Child Maltreatment Investigations and Family Well-being *

Katherine Rittenhouse†, David Simon‡, and Lindsey Lacey§

March 25, 2025

Abstract

Over one third of children in the U.S. are investigated by the child protection system (CPS) by age 18, but the effects of this interaction on families and children are largely unknown. Investigations aim to protect children from harm and direct services to families in need. However, they also bring the implicit threat of child removal, which can generate both stress and institutional distrust among affected families. Using a regression discontinuity design, we study the impacts of these interactions on child health. Children who are just above the threshold for being "defaulted" into an investigation see a reduction in injuries and an increase in preventative care in the two months following the referral. Effects are consistent across child gender, race and age. An exploration of mechanisms suggests that it is the investigation, rather than home removal, which improves child health. Finally, investigations increase the likelihood that a parent is jailed, suggesting that removal of criminal activity from the home, at least in the short term, may improve child health and safety.

Keywords: Economics of families and children, Child abuse and Neglect, Child Protection Systems

*Contact e-mail: katherine.rittenhouse@austin.utexas.edu. We are grateful to Julia Reuben, Justine Galbraith, Kyle Jennison, Alex Jutca, and the analysts at Allegheny County Department of Human Services for providing us with the data and answering countless questions. Thanks also to Ken Couch, Delia Furtado, Chris Mills, Analisa Packham, and seminar participants at APPAM, PPCW, University of Connecticut, UT Austin, the University of Missouri, and Texas Tech University. All errors are our own.

†University of Texas at Austin, LBJ School of Public Affairs

‡University of Connecticut and NBER

§Allegheny County Department of Human Services

1 Introduction

Every year over three and a half million families are investigated for child abuse and neglect, with 37% of all children experiencing an investigation before age 18 (Kim et al. 2017). Investigations involve a social worker visiting the child's home to make determinations about their safety and well-being, and may result in mandated services or removal of the child from the home (i.e., foster care placement). Investigations aim to protect children from harm and direct services to high-risk families. However, these interactions may also be burdensome and disruptive (Fong 2018). Detractors argue that investigations impinge on parental rights and family autonomy, with limited benefits to child safety. Severe racial disparities have led to critiques of the child protection system (CPS) by scholars and activists as a "family policing system" with particularly negative impacts on Black communities (Roberts 2022). As such, there is a growing movement to reduce investigations, limit CPS involvement in families' lives and, at the extreme, to abolish CPS entirely. Despite the widespread and controversial nature of this interaction, empirical evidence on the effects of CPS investigations is severely limited. We bring novel empirical evidence to this debate in a setting with plausibly exogenous variation in the likelihood of investigation among high-risk families.

We leverage a predictive risk model which defaults families above a set risk threshold into an investigation. Our regression discontinuity design compares outcomes for children just below vs. above this high-risk threshold. Using linked administrative data, including detailed Medicaid claims, we study effects of investigations on child physical health.

Our setting is Allegheny County, Pennsylvania (Pittsburgh). Allegations of maltreatment are made through referrals to the county's office of Children, Youth and Families (CYF). Screeners at CYF are charged with evaluating these referrals and determining whether to screen in the family for an investigation. Allegheny County has a rich Data Warehouse of linked administrative data across multiple public agencies, including from CYF and state

Medicaid claims. For every referral to CYF, we are able to link each individual (children and parents) to the Data Warehouse and follow them over time.

In 2016, Allegheny County implemented a predictive risk model to help workers decide which families to screen in for investigation. The Allegheny Family Screening Tool (AFST) uses information from the Data Warehouse on all individuals associated with a referral to predict the likelihood that a referred child will be removed from their home within two years. Call screeners are required to run the AFST before making a recommendation on whether to investigate or not. The AFST estimates a continuous risk score which is translated into a discrete referral-level score from 1 (lowest risk) to 20 (highest risk) observed by the screener. Referrals with risk scores above 17 fall under a high-risk protocol, and are "mandated" to be screened in for an investigation. In practice, this mandate is not binding, and high-risk protocol referrals may be screened out subject to supervisor approval. We therefore refer to referrals above the threshold as being "defaulted" into an investigation.

The cutoff for default investigation motivates our fuzzy regression discontinuity design, in which we compare outcomes for children directly above vs. below the threshold. We first document a strong first stage: receiving a score just above the cutoff for a default screen in increases the likelihood of an investigation by 19 percentage points. We then estimate reduced form effects of being over the threshold for a default investigation on a range of outcomes.

We focus on outcomes related to child health. The primary goal of investigations is to protect children from harm, and yet effects on child safety and well-being are largely unknown. Children above the threshold for a default investigation are 35% less likely to have an injury-related Medicaid claim in the 60 days following a referral.[1] Default investigations increase the likelihood of wellness visits and immunizations by 28% and 30%, respectively. Combining these and other measures into a "poor health index," we find that a default

[1]We focus on this time frame, as investigations are required to be completed within 60 days.

investigation reduces poor health by 0.07 standard deviations. We do not find evidence of longer-term impacts on child health, suggesting that investigations are most protective while they are ongoing.

Subgroup analysis shows that effects are similar across child age, race and gender. This suggests that investigations are protective across demographic groups. Results are robust to a wide set of specifications. We distinguish between changes in health vs. healthcare utilization by exploring heterogeneity across categories of reporters. Parents may be more likely to avoid taking their child to a doctor if they suspect a report came from a medical professional or other mandated reporter. We find similar effects when excluding these professionals, suggesting that reductions in injuries are reflective of improvements in child health. Additionally, increases in preventative care visits suggest that parents are not actively avoiding medical professionals during investigations.

Turning to mechanisms, we explore the possibility that child health improvements are the result of home removals. However, effects of being over the threshold for a default investigation on home removals are small in magnitude and statistically insignificant during the time period where we see impacts on child health. This suggests that investigations themselves are protective.

Finally, we study effects of investigations on parents' criminal justice system involvement. Parents just above the threshold for a default investigation are 1.5 percentage points (31%) more likely to be in jail in the two months following a referral. This suggests that investigations uncover criminal activity in the household. A reduction in criminal activity or the removal of bad actors from the child's home could help explain observed improvements in child health.

We contribute to a growing literature on the causal effects of child protection systems. The majority of this literature studies the effects of foster care, relying on the random assignment of cases to investigators and differential tendencies of investigators to remove a

child from their home (Doyle 2007, 2008, 2013; Bald et al. 2022; Gross and Baron 2022; Baron and Gross 2022; Roberts 2019). These papers have found both positive and negative effects of foster care for children on the margin of removal, depending on the setting, time period and child characteristics, with more recent work tending to show benefits of home removal. However, this approach can only assess the impact of removal, often a last resort for child welfare agencies. Grimon (2020) studies the effects of child welfare case openings on parents, finding that this intervention increases mental health and substance abuse treatment among mothers. We identify the impact of investigations, an early-stage interaction which affects one third of U.S. children. Additionally, we identify impacts for a different group of children - those at the margin of investigation. These children are likely to be better off and at lower risk than those at the margin of home removal. Our results show that child welfare interventions can improve outcomes for this group of children. Finally, while prior work on home removals has focused largely on educational outcomes and criminal justice system involvement, we study impacts of child welfare system involvement on direct measures of child health and safety.

We also add to the emerging literature on algorithms in the child welfare setting. Rittenhouse, Putnam-Hornstein, and Vaithianathan (2023) study the introduction of the AFST, finding that the algorithm reduced racial disparities in screening and removal decisions. Grimon and Mills (2023) study the effects of providing a risk score to call screening teams on child health, finding that the algorithm reduces measures of child harm. We abstract away from the direct effects of algorithms, and instead use a feature of the AFST implementation to study effects of investigations on child and family well-being.

Outside of economics, our findings contribute to related literatures in the social sciences. Qualitative work has posited that investigations create apprehension and distrust among families, disrupt institutional engagement, and may lead families to avoid interactions with mandated reporters such as doctors and therapists (Fong 2020, 2018). We build on this

qualitative work work by measuring the causal effect of an investigation on measures of child health, health care utilization, and parental involvement in the criminal justice system.

The paper proceeds as follows. Section 2 presents the institutional context. Section 3 describes the data we use, and Section 4 the empirical approach. Section 5 presents results related to child health, Section 6 discusses potential mechanisms for our main results and Section 7 concludes.

2 Background

2.1 Allegheny County Child Welfare Investigations

In Allegheny county, the Office of Children, Youth, and Families (CYF) is charged with protecting children from abuse and neglect. Alleged maltreatment is first brought to the attention of CYF by third-party reporters, including mandated reporters (e.g., educators or doctors) and non-mandated reporters (e.g. neighbors or family members). Reports can be made by phone or through an online service known as "Childline." A call screen worker then reviews and recommends whether to "screen in" each referral for an investigation.[2] After a screener makes their recommendation, a supervisor will make the final decision on whether to open an investigation.[3] The state of Pennsylvania distinguishes between "child protective services" referrals which involve allegations of physical or sexual abuse, and "general protective services" (GPS) which tend to involve allegations of neglect. Child protective service referrals are screened in 100 percent of the time by law and do not allow for screener discretion. As such, our analysis is limited to GPS referrals, which make up the

[2]See details at https://www.alleghenycounty.us/Human-Services/Programs-Services/Children-Families/When-a-Report-is-Made.aspx

[3]Anecdotally, screeners are more likely to recommend an investigation for younger children or when the specific allegation suggests a child's physical safety is at risk.

majority of reports in Allegheny county.[4]

If a report is screened in, an investigation is opened. Investigations aim to determine whether the allegations of maltreatment are credible, whether children are safe in their home, and whether the family requires various services or court interventions to ensure child safety.[5] To achieve these aims, the caseworker assigned to the investigation conducts one or more home visits, and conducts individual interviews with parents and children as well as the reporter and other individuals connected to the family. Home visits must occur within a maximum of three days from the time of referral, and every one to four weeks throughout the period of investigation. All home visits over the course of an investigation may be unannounced. At the initial visit, the caseworker informs the parents they are being investigated for maltreatment, discloses the allegations, walks through the home, and conducts one-on-one interviews with each household member. During home visits, caseworkers look for food, clothing, and appropriate toys, any environmental dangers, evidence suggesting criminal activity or drug use, weapons, and overall cleanliness.

According to state law and CYF policy, investigations are supposed to be closed within a maximum 60 days, after which one of several outcomes is possible. If there is not enough evidence to support the allegations, they are deemed unfounded and no further action is taken. The social worker may still connect the family with or provide information about optional services, but CYF will not have any continued interaction with the family. If the investigator instead finds reasons to be concerned about the child's well-being, they may open a CYF case for the family. A CYF case may involve periodic check-ins with the family, suggested or mandated services (including, e.g., mental health or substance abuse services), or, often as a last resort, removal of the child from the home. Both mandated services and

[4]Allegheny County provides a brief overview of the two types of referrals here: https://www.alleghenycounty.us/Human-Services/Programs-Services/Children-Families/Protective-Services.aspx.

[5]State regulations describing the county's rights and responsibilities for receiving and investigating GPS reports are included in PA Code § 3490.232. Information on investigation procedures in Allegheny County is derived from review of internal county documents detailing investigative practice standards.

child removal require a court order, and may be substitutes (i.e., parents are required to receive services as a condition of keeping the child at home) or happen concurrently (if, for example, parents are required to receive services as a condition for being reunited with their child).

2.2 Allegheny Family Screening Tool

In August 2016, Allegheny County, PA became the first jurisdiction in the country to use linked administrative data to implement a predictive risk model designed to assist in screening child welfare referrals. The Allegheny Family Screening Tool (AFST) draws on data from the county's Data Warehouse, which integrates human services data across multiple government agencies including the Department of Human Services, the County Housing Authority, Jail, Probation Office, and Health Department.[6] The AFST is meant to inform, rather than replace, human decision-making.

