuse: caregiver is imposing/threatening," and "neglect: caregiver unable to perform parental responsibilities."

B. Important predictors of outcomes

After fitting each of the random forest models, we computed variable importance for each case characteristic in our model, based on the variable permutation importance method. Permutation importance is based on measuring the decrease in predictive ability of the model when the value of the variable is randomly permuted. The more important the variable to the model, the larger the decrease in the resulting predictive power when it is permuted. To be clear, "importance" of a variable measured here is purely its importance to the predictive power of the model; it does not describe actual factors that DCS staff use to make decisions.

Table IV.2 lists the most important case characteristics associated with each of the five outcomes, as determined by permutation importance. Variables are listed in alphabetical order; visualizations of the top 30 characteristics identified with this method (and therefore most associated with each outcome) can be found in Appendix B.

Table IV.2 Ten most important case characteristics associated with each outcome

Screening in	Variables	Substantiation	Case opening	Removal	Composite measure
Adult age (mean)	ADI weighted average for ZIP code	✓	✓	✓	✓
Adult age (min)	Caretaker's race				✓
Adult age missing	County	✓			
Child age (max)	Criminal Conduct	✓			
Child age (mean)	Mandates	✓			
Child age (min)	Neglect: Caregiver is absent			✓	
Prior removals (vic)	Number of allegations on intake	✓	✓	✓	
Prior victim removal duration	Number of prior perpetrator open case	✓			
Rural Arizona flag	Number of victims on intake		✓	✓	✓
South Tucson flag	Perpetrator age	✓	✓	✓	✓
	Prior investigations (perp)		✓		✓
	Prior open cases (perp)		✓	✓	✓
	Prior open cases (victim)		✓	✓	✓
	Source type	✓	✓	✓	✓
	Victim age	✓	✓	✓	✓
	Victim's race	✓	✓	✓	✓

For screening in, six of the top ten characteristics are related to the ages of the children and adults involved in the hotline communication. Notably, one of the characteristics that predicts screening in is whether the family address was in South Tucson. The other important characteristics associated with screening in are Rural AZ and variables related to prior victim removals (the number and duration). The other four outcomes (substantiation, case opening, removal, and composite measure) are all defined at the same level (report/perpetrator/victim) and use the same case characteristics in the model. We see similarities in the important case characteristics across these outcomes. In fact, five case characteristics consistently appear among the top ten variables associated with all of these outcomes: victim's race, victim's age, source type, ADI weighted average for the ZIP code, and perpetrator's age. The ADI weighted average for ZIP code variable is a measure of socio-economic status. We explain and examine the relevance of this variable, and we explore the associations of these predictors with outcomes in more detail later in the report.

We intentionally do not address the specific ordering of characteristics within the top 10 characteristics here due to the presence of highly correlated variables necessitated by the current design (consider screening in and the inclusion of the minimum, mean, and maximum child age variables when there is only one child on the record). Highly correlated variables can easily "stand in" for one another at splits within the decision trees forming the random forest, causing the importance of the concept behind those variables (for example, child age) to be split across the correlated variables. For this reason, we do not put considerable weight on the specific ordering of the variables within the top ten characteristics, while at the same time feel confident that variables there are important. In future work, we plan to use a technique known as "predictive scales" to get a single measure of the importance of correlated variables such as

those for child or adult age, as suggested by a member of the DAAC. However, we were not able to do so for this report due to time constraints.

C. Assessment of demographic differences

To assess the potential for racial or ethnic bias in outcome-related decision making, we compared the rates of outcomes across racial or ethnic groups by calculating a relative risk (RR) for each racial/ethnic group, using White race as a reference group.[8]

For each outcome, we calculated relative risk in two ways. The *unadjusted relative risk* for a given racial or ethnic group is the rate of the outcome in the corresponding group, divided by the rate of the outcome in the White group. However, this comparison does not represent a direct assessment of racial or ethnic differences, because cases may differ across racial or ethnic groups in ways other than race or ethnicity. For example, as we saw in the descriptive results, rates of each type of allegation differ across racial or ethnic groups, as could other important case characteristics such as history, the source types, and socioeconomic status (measured at the zip code level using the ADI). Therefore, we calculated *adjusted relative risks* as marginal effects (also known as partial dependence). The adjusted relative risks hold constant all model characteristics other than race/ethnicity, so we can understand the direct impact of race or ethnicity across an identical case mix. Details for these calculations can be found in Appendix B.

Because race and socioeconomic status are highly correlated, we wanted to know if any differences across racial or ethnic groups could be attributed to differences in socioeconomic status. Therefore, we calculated two versions of the adjusted relative risks: one based on a model that does not include the socioeconomic status variable, and one based on a model that does include socioeconomic status. It is important to note that because DCS does not track any measures of socioeconomic status for the family under investigation, our measure of socioeconomic status (the ADI) was defined at the zip code level. This means we were not directly controlling for the socioeconomic status of the family. Instead, we controlled for the average socioeconomic status in the family's zip code. Still, if the adjusted relative risks of the two models with and without ADI included differ substantially, it is likely that the combination of a family's SES and their race/ethnicity impact their predicted outcomes. In this scenario, comparing families with a similar case mix and a similar SES but different race/ethnicities would lead to different outcomes because the impact SES has on predicted outcome depends on a family's race/ethnicity. In other words, White families with a low SES would have different outcomes than families with a low SES in a different race/ethnicity category. If the two adjusted relative risks are similar, there are two possible explanations: 1) SES has the same impact on outcomes for all families, regardless of their race/ethnicity, or 2) other variables or their interactions are functioning as a proxy for SES. In the first scenario, removing SES from the model would not impact the relative risk because SES has the same effect across all race/ethnicities. In

[8] We made the decision to use the White racial group as the reference group because the White racial group is the largest racial/ethnic group, so using it as the reference offers the most stably estimated reference group. More substantively, because our goal was in large part to examine racial bias towards historically marginalized racial/ethnic groups, it is easier to describe findings by comparing outcomes to the White group. That said, we are sensitive to the fact that this approach may not be preferred, and we are open to exploring different ways to present findings in future analysis.

the second scenario, removing SES from the model would not impact the relative risk because the other variable(s) forming a proxy can act in its place.

Table IV.3 shows the unadjusted and both versions of the adjusted relative risks for each outcome for seven demographic subgroups. We express the relative risks as the percent above or below the reference group, which is White for each of the racial or ethnic groups, "not South Tucson" for South Tucson, and "not rural Arizona" for rural Arizona. For example, an unadjusted RR of -4 percent for screening among Hispanic families indicates that hotline communications involving Hispanic families are screened in 4 percent less often than hotline communications involving White families (*Unadjusted*). These calls are screened in 3 percent less often than similar calls involving White families that may live in a different socioeconomic area (*Adjusted without SES*), and 2% less often than similar calls involving White families of that live in an identical socioeconomic area (*Adjusted with SES*).

Table IV.3 Unadjusted and adjusted relative risks by demographic subgroup, expressed as a percentage above or below the reference group

		Racial or ethnic Subgroups					Geographic subgroups	
Outcome	**RR**	Hispanic	African American	Native American	Asian/PI	Unknown	South Tucson	Rural Arizona
Screening	Unadjusted	-4%	+6%	+0%	+1%	-22%	-5%	+9%
	Adj (no SES)	-3%	+4%	+0%	+1%	+4%	-2%	+2%
	Adj with SES	-2%	+4%	-0%	+1%	+4%	-2%	+2%
Substantiation	Unadjusted	+14%	+12%	+37%	+11%	-33%	+55%	+23%
	Adj (no SES)	+3%	+7%	+9%	+10%	-23%	+11%	+11%
	Adj with SES	+2%	+7%	+8%	+10%	-22%	+6%	+8%
Case Opening	Unadjusted	+8%	+12%	+17%	+5%	-53%	+18%	-8%
	Adj (no SES)	+5%	+6%	+4%	+20%	-37%	+8%	+1%
	Adj with SES	+5%	+9%	+4%	+22%	-37%	+4%	+0%
Removal	Unadjusted	+4%	+13%	+29%	-16%	-66%	+67%	+11%
	Adj (no SES)	+5%	+12%	+11%	+17%	-50%	+23%	+7%
	Adj with SES	+3%	+11%	+09%	+15%	-51%	+11%	+2%
Composite measure	Unadjusted	+3%	+6%	+10%	+0%	-22%	+15%	+1%
	Adj (no SES)	+2%	+4%	+4%	+7%	-13%	+5%	+2%
	Adj with SES	+2%	+4%	+4%	+7%	-13%	+3%	+1%

NOTES: Reference group for racial or ethnic subgroups is "White", reference group for South Tucson is "not South Tucson," and reference group for rural Arizona is "not rural Arizona." RR = relative risk; Adj no SES = Adjusted for all case characteristics except socioeconomic status; PI = Pacific Islander.

The rates of calls being screened in are relatively constant across racial or ethnic groups, as evidenced by adjusted relative risks close to 0 percent relative to the reference groups. Calls involving Hispanic families are screened in 2 percent less often than similar calls involving White families, whereas those involving African American families are screened in 4 percent more often than similar calls involving White families. However, these magnitudes are relatively small, likely reflecting the fact that Arizona hotline workers use a relatively objective process to determine whether a call should be screened in, in that any call in which

there is a claim that meets the statutory definition of abuse or neglect is automatically screened in. We do observe a large unadjusted difference in the rate of screening in for calls in which race is unknown, but this observation does not reflect racial or ethnic bias – rather, it is a process issue, resulting from the fact that there is less detailed follow-up on reports that are not screened in, so they are more likely to be missing race or ethnicity information. The only time there is known race or ethnicity data for participants with more history is when a report is screened in, or where that data was already part of the child welfare system. This means that families that are reported and have no history are less likely to have race information recorded and less likely to be associated with subsequent child welfare outcomes (because history is so strongly associated with subsequent outcomes). This may explain, at least in part, why race information is more likely to be missing on cases that are not substantiated or opened for services.

For substantiation, case opening, removal, and composite measure, we observed large disparities across racial or ethnic groups. In fact, rates on all four outcomes tend to be higher among Hispanic, African American, Native American, and Asian/Pacific Islander families than among similar observations involving White families. The most extreme relative risks appear to be among Asian/Pacific Islander families. These observations are substantiated 10 percent more often, result in case opening 22 percent more often, and in removal 15 percent more often than we would expect for the same circumstances had they occurred among White families. However, less than 1 percent of observations occur among Asian/Pacific Islander families. This small sample size means that our sample statistics are more likely to be impacted by extreme values in the data and may not accurately represent population outcomes. This could in turn cause the extreme results seen for Asian/Pacific Islander families. We also see that the adjusted relative risks with SES and without SES are similar for these outcomes, implying that the impact a family's SES has on their outcome does not change depending on their race/ethnicity.

It is important to note that positive relative risks do not necessarily mean there is racial or ethnic bias in decision making. This is because our models only adjusted for information that was available at the time of intake. If there is information that is observed after the time of intake (such as during an investigation) that is known by the individuals making the outcome determination, it may justifiably impact outcomes For example, it is well known that poverty is associated with race/ethnicity (Semega et al. 2020), and the stresses associated with poverty are likely related to child welfare outcomes (Monahan et al. 2023, Weiner et al. 2021). We attempted to adjust for socioeconomic status at the zip code level using the ADI, but we did not have information regarding the specific family circumstances at the time of intake. Thus, adjusted differences across race groups could be the result of o differences in the family's socioeconomic status that we cannot account for, and not the result of biased decision making. Unfortunately, we have no way of disentangling these effects. Still, at the least, the disparity revealed by these results suggest we cannot discount the possibility that outcomes may be driven in part by biased decision making. Many of the CAC members shared experiences and insights about people in their communities experiencing biased treatment and suggested that study results be interpreted with an understanding about this bias. They also offered salient recommendations to combat bias such as suggesting that DCS be intentional about hiring diverse staff and training staff on cultural considerations. (See Appendix A.)

In addition to assessing racial or ethnic differences, we compared rates of outcomes for two geographic subgroups of interest: those in the South Tucson area (zip codes 85705 and 85706), and those in rural areas of Arizona. DCS is aware that these areas have experienced particularly large rates of DCS

involvement in recent years. We calculated relative risks for each of these groups compared to the rest of the state. In other words, the relative risk for South Tucson compares outcomes in South Tucson to those outside South Tucson, and likewise for rural Arizona. The rates of all outcomes except for screening in are higher in South Tucson than other parts of the state, even after controlling for other case characteristics. Observations are substantiated 6 percent more often, result in case opening 4 percent more often, and in removal 11 percent more often than would be expected for similar observations had they occurred outside South Tucson. CAC members from South Tucson discussed the high rates of poverty in this area as a contributing factor towards child welfare outcomes. We also see differences in the two adjusted rates for substantiation and removal in South Tucson, likely implying that a family's SES has different impacts on the outcome depending on if the family lives in South Tucson or not. In rural Arizona, observations are substantiated 8 percent more often than similar observations that occur in non-rural settings, but we observe very small differences in outcome rates for case opening, removal, and composite measure. Among the possible explanations offered by CAC members from rural Arizona are the high rates of poverty, mental illness, and substance use that contribute to more extreme cases, the lack of available, proximate, and easy-to-access resources in rural Arizona, and the lack of DCS staff capacity. Appendix A has notes detailing the input from CAC members.

D. Association between allegations and outcomes

To gain a better understanding of which allegations are most associated with outcomes, we compared rates of outcomes for all observations that included a particular allegation. As we did with the assessment of demographic differences, we express this comparison of rates as an adjusted relative risk, calculated as a marginal effect (partial dependence). Details for this calculation can be found in Appendix B.

Table IV.4 lists the relative risks for each outcome by allegation. Because the relative risks for infrequent observations are less likely to be meaningful due to the small amount of available information, we only show those for allegations with at least 1,000 observations. We sorted the table by the composite outcome measure, so allegations with low (negative) relative risks on the composite measure appear towards the top of the table, whereas allegations with high (positive) relative risks appear near the bottom. As before, we express each relative risk as the percentage above or below the reference group, which is the absence of the corresponding allegation. For example, we find that observations that contain the allegation "physical abuse: caregiver is imposing/threatening" are 14 percent less likely to be substantiated than otherwise similar observations that do not contain this allegation.

Table IV.4 Relative risks of each outcome by allegation, for allegations with at least 1,000 observations

Type	Allegation	Frequency	Substantiation	Case opening	Removal	Composite measure
Physical Abuse	Caregiver is imposing/threatening	8%	-8%	0%	-5%	-1%
Neglect	Unrealistic/negative view of child	6%	-6%	-3%	-1%	-2%
Neglect	Known sexual predator accesses child	4%	-13%	0%	4%	-1%
Neglect	Unwilling to meet child's medical needs	3%	-3%	2%	3%	-1%

Type	Allegation	Frequency	Substantiation	Case opening	Removal	Composite measure
Physical Abuse	Bodily injuries (bruises, etc.)	7%	7%	1%	1%	0%
Neglect	Exposure to domestic violence	21%	11%	5%	-4%	1%
Neglect	Close proximity to domestic violence	7%	13%	0%	1%	1%
Neglect	Injuries due to failure to supervise	2%	12%	0%	5%	1%
Emotional Abuse	Berating, name calling, targeting, and/or rejection	1%	-5%	8%	13%	1%
Neglect	Caregiver provides harmful substances	1%	11%	3%	8%	1%
Physical Abuse	Injured during domestic violence	6%	16%	4%	4%	3%
Physical Abuse	Caregiver provides harmful substances	1%	11%	5%	9%	2%
Neglect	Caregiver provides drugs/alcohol	1%	15%	3%	9%	2%
Physical Abuse	Placed in dangerous situation	3%	13%	6%	11%	4%
Neglect	Unsafe living environment	17%	10%	7%	15%	3%
Sexual Abuse	Evidence of sexual abuse	2%	23%	4%	7%	3%
Neglect	Substance exposed newborn	6%	36%	1%	-1%	5%
Neglect	Unable to control child's dangerous behavior	3%	0%	17%	20%	6%
Neglect	Caregiver unable to meet child's needs	11%	13%	9%	22%	5%
Physical Abuse	Face or head injury	2%	18%	12%	18%	5%
Neglect	Fearful due to threat/circumstances	1%	21%	16%	26%	5%
Physical Abuse	Multiple plane injuries	1%	22%	24%	31%	10%
Neglect	Unable to perform parental responsibilities	7%	64%	34%	59%	20%
Neglect	Caregiver is absent	5%	68%	57%	130%	37%

NOTE: The 10 most common allegations (those that are present in more than 5 percent of observations) are in bold text.

The allegations toward the top of this list are those that, on their own, are less likely to result in further DCS action in the form of the allegation being substantiated, a case being opened, or the victim being removed from the home. These include a mix of abuse and neglect allegations. One CAC postulated that some of these allegations, such as "physical abuse: caregiver is imposing/threatening," "neglect: unrealistic/negative view of child," and "emotional abuse: berating, name calling, targeting, and/or rejection" may be difficult to prove, whereas others, such as a substance-exposed newborn, are not. Another member explained that DCS cannot open a case for in-home services unless there are specific, relevant services available. Thus, case opening rates could be impacted by the nature of the concern relative to the services available, particularly in areas where those services are lacking.

Toward the bottom of Table IV.4 are allegations associated with an increased likelihood of DCS action. Of

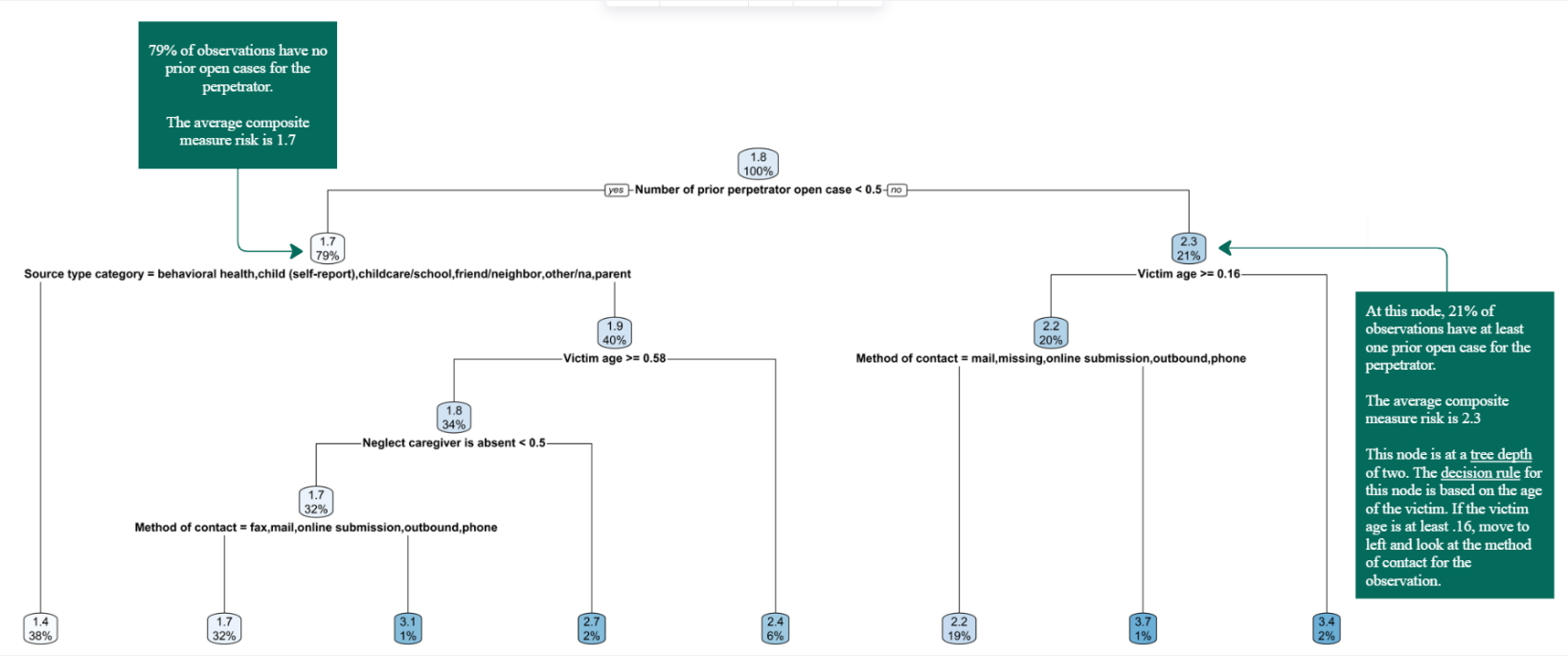
note, an observation with an allegation of "neglect: substance exposed newborn" is 36% more likely to be substantiated than similar observations that do not have this allegation. This is likely related to the state law that requires reporting of substance-exposed newborns and the policy that dictates substantiation when the allegation is confirmed by a health care professional's statement, typically based on a toxicology result. Other allegations that are highly associated with a DCS response include "Neglect: caregiver is absent" (especially associated with removal), "Neglect: unable to perform parental responsibilities," and "Physical abuse: multiple plane injuries." CAC members recognized that these allegations are most likely to be associated with circumstances with clear evidence and compelling reason for DCS intervention.

E. Identifying cases with low rates of outcomes

To dig deeper into the types of cases that have low rates of outcomes, we fit a decision tree model to the substantiation, case opening, removal, and composite outcomes. To avoid our results being impacted by potentially biased decision-making (which we noted as a possibility based on our results in Section III.C), we did not use the raw outcomes that were used to fit the random forest models. Instead, we used race-blind predictions of each outcome, which are the expected outcomes for each observation, averaged over potential distributions of race and ethnicity that could have occurred. Our procedure, described in detail in Appendix B, guarantees that two observations with identical case characteristics other than race will always have the same race-blind prediction. Our decision tree models use the same case characteristics that were used in the random forest models, other than excluding the race/ethnicity variables which will not impact the race-blind outcomes.