The AFST uses a wide range of variables from the Data Warehouse to predict a child's "risk," where risk is defined as the likelihood that the child will be removed from their home by CYF within two years conditional on investigation. The model includes information on previous investigations/allegations, child or parent involvement in the court system (such as time spent in jail), medical history (such as behavioral health issues), and demographics such as age, gender, and presence of biological mother and father.[7] From these variables, the model calculates a continuous latent risk score for each child on a referral.[8]

[6]The Data Warehouse was established in 2001 to provide integrated services to the county as well as information to the public (Allegheny County Analytics 2018). The warehouse is relatively unique in the extent to which it links data on families and makes it available to researchers, and has gained increasing visibility in the social sciences.

[7]For a complete list of the variables used in a recent version of the AFST, see Vaithianathan et al. (2019). Race was not included as a predictor.

[8]The training data set for the AFST uses all available previous data (up to two years before a given AFST version is introduced) to calculate this latent risk score. The strongest predictors of risk are measures of previous involvement with CYF (i.e., previous referrals, investigations, or home removals). Other variables used tend to be less singularly important in predicting future removals. Initially, the AFST used a probit model to predict these latent risk scores. The AFST has been updated several times since its introduction and

After calculating a risk score for each child in the household, the AFST then takes the maximum value across all children, such that there is one latent risk score per referral.[9] This referral-level latent score is mapped to discrete values between 1 and 20 based on predetermined cutoffs.[10] In most cases, call screeners are shown this discrete score along with a general category of low risk (AFST scores 1-9), medium risk (10-14), or high risk (15-17). While scores below 18 are meant simply to advise the screener's decision, referrals with scores of 18 and above are defaulted to be screened in for an investigation, and require supervisor approval to be screened out. In such cases, the call screener does not observe a numerical score, but rather a notification that the referral falls under the "high risk protocol." See Figure A1 for a visual example of what the screener would see.[11]

3 Data

We obtained administrative data from the Allegheny County Department of Human Services on the universe of child maltreatment referrals to CYF from December 2016 through Decem-

has used different input variables and estimators to make its predictions depending on the AFST version. In December 2018 the algorithm was revised to use a LASSO model and was re-trained on all available data. In July 2019 the lasso model was improved to refine how the model counted incoming criminal justice outcomes. Version 4 of the AFST was implemented in April 2022, and retired variables related to current allegations, Medicaid enrollment, and birth-records. See Appendix B for more details on the how the AFST score is constructed and its various versions.

[9]When the AFST was first introduced in August 2016, there was a separate latent risk score assigned to each child who was an alleged victim of abuse or neglect. While adult histories were part of individual child scores, the AFST before December of 2016 required that each child have previous administrative records to be scored. However, this generated many missing scores, particularly for younger children who had no history in the data warehouse. Therefore, in December 2016, this rule was relaxed and younger children could be scored solely off of the history of parents/other adults in the house. In addition to this, a fix was implemented that improved client matching and solved problems related to duplicate IDs in the warehouse. Because these were substantial changes in how the score was implemented, we follow Goldhaber-Fiebert and Prince (2019) and drop referrals before December 2016.

[10]The cutoffs are based on the risk ventiles in the training data.

[11]A "low risk" protocol was also introduced in November 2018 such that all children who got a score below 11 (or 12 in later versions) were defaulted to be screened-out for an investigation. However, we did not find any "first stage" effects of being below the threshold on having an investigation opened.

ber 2023.[12] For each referral, we observe the latent (continuous) risk score produced by the AFST. Each referral is associated with one or more individuals (alleged victims, other children in the household, parents linked to the alleged victim who are either in or outside the household, adults in the household, and alleged perpetrators) and one or more allegations. We track referrals through the child protection system, from initial report through screening decision and investigation outcomes. We also track individuals across referrals, observing their entire history of involvement with CYF during the span of our data.

Crucially, we also observe detailed Medicaid claims data from 2015 through 2023 for all individuals in our sample who are enrolled in Medicaid. For each claim we observe the exact date, and all associated ICD-10 diagnosis codes.

We create four main categories of Medicaid claims to measure child health in the months after a referral. The first category includes ICD codes related to injuries. The second includes codes which have been identified as suggestive of child abuse or neglect in prior work (Schnitzer et al. 2011; McConnell et al. 2020). The final two categories measure well-child visits and immunizations, respectively. Finally, we normalize and combine these categories into a "poor health" index, where injuries and abuse and neglect each increase the value of the index, while wellness visits and immunizations decrease the value of the index.[13] Figure 2 presents the average value of the poor health index in the twelve months after a referral for each risk score between 10 and 20. Note that poor health is in general increasing (i.e., health is worsening) with predicted risk, with the gradient between risk and health mostly driven by higher-risk referrals.

In our analysis of the longer-term health impacts of investigations, we include two additional categories in our index. The first includes ICD codes related to ambulatory care

[12]In our main analyses we limit our sample to referrals through December 2022, allowing a one-year follow-up period for each referral.

[13]To create the index we first normalize each of the four indicator variables using the entire sample of child-referrals by subtracting the sample mean and dividing by the sample standard deviation. We then sum the normalized indicators and divide by four, so that the poor health index takes values between -1 and 1.

sensitive conditions (ACSC). This category follows Grimon and Mills (2023) in identifying hospital visits for conditions which may be avoidable if proper preventative care is received. The second category includes infectious sicknesses such as the flu and lower respiratory tract infections (pneumonia and bronchitis). These may be prevented by immunizations or improved sanitary conditions in the home. While investigations may get children needed medical care to prevent these conditions, effects likely would not be immediate. As such we only include these outcomes in the long-term index.

Finally, we use several linked variables from the Allegheny County Data Warehouse. Namely, we connect parents to monthly indicators of their criminal justice system involvement, including jail stays and any court appearances. Court appearances include both criminal courts and magisterial courts, which handle lower-level cases such as small claims and minor drug possessions.

Our data is structured at the individual-referral level. One individual can appear multiple times if they are associated with multiple referrals. Treatment is assigned at the referral level, and all children living in the household at the time of an allegation are included on the referral. We observe outcomes at the individual level. Table 1 presents summary statistics at the child-referral level, separately for all children referred to CYF (Column 1), child-referrals which fall within our main bandwidth (Column 2), and compliers below the threshold (Column 3).[14]

As discussed above, we drop calls that are mandated by the state to be investigated (calls where the allegations include physical or sexual abuse). Finally, in later AFST versions the "high-risk" protocol is only applied to cases in which the referral includes a child under the age of 16 in the household, therefore we drop referrals (across all AFST versions for consistency) where there is not a child under the age of 16 in the household. Our main bandwidth (referrals with scores of 17 or 18) includes 18,857 child-referrals, including 12,724

[14]Complier means are discussed in greater detail in Section 4.1 below.

unique children and 7,810 unique referrals.

Of all child-referrals in our sample, approximately half are female and 39% are Black. Note, this indicates a significant over-representation of Black children, relative to their share of the population.[15] On average about three children are associated with the referral. In the six months prior to the given referral, 28% of children were included on a different referral, 13% of children had an investigation, and 72% of children were enrolled in Medicaid. The most common allegation categories are caregiver substance abuse, inadequate care or home, neglect, exposure to risk, and physical maltreatment.[16] Referrals are most likely to come from educational professionals (16%), social workers (15%), parents or relatives (13%), therapists (12%) and legal professionals (12%). 46% of child-referrals are screened in for investigation, and 5% are followed by the child being removed from their home within twelve months.

Children within our main bandwidth (with risk scores of 17 or 18) are more likely to be Black, and come from households with a larger number of children. They are more likely to have had a previous referral and investigation within the prior six months, and more likely to have been enrolled in Medicaid. Child-referrals within the bandwidth are 39% more likely to result in investigation, and 53% more likely to be removed from their home within twelve months.

Next, Table 2 presents summary statistics at the parent-referral level, separately for all parents associated with referrals to CYF (Column 1), parent-referrals which fall within our main bandwidth (Column 2), and compliers (Column 3).[17] Our main bandwidth includes 19,005 parent-referrals, including 12,581 unique parents and 7,795 unique referrals. We define an individual as a mother if their stated role is "parent" and their gender is female.

[15] See: https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/10/ACDHS_CYF-Race-Disproportionality-Brief_100217-final.pdf

[16] Physical maltreatment is differentiated from physical abuse in PA often by context. For example, a parent getting into an argument with a teenager where it escalates to them striking each other might by classified as physical maltreatment instead of abuse.

[17] Complier means are discussed in greater detail in Section 4.1 below.

We define fathers similarly. Note, there may be multiple mothers or fathers included in any given referral if, for example, there are multiple father figures associated with the referred children, or if the children in the household have different biological fathers. In our main bandwidth, 75% of parents are members of the same household as the alleged victim, and 43% of parents are listed as the alleged perpetrator. Within the bandwidth, 8% had a jail stay in the six months prior to the referral, and 45% appeared in criminal or magisterial court. 26% of parents within the bandwidth were involved in a CYF investigation within the previous six months.

4 Empirical Strategy

Our empirical strategy compares individuals with latent risk scores directly above relative to below the threshold for being "defaulted" into an investigation in a fuzzy regression discontinuity design. We estimate the reduced form effects using the following regression model:

$$Y_{ir} = \alpha + \beta_1 D_r + \beta_2 z_r + \beta_3 (D_r \times z_r) + \gamma_t + \epsilon_{ir} \tag{1}$$

Where Y_{ir} is the outcome for individual i who is associated with referral r. Recall that each referral r is assigned a single risk score, equal to the maximum risk score of all children on the associated referral. D_r is equal to one if referral r has a score higher than the default screen-in threshold, and z_r is the normalized algorithm score relative to the cutoff. β_1 is the coefficient of interest, capturing the effect of crossing the threshold. We allow for different slopes on either side of the threshold, and include month-year fixed effects γ_t. Since treatment is assigned based on the AFST score, which is identical for all children on the referral, our main specifications cluster at the referral level to account for auto-correlation within referrals of score assignment. We use a triangular kernel and a linear functional form.

Figure 1 presents the distribution of algorithm scores separately by AFST version. Note that the latent risk scores and relevant numerical cutoffs are distinct across models. For our main results, we pool and normalize across the four algorithm versions, to create one running variable. We include referrals within one discrete score of the threshold for a default investigation (i.e., 17-18, or within the red dashed vertical lines in Figure 1). We then normalize the score for each version by subtracting the relevant cutoff so that each is centered around zero, and dividing by the numerical size of the bandwidth on either side of the threshold. This leaves us with a running variable which ranges from (-1,1).

For each of our main results, we test robustness to alternative specifications including adding a vector of demographic controls and referral case characteristics, using a quadratic form of the running variable, and alternate bandwidths. Each version's respective MSE-optimal bandwidth is shown in Figure 1. The MSE-optimal bandwidth selection procedure recommended by Cattaneo, Idrobo, and Titiunik (2019) results in bandwidths that are uniformly larger than our preferred choice of 17-18. Optimal bandwidth selections reflect a bias-variance trade off, allowing larger bandwidths to reduce variance at the cost of increased bias. We thus argue the bandwidth choice for our main analyses is conservative, intuitive for our setting, and allows for consistent estimation across models.

4.1 Monotonicity

Our regression discontinuity design identifies the Intent-to-Treat (ITT) of a default investigation on child and family outcomes. As such, our results apply for families around the high-risk threshold who would not be investigated, *but for* the high-risk protocol. These are families the AFST identifies as high-risk, but call screeners would have chosen not to investigate.

It is worth investigating the characteristics of this group of compliers, to understand who we are identifying effects for. To do so, we use the method in Abadie (2002), modifying

Equation 1 to describe average characteristics of compliers directly below the threshold. Complier means are included in Column (3) of Tables 1 and 2. Among children, compliers are largely similar to children within the wider analysis bandwidth in terms of child demographics, lagged system involvement, and allegation categories. They differ somewhat in reporter categories, with compliers less likely to be reported by Anonymous or Other sources, which are typically considered less reliable.[18] Mechanically, compliers below the threshold do not have an investigation opened. 79% of these compliers are referred again within 12 months, and 2.5% are removed from their home within the following year. Turning to parents, compliers are again demographically similar to parents within the wider analysis bandwidth. Compliers are somewhat less likely to have been jailed or appeared in court in the prior six months.

Identification requires that the high-risk protocol weakly increases the likelihood of investigation for all children. This monotonicity assumption would be violated by "defiers," i.e., families who are less likely to be screened in when subject to the high-risk protocol. This type of behavior might arise if call screeners are more skeptical that cases which are identified as high-risk by the algorithm are in need of an investigation, or if call screeners actively rebel against the algorithm's recommendations. While it is difficult to empirically evaluate this assumption, anecdotal evidence and institutional details suggest that defier behavior is unlikely in our setting. Observation of and discussions with call screeners at CYF screeners revealed that they run the AFST only after examining the details of the referral and making a preliminary screening decision. Defiant behavior would require that screeners initially choose to investigate a family, but change their mind after seeing that the algorithm agrees that the family is high-risk. Further, screening out a high-risk protocol referral involves additional administrative burden, in that a supervisor is needed to override the default

[18]This implies that those below the threshold who are reported by Anonymous or Other sources are more likely to be "never takers," i.e., these referrals would be screened out even if they fall under the high risk protocol.

investigation, making it costly for call screeners to actively defy the AFST recommendation.

4.2 Validity

The validity of the regression discontinuity design rests on the assumption that potential outcomes are smooth around the cutoff. For example, this design would be invalid if call screeners could sort referrals to be directly below the threshold so they are easier to screen out. In our case, manipulation of the algorithm score around the cutoff would be difficult, if not impossible. The score is based off of administrative histories of the children and parents on the referral, and call screeners and families do not have the ability to change these inputs. Even if they did, they would need to understand the statistical and machine learning models well enough to precisely modify the key variables to guarantee the child went directly over (or under) the threshold. Running the algorithm multiple times does not change the score. Informal discussions with call screeners suggests that they do not attempt to change or manipulate the score in any way.[19]

We run three empirical tests to validate the research design. First, we test for manipulation in the running variable using local polynomial density estimators, following Cattaneo, Jansson, and Ma (2020). Figure A2 plots the the results of these tests for each AFST version, and notes the associated p-value. Visually we see no signs of bunching around either side of the "high risk" threshold. Across the four AFST versions, we fail to reject the null hypothesis of no manipulation around the cutoff, with p-values ranging from 0.200 to 0.852.

We next test for discontinuities in predetermined characteristics. Table 3 shows results from estimating Equation 1, where the outcome variable is set to a variety of predetermined child, referral and parent characteristics as well as six-month lags of each of our main outcome

[19] Our evidence from in-person observation revealed that the AFST is a relatively small part of the screening process, and checked largely to determine if a mandatory screen in is required or not. Our discussion with the IT specialists and data scientists who administer the DHS data confirmed that the risk score could not be manipulated by call screeners, parents, or other DHS workers.

variables of interest. Out of the 36 regressions, four are significant at the 10% level, in line with expected Type I error.

Finally, our main outcomes of interest are derived from Medicaid claims data, which we only observe for individuals enrolled in Medicaid. It would complicate the interpretation of our results if investigation led to changes in child Medicaid enrollment. We directly test for this by estimating Equation 1, setting the outcome variable to an indicator for childrens' Medicaid enrollment in the 1-2 months, and 3-12 months after the referral. Results are reported in Panel F of Table 3. Reassuringly, the estimated coefficients are statistically insignificant and small in magnitude, suggesting that investigations do not cause changes in Medicaid enrollment in the short or long term.

4.3 Relevance

In this section we verify that the default screen-in rule results in a strong first stage effect on the likelihood an investigation is opened. It is possible that the threshold could have little to no effect on the probability of investigation if most referrals with a score below the cutoff are screened in for an investigation or if referrals above the cutoff are screened out (through a supervisor override).

The regression discontinuity graph is shown in Figure 3. We plot the normalized risk score on the x-axis. The y-axis shows the percent of referrals investigated in each bin. Panel (a) plots the risk score bins within our "high-risk" bandwidth and shows a stark, discontinuous increase in the likelihood of investigation over the threshold. Panel (b) shows screen-in rates for the entire distribution of risks cores, where the x-axis shows the discrete numerical risk score as seen by the screener. Again, we see a clear and discontinuous increase around the threshold of a risk score of 18.

We estimate Equation 1, where Y_{ir} is a binary variable equal to one if an investigation is opened on referral r, and zero otherwise. Results are presented in Column (1) of Table

4. Children on referrals who scored just above the high-risk threshold are 19.2 percentage points more likely to be investigated, statistically significant at the 1% level. Figure A3 plots the first stage estimate separately for several subgroups. The first stage is similar and statistically indistinguishable across gender, race and age.

Next, we estimate Equation 1, setting Y_{ir} equal to an indicator for *any* investigation in the next twelve months. Families are often referred more than once, meaning it is possible that a child who is initially screened out is re-referred and investigated at a later date. Column (2) of Table 4 shows that children just above the threshold are 7.2 percentage points more likely to be investigated at any point in the next year (statistically significant at the 1% level).

The attenuation of the coefficient in column (2) relative to column (1) of Table 4 implies that many families below the threshold who are initially screened out are re-referred and investigated in the following year.[20] This implies our first stage may, at least in part, interpreted in part as a shift forward in the timing of investigation.

Figure A4 shows the dynamics of how the control group "catches up" to the treatment group in the year following a referral. In Panel (a) we report coefficients from seven separate regressions, setting Y_{ir} equal to an indicator for any investigation within different time periods from the date of reference referral. The first coefficient is equal to our first stage estimate reported in Column (1) of Table 4, while the last coefficient reported is equal to our twelve-month first stage estimate reported in Column (2) of Table 4. The discontinuity attenuates in the six months after the reference referral before stabilizing. Panel (b) reports raw means of the cumulative likelihood of investigation separately for children in our bandwidth below and above the threshold. The cumulative likelihood of investigation is increasing over time for both groups, but the rate of increase is higher for the control group. By twelve months

[20]The coefficients imply that at least 63% (1-$\frac{.0716}{.192}$) of children directly below the threshold who are initially screened out are investigated in the following year.

after the reference referral, 77% of children below the threshold and 86% of children above the threshold are screened in for an investigation.

In the next sections we show effects of being over the high-risk threshold on a variety of outcomes for children and parents. We present results in their reduced form.

5 Child Health

Protecting children from harm is the primary objective of investigations for maltreatment. However, the effects of such investigations on child health are *a priori* ambiguous. Investigations serve as a path to direct families to health services, provide counseling which could improve safety within the home and prevent maltreatment, and surveil the physical safety and medical care of children, possibly leading to improved health outcomes. However, by their nature investigations may be disruptive, increasing stress within the family or causing fractures between families and the medical system, leading to poorer health outcomes (Fong 2020). For example, investigations may lead parents to be suspicious of mandated reporters (such as nurses and doctors) making them less likely to seek preventative health care for their child.

We first analyze the impact of investigations in the 60 days after a call, when an investigation is likely still ongoing. During the investigation, families are under surveillance and subject to visits from a caseworker throughout the period of investigation, which may or may not be announced in advance. Investigations are mandated to last a maximum of 60 days. Figure A7 shows the distribution of investigation length in our main sample of referrals within the bandwidth, which are screened in for investigation. The average time between the date of the referral and the date of investigation closure is 52 days, with 87% of referrals lasting 60 days or less.

We separately estimate these short-run effects on each of the four main categories of Med-

icaid claims related to child well-being identified in section 3 above– injuries, abuse/neglect, wellness visits and immunizations. Results are reported in Table 5. Children above the threshold are 1.2 percentage points less likely to have an injury-related Medicaid claim (statistically significant at the 5% level). This represents a 35% decline in injuries relative to the control group mean of 3.5%. Investigations increase the probability of having a wellness visit by 28% (3.3 percentage points), and getting an immunizations by 30% (2.0 percentage points), statistically significant at the 1% and 5% level, respectively. There are no statistically significant effects on codes linked to abuse and neglect; though the sign on the coefficient is negative and the magnitude suggests small declines (5% of the control group mean). Investigations decrease the poor health index by 0.07 standard deviations in the first two months after a call, statistically significant at the 1% level.

We next estimate effects in the 3-12 months after the referral, i.e. after the investigation is likely over.[21] Here we study impacts on two additional measures of health – ACSC and infectious illness, which are most likely to be affected in the longer-term through changes in household sanitation or regular use of preventative health care. Results are reported in Table A1. For completeness, we also show effects on these additional health measures in the short run, and present effects on each poor health index in both the short and the long run. In the long run, we do not see a statistically significant impact on either of the poor health indices, and only one of the six components shows a statistically significant impact. Visits for ambulatory care sensitive conditions decline by 0.025 percentage points (10% of the control group mean), statistically significant at the 10% level. Although suggestive, this is consistent with investigations leading to more preventative healthcare in the short term, which in turn may reduce longer-term health problems.

Finally, we study short-run effects separately by child gender, race and age. Subgroup estimates for each of our main outcomes of interest are shown in Figure A6. Recall that

[21]Specifically, we study effects in the 61 to 360 days after the referral date.

the first stage is largely similar across subgroups, so heterogeneity is unlikely to be driven by differences in instrument strength. We do not find statistically distinguishable effects across subgroup, though we see suggestive evidence that reductions in injuries are larger for female children, Black children, and older children. As for preventative care, effects are larger in magnitude for White children and younger children. Overall, we find consistent improvements in child health across gender, race and age, suggesting that investigations are beneficial for all children in this high-risk sample.

5.1 Dynamics

We next consider how effects on child health change over time. For a given child health outcome Y_{ir} we estimate Eq. 1 separately for Y_{ir} measured in each 60-day period from 360 days before the referral to 360 days after the referral, for a total of twelve regressions. That is, for each of our outcomes of interest we create twelve variables Y_{ir}^{j}, measuring child health in 60-day intervals (indexed by j) relative to the referral, for $j \in [-6, -1]$ and $j \in [1, 6]$. We exclude Medicaid claims made on the day of the referral, as we cannot determine whether the report occurred prior to or after the Medicaid claim on that day. Regressions on Y_{ir}^{j} with $j < 0$ act as a falsification test, transparently testing for discontinuities in health outcomes prior to the referral. Statistically significant effects for outcomes measured in these time periods would indicate pre-existing differences around the high-risk threshold. Regressions on Y_{ir}^{j} with $j > 0$ allow us to investigate dynamics in observed effects on child health outcomes in the year following a referral.

We present this analysis for the outcome of child injuries in Figure 5, which shows β_1 and 95% confidence intervals associated with each of the 12 regressions. Reassuringly, we find consistently null effects on outcomes measured in each period prior to the referral date. We see a substantial decline in injuries in the two month period corresponding to the investigation, matching our findings in Table 5. However, in each subsequent period our estimated

effect of a default investigation on likelihood of injuries is close to zero and statistically insignificant. These results suggest that reductions in injuries do not persist after 60 days, when the investigation is most likely closed.

Figure A9 presents the same exercise for each of our child health outcomes.[22] We show effects both in our standard model with a linear running variable as well as quadratic as a robustness check. For injuries, wellness visits, immunizations, and the child health index (including both long and short run variables) we see an improvement in health during the 60 days after the referral. After 60 days, when the investigation is most likely closed, treatment estimates are in general close to zero and statistically insignificant. For the remaining outcomes (abuse/neglect, ACSC, and flu/LRTI), estimated effects are statistically insignificant in each 60 day period in the year before and after the referral.

5.2 Robustness

We further vet our child health results by showing robustness to a number of alternative specifications in Table 6. Panel A shows estimates for the first stage effect on investigation, and Panels B through F show estimates for each of our main child health outcomes.

The first column shows the baseline results for each outcome for ease of comparison. The second column includes a quadratic rather than linear form for the running variable. This more flexible functional form ensures we are not mistaking a discontinuity for a quadratic change in outcomes around the cutoff. Column 3 takes the baseline model with a linear running variable and adds additional controls.[23] If assignment to treatment is truly random at the cutoff then additional controls should not materially change the estimated coefficient.