The full decision trees that we fit are too large to show in this report. We include a *pruned* version of these trees (that is, the top of each tree) in Figure IV.1 (composite measure) and Appendix B (substantiation, case opening, and removal), to give the reader a sense of what they contain. We next briefly describe how to interpret these trees, using Figure IV.1 as an example. At each node (blue oval) in the tree, there are two numbers: (1) the average outcome at that node, and (2) the proportion of observations that reach that node. The top node includes 100 percent of observations, which have an average composite measure of 1.8. All nodes other than the bottom nodes also include a decision rule. A decision rule is a rule created by the model that will use a single predictor to split the data from the previous node into two groups. The decision rule at each node is optimized so that it best splits the remaining data into groups with similar outcomes. For the composite score decision tree, the first decision rule checks whether the number of prior open cases against the perpetrator is less than 0.5 (in other words, if it is zero). If so, the observation continues to the left; if not, to the right. We see that 79 percent of observations have no prior open cases against the perpetrator (average composite score of 1.7), whereas 21 percent have had prior open cases against the perpetrator (average composite score 2.3). For those with no prior open cases against the perpetrator, the next variable to check is whether the source type category is one of behavioral health, child (self-report), childcare/school, friend/neighbor, other/na., or parent. This list of sources consists of three of the four non-mandated reporter groups (it does not include relative/guardian), plus behavioral health and childcare/school. If an observation has a source type from one of the listed sources, we move the next node to the left, which tells us that the observations with these source types have an average composite risk of 1.4. We can see that this is a lower risk than observations that do not have source types from that list, as the next node to the right has an average composite risk of 1.9. Interpretation continues similarly throughout the rest of the tree.

Figure IV.1. Pruned version of the decision tree for the composite measure score, using race-blind predictions as the outcome

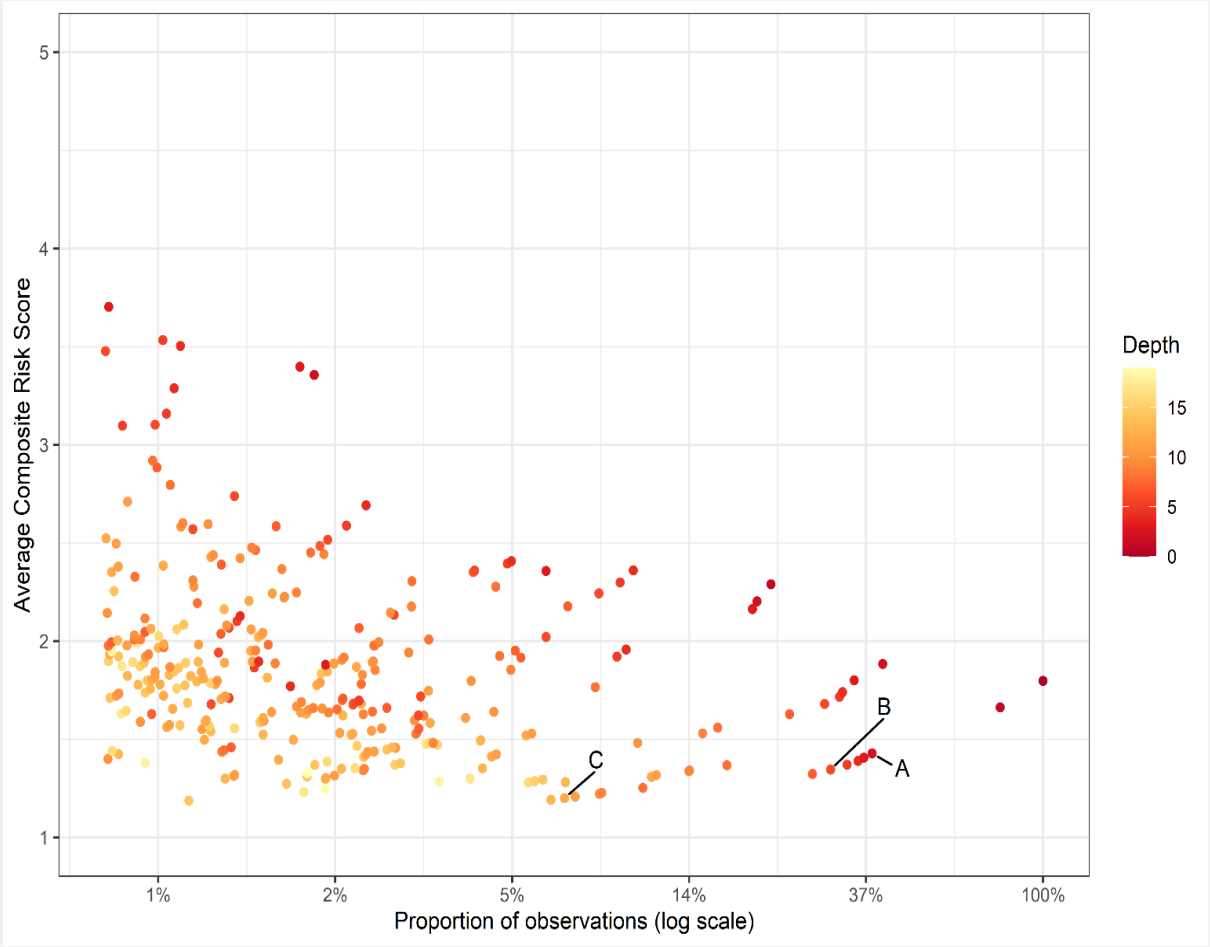


These trees are a rich source of information that can be used to identify clusters of observations with low outcomes. To understand the information contained in the trees, we first note that each node in the tree represents a cluster of observations that we can identify. Then, we plot three characteristics of each of these clusters:

1. The proportion of observations that reach the node
2. The average outcome at the node
3. The *depth* of the node, that is, how far down into the tree the node is

We use the next plot (Figure IV.2) to look for clusters of observations that have low risk of future harm to a child (low average composite measure score). The most useful clusters of this type will contain a relatively large proportion of the observations and require fewer decisions to identify (lower depth in the tree).

Figure IV.2. Node average outcome versus node proportion of observations, for each node in the composite score outcome decision tree



In Figure IV.2, we can see a group of several low-depth nodes clustered near points A and B in the plot. These points represent case mixes that are common (high proportion of observations), usually do not lead to a high risk outcome (low average composite risk), and are potentially easy to identify when examining a case (low tree depth and therefore less decisions rules to follow). For example, point A has an average composite risk of 1.43 and contains 38 percent of all observations in our data. Point A has a tree depth of only 2 and contains the following criteria:

- The number of prior open cases for a perpetrator is zero
- The source type is one of behavioral health, child (self-report), childcare/school, friend/neighbor, other/na, or parent

Putting it more plainly, the results tell us that when a perpetrator and victim in a report meet the criteria above, children are not typically removed in the course of an investigation or within 3 months, no in-home case is opened, and allegations are not typically substantiated. For observations that do meet this criteria, 6 percent lead to a substantiation, 11 percent lead to an open case, and 5 percent lead to removal. These outcome rates are much lower than the rates among all allegations, which are 17 percent for substantiation, 20 percent for case opening, and 11 percent for removal.

Point B on the plot reveals similar information as Point A. Though Point B contains serval more criteria than point A because it is deeper in the tree, but it still contains 30% of all observations. The additional criteria are:

- There are no allegations of "Neglect: caregiver is absent" or "Neglect: caregiver unable to perform parental responsibilities"
- There are no prior open cases that included the victim
- Victim older than 1.2 years old.

The average composite measure risk for this point is of 1.3. This point similarly represents a case mix of cases that is relatively low risk but occurs often. We can identify clusters with even lower risk by considering deeper nodes that correspond to a smaller proportion of observations. For example, Point C contains case mixes with an average risk score of 1.21. In contains all the criteria above, as well as the following additional criteria:

- No allegation of criminal conduct
- Source type is behavioral health or childcare/school
- The observation contains 5 or fewer allegations
- The victim was not involved in any prior investigations involving substance abuse
- Information on language of victim is not missing

Approximately 7 percent of all observations meet each of the above criteria. Among the observations in Point C, 3 percent have a substantiated allegation, 6 percent result in a case being opened, and 1 percent result in removal within three months of intake.

V. Discussion and next steps

This study offers important insights for policy makers, community members and other parties in Arizona who are interested in finding ways to reduce unnecessary CPS involvement with Arizona families. We hope that promising ideas will emerge from considering how state policies and community input relate to the key findings from this study, and where there may be opportunities to support families in a more helpful way. For example, given state policies surrounding reporting and substantiation of substance-exposed newborns, it is no surprise how strongly this allegation predicts child protective service intervention. However, community members have suggested that the state law is overly harsh by not considering the recency or severity of drug use, and that drug screening may happen more typically for pregnant women with a low socio-economic status. One recommendation is to adjust policy so that the severity and recency of drug use are taken into account in the reporting laws, and to consider when it may be more helpful to offer support for families that would not require child protective service involvement.

While the findings from this study (including the extremely valuable community input) are highly relevant to this examination, this report is likely to provoke more questions and ongoing discussion. This is expected given the complex nature of this subject and analysis, and because of the limitations and constraints of this study – time and resource limitations, the limited study and observation periods, and methodological factors[9]. Recognizing that, further analysis and engagement with invested members of the state is planned to help with interpretation of results, and to guide additional analysis and other next steps. We present ideas for follow-up analysis, and other recommendations for consideration, with the expectation that the conversation around this subject will help to prioritize what are the best next steps.

Below is a working list of ideas for additional analyses to inform policy making, including suggestions made by the DAAC. We expect other ideas will be added, and we will pursue the priority and feasible options:

- Additional analysis by County to assess variation in decision-making across state jurisdictions
- Examination into the type(s) of drug use and recency of use for substance-exposed newborns.
- Examination of the relationship between the alleged perpetrator and alleged victim child and how that relates to risk of future child welfare outcomes
- Further analysis to understand how socio-economic status impacts outcomes.
- Apply new statistical techniques to improve strength of findings:
  - Use a technique known as "predictive scales" to get a single measure of the importance of age (discussed in the Study Findings section.)
  - Use a technique known as "input shuffling" to help examine how one variable impacts others.

[9] For example, because it is not possible to know the ground truth need for protective service involvement, we examined system outcomes as proxies- substantiation, case opening and removal from home.

- Test the model, which is trained to predict the likelihood of future child protective service involvement, against more objective measures of child harm by externally validating the model if a viable data source can be obtained. (Similar models have validated similar models using medical records and critical events data.) (Vaithianathan et al. 2019)

CAC members offered rich input including recommendations to contextualize interpretation and suggestions to mitigate the need for child protective services. One significant theme particularly worth highlighting is the strong association of child protective services involvement with families in need due to living in poverty, in places with limited resources (such as South Tucson), and with limited information about the DCS system and supports available for them. CAC members emphasized the need to address these concerns. We indicated other suggestions in the main body of report but urge examination of the notes in Appendix A to capture all of them. It may also be worth looking at other research in this area to help inform how specific investments may be tied to reducing child protective service involvement. A relevant example is a compendium of work produced by Chapin Hall that describes the evidence-base for providing economic and concrete supports as a means for child welfare prevention. (Anderson et al. 2021)

References

Anderson, C., Grewal-Kök, Y., Cusick, G., Weiner, D., & Thomas, K. (2021). *Family and child well-being system: Economic and concrete supports as a core component*. [Power Point slides]. Chapin Hall at the University of Chicago. [Updated March 2023]

Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) Classification and Regression Trees. Wadsworth

Chouldechova, A., Putnam-Hornstein, E., 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, 134–148. https://www.cs.ubc.ca/~conati/522/532b2019/papers/chouldechovaCaseStudyPredictionFairnessEtc.pdf

Cromartie, J. (2023, March 22). Rural-urban commuting area codes. USDA ERS. https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx

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.

Drake, B., Jonson-Reid, M., Ocampo, M.G., Morrison, M., & Dvalishvili, 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.

Goldhaber-Fiebert, J.D., & Prince, L. (2019). Impact evaluation of a predictive risk modeling tool for Allegheny County's Child Welfare Office [Evaluation Report]. https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-6.pdf

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Marvin N. Wright, Andreas Ziegler (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, 77(1), 1-17. doi:10.18637/jss.v077.i01.

Monahan, E. K., Grewal-Kok, Y., Cusick, G., & Anderson, C. (2023). *Economic and concrete supports: An evidence-based service for child welfare prevention*. Chapin Hall at the University of Chicago.

Pope, Devin G., and Justin R. Sydnor. (2011). "Implementing anti-discrimination policies in statistical profiling models." American Economic Journal: Economic Policy 3.3: 206-31.

Samant, A., Horowitz, A., Xu, K., & Beiers, S. (2021). Family surveillance by algorithm. American Civil Liberties Union. https://www.aclu.org/fact-sheet/family-surveillance-algorithm

Semega, J., Kollar, M., Shrider, E. A., & Creamer, J. F. (2020). Income and Poverty in the United States: 2019. September 2020.

Statology. (2021, September 9). What is Considered a Good AUC Score? https://www.statology.org/what-is-a-good-auc-score/

Twala, Bheki ETH, M. C. Jones, and David J. Hand. (2008). "Good methods for coping with missing data in decision trees." Pattern Recognition Letters 29.7: 950-956.

Vaithianathan, R., Kulick, E., Putnam-Hornstein, E., & Benavides-Prado, D. (2019). Allegheny family screening tool: Methodology, version 2. Center for Social Data Analytics, 1-22.

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):e202770.doi:10.1001/jamapediatrics.2020.2770

Weigensberg, E., Cornwell, D., Leininger, L., Stagner, M., LeBarron, S., Gellar, J., ... & Pecora, P. J. (2018). Superutilization of Child Welfare, Medicaid, and Other Services (No. caaff77fa722452aa241ace4b218e353). *Mathematica Policy Research*.

Weiner, D. A., Anderson, C., & Thomas, K. (2021). *System transformation to support child and family well-being: The central role of economic and concrete supports.* Chicago, IL: Chapin Hall at the University of Chicago.

Wright, Marvin N., and Andreas Ziegler. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, vol. 77, no. 1, 2017, pp. 1–17. doi:10.18637/jss.v077.i01.\

Appendix A

Community Advisory Council Feedback

This page has been left blank for double-sided copying.

CAC Meeting Notes, Session One

African American Focus Group

April 23, 2022

Overview of focus group

A focus group of African Americans was held on Saturday, April 23, 2022, from 1:15–3:15 p.m., in person in Phoenix. There were a total of 12 focus group participants, including the community leader co-facilitator, Dr. Patricia Neff.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation about the purpose of and approach to the Next Event Analysis study was presented by Claire Louge of Prevent Child Abuse Arizona, with support from Dr. Patricia Neff, the group's community leader and co-facilitator.

Presentation Outline:

Purpose

- DCS receives about 84,000 reports of child neglect and abuse per year. Arizona law tells DCS what needs to be investigated.
- DCS has guidelines to determine which calls meet the definition of the law, and are taken as a report to receive an investigation.
- The purpose of the study is to answer the following questions: Can we determine which reports, based on information collected at intake, are least likely to need a DCS intervention, and might need something different (like a community service)? Is there a reasonable way we could change processes so families do not get unnecessarily investigated?

Approach

- We will examine data DCS is already collecting on reports and outcomes to learn about what information from reports is tied to later outcomes
- Our goal is to understand which report characteristics (or combinations of characteristics) are most and least associated with the need for DCS intervention
- We are <u>not</u> creating a "predictive risk model" that will be used by caseworkers or supervisors to profile people. We are <u>not</u> handing decision making over to a machine or algorithm.

Study Aims

- We hope to safely reduce the number of families who have DCS knock on their doors/investigate them
- We hope to identify prevention services designed to meet the needs of families who today receive a DCS investigation, but do not need DCS intervention
- We hope to use the information to expand funding for prevention services that help families stay safe and together

After the presentation, participants were asked to reflect on the information with the following questions:

- What do you believe is the role of DCS in your community?
- How does your community view DCS? Is DCS intervention helpful?
- When do you believe DCS involvement is needed? When do you believe it is not?
- As the team designs this study, what should we be thinking about?
- Do you have any thoughts about how this project could affect members of your community?
- Is there anything we haven't asked about that you would like to share?

Summary of questions, reflections and discussion from the African American focus group

Initial questions from focus group participants:

Is this study coming out of the situations regarding African American families being targeted unfairly, either being removed or investigated more often than White children? I saw articles that DCS is spending a lot of time on racial equity.

Is DCS keeping track of who is doing the reporting?—specific to the disproportionally (i.e. neighbors, schools, family) The ease at which someone can call to report even if it is just anger or a false allegation—there should be consequences for those that repeatedly call due to relationship contention and make false allegations.

Is DCS going to keep track of the rate of unsubstantiated vs. substantiated reports based on the person who is reporting the allegation? (i.e. professionals, community members, family members)—suggestion: some type of matrix that shows those trends

Will this study and the information collected eventually lead to modifications to AZ state law?

Why is this study happening right now? Participant noted that there is a lot of time and energy being spent currently about disproportionality—why is this issue becoming important right now? Who is leading the initiative on this project? Would like to know this information in order to know who to support in the future and going forward—who should they be backing at the state level that is making this happen?

As a participant in this process, how will we know what happens with the results of the study? Could DCS potentially implement in a community forum so that the community can report back based on the changes that are happening from this study?

At some point will there be conversation about the current way that parents and foster parents try to advocate for themselves when they are having challenges working with DCS? Ombudsman office is directly connected to DCS—feels it is a conflict of interest for escalating problems with the Department of Child Safety.

What do you believe is the role of DCS in your community?

DCS' role is to advocate for families who cannot advocate for themselves. They can help children who are not heard.

Their role is to keep kids with biological families. Sometimes that's not the best idea. Their stated priority is to get them safely reunified with their biological families.

Participant feels that DCS is here to disrupt families, over and over again, and feels the number of families that are affected by DCS in a negative way is high, but the number of children that actually need help is low.

Removals are more likely for children of color. **DCS is quicker to remove children of color from their families. Slower to remove for other children.** Feels Caucasian children and families tend to be given more resources in terms of prevention of entering the system.

Decisions to remove children from their families should be based on more than what the reports to DCS say. Other evidence needs to be gathered, and the family's history needs to be considered. What else is being done to make an educated decision to remove the child from an unsafe environment? Example: an unsubstantiated report where a child that is actually being abused, placed back in to the same environment without considering the history of that child—is DCS using supplemental information during their investigation?

Our biggest threat is the weaponizing of DCS.

There are callers who make "bad faith" reports. There need to be repercussions for callers who make a report on families simply because they don't like them. Reports endanger families' peace.

There should be a common-sense approach to investigation. It is obvious when a child is lying, or someone is reporting in bad faith.

DCS caseworkers don't have intercultural skills. Many caseworkers are not even parents and so do not have the experience to understand the reality of parenting.

DCS needs common sense. And cultural sensitivity.

DCS role is to investigate. Regardless of who makes the allegation, a lot of children who are investigated are on state insurance—why not have DCS use that connected system to review documents that could dispute or support an allegation of abuse?

Feels something that is not acknowledged by the Black community but that the community itself has weaponized DCS.

DCS needs a tool at the hotline level to determine whether or not to investigate. This should be a progressive matrix based on questions asked of the caller.

[When informed that DCS has a safety assessment tool at the hotline level]

Bias is built in to this tool. Who is assessing this tool?

There should be a box on the tool that says "That's just this kid talking" (the child is unhappy with something and wants to get their parent in trouble or otherwise lying).

It is so easy to report. Investigations are so disruptive.

DCS needs cultural competence training. They may interpret something said as a threat of violence when it is simply the way that certain families naturally speak to one another.

As a community, DCS involvement often happens and results in no substantiation. If a report is unsubstantiated there should be a formal closure process, as DCS history follows you. This is unfair especially in a community that is overrepresented in the child welfare system and with a DCS history you can be deemed ineligible to help other children and families in the future.

Feels as a community they turn to their own community support before making a report to DCS but on the flipside, for mandated reporters, there is a fear of not reporting and a tragedy resulting (Gabriel Fernandez mentioned).

How does your community view DCS? Is DCS intervention helpful?

DCS is traumatizing. They're supposed to be protecting, but it can make things worse. Removal is traumatic. What are DCS workers' education level? We have low experience workers who are not parents making life-changing decisions.

Feels there is a lack of common sense somewhere in the DCS process, whether it is the reporter or the investigator.

I do think DCS has the goal of helping families. DCS is overwhelmed. They should hire more people so they can each have fewer cases. Every level of decision making needs to be scrutinized– one individual should not be responsible for making these large life-changing decisions.

There is importance in lessening DCS caseload so that case managers are better able to manage the role, even if it means spending more money on the system—the community views DCS as underfunded, under supervised, and overburdened. This cannot create a strong or helpful system.

What is helpful is subjective. DCS is not helpful, but it can be useful. DCS is useful because they have data, and so they are useful in understanding what's happening.

There are inherent biases in the system. Workers may not understand what they're seeing.

Families need to code switch when interacting with DCS. Feels that children and families that cannot codeswitch are at a disadvantage with DCS involvement.

DCS participates in the community only as a police. If they want to be trusted, they should be in the community offering services and resources before intervention.

If you did not grow up in the African American community, it is understandable that you would not understand what you're seeing as an investigator or otherwise.

DCS should intervene when there are red flags. No one wants DCS in the neighborhood.

It's great we have a tool for assessing safety, but checking boxes takes humanness away from people. Check boxes can get in the way of common sense.