[22] We do not distinguish between short and long run variables in this exercise. This allows for the possibility that investigations could have more immediate effects on ACSC and flu / LRTI.

[23] We include as controls demographic variables (Black, White, Other, Female, age, and the number of children in the household), as well as a vector of dummy variables for the type of reporter who made the referral (anonymous, legal professional, medical professional, parent / relative, educational professional, social worker, therapist, or other), and allegation categories (caregiver substance abuse, provided inadequate care or housing, neglect, exposed child to risk, physical maltreatment, or other).

The fourth column additionally controls for a six month lag of the value of the outcome variable. The fifth column includes all the additional controls in columns 3 and 4, but specifies a quadratic rather than linear running variable. Finally, the sixth column includes the full set of controls with a linear form for the running variable and cuts the bandwidth in half. Larger bandwidths reduce variance but increase bias. We would be concerned that our estimates were picking up confounding factors if cutting the bandwidth in half caused the estimated coefficients to fall towards zero.

Across all these specifications and models, there are limited changes to the coefficients relative to the baseline. Effects are statistically indistinguishable across specifications and retain similar levels of statistical significance. On average, including a quadratic form of the running variable increases the estimated discontinuity. Likewise, the discontinuity tends to increase slightly when the bandwidth is cut in half.

5.3 Health vs. Healthcare Utilization

Reductions in Medicaid claims for injuries may represent improvements in health or reductions in healthcare utilization. Qualitative evidence has documented that investigations may create mistrust in mandated reporters, including medical professionals (Fong 2018). It is possible then that families who are investigated withhold medical care from their children out of increased fear medical professionals will report the injury to Child Protective Services. The welfare implications of such avoidance behavior are clearly different from those of a true reduction in injuries. We have several approaches for distinguishing between these two mechanisms, and argue that the evidence is most consistent with improvements in health.

First, our effects are focused on the 60 days after the referral, when an investigation is likely ongoing. During an investigation parents may be told explicitly to provide necessary medical care for their children, and know that the social worker could come to the home unannounced at any moment. If an investigator were to see an injured child who was

not given needed medical care, it could lead to a home removal on the grounds of medical neglect. Relatedly, we see *increases* in preventative care (wellness visits and immunizations), suggesting that parents are not actively avoiding medical professionals during this period.

Declines in healthcare utilization may be more likely if parents suspect the maltreatment report originated from a medical professional or other mandated reporter. While families are not informed of their reporter's identity, anecdotally they often guess the source of the report. We therefore investigate heterogeneity by reporter category in Table A2. We re-estimate effects on childrens' physical health outcomes for the full sample in Panel A, exclude referrals made by medical professionals in Panel B, and then limit the sample to referrals from non-mandated reporters in Panel C.

Column (1) presents estimates of the first stage effect on investigation separately for each subsample. The first stage is consistently significant at the 1% level, and is statistically indistinguishable across subsamples. Any differences in effects on health outcomes can thus not be explained by differential adherence to the high-risk protocol across reporter categories. Effects on injuries are reported in Column (2). Excluding medical professionals and mandated reporters does not significantly change the estimated effect on injuries, which remains statistically significant at the 10% level in each subsample. While we lose some precision on other health outcomes when excluding all mandated professionals, point estimates are statistically indistinguishable across subsamples for each outcome considered.

6 Mechanisms

Above we have documented that investigations lead to short-term improvements in child health and increases in preventative care. Investigations may directly cause these improvements through, for example, surveillance during the period of the investigation, counseling to families to improve the safety of the home or encouraging the use of preventative medical

care. In this section we consider three additional mechanisms through which investigations may impact child health: removing the child from the home, removing an abusive parent from access to the child through jail, or improving safety through decreasing parental criminal activity. While it is difficult to fully rule out or prove any one alternative mechanism, we provide suggestive evidence below.

6.1 Home Removal

After opening an investigation, CYF may remove a child from their home. As reported in Table 1, 5% of children in our sample and 7.5% of children in our analysis bandwidth are removed from their home within a year of being referred to CYF. Prior work has studied the impacts of home removals on child well-being, finding different effects in different time periods and settings. Earlier work in Illinois finds that children marginally placed in foster care have worse outcomes in terms of juvenile delinquency, teen motherhood, employment, adult crime and emergency room visits in the year after a referral (Doyle 2007, 2008, 2013).In contrast, later papers have found positive effects of removals on educational outcomes for young girls in Rhode Island (Bald et al. 2022), positive effects on educational outcomes and reductions in re-referrals,juvenile delinquency and adult crime in Michigan (Gross and Baron 2022; Baron and Gross 2022), and positive effects on educational outcomes and reductions in juvenile delinquency for children in South Carolina (Roberts 2019). To the best of our knowledge, no paper has studied the causal impacts of home removals on the child health outcomes we focus on. Still, it is possible that foster care has direct effects on child health. We explore whether our results can be explained by changes home removals, or if they are more likely the effects of investigations themselves.

We study the effect of being over the threshold for a default investigation on the cumulative likelihood of removal over time. We estimate separate regressions, setting Y_{ir} equal to an indicator for having any home removal over different time periods from the date of

referral. Figure 6 plots coefficients and 95% confidence intervals for six such regressions with cumulative time to removal on the x-axis. Effects in the 60 days following a referral are small in magnitude and statistically insignificant, suggesting that foster care placements cannot explain our short-term results on child health. While cumulative impacts increase slightly over time, they remain statistically insignificant at twelve months (360 days) after the referral date. The fact that cumulative effects increase over time is somewhat surprising, given that the first stage effect on investigations is declining over this time period. One possible explanation is that investigators consider the history of a child's involvement with CYF in making the decision to remove a child from the home. In this case, the likelihood of removal would increase with number of investigations over time. However, we suggest caution in interpreting this pattern as all but one coefficient is statistically insignificant.

In summary, short-run improvements in child health are driven by the investigation itself, as opposed to removal of the child from their home to a safer placement. Not only does this imply that investigations are protective, but also that child welfare interventions can impact outcomes for children at the margin of investigation.

6.2 Parents' Criminal Justice System Involvement

We next study the effects of investigations on parental involvement with the criminal justice system. Surveillance through investigations could increase the detection of crime, or act as a deterrent to criminal activity. If an allegation of maltreatment is substantiated that could further lead to prosecution of the involved parent. In some cases a parent could be jailed or legally restrained from visiting their children. Declines in criminal activity in the household, or the removal of an abusive parent could in lead to improvements in child health and safety.

We estimate Equation 1, where Y_{ir} is an indicator variable for each of jail stays and court appearances. Our data for these outcomes are at the monthly level. As such we estimate short-run effects in the one to two months after the referral month. We exclude the month

of the referral as we are not able to determine whether the jail stay or court appearance happened before or after the referral in that month. We estimate effects first for all parents, and then separately by parent race and gender. Recall, we define as a mother any female individual associated with the referral and given the role "parent," and fathers are defined similarly. As such there may be more than two parents associated with a referral.

Figure 7 reports results. Among all parents, receiving a risk score just over the threshold for a default investigation increases the likelihood of a jail stay in the following two months by 1.5 percentage points, or 31% relative to the mean (statistically significant at the 5% level). This effect is concentrated among Black mothers (2.3 percentage points, or 114% relative to the mean), and White fathers (3.6 percentage points, or 92% relative to the mean). White fathers also see a 6.0 percentage point increase in court appearances (17 percent relative to the mean), statistically significant at the 10% level, although there are null effects on court appearances in the full sample.[24]

We show a direct relationship between CYF investigations and parental jail time and prosecution. The removal of bad actors from the home could help explain the short term improvements in child health. If parents who are abusive or otherwise involved in criminal activity are removed from the household following an investigation, children may be living in a safer and more well-functioning environment. However, parental involvement in the criminal justice system may have additional impacts on child outcomes which we are unable to measure in our setting.[25]

[24]Note, in unreported results, we do not find any evidence that investigations affect childrens' criminal justice system involvement.

[25]Prior work studying the causal impacts of parental incarceration on children has found mixed evidence. In the U.S. context, Norris, Pecenco, and Weaver (2021) find positive effects of parental incarceration on children in Ohio in terms of reduced likelihood of incarceration and higher neighborhood quality as adults. Outside of the U.S., studies show negative impacts of parental incarceration on childrens' criminal and education outcomes in Sweden (Dobbie et al. 2018), null effects of parental imprisonment on childrens' criminal activity or educational outcomes in Norway (Bhuller et al. 2018), and positive impacts on child educational outcomes in Colombia (Arteaga 2023).

7 Conclusion

Over one third of children in the U.S. are investigated for child abuse and neglect. Despite the prevalence of such interactions, little is known about their impacts on children and families. While investigations aim to protect children from harm and direct services to families in need, they are also highly criticized for being stressful, burdensome, and in many cases unnecessary. We provide the first empirical evidence on the impacts of CPS investigations on child health and parent criminal justice system involvement.

We leverage an algorithm-generated risk score in a regression discontinuity design. Children and families above a given high-risk threshold are defaulted to an investigation, leading to a 19 percentage point jump in the probability of being screened in. By comparing outcomes for families on either side of this threshold, we identify the causal effects of investigations on child and parent well-being.

We find that investigations are protective of children's health and safety, reducing injuries and increasing preventative healthcare in the two months following a referral. Results are consistent across subgroups, and are robust to a wide set of alternative specifications. We do not find evidence of longer-term impacts on child health, suggesting that investigations may only be protective while they are ongoing.

We do not see short-term impacts on home removals for children on the margin, suggesting that investigations are protective in and of themselves. Finally, investigations increase the likelihood that some parents are jailed, suggesting a possible mechanism for improvements in child health and safety.

These findings add to the limited evidence on the impacts of early-stage interaction with the child welfare system. However, there are many potential outcomes of investigations which are not captured in our analysis. Future work is needed to expand our understanding of how investigations for maltreatment impact child and family well-being.

References

Abadie, Alberto. 2002. "Bootstrap tests for distributional treatment effects in instrumental variable models." *Journal of the American statistical Association* 97 (457): 284–292.

Allegheny County Analytics. 2018. *Allegheny County Data Warehouse.* Allegheny County, Pennsylvania: Allegheny County Department of Human Services, July.

Arteaga, Carolina. 2023. "Parental Incarceration and Children's Educational Attainment." *The Review of Economics and Statistics* 105, no. 6 (November): 1394–1410. ISSN: 0034-6535. https://doi.org/10.1162/rest_a_01129. eprint: https://direct.mit.edu/rest/article-pdf/105/6/1394/2178673/rest_a_01129.pdf. https://doi.org/10.1162/rest%5C_a%5C_01129.

Bald, Anthony, Eric Chyn, Justine Hastings, and Margarita Machelett. 2022. "The causal impact of removing children from abusive and neglectful homes." *Journal of Political Economy* 130 (7): 000–000.

Baron, E Jason, and Max Gross. 2022. *Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators.* Technical report. National Bureau of Economic Research.

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad. 2018. "Intergenerational effects of incarceration." In *AEA Papers and Proceedings,* 108:234–240. American Economic Association.

Cattaneo, Matias D, Nicolás Idrobo, and Rocıo Titiunik. 2019. *A practical introduction to regression discontinuity designs: Foundations.* Cambridge University Press.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma. 2020. "Simple local polynomial density estimators." *Journal of the American Statistical Association* 115 (531): 1449–1455.

Dobbie, Will, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. 2018. *The intergenerational effects of parental incarceration.* Technical report. National Bureau of Economic Research.

Doyle, Joseph J. 2007. "Child protection and child outcomes: Measuring the effects of foster care." *American Economic Review* 97 (5): 1583–1610.

———. 2008. "Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care." *Journal of political Economy* 116 (4): 746–770.

———. 2013. "Causal effects of foster care: An instrumental-variables approach." *Children and Youth Services Review* 35 (7): 1143–1151.

Fong, Kelley. 2018. "Concealment and Constraint: Child Protective Services Fears and Poor Mothers' Institutional Engagement." *Social Forces* 97:1785–1809.

Fong, Kelley. 2020. "Getting eyes in the home: Child protective services investigations and state surveillance of family life." *American sociological review* 85 (4): 610–638.