There are cases that should stop at a certain level, but don't. No stop gaps as cases advance. **How are decisions made at each level as cases advance? This should not be just about initial investigation—but decision making along the way.**

Children are able to report abuse, too—some children can exaggerate re: discipline. DCS intervention is not helpful then. Example: argument between mother and daughter—daughter went to neighbor, neighbor called authorities based just on the daughter's emotions after an argument.

When do you believe DCS involvement is needed? When do you believe it is not?

DCS should be involved when there is a clear and present danger, such as sexual abuse. DCS shouldn't be involved when they're just not sure.

Examples of DCS involvement not needed:

- A neighbor with 7 children, single father so the kids are a bit under-cared for, sometimes they're dirty, sometimes they're hungry, and from the outside looking in that could warrant a report, but when you know a family, know their children, you know that family is doing its best and as a neighbor. Participant would not call DCS, rather offer help.
- It's offensive to be investigated. I was investigated because a White teacher called after hearing us yelling in the background on Zoom calls. A White woman from DCS knocked on door. Investigation was stressful. DCS caseworker told me on the phone that my case was closed, but I got no letter/ email or anything from DCS to confirm that. I called a couple of times and still haven't received a written or otherwise professional response from DCS, which makes me feel uncomfortable and unsure of how true this information is. Perhaps, my family's case IS still under investigation? How would I prove otherwise in a court of law? How will we prove that our nonprofit is still credible and safe for family members? This has disrupted my entire life and business. The DCS office seems unaffected and doesn't care. Leaving us to build up again from almost a square one position, especially since the pandemic.

DCS should have a peer crisis interventionist. DCS needs a lower level that isn't an investigator.

As a Black counselor, I wouldn't interpret a family's behavior like a White counselor would.

Counter example—as a mandated reporter, you do not have the ability to get familiar with a family or situation prior to calling the report whether or not DCS intervention is needed.

Sometimes they get a report—like mom and dad not being able to afford day care—and that can be resourced.

[Participants described wanting what is essentially differential response, including having hotline workers decide if calls should be resourced or investigated, and sending out a professional that can resource families to the calls that need to be resourced, and an investigator for calls that need to be investigated.]

DCS needs to have the reputation to be a resource, not a tyrant.

As the team designs this study, what should we be thinking about?

They should be thinking about what families need. If DCS was looked at more as a resource vs. an investigative agency, then the community may welcome their involvement.

They need to consider biases and cultural sensitivity. Hopefully DCS can acknowledge bias to create a stronger safety assessment tool, but also consider those checkboxes can remove the humanness from interactions.

They should be asking more questions about what resources the families might need at the hotline level. DCS should be considering what the families that come to their attention really need involvement prior to opening a case. **DCS should be able to provide intervention that is not removal or an open case.**

Some people perceive DCS as the place to go for resources. Several years ago, DCS was encouraging that perception—that people should call DCS so that they can help families.

DCS should be a resource—should be perceived as helpful. If DCS cannot offer resources, they should be collaborating with the systems and programs that can offer referrals and resources.

DCS should not be able to claim that they focus solely on abuse and neglect as a child safety system.

DCS should not stop investigating but they should change the way they do it.

DCS could be like UMOM, or 211, a hub system. We need a navigator for families.

We know DCS is supposed to investigate, but then we get mad when they investigate. It is the manner that DCS investigate that's at fault.

DCS should be thinking about training. Look at mandated reporter training, and the pressure mandated reporters receive to report—is that making a difference in calls and data? Mandated reporter training needs to be very clear about what abuse/neglect looks like.

Likeability of parent can affect reports.

Look at the reporter and their background in data—who is making reports, and what boxes they check.

Consider interpretation of intake person and investigator, and their characteristics. It could be affecting data.

Break down data by zip code/region. They should be looking at data by zip codes—specifically the zip codes that have the highest numbers of calls, pre-intervention should be considered for those high rate zip codes.

There should be a family liaison in each school in high-removal zip codes whose job is to know families and partner with DCS when needed.

Consider that DCS workers (both hotline and investigators) don't "hit refresh" after each family. The previous case may be affecting their perception of the next one.

Does time of day, and number of calls that the hotline worker has taken so far, affect their decision making?

Consider race and the differences in findings.

Consider false positives and false negatives. How many reports are unsubstantiated and result with a negative event —example of the Scottsdale family where the child died, grandma had multiple reports but never led to re-removal.

Consider that there is a difference between need for investigation versus need for mediation.

We need to create a program in which people who are from the same culture as the family conduct the investigation.

Do you have any thoughts about how this project could affect members of your community?

The fact that they're doing the study is a positive thing—racial equity, human rights and liberties are important and it's good to know that those in power are starting to recognize that there are biases and lack of cultural awareness.

Hopefully this study can expose that there are some resources that are ineffective. We need to teach people how to fish, not just feed them. Asa member of the community, they have seen generational welfare dependence and the negative impacts of that.

I hope they take the information and that it exposes biases. I don't think that people are intentionally biased in the system.

With this data that is being evaluated, DCS should be moving forward by being transparent and sharing the information vs. seeking short term fixes.

DCS may be pouring funding into resources that are ineffective.

I think the data will suggest improvement that needs to be made for how protection is administered in the Black community.

Bias should be considered in every step of this study, i.e., how the information is being shared with DCS from these focus groups through the lens of the individuals facilitating the focus groups.

Hopefully this study will make DCS give the community more credit and recognize that they take care of their own and do not need DCS involvement to help with that.

Hire someone Black to write the narrative to this data study—we need this lens on the data.

There needs to be a cultural czar that oversees the interpretation of the results of the study.

Is there anything we haven't asked about that you would like to share?

Concerns shared about disconnect from actual feedback and the data that will be reviewed because it is being filtered from these focus groups to the people that are doing the study.

We want Mathematica and DCS at the next focus group so we can hear what the data has found directly from them.

Minority children and communities are built differently; abuse is subjective.

African Americans need to be trusted as a culture. We keep each other safe and accountable.

The person who reports might be interpreting a situation in a way that doesn't have the cultural context.

Cases should be extreme to call DCS. I wouldn't call on my neighbor.

Families should be kept together.

Think of bias. DCS should be okay to learn about culture and bias. It would be important to include training on cultural sensitivity, bias, and diversity, DCS should create a culture where it is okay to talk about biases. Learning should be safe vs. putting individuals on the defense. It's important for DCS to learn how to create that safe space for their staff.

Native American Focus Group

April 23, 2022

Overview of Focus Group

A focus group of African Americans was held on Saturday, April 23, 2022, from 4:00–6:00 p.m., virtually over Zoom. There was a total of eight focus group participants, including the community leader co-facilitator, Holly Figueroa.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation about the purpose of and approach to the Next Event Analysis study was presented by Claire Louge of Prevent Child Abuse Arizona, with support from Holly Figueroa, the group's community leader and co-facilitator.

Presentation Outline:

Purpose

- DCS receives about 84,000 reports of child neglect and abuse per year. Arizona law tells DCS what needs to be investigated.
- DCS has guidelines to determine which calls meet the definition of the law, and are taken as a report to receive an investigation.
- The purpose of the study is to answer the following questions: Can we determine which reports, based on information collected at intake, are least likely to need a DCS intervention, and might need something different (like a community service)? Is there a reasonable way we could change processes so families do not get unnecessarily investigated?

Approach

- We will examine data DCS is already collecting on reports and outcomes to learn about what information from reports is tied to later outcomes
- Our goal is to understand which report characteristics (or combinations of characteristics) are most and least associated with the need for DCS intervention
- We are not creating a "predictive risk model" that will be used by caseworkers or supervisors to profile people. We are not handing decision making over to a machine or algorithm.

Study Aims

- We hope to safely reduce the number of families who have DCS knock on their doors/investigate them
- We hope to identify prevention services designed to meet the needs of families who today receive a DCS investigation, but do not need DCS intervention
- We hope to use the information to expand funding for prevention services that help families stay safe and together

After the presentation, participants were asked to reflect on the information with the following questions:

- What do you believe is the role of DCS in your community?
- How does your community view DCS? Is DCS intervention helpful?
- When do you believe DCS involvement is needed? When do you believe it is not?
- As the team designs this study, what should we be thinking about?

- Do you have any thoughts about how this project could affect members of your community?
- Is there anything we haven't asked about that you would like to share?

Summary of questions, reflections and discussion from the Native American focus group

Initial questions from focus group participants:

Will there be an opportunity to share thoughts in written form outside of the Zoom call today? [Facilitator shared that additional thoughts can be emailed.]

What do you believe is the role of DCS in your community?

To protect children, but to not be punitive towards families. DCS can be punitive towards birth parents. How we collect urine drug screens without consent is a problem.

DCS' role is to protect children.

DCS could be doing a better job at protecting children, believes DCS should investigate every call.

DCS uses data collection without consent, particularly urine drug screenings for newborns.

The role of DCS is to investigate and look at reports. Caseworkers sometimes give up because they cannot get into contact with parents. There are complications resulting from crossing tribal lines.

If it's possible, DCS should research which parent is more stable, such as who has a job, who is taking children to doctor's appointment, and signatures on school documents. Look at background of parents to make decisions about which parent the child should be placed with.

When receiving a call, DCS should think about the stability of the individual calling. A lot can get overlooked.

How does your community view DCS? Is DCS intervention helpful?

In my small community, DCS is viewed very negatively—they are not coming to help, they are only coming to cause harm. We are not open to dealing with DCS workers. Parents who have interactions with DCS are shunned and viewed as bad parents.

DCS involvement can be helpful especially for the children and sometimes for the parent, but overall it is not viewed as a good thing even with positive outcomes.

DCS should respond to abuse, but they shouldn't intervene for certain situations. Urine drug screens in pregnancy are taken without consent. Then DCS investigates at birth. It's not helpful for DCS to investigate those.

There is a sense of distrust of DCS.

If a child is being neglected or abused, then there has to be an investigation. It's not easy to answer if DCS intervention is helpful or not.

Association with DCS/fear, Kachina (spirit) example—if you're naughty the Kachina will come take you away, if you're naughty DCS will come take you away.

I think my community views DCS in a negative way. Especially when everyone starts to gossip and create their own narratives.

In my culture it is considered bad luck to buy or make things for your baby before your baby is born. When I was in the hospital for labor, the staff asked what I had brought for the child and when I told them nothing, they made the assumption that I was unprepared rather than asking for clarification. They recorded that I was indifferent toward my child. Lack of cultural competence—that was not helpful especially as a new parent.

Feels that cultural competence should be part of the role of DCS but it is not. They should always be learners because they are dealing with so many different types of people.

DCS is perceived as having lack of cultural awareness.

DCS can be used as a threat from family and community —false allegations are made to get parents in trouble.

I feel like DCS are doing the minimum to act like they care.

When do you believe DCS involvement is needed? When do you believe it is not?

If you are worried about a child being harmed physically or sexually then there should be an intervention—but intervention should be respectful because not every report is based from the truth.

This is a hard question to answer. Children may not know they are being abused or neglected.

If there is a call, DCS should get involved.

If the school or a friend's parent shows concern, it seems like DCS should be involved.

I think when children are getting neglected, abused or sexually abused. Or if there are drugs involved.

This is a hard question because as a child you do not know you are being neglected. You think your parents love you, and a child may not have the knowledge or the words to be able to tell someone they are being abused. They may be confused when someone comes into their lives and tells them they're being abused, or they are removed from their family.

DCS should intervene when there is more than one report of concern for the child's wellbeing. Reports of physical/sexual abuse or neglect. Maybe to even help struggling families with resources.

I think when you have one or more people concerned for a child and have made phone call on behalf of a child, DCS should be involved. Especially when a child is seeking help.

The threat of child removal discourages parents from reaching out and asking for help from DCS. DCS could give resources instead of threatening removal. There are so many types of abuse and when it is not physical or sexual, it can be harder to identify and recognize the abuse (especially things like mental and financial abuse). This can also create fear in a parent that the children will be removed from both parents if one parent comes out about the other parent's abuse.

Body language of children should be considered. Children can pick up on what parents want them to do. Different factors need to be considered in investigations, such as school attendance, visits to ER, and if children are showing signs of fear.

DCS workers can point us to resources. I learned about a crisis center when I needed a break for several days. DCS has the ability to point you in a direction to get help.

DCS could be more helpful by pointing us to more resources, investigating and following through, and spending more time and energy on serious child abuse cases.

Family members can use DCS as weapon, who can make false accusations—calls are weaponized. DCS is used as threat.

In parental disagreements, DCS is weaponized.

If teachers see bruises or get a disclosure of sexual abuse, DCS should investigate.

There are a lot of terrible cases. Many kids suffer and no one reports it. There needs to be more intervention.

DCS should do unannounced visits to follow up about concerns about child abuse or neglect.

They should have another caseworker look and may see something that the first caseworker missed.

There are not enough workers to follow up on every case.

As the team designs this study, what should we be thinking about?

Participant questions: Does DCS investigate every call? How does DCS make a decision to investigate?

[facilitator explained that DCS does not investigate every call, and that they have a decision making tool to determine whether or not to investigate.]

Participant question: Do hotline calls reported by mandated reporters get more follow up (or are they more likely to be investigated) than a call from an individual who is not a mandated reporter?
DCS should consider culture.

They should be trying to find if the child is safe to remain where they are.

DCS should be considering that marijuana is now legal. Unsubstantiated early pregnancy drug screens should not be investigated; especially marijuana. **For substance-exposed newborn investigations, consider the severity of the drug and recency of use.**

Consider that not everyone is screened for drugs when they are pregnant—some people, especially those with lower SES may be more likely to be screened.

DCS should consider the health needs of people in the home, and how that interfaces with the kinds of calls that are received.

Taking children away can be harmful to parents and children. It shouldn't be done lightly or punitively.

The researchers should **consider the impact of having another caseworker look at a case vs. just one individual.**

Participant questions: Does DCS ever question the mental stability of the person calling? Do they ask them if they have been drinking or using drugs, or ask them how they know what they are reporting? How does DCS know that the reporter knows what they know, or is telling the truth?

Consider the motivation to report.

Harm reduction model is important. For substantiated reports that only happened one time, could DCS look at parenting classes? Or substance abuse classes?

DCS should be aware of cultural differences. Some cultures may not look you in the eyes.

Consider differences in how sound reports are, differences in how culturally sensitive DCS workers are, and how differences in how DCS treats people could affect the data.

Do you have any thoughts about how this project could affect members of your community?

It's commendable that DCS is considering different cultural groups before embarking on this study. I commend DCS for looking at different cultural lenses.

This study can increase trust by reducing unnecessary investigations.

DCS will increase trust by giving resources in behavioral health and parenting classes, not by taking children away.

If DCS builds a trusting relationship, then they are in a good position to refer families to resources.

DCS could make a difference by referring families to resources through a warm hand off.

Is there anything we haven't asked about that you would like to share?

As soon as I told DCS I was Native, they said they would need to pass me off to Tribal CPS, and it broke my heart. It's frustrating to me. Could DCS be a bridge between family and the Navajo Nation CPS? Consider collaboration between agencies.

Tribal CPS system doesn't work and does not respond. DCS and tribal jurisdiction adds an additional level of complexity. They do not collaborate well together. It is wrong of DCS to be able to turn someone away who is seeking help especially when a tribe is not willing to help.

Some people are good at 'playing the part' and **can get DCS to believe what they are saying. Others are unable to hide their emotions and may get angry when being investigated and be perceived differently because of their emotional reactions.**
If caseworker/hotline worker could have connected me to legal resources, that would have helped a lot.

Can we let a Native American person choose which agency they want to work with (DCS or Tribal CPS)?

Participant question: What determines if DCS takes a case or hands it off to Tribal CPS?

Look at how gaps in resources may be affecting the calls received in certain areas.

On the reservation, it is difficult to do certain things because the community is so entwined and that can limit your access to resources and help—in that case, DCS should be able to help if a parent cannot access what they need through their reservation.

Rural Arizona Focus Group

April 23, 2022

Overview of Focus Group

A focus group of rural Arizonans was held on Saturday, April 23, 2022, from 9:00–11:00am, virtually over Zoom. There were a total of eight focus group participants, including the community leader co-facilitator, Angie Burleson.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation about the purpose of and approach to the Next Event Analysis study was presented by Claire Louge of Prevent Child Abuse Arizona, with support from Angie Burleson, the group's community leader and co-facilitator.

Purpose

- DCS receives about 84,000 reports of child neglect and abuse per year. Arizona law tells DCS what needs to be investigated.
- DCS has guidelines to determine which calls meet the definition of the law, and are taken as a report to receive an investigation.
- The purpose of the study is to answer the following questions: Can we determine which reports, based on information collected at intake, are least likely to need a DCS intervention, and might need something different (like a community service)? Is there a reasonable way we could change processes so families do not get unnecessarily investigated?

Approach

- We will examine data DCS is already collecting on reports and outcomes to learn about what information from reports is tied to later outcomes
- Our goal is to understand which report characteristics (or combinations of characteristics) are most and least associated with the need for DCS intervention
- We are <u>not</u> creating a "predictive risk model" that will be used by caseworkers or supervisors to profile people. We are <u>not</u> handing decision making over to a machine or algorithm.

Study Aims

- We hope to safely reduce the number of families who have DCS knock on their doors/investigate them
- We hope to identify prevention services designed to meet the needs of families who today receive a DCS investigation, but do not need DCS intervention
- We hope to use the information to expand funding for prevention services that help families stay safe and together

After the presentation, participants were asked to reflect on the information with the following questions:

- What do you believe is the role of DCS in your community?
- How does your community view DCS? Is DCS intervention helpful?
- When do you believe DCS involvement is needed? When do you believe it is not?
- As the team designs this study, what should we be thinking about?
- Do you have any thoughts about how this project could affect members of your community?
- Is there anything we haven't asked about that you would like to share?

Summary of questions, reflections and discussion from the Rural Arizona focus group

Initial questions from focus group participants:

Will this study be for all Arizona data, or just rural populations?

Who is Mathematica?

How will they incorporate feedback, and to what extent will they use it in this study?

What do you believe is the role of DCS in your community?

To protect the children that may not be normally protected in a different manner through schools or other community functions.

The people that come in and out of the house to check in on placements and sometimes bring things the children may or may not need (foster family perspective). One time, a DCS caseworker brought school supplies, including scissors, to a foster child in our home who had behavioral issues and should not have scissors.

Not just to protect kids but also to support their families and give them resources that they may need to better care for their children.

DCS protects children and intervenes when protection is needed but also assists in finding placement for children—"forever families." Feels their role is noteworthy.

DCS acts as a buffer between foster families and biological families—DCS protects foster families... from potential negative interactions with the biological families.

DCS takes care of all the people involved in a case—managing all the moving pieces that come with a case—children's' emotional/physical needs, parents' needs, case management, health care, and any additional services.

How does your community view DCS? Is DCS intervention helpful?

In general and in discussion with community members—they view DCS as an important aspect and the agency to turn to when they see what they feel is abuse, however, there is a feeling that sometimes they go too far and sometimes they don't get it quite right and they drag cases out too long. Other times they are not quite as observant as they should be to children that could be considered to be languishing in foster care—there is mixed viewpoint on DCS and whether or not they are helpful.

There is a mixed view of DCS. Some are glad that DCS intervenes. Some view DCS negatively, that "DCS is here just to take our kids away." Some believe that they receive calls because their neighbors dislike them. When DCS is called, is it that there is a new neighbor who doesn't like the family, or is it for the safety and protection of the children?

DCS is an agency to report concerns, but it's also an agency that some parents turn to in order to turn in the other parent because they have hard feelings against that parent. When DCS does step in to

remove children, they don't communicate to schools that they have been taken—the schools call the parent to ask where their kids are.

From the perspective of a foster parent, there are a lot of DCS personnel who are inadequate for the job and who don't know what they're doing. They are in and out of the job.

DCS gives biological families a lot of chances, which "leaves the kid hanging."

DCS opens a lot of cases that involve false accusations of parents trying to get the other parent in trouble.

Assumption of DCS is that often the reports come from children themselves, and **children will call DCS because they don't like being disciplined**. This can cause an unnecessary intervention because discipline differs so much culturally and by community—DCS does not take this into consideration– they can be too enthusiastic about proving that they are doing their job. Heard from neighbors that there was a DCS case based on—a 12-year-old having TV time taken away.

Feelings that children lie and that can lead to unnecessary cases and investigations.
Discipline is different for every family. **What is the boundary between discipline and abuse?** If a child is having a tantrum in a grocery store and the parent physically lifts them off the floor, that may be reported as physical abuse. We need to separate these kinds of things from abuse. We must also separate things like children falling while playing. Many times, DCS investigates when it is not necessary. They are overenthusiastic in their investigations.

DCS doesn't know about available resources in community that they can refer families to. For example, DCS thought that Head Start was only for kids with disabilities.

DCS comes in with their own perception of the child—foster families have to deal with the bad behaviors. DCS will try to tell foster families how to parent children and do not look at the whole picture—DCS does not know the children—foster families feel unsupported.
Many caseworkers don't use common sense.

DCS may be biased towards mothers' opinions.

Communication differs across cultures—DCS may misinterpret because they may not understand cultural differences. DCS has large caseloads. They don't take time, or they may not have time, to understand certain situations. They may have prejudice against parents who get angry easily, and recommend removal of children. But displaying anger may be part of their culture. The removal would be worse for the children.

Some caseworkers are good and work to get children back with their families.

Many DCS case managers are very busy and have large caseloads and do not take the time to fully engage with families, and then make recommendations to separate families without considering the impact.

Recommendation that DCS workers have round table discussions to learn from foster parents.

When do you believe DCS involvement is needed? When do you believe it is not?

Reports that come from schools should be investigated because they have a relationship with the child. DCS should communicate with schools when they remove children. Reports that come from law enforcement and medical professionals should also be investigated.

DCS should communicate with the school when children are removed and placed somewhere else so that the school can feel "at peace" that the child is safe.

DCS does not need to be involved with cases for most one-time occurrences such as an accident, or moments of poor behavior of the parent.

DCS should not investigate families who live off the grid and who don't have things like normal electricity or water. Just because they live off grid does not make it abuse or neglect.

DCS investigated a boy whose family was living off grid and accidentally tripped into a fire. DCS questioned mom at the hospital as she was trying to comfort her son. DCS should be sensitive to these things and wait for the mother to comfort her son before questioning her.

DCS needs to use common sense. They go by their checklists only, that's where they mess up.

DCS is often not aware of culture differences. DCS has a one-size-fits-all approach—this ignores cultural differences. We need to have a common sense approach.

As the team designs this study, what should we be thinking about?

They need to see if calls are reoccurring—reoccurring callers or families who get reoccurring calls. **Some callers may be racially profiling people. Some investigators may be racially profiling people. DCS needs to look at who is reporting families, and if there are people who are frequent reporters.**

DCS should not investigate calls regarding adults who are dealing with a child having a tantrum.

They need to assess where calls are coming from.

Foster care parent—I have called DCS three times, and I hate calling the hotline. I feel like I am wasting my time. They don't ask the right questions. There is misinformation about where to call for foster parents.

The hotline should be better. The hotline does not ask enough questions and they do not use common sense.

Hotline workers don't use common sense when answering calls from foster parents.

[Many participants gave feedback about the hotline, rather than the study.]

DCS should consider that there are not a lot of mental health resources, or other resources, in rural areas. Sometimes children are receiving everything that they can. This does not mean that they are being neglected. Resource availability should be taken into account.

Marijuana is legal and parents should not be investigated for marijuana use. This should also be considered for substance exposed newborns.

They should follow up with callers.

There also should be a balance when considering harm to the child from the experience of being taken away from their parents as well as the action of the parents.

Do you have any thoughts about how this project could affect members of your community?

I believe that if we're able to cut down on unnecessary calls, it'll help DCS focus on higher priority cases. If they didn't need to divert their time, they will be able to focus on things that are more priority.

Hopes the data can identify repeat offenders (repeat callers) for unsubstantiated calls. There need to be repercussions for people who make false reports. When repeat false reporters are found, they need to be warned. Some parents are in the habit of surveilling parents. Take into account the reporter—if there are people who are repeatedly reporting of cases that are unnecessary, they need to do something about that.

I'm hoping the study will have positive impact and create a lighter case load for DCS. And for parents to feel that they aren't being discriminated against because of culture, and that children are protected.

With this information, DCS can truly investigate cases that need to be investigated.

I hope that DCS will be able to crack down on false allegations. Reduce investigations of repeat offenders that falsely report.

Hope that the data will create more consistency between case managers. **Take into account that plans for children change when DCS caseworker changes.**
We should feel that DCS is working with us, as foster parents. If every time DCS comes, it seems they are trying to find fault with foster parents, it feels like they are against us.

I'm hoping this study, and the trust it will build, will help keep foster families longer, and help recruit new families. Build trust in DCS will increase willingness to be foster parents.

If we apply a common-sense approach, and we can reduce investigation of foster families based on false accusations, we will increase foster placements. There is the potential to lessen the experience of siblings being split up because there are no foster family placements available to keep children together. More harm is caused by DCS when children are removed abruptly and separated.

Participant question: In rural areas, are we leaving kids in homes because we don't have foster placements for them?

We need community specific data-consider where the calls are coming from, and what is available in each community, and why children in certain communities are being removed.

I hope that this will make DCS a partner with parents—in rural areas, we try to make most of resources.

DCS needs to communicate with the community.

DCS should have an eye on the bigger picture, and to weigh all the factors, basing their choices on what's best for the child.

Don't put foster parents down. Permit foster parents to discipline parents. Don't lavish foster kids with gifts.

DCS should stop investigating the little things, and focus on the big ones. One parent called DCS on the other parent because other parent didn't want child to celebrate Christmas. But there are other cases that the children die before investigation is opened. **They should be more involved in children being abused, rather than disagreements between parents.** Pay more attention to cases that really need help.

Is there anything we haven't asked about that you would like to share?

Transportation would improve connectivity in rural areas. It is very inconsistent and needing improvement.
Consider the trauma of removal.

Participant question: Do they track calls to the hotline from foster parents, what happens to those? Are foster parents told to use the hotline? Are they told to call the hotline?

Participant question: Is the hotline a catch-all for inquiries about ongoing cases, foster placements, as well as suspicions of abuse?

Participant question: Are they going to determine where these calls are coming from? Specifically, will they be looking at how many calls are coming from law enforcement?

It would have a huge impact if DCS and law enforcement worked together.
Does the volume of other calls (ones that are not about suspicions of child abuse and neglect) affect the outcomes of calls and the rates of investigation?

The distrust of DCS in rural communities is higher. When people call on each other, they are more likely to know each other. DCS may tell them they need to do something, but the resource isn't available, like mental health.

If DCS would stop focusing so much on the little things where there is no need for a case and focus more on the issues where they really should be involved, they would not be considered so ineffective by the community.

Does DCS investigate more in rural areas? Why are rural communities being investigated more? Are rural communities being investigated more for certain reasons? And are those things associated with a different culture of being rural?

Participant question: I would like to see the outcome of hotline calls—who monitors the outcomes of reports?

DCS should develop a user-friendly training for case managers/investigators with cultural differences considered as well as individual situations. There also should be a balance when considering harm to

the child from the experience of being taken away from their parents as well as the action of the parents.

The abrupt change to Guardian was VERY painful and poorly implemented.

They should increase the salary of the case managers who demonstrate competency in their abilities to identify the levels of threats of harm occurring. Increase training to include numerous case studies. Increase the pool of case managers and reduce their caseloads.

Think about whether the children are in danger versus the harm generated by the DCS intervention. Consider culture. Think of how to adequately screen and train individuals to make these determinations. This could build confidence toward DCS from the community, reduce unnecessary investigations and reduce stress upon case managers and DCS staff.

Make sure DCS communicates about the changes that this study will create.

Consider differences in decision making processes when looking at case outcomes. Were some decisions made independently by one person, and some by several?

South Tucson Focus Group

April 27, 2022

Overview of Focus Group

A focus group of South Tucson residents was held on Wednesday, April 27, 2022, from 6:00–8:00 p.m., virtually over Zoom. There was a total of eight focus group participants, including the community leader co-facilitator, Erika Diaz. This group was a mix of Spanish and English speakers. Four focus group participants were Spanish-speaking only, two were English-speaking only, and one was bilingual. The Spanish-speaking community leader co-facilitator served as a translator during the group. Notes were taken by English speakers based on the live translations relayed by the community leader co-facilitator.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation about the purpose of and approach to the Next Event Analysis study was presented by Claire Louge of Prevent Child Abuse Arizona, with support from Erika Diaz, the group's community leader and co-facilitator.

Presentation Outline:

Purpose

- DCS receives about 84,000 reports of child neglect and abuse per year. Arizona law tells DCS what needs to be investigated.
- DCS has guidelines to determine which calls meet the definition of the law, and are taken as a report to receive an investigation.
- The purpose of the study is to answer the following questions: Can we determine which reports, based on information collected at intake, are least likely to need a DCS intervention, and might need something different (like a community service)? Is there a reasonable way we could change processes so families do not get unnecessarily investigated?

Approach

- We will examine data DCS is already collecting on reports and outcomes to learn about what information from reports is tied to later outcomes
- Our goal is to understand which report characteristics (or combinations of characteristics) are most and least associated with the need for DCS intervention
- We are <u>not</u> creating a "predictive risk model" that will be used by caseworkers or supervisors to profile people. We are <u>not</u> handing decision making over to a machine or algorithm.

Study Aims

- We hope to safely reduce the number of families who have DCS knock on their doors/investigate them
- We hope to identify prevention services designed to meet the needs of families who today receive a DCS investigation, but do not need DCS intervention
- We hope to use the information to expand funding for prevention services that help families stay safe and together

After the presentation, participants were asked to reflect on the information with the following questions:

- What do you believe is the role of DCS in your community?
- How does your community view DCS? Is DCS intervention helpful?
- When do you believe DCS involvement is needed? When do you believe it is not?
- As the team designs this study, what should we be thinking about?
- Do you have any thoughts about how this project could affect members of your community?
- Is there anything we haven't asked about that you would like to share?

Summary of questions, reflections and discussion from the South Tucson focus group

What do you believe is the role of DCS in your community?

DCS is here to take reports for the security of children.

To protect children from abuse, especially when it's hidden.

I think their role is to monitor family and children's well-being.

DCS's role is protection and prevention of child abuse and is there for their safety —to support their well-being and development.

To protect how children are being treated and being cared for.

Checking in on families and child welfare.

They are first responders for child abuse reports.

How does your community view DCS? Is DCS intervention helpful?

They view them as the bad guys—they think they are coming to take their kids, and they won't give them back. They hear the stories on the news that reinforce those beliefs that DCS is bad.

Growing up, I was taught that CPS takes you from your parents; it's something we feared. In the educator community it is something we are obligated to report to. For myself I have had bad/unresponsive interactions with DCS so they are someone I am obligated to report to but unfortunately don't trust fully.

They are scary and just want to take children away.

DCS is viewed as the result of calling the police. If you call the police, they will call the Department of Child Safety and ultimately that will affect children.

There's a lot of fear and misinformation in the community about what DCS does. There is lack of information on how they define abuse/neglect. Community and providers do not know what to report. Where is the line? When does something become neglect? If a child does not shower for a month and their parents are fine with it, is that neglect?

It is useful to have this department to improve the development and stability of the children but it is necessary that they have sensitivity.

DCS is helpful when we get free child care.

When do you believe DCS involvement is needed? When do you believe it is not?

I feel like if there is a bruise or something on a child and it is obvious that a child is being abused, DCS is necessary. However, there are times when there are **reports made because someone is mad at you or for revenge/no proven reason.**

When there is physical abuse or neglect it is necessary. Maybe unnecessary when there is a child being spanked because that it how their parents choose to discipline.

Participant is not sure when DCS actually does investigate.

I think it is necessary for DCS to be involved when there is violence, but I don't know, for example when a child is not bathed regularly, when is that seen by DCS as child abuse? When community members have conversations with families where the child hasn't been bathed in 30 days or the child hasn't been attending school, at what point does that constitute a report? The community is fearful to report because of a distrust of what DCS involvement leads to for the family and child—is it worse for a child to end up in foster care than not be bathed?

Parents need information on what a child's rights are, but also where parents are responsible for taking action. Can we force a child to shower?

When there is very strong evidence about a minor's condition.

It is needed when an intervention has to be carried out due to unsafe conditions for children.

In my community, when I want to help, if I call DCS, is it going to be worse? We don't know what DCS is going to do. Community tries to help each other and determine if they should call. DCS either doesn't intervene or does something extreme, like removing a child. There is lack of understanding about what DCS does, and how abuse and neglect is actually defined.

DCS is perceived as either not investigating when it's probably needed, or removing children when that seems too extreme for the situation. Causes a lot of confusion about what should be reported to DCS. People fear calling DCS because they know it could get a child removed and they are not sure if the situation warrants such an extreme measure. On the other hand, sometimes the calls receive no intervention. **It is not clear what gets an intervention and what does not, or what warrants child removal.**

There is a difference between discipline and abuse, but it's not often a clear line.

Discipline practices are cultural—'la chancla' or pinching is common in Latinx culture—does that constitute abuse? DCS should provide information about how abuse is defined. **Providers need to know what's reportable and so do parents.**

Examples—Child eating only frozen food. Children walking around because their parents are at work or don't have a vehicle. A child left in child care for 8-10 hours in a day. A 17-year-old sending nude pictures of herself and the parents not doing anything about it. Depending on perspective, any one of these things could be considered neglectful by a community member. At what point do you report?

What do you do with a child that doesn't follow the rules? Does that constitute a DCS report?

Permissions with technology—children using technology inappropriately but what else can you do besides taking away technology and what is a kids right vs. parental rights?

Are kids really better off in foster care? Running a child care center, I have foster children who are left in my care for the whole day. Aren't they with foster parents so they could get more love, time, and attention from them?

Community and providers are really looking to DCS for clear direction and guidance. Knowing what DCS considers abuse and neglect would help providers keep parents accountable.

Prevention programs/home visiting programs should give parents information to stay within the "correct limits" that DCS believes is abuse. The community needs guidance from DCS to know how to 'follow the rules.'

It's important for DCS to work with the school and community and understand the needs of the community they are working with; it's important for DCS to consider a community's needs prior to intervention and investigation.

DCS is needed for true neglect or abuse—the community does not understand the limits that DCS enforces. It would be important to have more specific guidelines for the community provided by DCS.

A participant's sister was in a group home, and the group home reported her for feeding the child chips (perceived as only feeding the child chips and a misunderstanding surrounding nutrition), ultimately the child was removed and placed with a family member and her sister/the child's mother was also living in the placement. There was contention surrounding parenting and who was "in charge" and her sister was kicked out, then the family member gave the baby back to DCS. The child was placed with a foster family and ultimately adopted out after her sister's rights were severed—her sister died 2 years later but the family was able to reunite with her niece after her sister's death through church, despite the child being severed and adopted out years ago.

As the team designs this study, what should we be thinking about?

DCS should be thinking about what children are saying when they investigate. Personal example—when DCS was called on a family with a young child who was fearful of going home, DCS did not determine child was unsafe —said kid was just being 'whiney.' But the kid had bruises, and the child had stories that were scary—like being put on the ceiling fan, or in the oven as a game. Older brother said he was just being whiney—probably echoing the parents. But later on, the child came to school with a broken arm. What the child says is sometimes not taken into consideration—the parents' word is sometimes taken over the child's.

There are false negatives—listen to the children. Investigation may find the child to be safe, but are they really?

Culture influences how things are perceived, and what is normal.

Culture influences the care for children.

Times have changed because of the pandemic. Especially over the last 2 years, from the perspective of someone who has volunteered in a domestic violence shelter, things that happen in the Latinx community differ from the American community. There are certain things that DCS responds to that the community does not perceive is abuse. The community would like to work together with DCS but they would also like DCS to understand cultural differences. Particularly because there are lots of families in survival mode right now because of the pandemic, it feels there are more unnecessary calls being made to report things that they may not have reported prior to the pandemic. **Take into account the circumstances, like the pandemic, when looking at data.**

DCS should be more considerate of the culture, what's happening in all of these communities, and how to help the communities that have been impacted by the pandemic.

DCS does not consider other points of view. It's important to be sensitive/sensible about what type of families you're working with.

Community providers want information about what DCS does and what it is responsible for.

Personal story: When I was a child, I was in and out of foster care from ages 1-14. I was molested in foster care. In adulthood, I have had DCS called on me. I was worried they would take my child, but when I said I didn't have money for gas, they helped. They helped get me back on my feet.

Participant wonders why DCS wants to reduce intervention since they give families resources and connect them to resources. They show up when families need assistance. **DCS is able to help—so why are they wanting to reduce intervention?**

DCS is the most well-known organization that works with families and children who need help. They should take responsibility for their role in the community and should be supporting families. **If you are trying to decrease calls into the hotline or decrease investigations, what are you doing to make up for those lack of calls—how are you teaching the community to look for the right help if you don't want them turning to DCS?** Example: It's impossible to get free child care services, but if DCS is called you get it immediately.
This needs to be considered before making changes in which they investigate less. Whether or not it is true, they have the reputation of being the agency to call when you want families to get help. If you're trying to reduce investigations, how will you make sure that families will still receive help?

Participant example: I couldn't get assistance with child care, and I asked to be reported to DCS. When DCS investigated, I got child care assistance immediately. **DCS provides resources we can't get easily anywhere else.**

Questioning children in front of their parents is essentially useless. Children will lie to make sure that they don't get in trouble with their parents. Question children separately.

Do you have any thoughts about how this project could affect members of your community?

I think it's a good project but it's important that DCS works as a team, and they should be collaborating with schools, DDD, etc. to figure out if a child is in danger prior to removal. They should be collecting more information about how a family is functioning prior to opening an investigation.

DCS should really be putting more time into understanding families and cultures. Participant appreciates that the focus group is allowing an opportunity for this community to share their thoughts, but DCS really needs to get feedback from each community to make positive changes and consider culture when they are making changes.

Participant appreciates that DCS is asking community, and being part of this focus group is a big role. **Recommends that DCS gets more input from the Latinx community** as they accomplish what they are working to accomplish in this study.

Whatever DCS is trying to accomplish—they need to work with community to implement.

Co-facilitator/Translator's Summary of Focus Group Themes:

People within the community don't know what to do—they are being themselves, living within their culture, then will hear from another person that they are going to have their children removed because of XYZ—the misinformation is creating fear, so causes hesitation to call DCS at all because of a deep distrust.

The result of DCS intervention is unknown, sometimes a parent just needs help and instead their children are taken away from them and they cannot get them back, and so they are constantly trying to determine as a community if a child is safe or not, what constitutes a need for intervention. There are instances where the community thinks DCS should be involved, but still will not report because of the lack of understanding of the process and the purpose of DCS.

DCS should be collecting more information from teachers, doctors, neighbors, etc. when a report is made vs. having one person make a decision that impacts an entire family and community.

DCS should have more bilingual information available for education and resources.

CAC Meeting Notes, Session Two

African American Focus Group

May 6, 2023

Key context from participants

- Differences in substantiation, case opening and removal may be explained by **misinterpretations and misunderstandings of African American cultural norms** by callers and investigators.
- The **potential bias of hotline workers** may affect the interpretation of allegations.
- The **DCS system is opaque** —participants wondered about the processes in the DCS system, and
- who makes decisions and how.

Key recommendations for report

- Data regarding **types of reporters** would be helpful information (mandated reporter, neighbor, family member, etc.).
- It was strongly suggested that **definitions of each allegation are included in the report**, not just the truncated label.
- It was recommended that this information, when shared with the public/communities, is made as **digestible and understandable as possible**. "Delivery matters."
- It was suggested that **DCS is given more access to data from other family-serving systems** to help them make better decisions.

Overview of Focus Group

A focus group of African Americans was held on Saturday, May 6, 2023, from 1:00–3:00 p.m., in person in Phoenix. There was a total of six focus group participants, including the community leader co-facilitator, Dr. Patricia Neff. All six focus group members had been a part of the first focus group held in April 2022, which was asked for guidance on how the Next Event Analysis study should be designed.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation on the initial findings of the Next Event Analysis study was presented by Allon Kalisher and Jonathan Gellar of Mathematica, and Katherine Guffey of DCS. Specifically, the group was shown:

- the most common types of allegations reported to the DCS hotline
- the most common allegations on investigated reports as differentiated by race
- the rates of substantiation outcomes of allegations by race (how often allegations are substantiated for each race as compared to White people)
- the unadjusted and adjusted rates of each outcome by race
- the "relative risk" of most common allegations, meaning how likely substantiation, case opening and removal happens for each allegation type
- the types of allegations with the lowest and highest relative risk

Participants were asked to reflect on the information with the following questions:

- What are your thoughts on these results?
- How do you make sense of this?
- How do you think people will react when they see these numbers?
- What would you want people and policy-makers to know about these results?

Summary of questions, reflections and discussion from the African American focus group

Participants asked for clarification about the difference between internal and external data.

It was asked if DCS would be using this information to predict what children were at a higher risk of negative outcomes.

After it was clarified that a case's race is determined by the race of the child's primary caregiver, some participants believed it would be better to have the child's race determine the case's race.

It was asked if the data would be broken down by the kinds of callers. Data regarding where the reports are coming from would be helpful information (mandated reporter, neighbor, family member, etc.).

It was asked how the bias of hotline staff could be considered.

When asked what they would want policymakers to know, participants shared:

- They encourage DCS to have units for each culture, and/or make a concerted effort to hire staff members that can understand each culture on a different level.
- There may be misinterpretations and misunderstandings by callers and investigators due to cultural norms. Example: An African American could say that they will 'bust someone in the head' but it's not literal. A White person or a person of a different culture could misinterpret it as literal.

When asked to give context on why African Americans have a higher substantiation, case opening, and removal rate, it was asked if the characteristics of the hotline worker are being considered.

"Is there a bias in the interpretation of what the hotline people hear?"

It was suggested that the ease of substantiating an allegation is discussed in the report.

It was suggested that policy is changed so that not every substance exposed newborn allegation is required to be investigated.

It was strongly suggested that definitions of each allegation are included in the report, not just the truncated label.

It was mentioned that a caseworker's perception has disproportionate influence in case decisions. Participant feels hotline workers' perception of the reporter is important to how reports end up being classified in the system.

The participants had a lot of questions about how decisions were made at different points, and noted that the leadership of DCS Directors matter. Participants wondered about the variables in the DCS system —who makes decisions and how.

It was suggested that mandated reporters are given a 'decision tree' to help them determine whether or not to make a report.

A participant asked if AI could be used to eliminate the 'human drama' of decision making.

It was recommended that this information, when shared with the public/communities, is made as digestible and understandable as possible. "Delivery matters."

It was suggested that DCS is given more access to data from other family-serving systems, like "McKinney Vento" (data on homelessness), which could help DCS make better decisions.

"DCS doesn't come as a source of help. It's not a resource agency. It's scary. It's something that Black families use against each other. It's child policing. It hasn't had a positive face. We need to change that. We need to change the face of DCS. The messaging is negative. Weaponization. It's negative because they're intervening when they don't need to."