Goldhaber-Fiebert, Jeremy D., and Lea Prince. 2019. *Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County's Child Welfare Office.* Report prepared for Allegheny County Department of Human Services. Stanford University, March. https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-6.pdf.

Grimon, Marie-Pascale. 2020. *Effects of the Child Protection System on Parents.*

Grimon, Marie-Pascale, and Chris Mills. 2023. "The Effect of an Algorithmic Tool on Child Welfare Decision Making: A Preliminary Evaluation."

Gross, Max, and E Jason Baron. 2022. "Temporary stays and persistent gains: The causal effects of foster care." *American Economic Journal: Applied Economics* 14 (2): 170–99.

Kim, Hyunil, Christopher Wildeman, Melissa Jonson-Reid, and Brett Drake. 2017. "Lifetime prevalence of investigating child maltreatment among US children." *American journal of public health* 107 (2): 274–280.

McConnell, Margaret A, R Annetta Zhou, Michelle W Martin, Rebecca A Gourevitch, Maria Steenland, Mary Ann Bates, Chloe Zera, Michele Hacker, Alyna Chien, and Katherine Baicker. 2020. "Protocol for a randomized controlled trial evaluating the impact of the Nurse-Family Partnership's home visiting program in South Carolina on maternal and child health outcomes." *Trials* 21:1–21.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver. 2021. "The effects of parental and sibling incarceration: Evidence from ohio." *American Economic Review* 111 (9): 2926–2963.

Rittenhouse, Katherine, Emily Putnam-Hornstein, and Rhema Vaithianathan. 2023. "Algorithms, Humand and Racial Disparities in the Child Protection System: Evidence from the Allegheny Family Screening Tool."

Roberts, Dorothy. 2022. *Torn apart: How the child welfare system destroys Black families–and how abolition can build a safer world.* Basic Books.

Roberts, Kelsey V. 2019. "Foster care and child welfare." *PhD dissertation, Clemson University.*

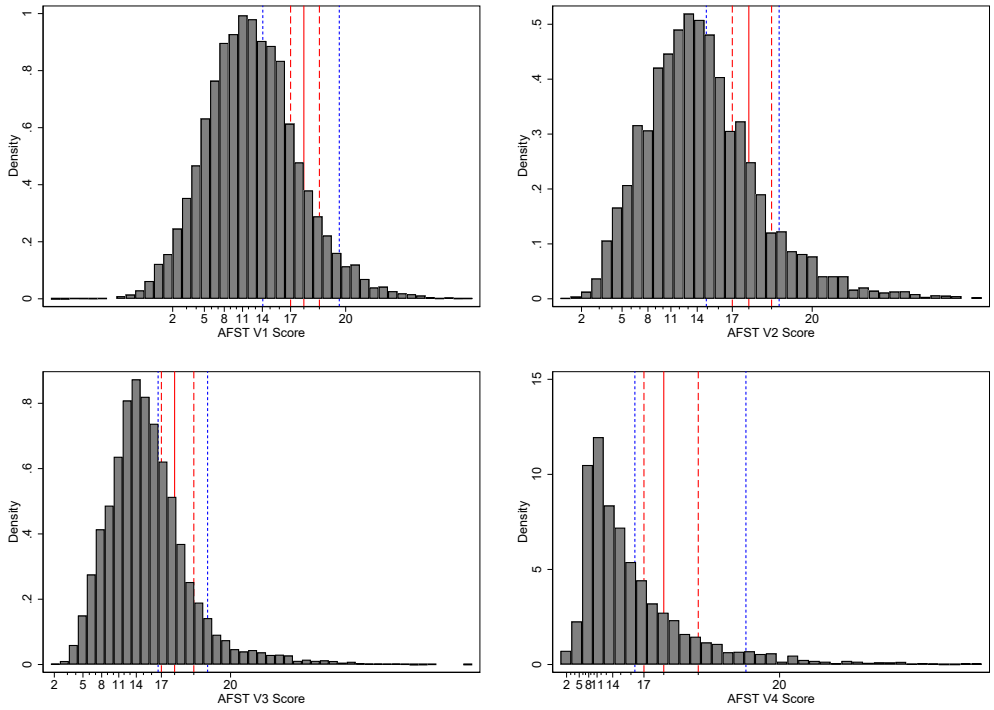
Schnitzer, Patricia G., Paula L. Slusher, Robin L. Kruse, and Molly M. Tarleton. 2011. "Identification of ICD codes suggestive of child maltreatment." *Child Abuse Neglect* 35 (1): 3–17. ISSN: 0145-2134. https://doi.org/https://doi.org/10.1016/j.chiabu.2010.06.008. https://www.sciencedirect.com/science/article/pii/S0145213411000081.

Vaithianathan, Rhema, Emily Kulick, Emily Putnam-Hornstein, and D Benavides-Prado. 2019. "Allegheny family screening tool: Methodology, version 2." *Center for Social Data Analytics,* 1–22.

Vaithianathan, Rhema, Emily Putnam-Hornstein, Nan Jiang, Parma Nand, and Tim Maloney. 2017. "Developing predictive models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation." *Center for Social data Analytics.*

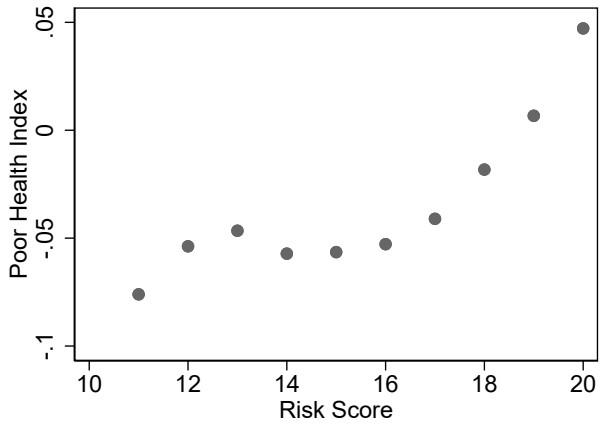
Figures

Figure 1: Density and Bandwidths by AFST Version



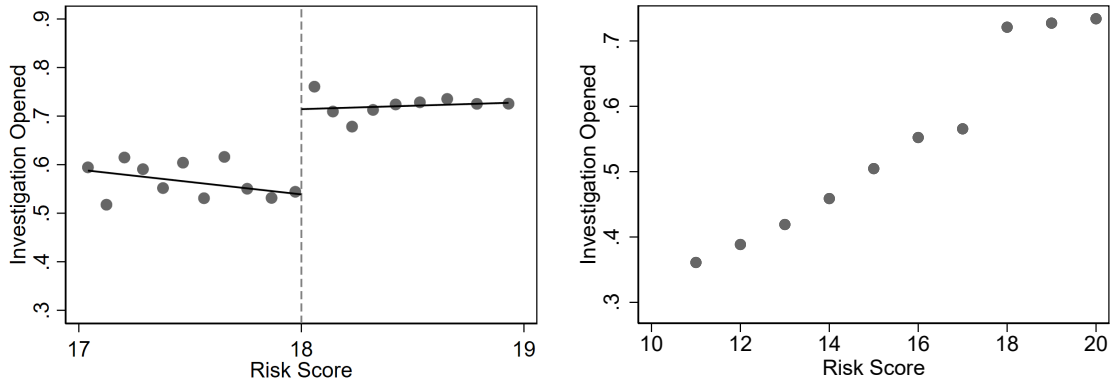
For each of the four AFST versions, this figure plots the density of latent risk score. Latent risk score is on the x-axis, with discrete scores marked by ticks. The solid vertical red line shows the cutoff for a high-risk protocol at the discrete risk score of 18. The long-dashed vertical red lines at discrete scores of 17 and 19 show the endpoints for our main bandwidths, while the short-dashed blue lines show the endpoints for the MSE-optimal bandwidth for each AFST version.

Figure 2: Poor Health Index by Risk Score



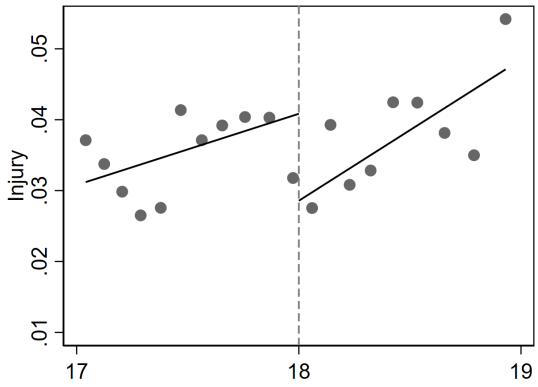
In this figure, the x-axis shows the discrete risk score for a given child-referral and the y-axis shows the average of our calculated 12-month poor health index in that risk score bin. That is, the index captures health in the year following the referral. The sample is all child-referrals with a discrete risk score greater than 10, who were enrolled in Medicaid in the month of the referral. Note that the risk score is a referral-level variable, while the poor health index is calculated at the child-referral level. The poor health index is calculated as follows: each of the binary variables for injury, abuse/neglect, lack of wellness visit and lack of immunization within 12 months from the referral date are normalized in the full sample of child-referrals. For each child-referral, we then take the average of the four normalized values, so that the index takes a value between -1 and 1 for each child-referral.

Figure 3: First Stage

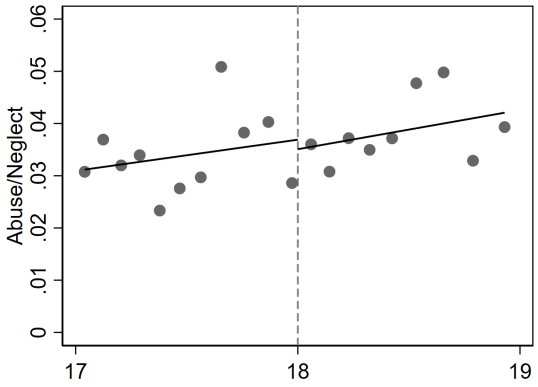


Each of these figures plots the first stage of being over the threshold for a "default" investigation on the likelihood of having an investigation opened. In panel (a), the x-axis is pooled and normalized risk score, ranging from 17 to 19 (our main RD bandwidth) The dashed vertical line shows the threshold value for a default investigation. In panel (b), the x-axis is pooled discrete risk score, ranging from scores of 11 to 20. Referrals with scores above 17 are subject to a default investigation.

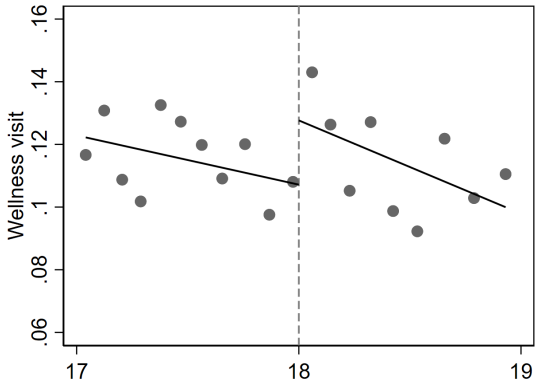
Figure 4: Regression Discontinuity Plots: Child Health



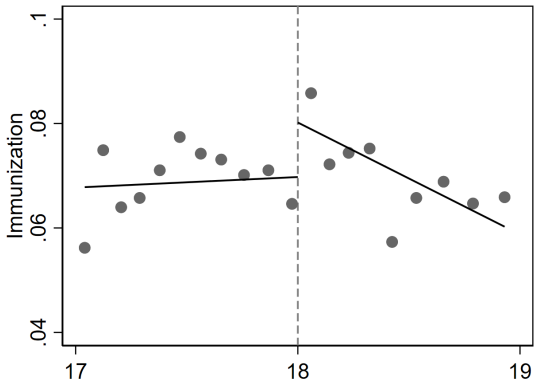
(a) Injury



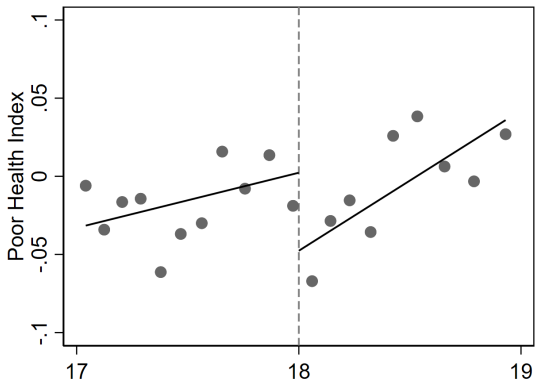
(b) Abuse/Neglect



(c) Wellness Visit



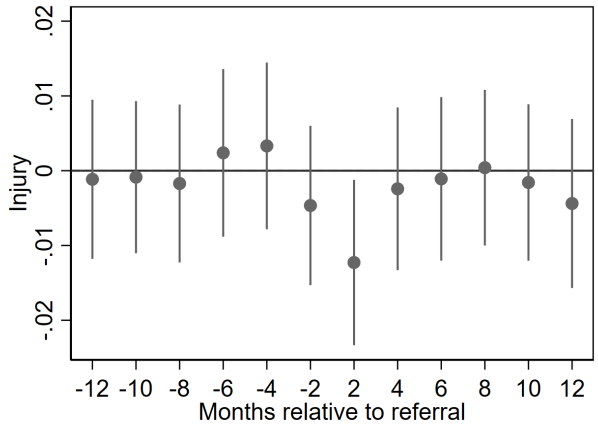
(d) Immunization



(e) Index

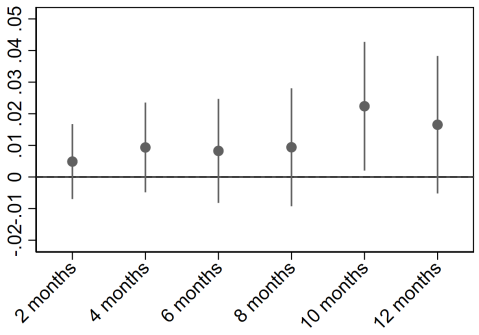
For each outcome variable reported in Table 5, this figure plots the average value in each of twenty risk score bins within the bandwdith. Risk scores to the right of the vertical line represent those which are subject to a default investigation under the high-risk protocol.

Figure 5: Dynamic Effects of Child Injuries Relative to Time of Referral



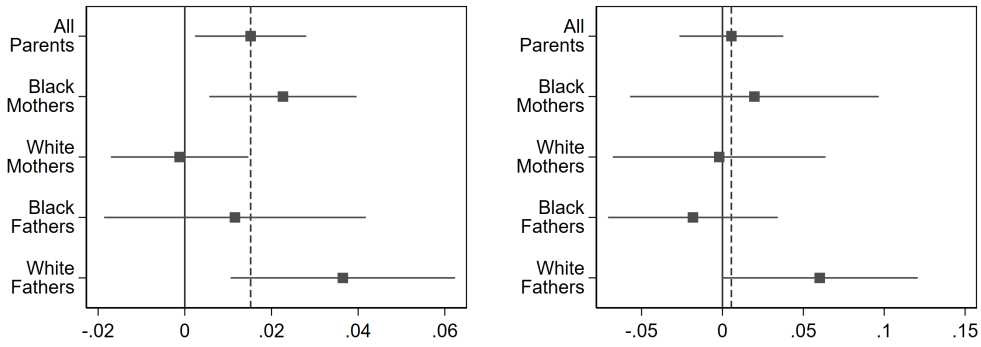
This figure plots coefficients and 95% standard errors for twelve separate regressions estimating Equation 1, where Y_{ir} is equal to injury in twelve different time periods. The x-axis shows the 60-day time period in which Y_{ir} is measured relative to the referral. The day of the referral is excluded, so that month 2 relative to the referral includes days 1-60 after the referral, while month -2 includes days 1-60 before the referral. The sample is all child-referrals in our main bandwidth (risk score of 17-18).

Figure 6: Cumulative Effect on Home Removals



This figure shows the cumulative effect of a default investigation on the likelihood of home removal in the year following a referral. Coefficients and standard errors for seven different regressions estimating Equation 1 are reported, where Y_{ir} is set to home removal within a given cumulative time time frame after the referral, reported on the x-axis.

Figure 7: Parent Criminal Justice System Involvement



(a) Jail (1-2 months)

(b) Court (1-2 months)

This figure shows coefficients and 95% confidence intervals from estimating Equation 1 for different subgroups of parents within our main analysis sample. Y_{ir} is equal to the short-run outcome variable described in the subplot label. The vertical dashed line shows the estimated coefficient for the full sample.

Tables

Table 1: Summary Statistics (Child-Referral)

	(1) Full Sample	(2) Bandwidth	(3) Complier Mean
Panel A: Child Demographics			
Female	0.494	0.506	0.507
Black	0.393	0.467	0.592
White	0.429	0.354	0.243
Age	8.290	8.765	9.096
Number of Children	2.882	3.176	3.172
Panel B: Lagged System Involvement			
Referral (Prev. 6 months)	0.278	0.430	0.422
Investigation (Prev. 6 months)	0.135	0.245	0.226
Medicaid (Prev. 6 months)	0.722	0.870	0.992
Panel C: Allegation Categories			
Caregiver Substance Abuse	0.235	0.283	0.280
Inadequate Care or Home	0.253	0.257	0.299
Neglect	0.112	0.108	0.067
Exposure to Risk	0.184	0.181	0.193
Physical Maltreatment	0.202	0.164	0.222
Panel D: Reporter Categories			
Anonymous	0.082	0.083	0.007
Legal	0.116	0.123	0.087
Medical	0.071	0.067	0.070
Parent or Relative	0.131	0.151	0.206
Educational Professional	0.155	0.135	0.271
Social Worker	0.154	0.176	0.153
Therapist	0.120	0.109	0.175
Other	0.170	0.155	0.036
Panel E: Outcomes			
Investigation Opened	0.459	0.637	0
Referral (12 months)	0.468	0.581	0.793
Home Removal (12 months)	0.049	0.075	0.025
Observations	128423	18857	

This table presents summary statistics for our main analysis sample, at the child-referral level. (I.e., one child may be included multiple times if she appears on multiple referrals.) Column (1) presents average values for the full sample, across the distribution of risk score. Column (2) presents average values for child-referrals within our main RD bandwidth (risk score of 17-18). Column (3) presents the control complier mean, calculated using the complier-describing methods of Abadie (2002). Note that this value shows average characteristics for compliers below the threshold.

Table 2: Summary Statistics (Parent-Referral)

	(1) Full sample	(2) Bandwidth	(3) Complier Mean
Panel A: Parent Demographics			
Mother	0.451	0.412	0.392
Father	0.545	0.587	0.609
Black	0.378	0.461	0.531
White	0.472	0.411	0.383
Age	36.2	36.7	37.7
Household Member	0.762	0.754	0.783
Alleged Perpetrator	0.471	0.433	0.435
Panel B: Lagged System Involvement			
Jail (Prev. 6 months)	0.047	0.079	0.048
Court (Prev. 6 months)	0.315	0.449	0.378
Referral (Prev. 6 months)	0.288	0.448	0.417
Investigation (Prev. 6 months)	0.146	0.262	0.206
Observations	130864	19005	

This table presents summary statistics for our main analysis sample, at the parent-referral level. (I.e., one parent may be included multiple times if she appears on multiple referrals.) Column (1) presents average values for the full sample, across the distribution of risk score. Column (2) presents average values for parent-referrals within our main RD bandwidth (risk score of 17-18). Column (3) presents the control complier mean, calculated using the complier-describing methods of Abadie (2002). Note that this value shows average characteristics for compliers below the threshold.

Table 3: Validity Tests

	RD Estimate	Outcome mean	Obs.
Panel A: Demographics			
Female	0.00218	0.505	18857
Black	-0.0214	0.467	18857
White	0.0226	0.353	18857
Age	-0.187	8.555	18840
Panel B: Allegation Categories			
Caregiver Substance Abuse	-0.00719	0.297	18857
Inadequate Care/Home	-0.0215	0.258	18857
Neglect	0.00740	0.105	18857
Exposure to Risk	0.0307	0.181	18857
Physical Maltreatment	-0.00749	0.167	18857
Panel C: Reporter Categories			
Anonymous	0.0264*	0.0805	18857
Legal Professional	-0.00474	0.120	18857
Medical Professional	-0.000741	0.0669	18857
Parent or Relative	-0.0169	0.144	18857
Educational Professional	-0.0166	0.131	18857
Social Worker	0.0146	0.189	18857
Therapist	-0.00320	0.111	18857
Other Reporters	0.00291	0.156	18857
Panel D: Lagged Outcomes			
Investigation (Prev. 6 months)	-0.0103	0.219	18857
Removal (Prev. 6 months)	0.00164	0.00247	18857
Medicaid (Prev. 6 months)	-0.0109	0.852	18857
Injuries (Prev. 6 months)	-0.00379	0.0795	18857
Abuse/Neglect (Prev. 6 months)	-0.0106	0.0650	18857
Wellness Visit (Prev. 6 months)	-0.0257*	0.216	18857
Immunization	-0.0105	0.153	18857
LRTI (Prev. 6 months)	-0.0000157	0.0115	18857
ACSC (Prev. 6 months)	-0.0259*	0.180	18857
Poor Health Index (Prev. 6 months)	-0.00938	0.0118	18857
Panel E: Parent Demographics and Lagged Outcomes			
Mother	0.0127*	0.415	19005
Black	-0.0287	0.460	19005
White	0.0291	0.408	19005
Age	0.204	36.30	18797
Household Member	0.00686	0.750	19005
Alleged Perpetrator	-0.00230	0.438	19005
Jail	-0.000444	0.0790	19005
Court	0.00227	0.441	19005
Investigation	-0.00326	0.235	19005
Panel F: Medicaid Enrollment			
Medicaid (1-2 months)	-0.00958	0.876	18857
Medicaid (3-12 months)	-0.00676	0.892	18857

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .001$

Column (1) of this table presents coefficients and standard errors from estimating Equation 1, setting Y_{ir} equal to each of a variety of child, referral or parent characteristics. The sample for Panels A - D and Panel F is all children on referrals with a risk score of 17-18. The sample for Panel E is all parents on referrals with a risk score of 17-18. Column (2) presents outcome means for observations to the left of the cutoff (i.e., not subject to a default investigation).

Table 4: First Stage

	(1) Investigation	(2) Any Investigation 12 months
Default Investigation	0.192***	0.0716***
	(0.0252)	(0.0200)
Mean	0.566	0.771
RightBW	8694	8727
LeftBW	10112	10128

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table reports reduced-form results from estimating Equation 1, setting Y_{ir} equal to each of the outcome variables described in the column titles. Average values of the outcome variable below the threshold, as well as the number of observations in each of the left and right bandwidth, are reported below the coefficients and standard errors. The sample is all child-referrals with a referral risk score of 17-18.

Table 5: Results: Child Physical Health

	(1) Injury	(2) Abuse/Neglect	(3) Wellness visit	(4) Immunization	(5) Index
Default Investigation	-0.0123**	-0.00174	0.0327***	0.0203**	-0.0710***
	(0.00564)	(0.00594)	(0.0118)	(0.00921)	(0.0229)
Mean	0.0355	0.0338	0.115	0.0687	-0.0164
RightBW	8723	8723	8723	8723	8723
LeftBW	10132	10132	10132	10132	10132

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table reports reduced-form results from estimating Equation 1, setting Y_{ir} equal to each of the outcome variables described in the column titles, measured in the 60 days after a referral. Average values of the outcome variable below the threshold, as well as the number of observations in each of the left and right bandwidth, are reported below the coefficients and standard errors. The sample is all child-referrals with a referral risk score of 17-18.