Native American Focus Group

May 2023

Key context from participants

- **Higher levels of stress** in Native American communities and families may affect rates of substantiation, case opening, and removal.
- The **lack of understanding of the right to ask for due process may be affecting substantiation rates** for Native American communities, and other marginalized communities.
- The Native American **cultural norm of "unassertiveness" may affect substantiation rates**, as Native Americans, especially in older generations, may not feel comfortable asking 'authority' for due process.
- There may be **bias against Native Americans** in the DCS system.
- It was noted that the **ease of substantiating an allegation may affect how frequently it leads to substantiation.**
- It was mentioned that **Native Americans lack education on healthy adult relationships, and what constitutes abuse**, and this could affect data.
- Factors such as the **cost of living and poverty** may be affecting the data in the Native American community.

Key recommendations for report

- Make it clear in the report that the data is DCS data only, meaning that Native Americans represented in this data are living off reservations.
- Explain that the race data pertains to the race of the primary caregiver of the child, not the child.
- Explain that caseworkers sometimes guess the race of the parent. When presenting the data, it should be clarified that there is no guarantee that the race/ethnicity data is accurate.
- Explain that though race data is not captured at the hotline call, it is known for some families due to past investigations, and this is how data on the difference in screened in rates between races is determined.

Overview of Focus Group

A focus group of Native Americans was held on Saturday, May 6, 2023, from 10:00 a.m.–12:00 p.m., in person in Phoenix. There was a total of five focus group participants, including the community leader co-facilitator, Holly Figueroa. All five focus group members had been a part of the first focus group held in April 2022, which was asked for guidance on how the Next Event Analysis study should be designed.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation on the initial findings of the Next Event Analysis study was presented by Allon Kalisher and Jonathan Gellar of Mathematica, and Katherine Guffey of DCS. Specifically, the group was shown:

- the most common types of allegations reported to the DCS hotline
- the most common allegations on investigated reports as differentiated by race
- the rates of substantiation outcomes of allegations by race (how often allegations are substantiated for each race as compared to White people)
- the unadjusted and adjusted rates of each outcome by race
- the relative risk of most common allegations, meaning how likely substantiation, case opening and removal happens for each allegation type
- the types of allegations with the lowest and highest relative risk

Participants were asked to reflect on the information with the following questions:

- What are your thoughts on these results?
- How do you make sense of this?
- How do you think people will react when they see these numbers?
- What would you want people and policymakers to know about these results?

Summary of questions, reflections and discussion from the Native American focus group

A participant asked if false allegations are ever made to DCS, and if there were "bad faith" reports. The answer is yes, but it is challenging to determine who makes these bad faith reports, as they are often anonymous.

A participant asked if hospital and school attendance records are used to validate the information received by the DCS hotline. Mathematica answered that hospital data could be used to validate data, but it is not currently reflected in the analysis, and that school attendance records are not currently used to validate DCS data. It was suggested that school attendance records, including the Native American Education Department and ADE could be a source of data validation. There is a challenge to linking data from hospitals and schools to the same people in a DCS case.

It was clarified that a family's race in this study is determined by the race of the primary caretaker, not the child. It was suggested that this is made clear in the study report.

When presented the data on the most common allegations on investigated reports, a participant noted that it "doesn't seem accurate." When asked to explain, participant noted that it seems like things were missing from this list. It was clarified to the participant that what was shown on the screen were the top 10 most common allegations.

Participants were surprised that sexual abuse allegations did not make it into the top 10 most frequent list. DCS/Mathematica clarified that this data is not including any reports made to tribal social services.

A participant felt that the calls made about substance-exposed newborns should be a higher number than shown because of calls due to mandated reporting from hospitals.

A question was asked if mandated reporters need to report suspected substance exposure during pregnancy, or if reports can only be made at birth. DCS answered that reports can be made at any time, but DCS can only act to investigate substance exposed newborns after the birth event.

It was clarified that these data were about allegations captured in calls, not about what is "ground truth."

A participant asked about the time span in which the data were captured. The answer is about a year and a half.

There was a question asked if this data included Tribal CPS data. This data is DCS data only, meaning that Native Americans represented in this data are living off reservations. It was noted that this should be made clear in the report.

It was asked how hotline workers know the race of the child or family if they are not asking in the screening tool. The answer given was that race is known if prior reports were made on the family and if they were investigated in the past.

It was asked how Native American status is determined, and if tribal affiliation is verified. It was answered that it is self-reported.

It was recommended that the data is broken down further by cultural group, if possible, as there are many cultures within a racial group.

It was recommended that the report explain that the race data pertains to the race of the primary caregiver of the child, not the child.

It is recommended that the report state that caseworkers sometimes guess the race of the offending parent. Participant offered feedback that when presenting the data, it should be clarified that there is no guarantee that the race/ethnicity data is accurate.

It was clarified that DCS does not take a report when the allegations of abuse are against an adult that is not the primary caretaker of the child—that falls into law enforcement jurisdiction.

A participant asked about substantiation, and how an allegation becomes substantiated. It was asked if parents know that they can request due process by a court to potentially avoid substantiation.

When asked to give context for the higher rates of substantiation, case openings, and removals, participants mentioned:

- Higher levels of stress in Native American communities and families
- Potential bias in the DCS system
- Native Americans frequently have multiple families living in homes

It was noted that the ease of substantiating an allegation may affect how frequently it leads to substantiation.

A participant said that the Native American cultural norm of 'not being assertive' and 'compliance' may affect substantiation rates, as Native Americans, especially in older generations, may not feel comfortable asking "authority" for due process.

It was suggested that when investigations happen, there should be information left with families about the court process, resources for navigating the child protection system, and resources for addressing the needs of the family, such as getting orders of protection. "We need instructions for parents. There's no knowledge."

A participant stated that policy makers should be thinking about factors such as the cost of living, and poverty. One participant with experience of the child protection system noted, "If I knew resources were available, I would have been able to solve the problem."

The cost of requirements for reunification should also be considered, such as parenting classes. Court mandated parenting classes should be free.

A participant mentioned that Native Americans lack education on healthy adult relationships, and what constitutes abuse.

A participant expressed gratitude that DCS was sitting down to talk with them and gain their perspective.

The discussion ended with a question: how do we connect people with what they need?

It was asked if the Next Event Analysis Study report would be shared with reservations, and suggested that tribal CPS conduct similar analysis to learn about how they could do better.

Rural Arizona Focus Group

May 20, 2023

Key context from participants

- Differences in rates of case openings in rural Arizona could be explained by **the lack availability, proximity, and ease of access to resources in rural Arizona**.
- Rural Arizonans likely live in rural Arizona for an increased desire for privacy—they may have a **higher distrust of strangers**. They may have a different reaction to investigators coming to their homes, and may be perceived as more defensive. This **may explain differences in substantiation and case opening** for rural Arizonans**.**
- Participants noted that **DCS staff capacity in rural Arizona** may be affecting the rates of cases being opened.
- Participants noted that rates of **poverty, mental illness, and substance use in rural Arizona** may be making cases more extreme.

Key recommendations for report

- Participants recommended that **all allegations are specifically defined** in the report.
- It was recommended that the report include the **definition of the terms** "allegation," "substantiation," "case opening" and "removal," as well as the **decision criteria** for making these determinations.
- It was noted that the **severity of allegations** is not measured at the hotline level. It is suggested to mention this in the report.
- It was discussed that removal may not be the best proxy for child protection. Example: Sex abuse allegations may not lead to high rates of removal, but this does not necessarily mean DCS is not working to protect child victims of sex abuse. The **limitations of using substantiation, case opening, and removal as proxies for child protection** should be discussed in the report.
- It was noted it would be useful to include data on the **kinds of callers** to the hotline and the outcomes by type of reporter.
- It was recommended that this **data is broken down by county** in the report.
- It was suggested that **narrative examples are added to explain the data**.
- It was strongly recommended that the **results are presented as clearly and simply as possible** so that the general public is able to understand.
- It was noted that the **differences between races shown in the data may exacerbate bias or misguide decision making,** and that this should be considered when writing the report.

Overview of focus group

A focus group of rural Arizonans was held on Saturday, May 20, 2023, from 10:00 a.m.–12:00 p.m., virtually over Zoom. There was a total of seven focus group participants, including the community leader co-facilitator, Angie Burleson. All seven focus group members had been a part of the first focus group held in April 2022, which was asked for guidance on how the Next Event Analysis study should be designed.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation on the initial findings of the Next Event Analysis study was presented by Allon Kalisher and Jonathan Gellar of Mathematica, and Katherine Guffey of DCS. Specifically, the group was shown:

- the most common types of allegations reported to the DCS hotline
- the most common allegations on investigated reports as differentiated by race
- the rates of substantiation outcomes of allegations by race (how often allegations are substantiated for each race as compared to White people)
- the unadjusted and adjusted rates of each outcome by race
- the "relative risk" of most common allegations, meaning how likely substantiation, case opening and removal happens for each allegation type
- the types of allegations with the lowest and highest relative risk

Participants were asked to reflect on the information with the following questions:

- What are your thoughts on these results?
- How do you make sense of this?
- How do you think people will react when they see these numbers?
- What would you want people and policy-makers to know about these results?

Summary of questions, reflections and discussion from the Rural Arizona focus group

A participant stated that some people have anger management issues, but it may not mean the child is unsafe in their presence, and that it may be more harmful to remove the child. The perception of anger is subjective.

Participants recommended that all allegations are specifically defined in the report.

A participant asked what is being done to prevent substance exposure to children whose parents have a pattern of substance use, and continue to have children.

When asked to give context for the data about substantiation, case opening, and removal differences between urban and rural populations, a participant suggested that people who live in rural areas may have a greater sense of distrust of people "knocking on their door" than urban areas. Distrust is more pervasive in rural areas.

A participant reported that people tend to live in rural areas because they like their freedom, and they may have different acceptable norms, such as spanking or yelling. Somebody that represents a system might see them in a different light. There may be different standards that rural people live by. They may be perceived differently. Though there is a general respect for authority figures such as police, rural people may believe that they are being misunderstood when being investigated.

"If you're living off grid, there's a reason. If you're rural, it's likely that you don't want someone in your space."

When reflecting on why substantiation is higher for rural Arizonans but case opening rates are less, a participant noted that substantiation may be higher because if a person comes onto the property of someone living in a rural area, they may brandish a weapon, which may affect the perception of that person. That initial perception may be different than reality, which is why substantiation is higher and case opening is lower.

A participant noted that the DCS staff capacity in rural Arizona may be affecting the rates of cases being opened.

A participant mentioned the limited resources available in rural communities. “In order to open a case, there needs to be resources available to make the case possible.” The participant also noted that rates of poverty, mental illness, and substance use in rural Arizona may be making cases more extreme.

A participant noted that bias exists, and may be leading to higher substantiation, case opening, and removal. A Black person may be perceived as ‘unsafe’ for the same behavior exhibited by a White person. There may also be cultural norms differences. “What we may call discipline may be misinterpreted as abuse.”

A participant asked for clarification on the difference between a case being open and a removal.

It was recommended that the report include the definition of the terms “allegation,” “substantiation,” “case opening” and “removal,” as well as the decision criteria for making these determinations.

A participant noted that it might be triggering or ‘activating’ to see that allegations of sex abuse are less likely to result in removal, and that an explanation for why should be given. It may be perceived as DCS neglecting to keep children experiencing sexual abuse safe.

A participant noted that allegations are based on interpretation and can be affected by bias.

A participant mentioned that a lot of allegations are difficult to prove, and the ease of proving allegations may affect substantiation numbers.

It was clarified by DCS that the majority of sex abuse is underreported, and these numbers do not reflect the prevalence of actual sexual abuse.

It was discussed that removal may not be the right proxy for child protection. When considering the sex abuse allegation example, because a child is not being removed after sex abuse allegation doesn’t mean that nothing is being done to protect the child.

It was suggested that narrative examples are added to explain the data.

It was noted that DCS may not be the right answer for the things that are shown by the data to be the lowest risk of removal—it may be showing that there’s a need for something else.

It was noted that there are allegations that are happening frequently, but don’t often require child removal. This has policy implications for what can be done to meet the needs of families and children.

It was noted that allegation doesn’t measure severity. It was suggested that the screening tool is revised to account for allegation severity.

A participant asked if the kinds of mandatory reporters are being accounted for in this study, and that it would be useful to include data on the kinds of callers to the hotline.

Participants asked about the definitions of certain kinds of allegations presented, such as “Caregiver is absent and unable to perform responsibilities.” It was noted that lack of affordable child care and housing could contribute to the frequency of this allegation.

DCS clarified that “caregiver absence” often means the caregiver has been arrested or is deceased. A participant mentioned that mortality rates with substance use and arrest rates could help interpret “caregiver is absent” allegation.

A participant asked if DCS assesses the parent’s ability and desire to meet the needs of the child and address the allegation.

It was noted that the reason that rural people may have a higher rate of substantiation but not open cases may be because of lack of resources.

Participants gave examples of parents struggling to make ends meet and the lack of available, proximate, affordable resources to help in rural Arizona.

A participant asked if it was possible to break down the data by cases that have a substantiation and a removal and cases that are substantiated and not removed.

It was recommended that this data is broken down by county in the report.

It was strongly recommended that the results are presented as clearly and simply as possible so that the general public is able to understand.

Participants spoke about the fears of having this data misinterpreted. “People will confirm their bias.” As a result of viewing this data, participants noted that people may feel afraid of removing children from families of races overrepresented in the data because they do not want to contribute to disproportionality, or, alternatively, this data may confirm and exacerbate the bias against certain people or communities. “We don’t want to put a target on anyone’s back.”

The participants thanked DCS and the researchers for the discussion, and thanked them for “getting human perspective on these numbers.”

South Tucson Focus Group

May 23, 2023

Key context from participants

- South Tucson's **immigrant population may lack understanding of the acceptable standard for raising children in regard to living conditions and discipline**, versus what constitutes child maltreatment.
- South Tucson residents **may not know who to ask for what help, and the role of DCS, police, and community resources.**
- There is the perception of **lack of police response to violence** in South Tucson.
- South Tucson is **under-resourced**.

Key recommendations for report

- Participants recommended that all **allegations are specifically defined** in the report.
- Most people do not have a basic understanding of the DCS process. It would be **helpful to define the process—from hotline call to potential outcomes—as simply as possible**, as many focus group participant questions were about the DCS process.
- Strong recommendation that the **study report is communicated as simply and clearly as possible** to facilitate community understanding.

Overview of focus group

A focus group of South Tucson residents was held on Tuesday, May 23, 2023, from 5:00–7:00 p.m., virtually over Zoom. There was a total of six focus group participants, including the community leader co-facilitator, Erika Diaz. All six focus group members had been a part of the first focus group held in April 2022, which was asked for guidance on how the Next Event Analysis study should be designed.

This focus group was made up of four Spanish-speaking participants, one English-speaking participant, and one bilingual English—and Spanish-speaking community leader co-facilitator. Also present was Paola Lopez, a Spanish-speaking research assistant with Mathematica, who helped translate concepts presented and took notes on participant feedback. Throughout the focus group, people speaking paused to allow Erika Diaz or Paola Lopez to translate English to Spanish or Spanish to English.

The group began with introductions and an icebreaker question, followed by the reading of a verbal consent form about the focus group. Following that, a presentation on the initial findings of the Next Event Analysis study was presented by Allon Kalisher, Jonathan Gellar, and Paola Lopez of Mathematica, and Katherine Guffey of DCS. Specifically, the group was shown:

- the most common types of allegations reported to the DCS hotline
- the most common allegations on investigated reports as differentiated by race
- the rates of substantiation outcomes of allegations by race (how often allegations are substantiated for each race as compared to White people)
- the unadjusted and adjusted rates of each outcome by race
- the "relative risk" of most common allegations, meaning how likely substantiation, case opening and removal happens for each allegation type

Information on the allegations most and least likely to result in child removal was included in the slide deck for this focus group, but due to time, this was not presented to the focus group participants.

Participants were asked to reflect on the information with the following questions:

- What are your thoughts on these results?
- How do you make sense of this?
- How do you think people will react when they see these numbers?

- What would you want people and policy-makers to know about these results?

Summary of questions, reflections and discussion from the South Tucson focus group

A participant noted that it was promising to hear about the study aims and what it could do for families.

The conversation demonstrated that participants did not have much knowledge about the DCS process, or who had the authority to remove children from their families.

A participant asked about the difference between calling the hotline versus calling DCS directly. DCS clarified that this data refers to any call made to DCS.

Participants asked about the definitions of allegations, such as "threatening living environment." It is suggested that definitions of each allegation are clearly stated.

Researchers noted that there is overlap between the racial demographic categories and the South Tucson group, as there are people of diverse races in South Tucson.

One participant voiced that she was analyzing the data and giving feedback as best she could. "Living conditions being a threat" could be due to cultural differences, as South Tucson has large number of immigrants. The participant theorized that because of the large immigrant population, this may lead to a lack of knowledge about available resources, as well as standards and norms. These immigrants may not know what is acceptable to DCS or other authority figures in terms of living conditions. They may also be residing in housing that is not well upkept by landlords, which could affect perception of families. They may also not be able to invest in the maintenance of their homes because they do not own them.

One participant noted, "After [DCS] takes our kids, they need to stop scaring the parents, and help them."

A participant, reflecting on the domestic violence data, said that the pandemic may be affecting rates of domestic violence. In the pandemic, many people were faced with different living conditions and had tensions in household rise, which could have affected the number of reports. The participant also said that South Tucson has a lack of police and safety personnel. Police may investigate drug use, but not violence.

Several participants shared examples of police being called but never arriving to help. Police may not be responding to the safety needs of children in families.

A participant mentioned that individuals in South Tucson don't know who to call for what, and so they end up calling everyone/anyone for help. There is a perception that police do not respond to this community. There is a lot of poverty. Violence is higher because police doesn't answer.

The group had challenges understanding what the term "substantiation" meant.

When asked for their perspective on the large difference in substantiation, case opening, and removal, participants noted:

- Lack of investment in libraries and education in South Tucson. "There is more investment in prisons than in education for south Tucson."
- Financial difficulties. Lack of resources. No access for medical or behavioral health. There are no resources or funding for items residents have been asking for. All investments seem to go to North Tucson. We all pay taxes however get no funding or investments put back into our community.
- Parks have been built, but they are inaccessible, you must pay to get in, which means children and others don't have places for recreational activity thus, turning to more dangerous or illegal activities, creating a cycle.

- People travel to North Tucson or Sahuarita for resources and recreation. "More funds are invested in North Tucson. No one wants to build libraries in South Tucson. It's like we're throwing them away. There is less money in South Tucson, and less resources, and it's causing damage."
- There is a lot of drug use and selling in South Tucson.
- To compensate for the poor educational institutions in South Tucson, participants note that many residents have to send children to schools outside South Tucson (more East). However, even those schools need more resources and funding.
- Residents know that there is an opportunity for funding and resources, but they have not been given attention by politicians or the state. They feel as though it is primarily being given to North Tucson and other areas, ignoring them and their struggles.
- It is evident that politicians want to keep the rich, rich and keep the poor, poor.
- A participant called for action to raise voices and advocate for investments in South Tucson.

A participant is interested in knowing what kinds of people make calls to the DCS hotline about children in South Tucson. Are they from inside the community? Are they teachers in the schools in North Tucson that have students from South Tucson? Are they police?

A participant shared that South Tucson residents may not know what the standards are for abuse and neglect. There is the desire for clear information on what is acceptable and what is not when it comes to discipline and maltreatment.

It is recommended that residents of South Tucson are given opportunities to be educated on the role of police, the role of DCS, and what is acceptable in American society.

A participant theorized that South Tucson may not be as close knit as other communities. In other communities, people may avoid calling DCS/authorities on each other to protect one another from intervention.

It is recommended that leaders work to better understand the culture of South Tucson. "When you understand what they're going through, you understand that they aren't wrong, they just don't know better."

Residents of South Tucson lack resources, funding and knowledge. They are scared to ask for help.

It was strongly recommended by the community leader co-facilitator that the study report is presented as clearly and simply as possible. If community presentations about the study are conducted, they should be done at a third grade reading level, because much of the data shared during the focus group was perceived to be too complex to be understood.

"If they want real responses from South Tucson, we need to simplify our communication of these stats. It's the community's stories and understanding of the data that will generate real solutions. If you want to include real community voices and get solutions, we need to simplify the data."

Appendix B

Technical details and additional data output

This page has been left blank for double-sided copying.

Data preparation

Case characteristics

Table B.1. Case characteristic variables used in the random forest models

Domain	Characteristic variables description: Screening in	Characteristic variables descriptions: All other outcomes
Demographics	• Number of children and adults on report • Ages of children and adults (mean, min, max) • Gender of children and adults (% male) • Race of children and adults (% White, % Hispanic, % African American, % American Indian, and % Asian/Pacific Islander) • Race of the primary caretaker	• Number of perpetrators and victims on intake • Age of perpetrator and victim • Gender of perpetrator and victim • Race of perpetrator and victim • Whether the perpetrator and victim speak English • Race of the primary caretaker
Prior history	• Prior child removals (count and total duration) • Number of prior investigations that included adults as perpetrator (overall, and separately for investigations that involved substance abuse and domestic violence) • Number of prior investigations that included children as victims (overall, and separately the number that included allegations of substance abuse and domestic violence) • Number of prior open cases involving adults as perpetrators, or (separately) children as victims	• Prior victim removals (count and duration) • Number of prior investigations involving the perpetrator (overall, and separately for investigations that involved substance abuse and domestic violence) • Number of prior investigations involving the victim (overall, and separately for investigations that involved substance abuse and domestic violence) • Number of prior open cases involving the perpetrator and (separately) victim
Allegations	N/A	• Number of allegations on the intake involving the perpetrator/victim combination • Flags for each of the 54 allegation types that are tracked in the Guardian system • Flag for whether there is any allegation of criminal conduct
Geography	• County • Rural Arizona (yes/no) • South Tucson (yes/no) • Area deprivation index (zip code level)	
Source	• Source type category • Mandated reporter (yes/no) • Method of contact	
Other	• High profile (yes/no) • Potential ICWA (yes/no) • Provider type (foster home, group home, or kinship) • Source safety concerns (yes flag for any listed)	

Missing data

There are two kinds of missing data in our analysis: missing outcome data, and missing predictor data. Outcome data can be missing for three of the five outcomes: substantiation, removal, and the composite

score. These outcomes are most commonly missing when an investigation is still ongoing at the time we received data (03/04/2023). Screening in data are never missing, because they are based on a decision made at the time of the call and we have a record of each decision. Likewise, removal is data are never missing because we define the outcome to be removal within three months of intake, and more than three months elapsed between the end of our study period and the time of data collection.