Table 6: Results: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Investigation						
Default Investigation	0.192***	0.230***	0.194***	0.194***	0.229***	0.221***
	(0.0252)	(0.0366)	(0.0245)	(0.0245)	(0.0357)	(0.0337)
# of Observations	18857	18857	18840	18840	18840	9746
Panel B: Injury						
Default Investigation	-0.0123**	-0.0139*	-0.0119**	-0.0116**	-0.0137*	-0.0142*
	(0.00564)	(0.00798)	(0.00564)	(0.00553)	(0.00785)	(0.00743)
# of Observations	18857	18857	18840	18840	18840	9746
Panel C: Abuse/Neglect						
Default Investigation	-0.00174	0.00168	-0.00162	-0.000664	0.00293	0.00771
	(0.00594)	(0.00830)	(0.00592)	(0.00583)	(0.00818)	(0.00775)
# of Observations	18857	18857	18840	18840	18840	9746
Panel D: Wellness Visit						
Default Investigation	0.0327***	0.0542***	0.0300***	0.0301***	0.0490***	0.0404**
	(0.0118)	(0.0175)	(0.0116)	(0.0116)	(0.0172)	(0.0163)
# of Observations	18857	18857	18840	18840	18840	9746
Panel E: Immunization						
Default Investigation	0.0203**	0.0364***	0.0189**	0.0197**	0.0342**	0.0304**
	(0.00921)	(0.0136)	(0.00909)	(0.00907)	(0.0134)	(0.0128)
# of Observations	18857	18857	18840	18840	18840	9746
Panel F: Index						
Default Investigation	-0.0710***	-0.104***	-0.0666***	-0.0678***	-0.0983***	-0.0807**
	(0.0229)	(0.0336)	(0.0227)	(0.0226)	(0.0331)	(0.0315)
# of Observations	18857	18857	18840	18840	18840	9746
Polynomial Degree	Linear	Quadratic	Linear	Linear	Quadratic	Linear
Demographic and Reporter Controls	no	no	yes	yes	yes	yes
Control for 6 month lag of outcome	no	no	no	yes	yes	yes
Bandwidth	baseline	baseline	baseline	baseline	baseline	half baseline

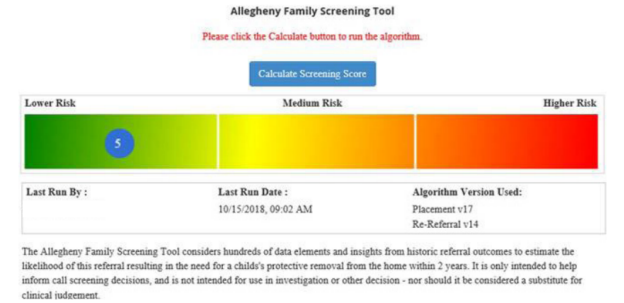
Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

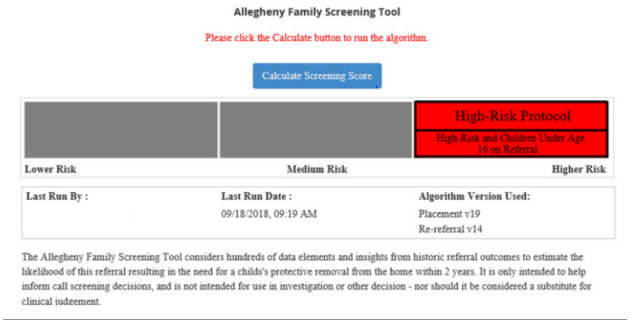
This table reports reduced-form results from estimating Equation 1, setting Y_{ir} equal to each of the outcome variables described in the panel titles. Modifications to the main analysis specification are ntoed in the table notes, where modifications include quadratic polynomial, addition of demographic and reporter controls, addition of controls for a 6-month lag of the outcome variable, and cutting the bandwidth in half. Total number of observations in the bandwidth are reported below the coefficients and standard errors.

Appendix A: Additional Figures and Tables

Figure A1: Screener View of AFST Output



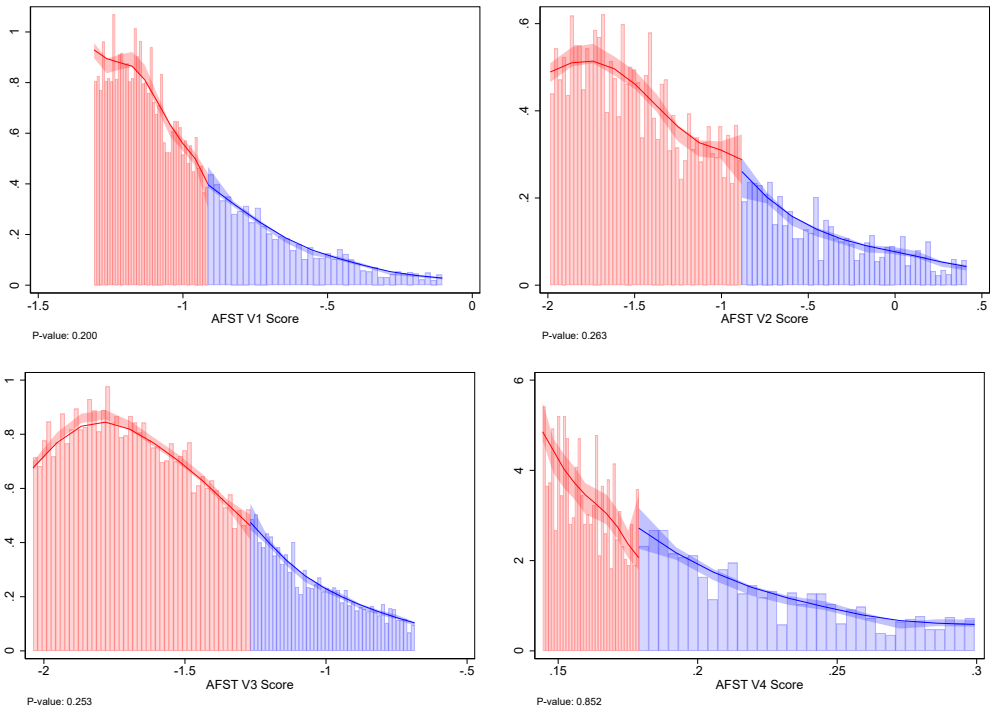
(a) No protocol



(b) High-risk protocol

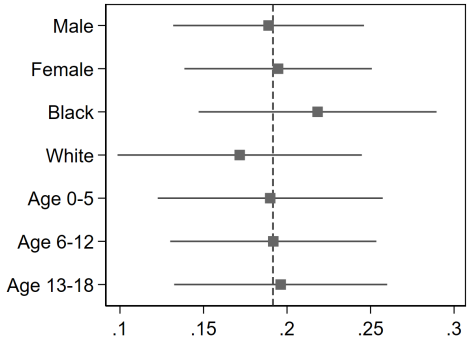
This figure shows two versions of what a screener might see after running the AFST. Panel (a) shows the AFST result for a risk score of 5, and panel (b) shows the AFST result for a high-risk protocol referral (risk score above 17).

Figure A2: Manipulation Testing



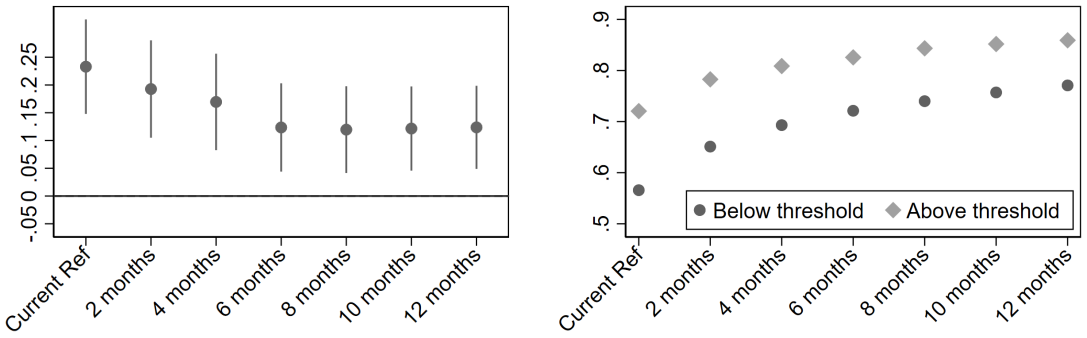
This figure reports results from manipulation testing, following Cattaneo, Jansson, and Ma (2020). Manipulation testing, allowing for separate bandwidths on each side of the cutoff, is conducted for each AFST version, and p-values are noted below each subfigure.

Figure A3: First Stage by Age and Race



This figure shows coefficients and 95% confidence intervals from estimating Equation 1 for different subgroups of our main analysis sample, where Y_{ir} is equal to one if a referral is investigated, and zero otherwise. The vertical dashed line shows the estimated coefficient for the full sample, as reported in Table 4.

Figure A4: Cumulative Effect on Investigation

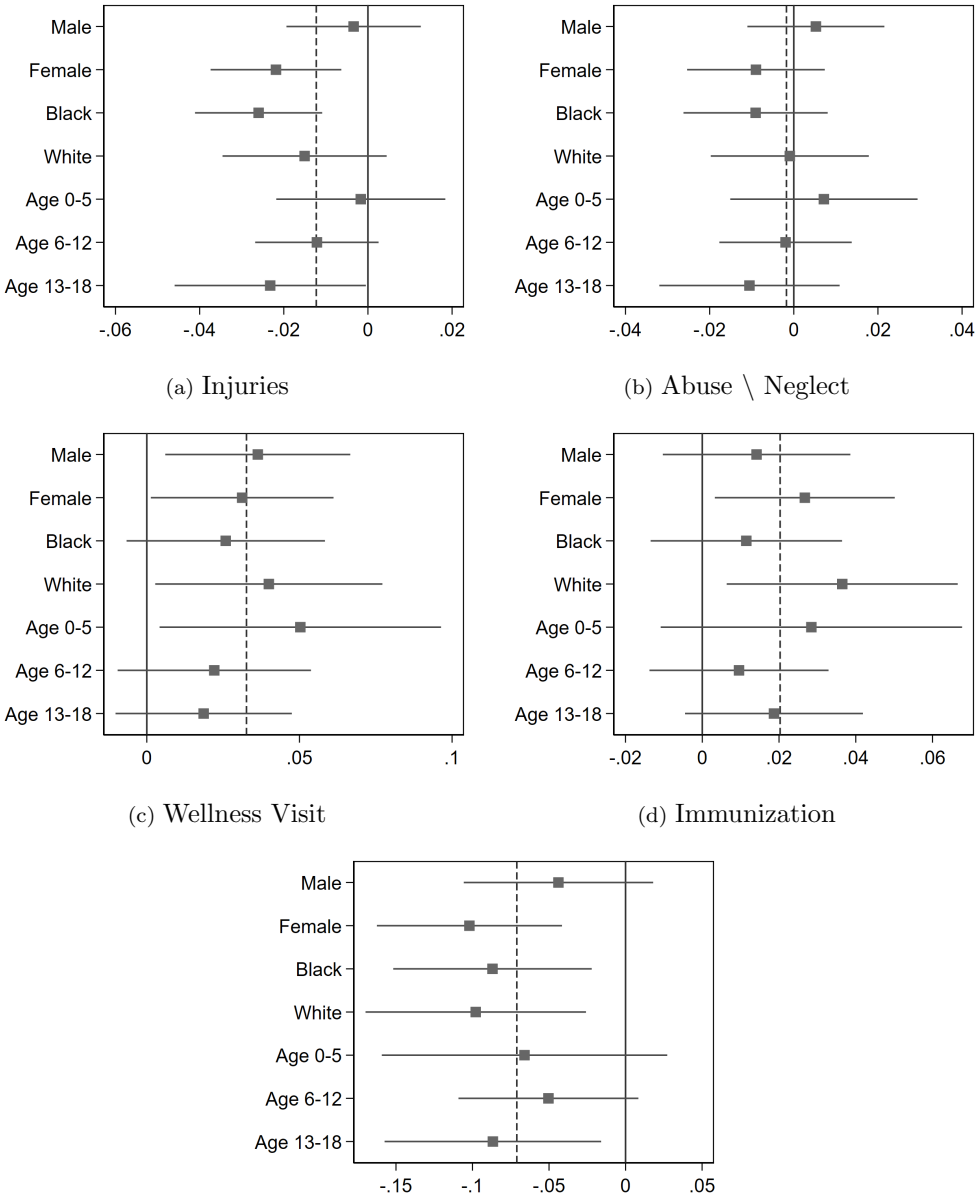


(a) RD Estimate

(b) Raw Means

This figure shows the cumulative likelihood of investigation for our treatment and control groups in the year following a referral. In panel (a), coefficients and standard errors for seven different regressions estimating Equation 1 are reported, where Y_{ir} is set to investigation within a given cumulative time time frame after the referral, reported on the x-axis. The first coefficient corresponds to Column (1) of Table 4, while the last coefficient corresponds to Column (2) of Table 4. Panel (b) reports raw means of cumulative investigation likelihood for the entire group of referrals above the threshold (light gray diamonds) and below the threshold (dark gray circles), respectively. Again, cumulative time since the referral is reported on the x-axis. The sample is all child-referrals within our main bandwidth (risk score of 17-18).