There are a variety of reasons why data on case characteristics would be missing. For demographic variables, data can be missing when less information is known about all the individuals (perpetrators and victims) involved in a hotline communication; this is most common for hotline communications that were not screened in and therefore not investigated. For other case characteristics, it is not always clear why data are missing. To handle missing data in our models, we use an approach called missingness in attributes, or MIA (Twala et. al. 2008). We implement MIA in two steps:

1. For any variable with missing values, we create a flag that indicates that the variable is missing. For binary or categorical variables, this is equivalent to adding a "missing" level to the original variable.
2. For continuous variables, we replace missing values with the median value of the predictor

MIA is particularly useful in capturing informative missingness. This approach has the advantage of not having to exclude observations with missing predictor data, while still being able to handle observations for which data is not missing at random.

Statistical methods

Calculating relative risk

In two parts of our analysis, we express the effects of individual predictors as relative risks: (1) when assessing differences across demographic groups (race and geography), and (2) to express the effects of individual allegation types. In this section, we describe the calculations in detail.

1. Relative risk for demographic groups

The relative risk for each race/ethnicity is a comparison of the rates of outcomes for observations involving families of that race or ethnicity to rates in the reference group (White families). Before doing this comparison, we first had to classify each observation into a single racial/ethnic category. Because each observation involves at least two individuals, and those individuals may not be of the same race/ethnicity, this process was not immediately straightforward. To be consistent with prior work conducted by DCS, we decided to classify observations according to the race of the primary caretaker at the time of intake. If this information was not available, we used the race of the perpetrator, and if the perpetrator's race was missing, we used the race of the victim.

The **unadjusted relative risk** for each racial or ethnic group is the rate (or mean for the composite measure) of the outcome for observations classified into the corresponding racial/ethnic group, divided by that of observations classified into the reference group (White race). However, as described in the report, rates of outcomes may differ across racial/ethnic groups for reasons other than race/ethnicity. To control for other factors that could influence rates, we calculate the **adjusted relative risk** for each racial/ethnic group as a marginal effect, also known as partial dependence (Hastie et. al. 2019).

Specifically, we illustrate the procedure when calculating the relative risk when race is African American, relative to White:

1. Subset the data set to observations classified as being among African American families
2. Predict the outcome among these observations and average the resulting predictions. This is the marginal rate (or mean for the composite measure) outcome among African American observations.
3. Change all race variables in the data subset to represent "White." For the screening-in outcome, this means that we set the child race to 100 percent White (and 0 percent other races) and the adult race to 100 percent White (and 0 percent other races) and set the caretaker race to "White." For all other outcomes, we set the race for the perpetrator, victim, and caretaker all to "White."
4. Predict the outcome among the modified observations and average the resulting predictions. This is the marginal rate (or mean for the composite measure) outcome, among African American observations, if the observations had instead been made among White families.
5. Divide the marginal rate from Step 2 by the marginal rate from Step 4. This is the adjusted relative risk for African Americans.
6. In our tables, we subtract 1 from the relative risks, so that positive numbers indicate an increase in the rate of the outcome (relative to White families), and negative numbers indicate a decrease in the rate of the outcome (relative to White families).

Note that we define the marginal effect to be among the observations originally classified as belonging to the particular race. This provides a more targeted interpretation of the adjusted relative risk: it is the expected rate of the outcome among observations classified according to the particular race, relative to the expected rate of the outcome among the same observations if those observations involved a White family. This approach is analogous to estimating an average treatment effect on the treated (ATT) in the causal inference literature.

The relative risks for the two geographic subgroups are calculated similarly, but each uses a different reference group. For South Tucson, the reference group is "not South Tucson." In other words, we estimate the rate of the outcome among observations in South Tucson, relative to the same observations had they occurred outside of South Tucson. For rural Arizona, we similarly define a reference group that is "not rural Arizona." For both case characteristics, Step 3 of the process above involves changing the corresponding flag in the data from "1" to "0."

2. Relative risk for allegation types

The relative risk calculation for allegation types is similar to the calculation for the geographic subgroups (South Tucson and Rural Arizona) described above. For each allegation type, we define a reference group as "not having that allegation type" for the observation. The procedure for calculating these relative risks are as follows:

1. Subset the data to all observations that include a particular allegation type.
2. Predict the outcome for these observations and average the predictions. This is the marginal rate of the outcome, for the corresponding allegation type.

3. Change the flag in the data subset that corresponds to that allegation type from "1" to "0".
4. Predict the outcome among the modified observations and average the resulting predictions. This is the marginal rate (or mean for the composite measure) outcome, among observations of the corresponding allegation type, had that allegation type not appeared on the observation.
5. Divide the marginal rate from Step 2 by the marginal rate from Step 4. This is the adjusted relative risk for the corresponding allegation type.

Note that, because each observation can have more than one allegation type, that the subsets created in Step 1 are not mutually exclusive across allegation types.

Race-blind prediction

Recent research has highlighted how subjective decision-making within the child welfare system can lead to misleading data, resulting in predictive models that produce biased results (Pope and Sydnor 2011). The most common example cited in the literature is that of racial or ethnic bias. Although our goals for this analysis are not to develop a predictive modeling tool for real-time decision making, we have already seen (in Section III.C) that there seem to be differences across racial/demographic groups that cannot be explained by other case characteristics, and we would like to control for these differences, so they do not affect our inferences. We use a procedure known as race-blind prediction.

Race-blind prediction guarantees that, for two observations that share all the same case characteristics other than race, the prediction will be the same. This property could be obtained from a model that excludes the race variables altogether, but if one chooses this approach, it would likely result in omitted variable bias that differentially affects racial or ethnic groups, because many of the variables included in the model are correlated with race. Variables that could potentially be correlated with race include geographic variables (county as well as indicators for South Tucson and rural Arizona), zip code-level socioeconomic status (as measured by the area deprivation index), and the presence of certain allegation types. Our procedure ensures that race does not directly impact predictions, while also removing the potential for indirect effects through omitted variable bias.

We illustrate the procedure for calculating race-blind predictions for observations at the report/perpetrator/victim level. These observations contain three race variables: perpetrator race, victim race, and caretaker race. Our procedure is as follows:

1. Fit the random forest model with all case characteristics, including the three race/ethnicity variables.
2. In the data, drop the three race variables. Then expand the dataset, so that for every observation in the original data set, we include a record for all possible combinations of the three race variables. Because each race variable has six possible values (White, African American, Hispanic, American Indian, Asian/Pacific Islander, and unknown), the new modified dataset will have $6^3 = 216$ times as many rows as the original data set.
3. Predict the outcome for each of the rows in the new dataset.
4. For every one of the original observation IDs, take a weighted average of the 216 predictions. The weights are determined by the relative frequency of the corresponding combination of race variables in the raw data. We refer to this weighted average as the **race-blind prediction.**

In practice, we did not predict the model for all 216 combinations of the three race/ethnicity variables because most of these combinations are rare and thus would be assigned very small weight (and would therefore not affect results). Instead, we considered the following combinations: (a) all combinations where the three race variables were the same, and (b) any combination that occurred in the data at least 1.5 percent of the time. The combinations of the variables that we use, along with their weights, are listed in Table B.2.

Table B.2. Combinations of race/ethnicity used for the race/ethnicity-blind prediction

Caretaker race	Perpetrator race	Victim race	Weight
White	White	White	0.225
Unknown	Unknown	Unknown	0.223
Hispanic	Hispanic	Hispanic	0.171
African American	African American	African American	0.080
White	White	Unknown	0.066
Hispanic	Hispanic	Unknown	0.050
Native American	Native American	Native American	0.034
White	White	Hispanic	0.030
Unknown	White	White	0.027
White	Unknown	White	0.025
Unknown	Unknown	White	0.025
African American	African American	Unknown	0.020
Unknown	Hispanic	Hispanic	0.020
Asian/Pacific Islander	Asian/Pacific Islander	Asian/Pacific Islander	0.005

Model implementation

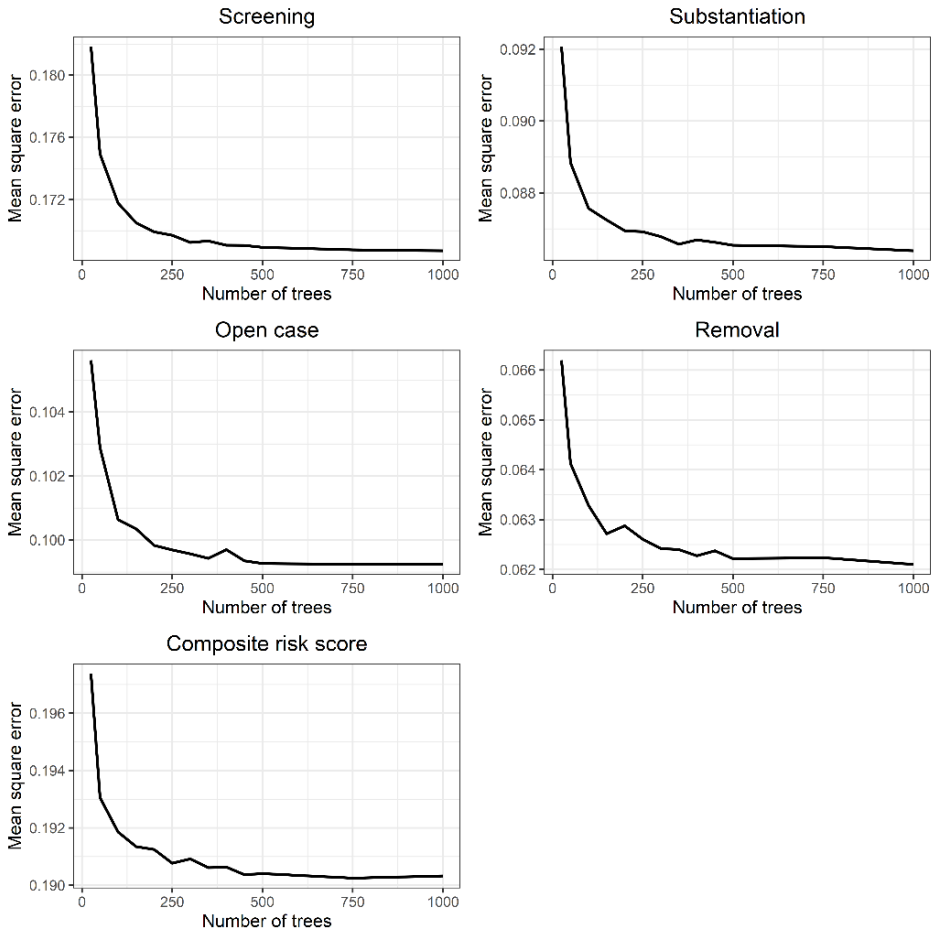
Random forest models

We fit a random forest model to five outcomes: four binary outcomes (screening in, substantiation, case opening, and removal), and one continuous outcome (composite measure). We treat the composite measure outcome as continuous even though it only takes integer values, for two reasons. First, treating the outcome as continuous preserves the ordering of the different levels of the outcome (for example, a 5 indicates higher risk of harm than a 4). And second, because we are treating the ordinal scale as a level of risk, predictions between integer values are interpretable – they indicate a level of risk that is between the two adjacent integer values. Note that because the random forest is nonparametric, the model does not assume that the risk scale is linear. In other words, we do not assume that the difference in risk between a 1 and a 2 is the same as between a 2 and a 3, or between a 4 and 5.

To fit the models, we use the `ranger` package in R (Wright and Ziegler, 2017). ranger is a fast implementation of random forests that is suitable for both classification (binary outcomes) and regression (continuous outcomes). We fit the model using 1,000 trees and used the default settings for all other model parameters. This includes an "mtry" value (the number of parameters to consider at each node split) equal to the square root of the number of parameters, using the Gini index to choose splitting rules for binary outcomes, and minimizing response variances to choose splitting rules for continuous

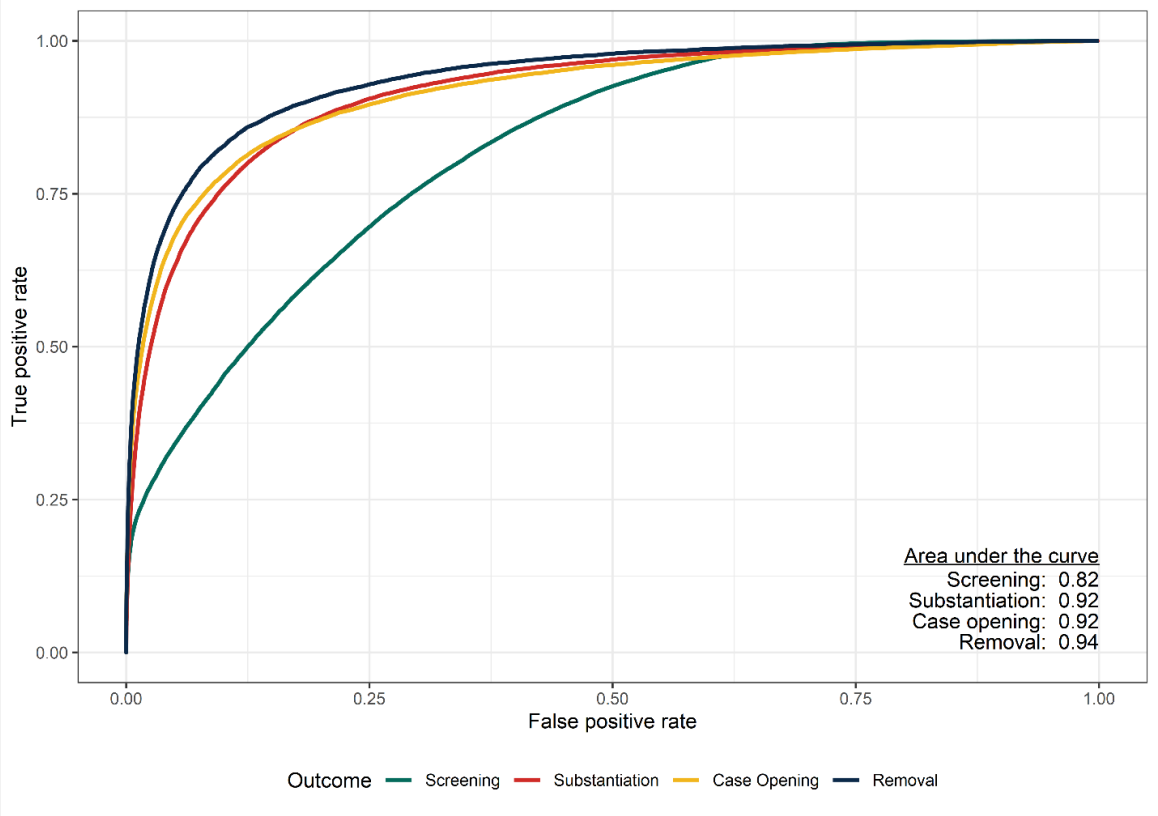
outcomes. When using a random forest model for prediction, tuning of the parameters is required to get the absolute best performance possible from the model. However, since we are using the model for inference, when the predictions of the default models were consistently high (AUC > 0.8 for screening in and AUC > 0.9 for the others), due to time constraints we chose not to further tune the models at this time in favor of deepening the analysis elsewhere. Tuning the models might result in changes to the variable importance values (and thus the ordering of variables), but due to the lack of headroom to improve the model, we anticipate that the overall story would remain similar to that of the models using default parameters. We fit one version of each model and compare the prediction error as a function of the number of trees to ensure the model is fully converged. Figure B.1 shows these results, which indicate that the model appears to be sufficiently converged using 1,000 trees.

Figure B.1. Prediction error vs. the number of trees for each of the five random forest models.



Despite prediction not being the focus of our analysis, we show the receiver operating characteristic (ROC) curves for each of the four binary outcomes in Figure B.2. The ROC curve measures the ability of the model to distinguish between cases with each outcome in a binary model, for example, to distinguish between cases that were and were not substantiated. A common summary measure of the ROC curve is the area under the curve (AUC); models with perfect ability to distinguish outcomes will have an AUC of 1. AUC values above 0.7 traditionally indicate good predictive performance; above 0.8 strong predictive performance; and above 0.9 excellent predictive performance (Statology, 2021). AUC will increase when models that have highly predictive variables and are therefore better able to distinguish if an outcome has occurred. All 4 binary models have AUC values above .80, indicating that they can distinguish between an outcome occurring and not occurring.

Figure B.2. Receiver operating characteristic curves for the four models with binary outcomes



Finally, we show plots of the 30 most important variables in Figures B.3 through B.7. We measured variable importance through variable permutation. Permutation importance is based on the decrease in predictive ability of the model when the value of the variable is randomly permuted. If the variable is important, permuting its values will decrease the predictive power of the model and the more important the variable, the larger the effect on the predictive power of the model it's permutation will have. The following plots show the 30 most important predictors from each of the five models. Because the x-axis is only interpretable in a relative sense, we scaled the axis to range from 0 to 100, which is common practice.

Figure B.3. Variable importance plot for screening

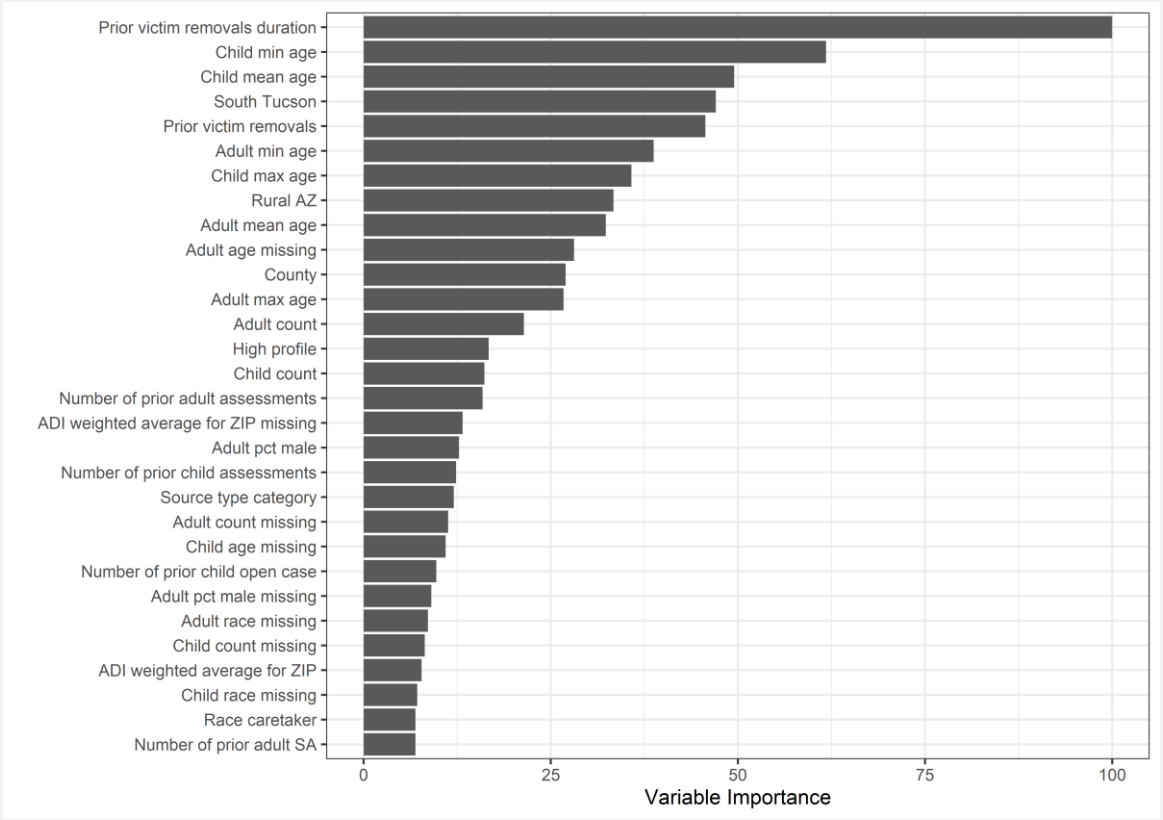


Figure B.4. Variable importance plot for substantiation

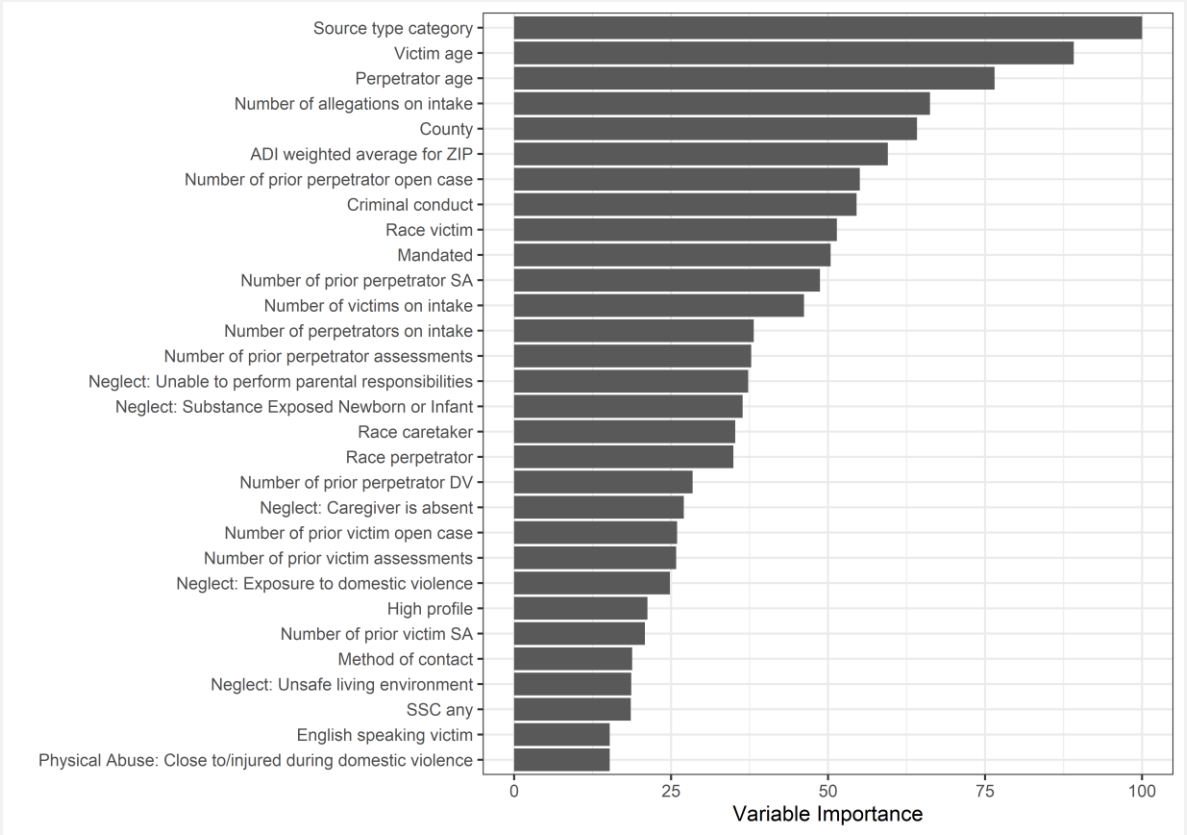


Figure B.5. Variable importance plot for case opening

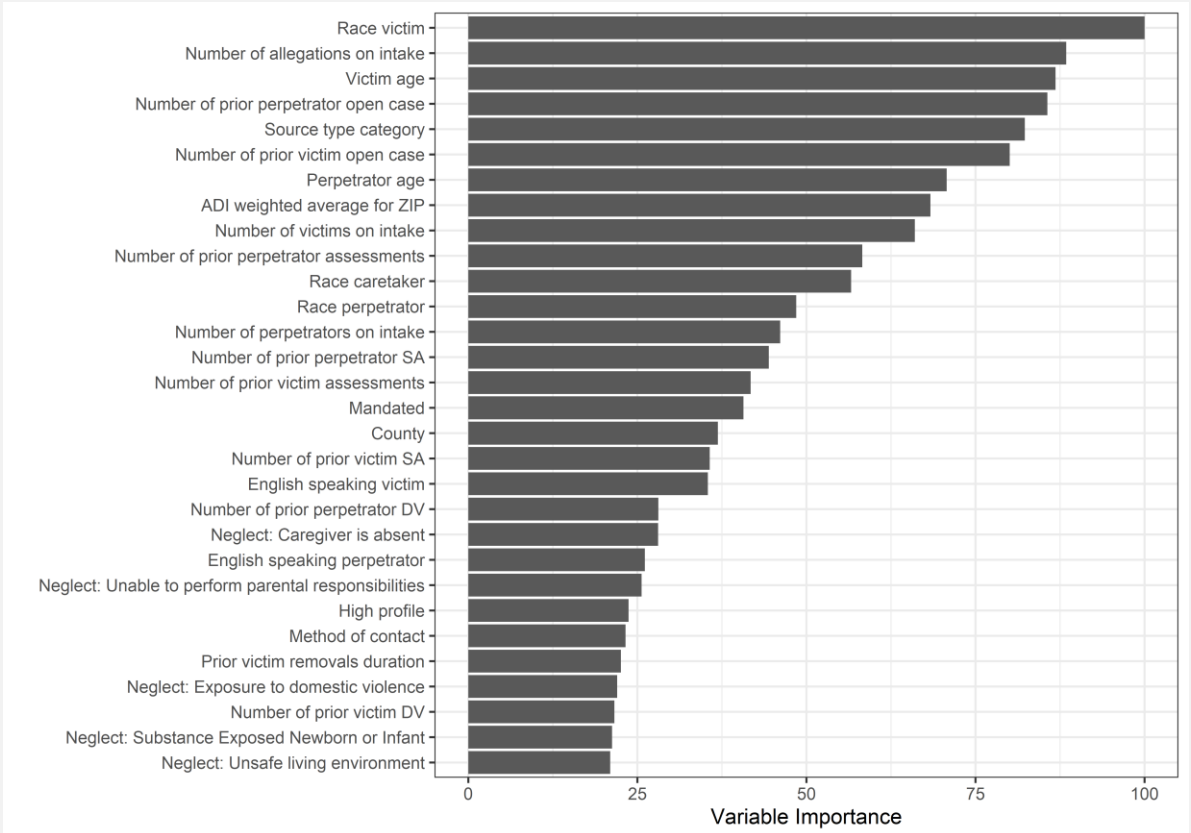


Figure B.6. Variable importance plot for removal

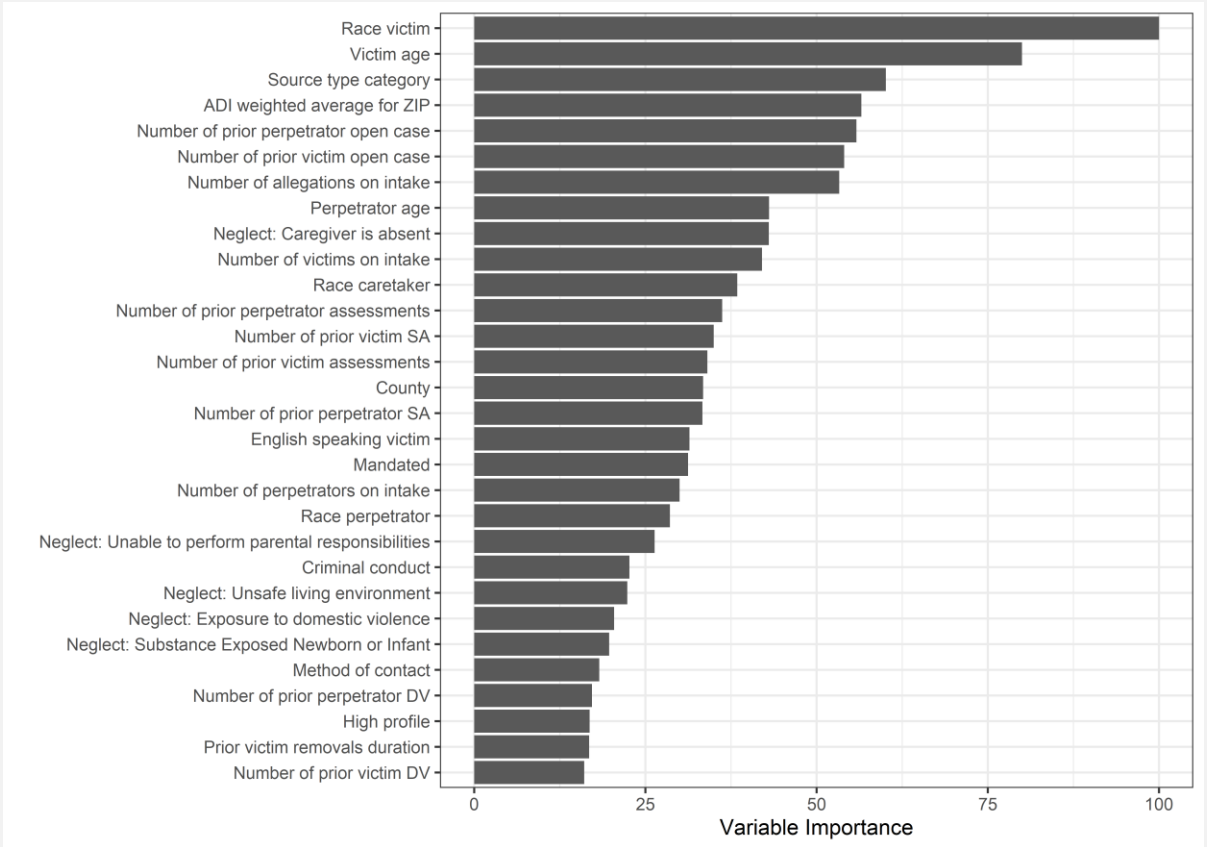
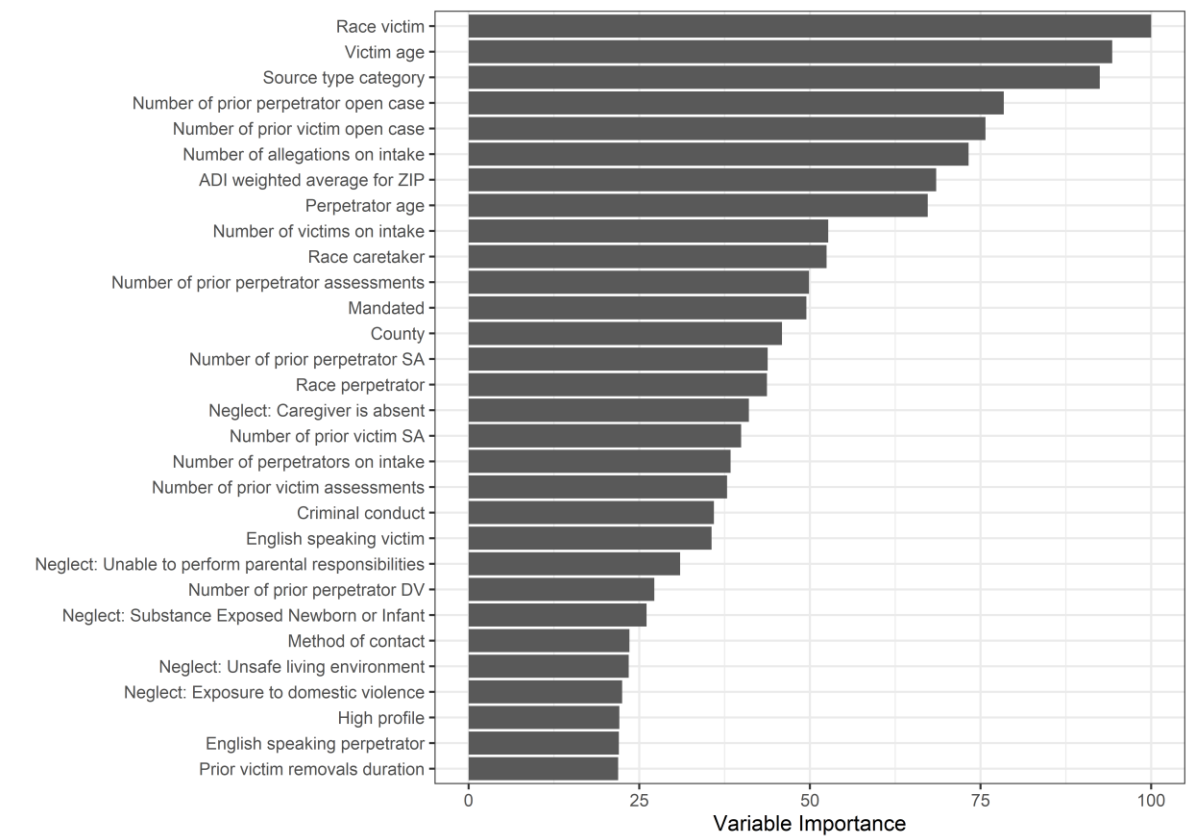


Figure B.7. Variable importance plot for composite measure

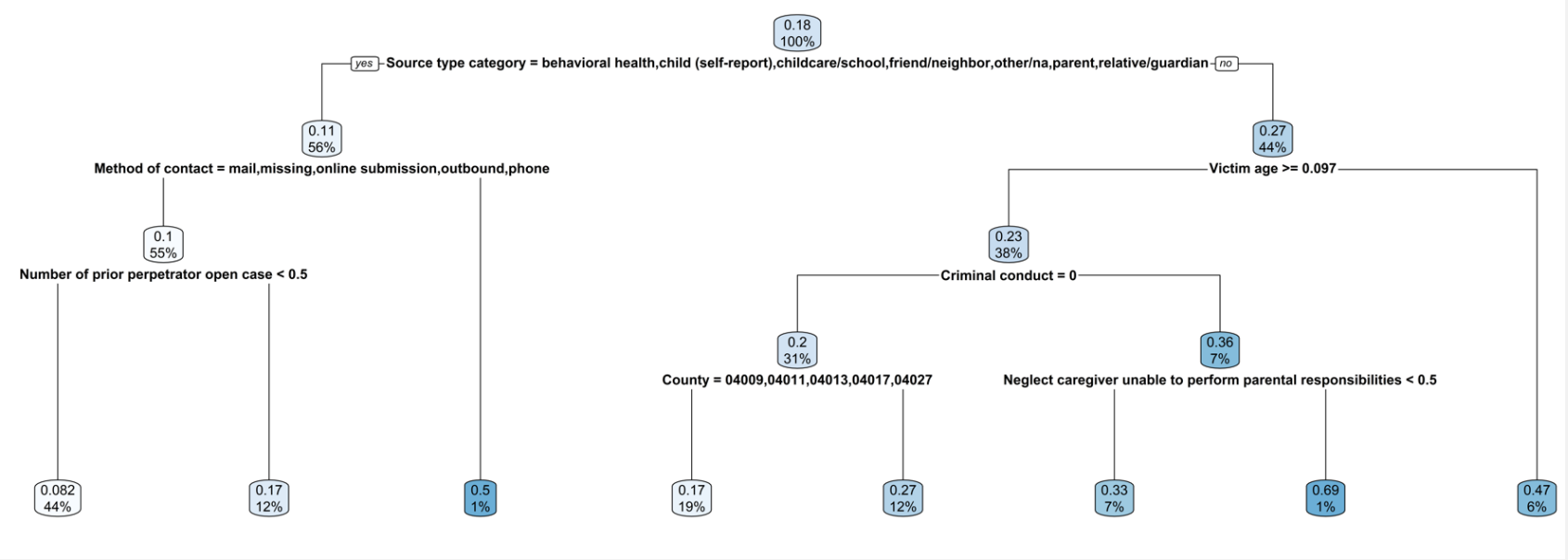


Decision tree models

To identify combinations of case characteristics that are associated with low rates of outcomes, we fit a decision tree to each of the outcomes that include allegations. Those outcomes are substantiation, case opening, removal, and the composite outcome. As discussed in the report, rather than use the raw outcome, we use the race-blind prediction of each outcome based on the random forest model (see Appendix B, Model Implementation for a description of these calculations). In Figures B.8, B.9, and B.10, the top number is the outcome rate for each node. For example, ".18" in the first node of Figure B.8. means 18 percent of observations meeting the source type category conditions were substantiated, using a race-blind prediction.

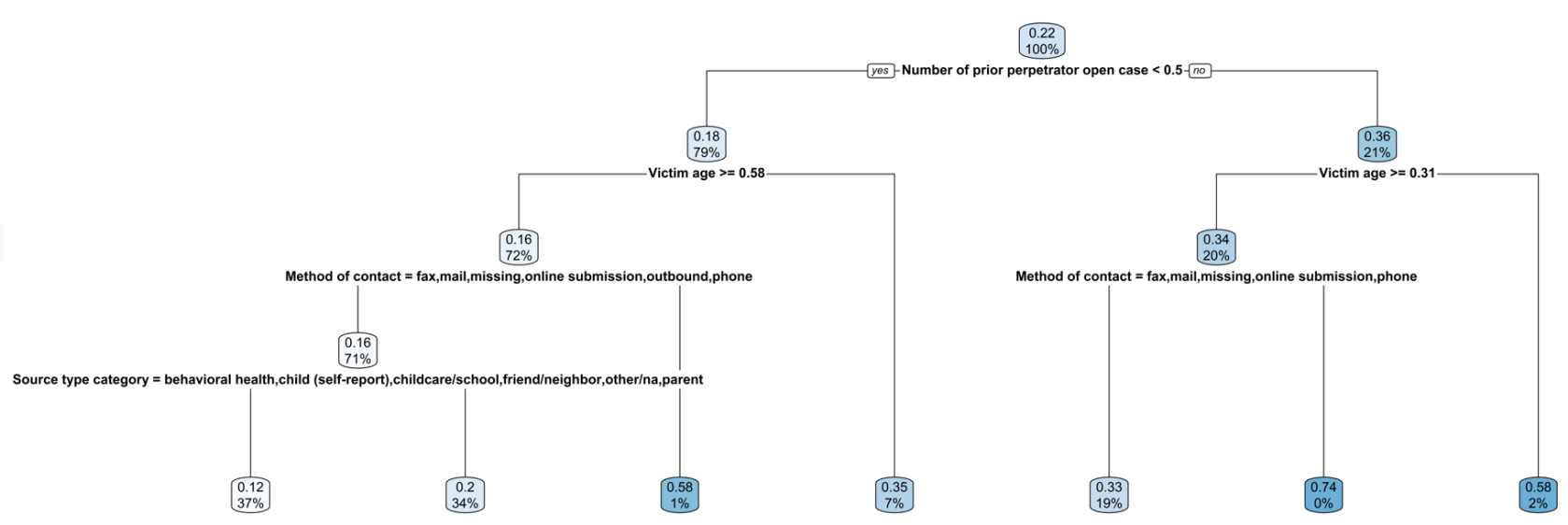
We implement the decision trees using the `rpart` package in R, which is similar to the algorithm described in the original paper on classification and regression trees by Breiman et. al. (1984). As noted in the report, we do not tune this model as our goals are to make inferences rather than optimize predictions and the default model has satisfactory prediction accuracy.

Figure B.8. Pruned version of the decision tree for substantiation, using race-blind predictions as the outcome



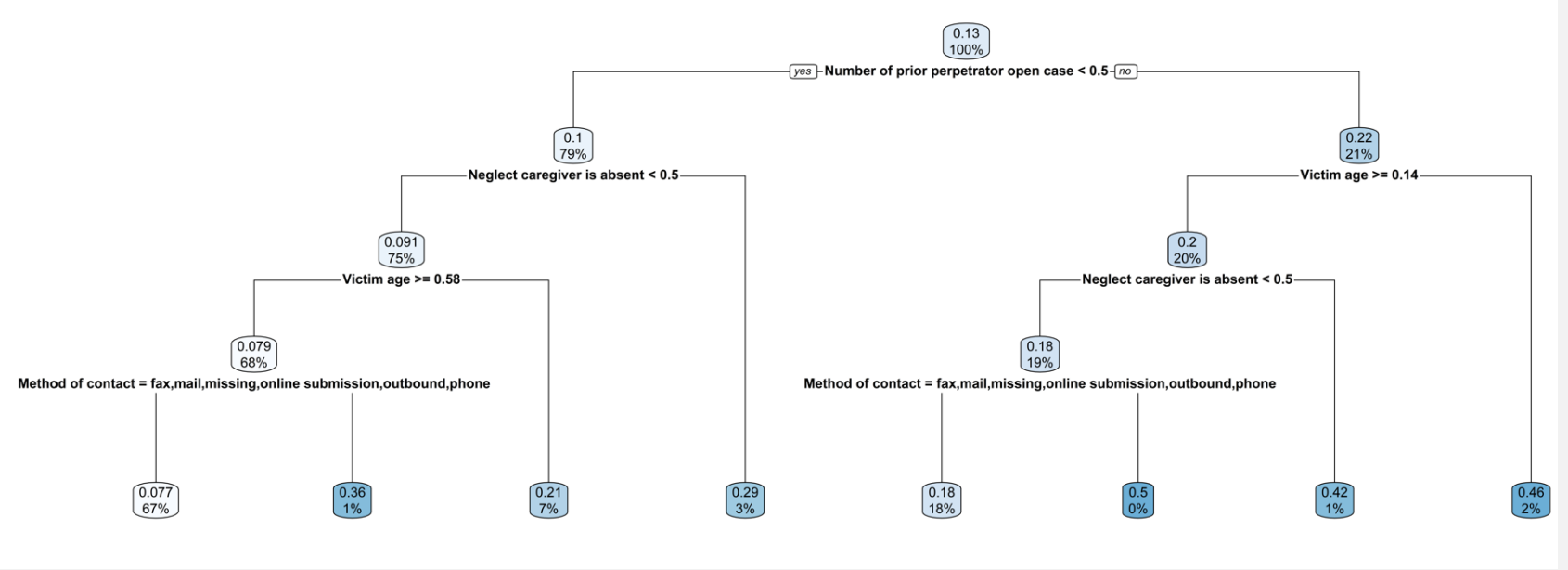
Source:

Figure B.9. Pruned version of the decision tree for case opening, using race-blind predictions as the outcome



Source:

Figure B.10. Pruned version of the decision tree for removal, using race-blind predictions as the outcome



Source:

Descriptive results

Table B.3 shows the distribution of all allegations at the state level, labeled with the full-text descriptions from Arizona DCS's Guardian data system. Appendix C is DCS's Hotline Report Decision Tool, which displays the full-text questions associated with these allegations.