Figure A6: Child Health Results by Subgroup



(a) Injuries

(b) Abuse \ Neglect

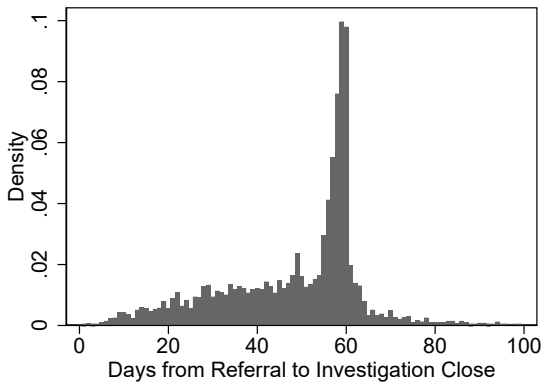
(c) Wellness Visit

(d) Immunization

(e) Poor Health Index

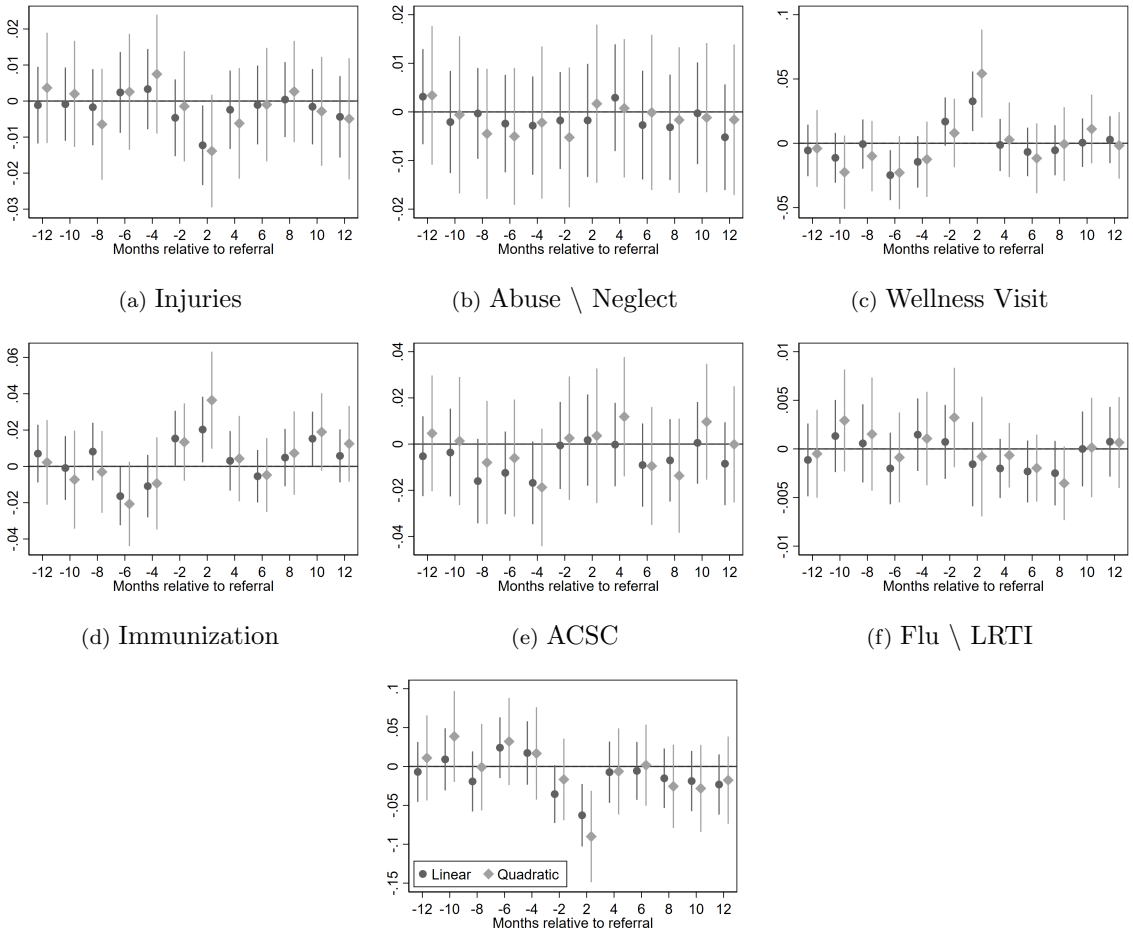
This figure shows coefficients and 95% confidence intervals from estimating Equation 1 for different subgroups of our main analysis sample, where Y_{ir} is equal to the short-run outcome variable described in the subplot label. The vertical dashed line shows the estimated coefficient for the full sample, as reported in Table 5.

Figure A7: Investigation Length



This figure shows the distribution of investigation length, defined as the days between a referral date and investigation close date, for the sample of screened-in referrals with a risk score of 17-18. One observation is included per referral. Referrals with investigation length above 100 days (2% of referrals) are excluded from the figure for readability.

Figure A9: Dynamic Effects of Child Health Outcomes Relative to Time of Referral



(a) Injuries (b) Abuse \ Neglect (c) Wellness Visit

(d) Immunization (e) ACSC (f) Flu \ LRTI

(g) Poor Health Index

Each subfigure plots coefficients and 95% standard errors for twenty four separate regressions estimating Equation 1, where Y_{ir} is equal to a child health outcome in twelve different time periods. The child health outcome is described in the subplot captions. In panel (g), the poor health index is calculated using each of the six health variables. The x-axis shows the 60-day time period in which Y_{ir} is measured relative to the referral. The day of the referral is excluded, so that month 2 relative to the referral includes days 1-60 after the referral, while month -2 includes days 1-60 before the referral. The sample is all child-referrals in our main bandwidth (risk score of 17-18). Dark gray circles reference regressions with a linear control, while light gray diamonds reference regressions which include a quadratic control.

Table A1: Child Health: Short and Long Run

	(1) Injury	(2) Abuse/Neglect	(3) Wellness visit	(4) Immunization	(5) Index (SR)	(6) ACSC	(7) Flu/LRTI	(8) Index (LR)
Panel A: 1-60 days after referral								
Default Investigation	-0.0123**	-0.00174	0.0327***	0.0203**	-0.0710***	0.00174	-0.00158	-0.0505***
	(0.00564)	(0.00594)	(0.0118)	(0.00921)	(0.0229)	(0.0101)	(0.00221)	(0.0175)
Mean	0.0355	0.0338	0.115	0.0687	-0.0164	0.0875	0.00474	-0.00133
Panel B: 61-360 days after referral								
Default Investigation	-0.00211	-0.00675	-0.0240	0.00850	0.000149	-0.0251*	-0.00587	-0.0178
	(0.0109)	(0.0107)	(0.0172)	(0.0144)	(0.0212)	(0.0151)	(0.00380)	(0.0165)
Mean	0.117	0.112	0.331	0.215	-0.0173	0.258	0.0207	0.0125
RightBW	8723	8723	8723	8723	8723	8723	8723	8723
LeftBW	10132	10132	10132	10132	10132	10132	10132	10132

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table reports reduced-form results from estimating Equation 1, setting Y_{ir} equal to each of the outcome variables described in the column titles. Average values of the outcome variable below the threshold, as well as the number of observations in each of the left and right bandwidth, are reported below the coefficients and standard errors. The sample is all child-referrals with a referral risk score of 17-18.

Table A2: Child Health: Excluding Medical and Mandated Reporters

	(1) First Stage	(2) Injury	(3) Abuse/Neglect	(4) Wellness visit	(5) Immunization	(6) Index
Panel A: Base						
Default Investigation	0.192***	-0.0123**	-0.00174	0.0327***	0.0203**	-0.0710***
	(0.0252)	(0.00564)	(0.00594)	(0.0118)	(0.00921)	(0.0229)
Mean	0.566	0.0355	0.0338	0.115	0.0687	-0.0164
RightBW	8690	8723	8723	8723	8723	8723
LeftBW	10116	10132	10132	10132	10132	10132
Panel B: Exclude Medical Reporters						
Default Investigation	0.193***	-0.0112*	0.00149	0.0326***	0.0188*	-0.0628***
	(0.0260)	(0.00581)	(0.00610)	(0.0124)	(0.00966)	(0.0238)
Mean	0.568	0.0355	0.0335	0.116	0.0692	-0.0173
RightBW	8102	8135	8135	8135	8135	8135
LeftBW	9438	9454	9454	9454	9454	9454
Panel C: Exclude Mandated Reporters						
Default Investigation	0.174***	-0.0167*	0.00139	0.0272	0.0196	-0.0673
	(0.0440)	(0.00940)	(0.0101)	(0.0205)	(0.0167)	(0.0416)
Mean	0.525	0.0322	0.0274	0.111	0.0750	-0.0339
RightBW	2864	2891	2891	2891	2891	2891
LeftBW	3131	3134	3134	3134	3134	3134

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table reports reduced-form results from estimating Equation 1, setting Y_{ir} equal to each of the outcome variables described in the column titles. Average values of the outcome variable below the threshold, as well as the number of observations in each of the left and right bandwidth, are reported below the coefficients and standard errors. In Panel A, the sample is all child-referrals with a referral risk score of 17-18. In Panel B, the sample is further limited to reports not made by a medical professional, and in Panel C, the sample is limited to reports made by non-mandated reporters.

Appendix B: Allegheny Family Screening Tool

Here we discuss additional details about the implementation and use of the AFST.

Development: First implemented in August 2016, Allegheny County Analytics contracted with external academics to create the tool. Coefficients were originally estimated (v1) using a probit model on all available referrals in Allegheny county from April 2010 and April 2014 (76,964 observations). Coefficients are only estimated a single time, and are updated (using available new data) only when each new version of the model comes out.

Allegheny County DHS transitioned to using version 2 (V2.0) of the AFST in November of 2018. The tool underwent several changes, including a transition from a probit model to a LASSO model. Both versions 1 and 2 of the AFST use administrative data on child welfare records, criminal involvement, behavioral health records, and service usage. However, version 2 of the AFST eliminated public benefit variables, such as TANF And SNAP usage. Additionally, version 2 included variables from birth record sources. See Vaithianathan et al. (2017) for details on the methodologies, including variable selection for version 1 and Vaithianathan et al. (2019) on version 2.

The third version of the model (V2.1), implemented in July of 2019, incorporated data on HealthChoices (Pennsylvania's Medicaid) enrollment, and made minor changes to lag times on JPO/ACJ variables. The fourth iteration of the AFST (V2.2) went into affect in April of 2022. In this build, the model moved to an entirely different data infrastructure, managed internally by the Data Warehouse Team. The universe of variables utilized by the model were reconstructed within the new data infrastructure, and some variable definitions were redefined. Additionally, this model retired variables on HealthChoices enrollment, birth-record-derived variables, and variables related to current allegations.

High Risk Protocol: the default screen-in has evolved over time. Version 1 applies a default screen in to investigation for anyone over the risk score threshold of 18. Versions 2, 3, and 4 additionally require that at least one of the children on the referral is younger than 16. Having at least one child under the age of 16 is common for the families in our sample.