Table B.3. List of allegations (full text), and the overall proportion of observations (report/perpetrator/victims) that contain each allegation

	Type	Allegation	Count (N=141,483)	Percent
1	Neglect	Significant incident or repeated **exposure to domestic violence**	30,024	21.2%
2	Neglect	Living environment is a threat to child's safety **(unsafe living environment)**	24,739	17.5%
3	Neglect	**Caregiver** unwillingly or **unable to meet child's needs**	16,111	11.4%
4	Physical Abuse	**Caregiver is** physically or verbally **imposing/threatening**	10,920	7.7%
5	Neglect	**Caregiver unable to perform parental responsibilities**	10,470	7.4%
6	Neglect	Child in **close proximity** and/or injured during incident of **domestic violence**	9,469	6.7%
7	Physical Abuse	Other **Bodily Injuries (Bruises**, abrasions, swelling, welts, **etc**.)	9,461	6.7%
8	Neglect	**Substance Exposed Newborn** or **Infant**	8,212	5.8%
9	Physical Abuse	Child in close proximity and/or **injured during** incident of **domestic violence**	8,150	5.8%
10	Neglect	**Unrealistic** expectations, predominately **negative** terms or distorted **view of child** by caregiver	8,006	5.7%
11	Neglect	Caregiver is absent	6,782	4.8%
12	Neglect	Caregiver allows known sexual predator access to the child	5,055	3.6%
13	Physical Abuse	Caregiver failed to protect/placed in dangerous situation	4,344	3.1%
14	Neglect	Unwilling or unable to meet child's needs for medical health care	4,149	2.9%
15	Neglect	Unable/unwilling to control child's behavior that threatens serious or severe harm to self or others	3,801	2.7%
16	Sexual Abuse	Evidence or disclosure of sexual abuse	3,326	2.4%
17	Physical Abuse	Face or head injury	2,390	1.7%
18	Neglect	Injuries due to neglect or failure to supervise	2,142	1.5%
19	Emotional Abuse	Incident or pattern of behavior directed toward child/interferes with normal functioning (berating, name calling, targeting, rejection)	2,019	1.4%
20	Neglect	Caregiver allows/provides substances that may cause/caused harm (non-prescribed medication, inappropriate dosages of over the counter or prescribed medications, adult medication, alcohol, illegal drugs, synthetic drugs)	1,702	1.2%

	Type	Allegation	Count (N=141,483)	Percent
21	Physical Abuse	Caregiver allows/provides substances that may cause/caused harm (non-prescribed medication, inappropriate dosages of over the counter or prescribed medications, adult medication, alcohol, illegal drugs, synthetic drugs)	1,438	1.0%
22	Neglect	Caregiver allows/provides substances that may cause/caused harm (non-prescribed drugs, illegal drugs, adult drugs, alcohol)	1,262	0.9%
23	Physical Abuse	Multiple plane injuries	1,179	0.8%
24	Neglect	Child extremely fearful due to threat or present circumstance	1,130	0.8%
25	Neglect	Recklessly or deliberately exposed to sexually explicit materials or acts	978	0.7%
26	Physical Abuse	Caregiver allows/provides substances that may cause/caused harm (non-prescribed drugs, illegal drugs, adult drugs, alcohol)	941	0.7%
27	Physical Abuse	Caregiver restricted/confined child; used threat of harm/intimidation	883	0.6%
28	Physical Abuse	Injury inconsistent with explanation	840	0.6%
29	Physical Abuse	Unknown injuries but observed to be forcefully struck	834	0.6%
30	Physical Abuse	Child extremely fearful due to threat of abuse or neglect, home situation, present circumstance	622	0.4%
31	Physical Abuse	Fractures	538	0.4%
32	Emotional Abuse	Extremely fearful because of home situation, present circumstance or threat of abuse	421	0.3%
33	Physical Abuse	Burns	341	0.2%
34	Neglect	Death of a child/Suspicious Death	327	0.2%
35	Neglect	Caregiver permits a child to enter/remain in a structure/vehicle used to manufacture drugs	180	0.1%
36	Physical Abuse	Manufacturing of drugs	139	0.1%
37	Physical Abuse	Caregiver permits child to be in structure/vehicle used for manufacturing drugs	134	0.1%
38	Physical Abuse	Brain Injury	132	0.1%
39	Sexual Abuse	Persistent, highly sexualized behavior (predatory) outside of age-appropriate sexual exploration	110	0.1%
40	Neglect	Medical diagnosis of malnutrition/failure to thrive without previously diagnosed health condition	104	0.1%
41	Neglect	Manufacturing of Drugs	95	0.1%
42	Physical Abuse	Death of a child due to physical abuse/suspicious death	92	0.1%
43	Physical Abuse	Caregiver takes, entices or keeps a court ward from the lawful custody of DCS or the Department of Juvenile Corrections (including intentional failure, refusal, or impeding the child's return at end of visitation)	85	0.1%

	Type	Allegation	Count (N=141,483)	Percent
44	Sexual Abuse	Sex Trafficking	76	0.1%
45	Physical Abuse	Medical diagnosis of malnutrition/failure to thrive without previously diagnosed health condition	75	0.1%
46	Sexual Abuse	Pain in the genital or anal areas and there are indicators of sexual abuse	50	0.0%
47	Sexual Abuse	Child has a Sexually Transmitted Infection (STI) that may indicate sexual abuse	35	0.0%
48	Neglect	Caregiver failed to protect/placed in dangerous situation	8	0.0%
49	Physical Abuse	Death of a child/Suspicious Death	3	0.0%
50	Emotional Abuse	Child extremely fearful due to threat or present circumstance	1	0.0%
51	Neglect	Child extremely fearful due to threat of abuse or neglect, home situation, present circumstance	1	0.0%
52	Physical Abuse	Living environment is a threat to child's safety	1	0.0%
53	Physical Abuse	Significant incident or repeated exposure to domestic violence	1	0.0%

Appendix C

DCS Hotline Report Decision Tool

This page has been left blank for double-sided copying.

DCS-1968
(09/21)

ARIZONA DEPARTMENT OF CHILD SAFETY

HOTLINE REPORT DECISION TOOL

Statutory Authority

Does the information collected justify accepting the report based on Arizona law?

1. Is the alleged victim presently under the age of 18? ☐ Yes ☐ No
   a. Are there other children in the home under the age of 18? ☐ Yes ☐ No ☐ Insufficient Information
2. Is the alleged victim a resident or present in Arizona? ☐ Yes ☐ No
3. Is the alleged perpetrator a caregiver? ☐ Yes ☐ No
   (Caregiver means a parent, guardian, custodian of the child or adult member in the child's household)
4. Can the location or identity of the child, family or perpetrator be reasonably ascertained? ☐ Yes ☐ No

Types of Abuse and Neglect

Physical Abuse ***~Please note that bolded allegations with an "*" need a corresponding Neglect allegation.***

A.R.S. § 8-201: "Abuse" means the infliction or allowing of physical injury, impairment of bodily function or disfigurement or the infliction of or allowing another person to cause serious emotional damage as evidenced by severe anxiety, depression, withdrawal or untoward aggressive behavior and which emotional damage is diagnosed by a medical doctor or psychologist and is caused by the acts or omissions of an individual who has the care, custody and control of a child. Abuse includes :...(b) Physical injury that results from permitting a child to enter or remain in any structure or vehicle in which volatile, toxic or flammable chemicals are found or equipment is possessed by any person for the purpose of manufacturing a dangerous drug.(c) Unreasonable confinement of a child.

☐ 1. Death of a child due to physical abuse or suspicious death (excluding situations where a death certificate will be signed, vehicular accidents, and suicides unless committed by a firearm)

☐ **2. *Medical diagnosis of malnutrition / failure to thrive without previously diagnosed health condition**

☐ 3. Injuries including, but not limited to:
- — Fractures
- — Brain injury (e.g. subdural hematoma, "shaken baby syndrome")
- — Multiple plane injuries
- — Burns (e.g. immersion burns, cigarette burns, unexplained burns)
- — Face or head injury (e.g. bruises, cuts, abrasions, swelling)
- — Physical injury resulting from permitting a child to enter or remain in a structure or a vehicle that is used for the purposes of manufacturing dangerous drugs
- — Other bodily injuries (e.g. scratches, bruises, cuts, abrasions, swelling, welts)

☐ 4. Injury inconsistent with explanation

☐ 5. Caregiver placed child in a dangerous situation or failed to protect the child from abuse

☐ **6. *Caregiver permits a child to enter /remain in a structure or vehicle that's used for the purposes of manufacturing drugs**

☐ **7. *Manufacturing of drugs** (includes Marijuana grow if unsecured and accessible to a child, or if the number of plants exceed six per adult or more than twelve in a household).

☐ 8. Unknown injuries to a child observed or reported to be forcefully struck in the face, head, neck genitalia or abdomen which could likely cause an injury

☐ 9. Child injured during an incident of domestic violence

☐ **10. *Child was in close proximity to an incident of domestic violence and could have been injured (this includes being held by one of the adults during the incident)**

☐ 11. Caregiver is described as physically or verbally imposing and threatening that may include, but not limited to:
- — Brandishing weapons
- — Throwing objects at the child that may cause harm
- — Behaving in attacking or aggressive way

☐ 12. Caregiver restricted or confined child to an enclosed area and/or used a threat of harm or intimidation to force a child to remain in a location or position which may include, but not limited to:
- — Binding a child's arms or legs together
- — Binding a child to an object
- — Locking a child in a cage or confined space

DCS-1968
(09/21)
Page 2

ARIZONA DEPARTMENT OF CHILD SAFETY
HOTLINE REPORT DECISION TOOL

Physical Abuse -*continued*

☐ 13. Child is extremely fearful because of their home situation, present circumstance or because of a threat of additional abuse or neglect (this does not refer to fear of legal disciplinary practice or generalized fear)

☐ **14. *Caregiver knowingly allows or provides the child substance(s) that caused or may cause the child harm including, but not limited to:**
- — Non-prescribed medication that caused or may cause harm
- — Inappropriate dosages of over the counter or prescribed medication that caused or may cause harm
- — Adult medication that caused or may cause harm
- — Alcohol that caused or may cause harm
- — Illegal drugs (as defined in §13-3401)
- — Synthetic drugs

☐ 15. Caregiver takes, entices or keeps a court ward from the lawful custody of the department of child safety or the department of juvenile corrections, including the intentional failure, refusal, or impeding of the child's return at the end of visitation. *Excludes tribal wards.

Sexual Abuse

A.R.S. § 8-201: Inflicting or allowing sexual abuse, sexual conduct with a minor, sexual assault, molestation of a child, commercial sexual exploitation of a minor, sexual exploitation of a minor, incest or child prostitution

☐ 16. Evidence or disclosure of sexual abuse including:
- — Sexual conduct
- — Sexual assault
- — Child molestation
- — Incest
- — Sex trafficking
- — Creating child pornography
- — Child seduction or grooming

☐ 17. Indicators consistent with sexual abuse without disclosure including, not limited to:
- — Child has a Sexual Transmitted Infection (STI) that may be an indicator of sexual abuse
- — Child has complained of pain in the genital or anal areas and there are other indicators of sexual abuse
- — Persistent, highly sexualized behavior in a child under the age of 5 chronologically or developmentally that is grossly age inappropriate and there is reason to believe that the most likely manner in which this was learned is having been sexually abused
- — Child displays highly sexualized behavior outside of age appropriate sexual exploration and the behavior is predatory in nature and not an isolated incident and there is reason to believe that the most likely manner in which this was learned is having been sexually abused

Neglect ~*Please note that bolded allegations with an "*" need a corresponding Physical Abuse allegation.*

Neglect concerns involving congregate care facilities shall be screened in as a licensing issue.

A.R.S. § 8-201:"Neglect" or "neglected" means: (a) the inability or unwillingness of a parent, guardian or custodian of a child to provide that child with supervision, food, clothing, shelter or medical care if that inability or unwillingness causes unreasonable risk of harm to the child's health or welfare, except if the inability of a parent, guardian or custodian to provide services to meet the needs of a child with a disability or chronic illness is solely the result of the unavailability of reasonable services (b) permitting a child to enter or remain in any structure or vehicle in which volatile, toxic or flammable chemicals are found or equipment is possessed by any person for the purposes of manufacturing a dangerous drug (c) a determination by a health professional that a newborn infant was exposed prenatally to a drug or substance and that this exposure was not the result of a medical treatment administered to the mother or the newborn infant by a health professional…the determination by the health professional shall be based on one or more of the following: (i) clinical indicators in the prenatal period including maternal and newborn presentation.; (ii) history of substance use or abuse.; (iii) medical history.; (iv) results of a toxicology or other laboratory test on the mother or the newborn infant.; (d) diagnosis by a health professional of an infant under one year of age with clinical findings consistent with fetal alcohol syndrome or fetal alcohol effects.; (e) deliberate exposure of a child by a parent, guardian or custodian to sexual conduct or to sexual contact, oral sexual contact or sexual intercourse, bestiality or explicit sexual materials (f) any of the following acts committed by the child's parent, guardian or custodian with reckless disregard as to whether the child is physically present: (i) sexual contact (ii) oral sexual contact (iii) sexual intercourse (iv) bestiality.

A.R.S. § 8-201.01; B. A parent may not be considered as having abused, neglected or abandoned or charged with abuse, neglect or abandonment of a biological, foster or adoptive child solely for seeking inpatient treatment or an out-of-home placement if the child's behavioral health needs pose a risk to the safety and welfare of the family. C. A parent may not be considered as having abused or neglected or charged with abuse or neglect of a child solely for bringing into the home a biological, foster or adoptive child whose behavioral health needs pose a risk to the safety and welfare of the family.

Neglect - *General*

☐ 18. Death of a child due to neglect or suspicious death (excluding situations where a death certificate will be signed, vehicular accidents, and suicides unless committed by a firearm)

DCS-1968
(09/21)
Page 3

ARIZONA DEPARTMENT OF CHILD SAFETY
HOTLINE REPORT DECISION TOOL

Neglect - *continued*

Neglect - *General continued*

☐ **19. *Caregiver knowingly allows or provides the child substance(s) that caused or may cause the child harm including, but not limited to:**
- — Non-prescribed medication that caused or may cause harm
- — Inappropriate dosages of over the counter or prescribed medication that caused or may cause harm
- — Adult medication that caused or may cause harm
- — Alcohol that caused or may cause harm
- — Illegal drugs (as defined in §13-3401)
- — Synthetic drugs

☐ 20. A newborn or infant (up to 12mos) who has been exposed prenatally to drugs (including marijuana), or alcohol, including an infant who is exhibiting symptoms consistent with fetal alcohol syndrome or fetal alcohol effects. (i.e. substance exposed newborn).

☐ 21. Caregiver is absent including, but not limited to:
- — Child is alone and is not capable of caring for self or other children
- — No parent, legal guardian or custodian willing or able to provide care, custody or control
- — Child is left with person with no legal authority to care for the child and the person cannot care for the child
- — Parent leaves child with person unable to provide adequate care
- — Leaving child unattended in car

☐ 22. Injuries due to neglect or failure to supervise including, but not limited to:
- — Fractures
- — Brain injury or serious head injury (e.g. subdural hematoma)
- — Burns (e.g. immersion burns, cigarette burns, unexplained burns)
- — Drowning or near drowning
- — Leaving child unattended in a car causing injury

☐ 23. Child is extremely fearful because of their home situation, present circumstance or because of a threat of additional neglect (this does not refer to fear of legal disciplinary practice or generalized fear)

☐ 24. Caregiver is unable to perform parental responsibilities consistent with basic needs leaving the child in a threatened state due to the caregiver's:
- — Substance use/abuse
- — Behavioral/mental illness or condition (e.g. depression, situational stress)
- — Physical impairment
- — Cognitive functioning

☐ 25. Caregiver describes or acts towards the child in predominately negative terms, has a distorted view of the child, or has extremely unrealistic expectations given the child's age or level of development

☐ 26. Caregiver recklessly or deliberately exposes the child to sexually explicit material or acts

☐ 27. Caregiver is unwilling or unable to meet the child's needs for supervision, food, clothing, shelter or medical care

☐ 28. Caregiver allows known sexual predator access to the child

Neglect - *Failure to Protect*

☐ **29. Caregiver permits a child to enter /remain in a structure or vehicle that's used for the purposes of manufacturing drugs**

☐ **30. *Child was in close proximity to an incident of domestic violence and could have been injured (this includes being held by one of the adults during the incident)**

☐ 31. Child's behavior in the home threatens serious or severe harm to self or others and the caregiver is unwilling or unable to control the behavior

☐ 32. Significant incident or repeated exposure to domestic violence that causes risk of harm to the child

DCS-1968
(09/21)
Page 4

ARIZONA DEPARTMENT OF CHILD SAFETY
HOTLINE REPORT DECISION TOOL

Neglect -*continued*

Neglect - *Environment*

☐ **33.** Living environment is a threat to the child's safety including, but not limited to:
- — Building structure or contents capable of falling in
- — Exposure to elements in extreme weather
- — Fire hazards
- — Electrical wiring exposure
- — Weapons accessible and available
- — Access to dangerous objects or harmful substances
- — Access to drugs or paraphernalia
- — ***Manufacturing of drugs** (includes Marijuana grow if unsecured and accessible to a child, or if the number of plants exceed six per adult or more than twelve in a household).
- — Housing is unsanitary, filthy, infested, a health hazard (e.g. human/animal feces, undisposed garbage, access to dangerous objects or harmful substances, etc.)

Neglect - *Medical*

☐ 34. Caregiver is unwilling or unable to meet the child's needs for medical health care placing child at imminent risk of harm including but not limited to:
- — Child requires medical care and caregiver is unwilling or unable to seek treatment
- — Caregiver does not administer prescribed medical/psychiatric medication
- — Caregiver leaves medical provider against medical advice (AMA)

☐ 35. Medical diagnosis of malnutrition / failure to thrive without previously diagnosed health condition

Emotional Abuse

A.R.S. § 8-201: "Emotional abuse" means the infliction or allowing… or the infliction of or allowing another person to cause serious emotional damage as evidenced by severe anxiety, depression, withdrawal or untoward aggressive behavior and which emotional damage is diagnosed by a medical doctor or psychologist and is caused by the acts or omissions of an individual who has the care, custody and control of the child.

☐ 36. An incident or pattern of behavior by a caregiver directed toward a child that interferes with child's normal daily functioning including, but not limited to:
- — Berating
- — Name calling
- — Targeting
- — Domestic violence
- — Rejection

☐ 37. Child is extremely fearful because of his/her home situation, present circumstance or because of a threat of additional abuse or neglect (this does not refer to fear of legal disciplinary practice or generalized fear)

Safe Haven

An unharmed newborn infant (30 days old or younger) that is left with a safe haven provider as defined in §13-36.23.01 will be screened as a Screened Out Intake, with the screened out indicator of Safe Haven. If the newborn infant is currently the alleged victim in an open DCS Report, or is alleged to be abused or neglected at the time of the phone call, then Safe Haven will not apply.

Insufficient Information

If there is insufficient information to make a confident screening decision, consider a collateral contact. Statute requires a report to be taken if the location or identity of the victim or perpetrator can be reasonability ascertained.

DCS-1968
(09/21)
Page 5

ARIZONA DEPARTMENT OF CHILD SAFETY

HOTLINE REPORT DECISION TOOL

Vulnerability

Factors to consider when assessing child vulnerability, e.g. any condition that results in child's inability to protect self or seek protection from others, check any factor that applies to any child:

- ☐ Age 5 and under
- ☐ Diminished physical capacity (e.g. unable to protect themselves or seek protection due to a physical disability)
- ☐ Diminished mental capacity (e.g. unable to protect themselves or seek protection due to a cognitive disability)
- ☐ Medical or emotional needs (e.g. unable to protect themselves or seek protection due to medically fragile or serious mental illness)
- ☐ Lacks visibility in the community (e.g. not in school or child care)

Did the abuse or neglect happen less than 3 years ago, or is the child currently vulnerable, or is the child currently unsafe, or is there criminal conduct?

○ Yes ○ No - if no, consider screening as a Screened Out Intake with a Historical Screened Out Indicator.

Response Time

Refer to Vulnerability Factors. The response time is determined based on the vulnerability factors of the alleged victim child(ren).

☐ **Priority 1**

- ☐ Death of a child
- ☐ Near fatality
- ☐ Abuse or neglect that threatens to immediately cause, or has caused, serious harm or death
- ☐ Serious physical injuries to a child (including but not limited to fractures, burns, multiple plane injuries, acceleration/deceleration injuries [shaken baby syndrome], injury to internal organs, etc.)
- ☐ Child is alone and is not capable of caring for self or other children
- ☐ Evidence or disclosure of sexual abuse toward a child and the perpetrator has access to the child or the perpetrator is unknown
- ☐ Substance exposed newborn (SEN) who is expected to be discharged from the hospital within 24 hours

☐ **Priority 2**

- ☐ Abuse or neglect of a child age 0-3
- ☐ Abuse or neglect of a vulnerable child, and the child or perpetrator has been the subject of a prior report (this includes the child as a victim in a prior report or the adult as a perpetrator in a prior report)
- ☐ All criminal conduct allegations not requiring a Priority 1 response

☐ **Priority 3**

- ☐ Abuse or neglect of a child that occurred within the last 12 months and does not require a Priority 1 or 2 response

☐ **Priority 4**

- ☐ Private Dependency Petition (PDP)
- ☐ Abuse or neglect that has occurred over one year ago and does not require a Priority 1, 2 or 3 response

Mathematica Inc.

Our employee-owners work nationwide and around the world.

Find us at **mathematica.org** and **edi-global.com**.