NBER WORKING PAPER SERIES

CHILD MALTREATMENT INVESTIGATIONS AND FAMILY WELL-BEING

Katherine Rittenhouse
David Simon
Lindsey Lacey

Working Paper 34334
http://www.nber.org/papers/w34334

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2025

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 Jason Baron, Lonnie Berger, Janet Currie, Eric Chyn, Ken Couch, Joe Doyle, Delia Furtado, Chris Mills, Margaret McConnell, Analisa Packham, Marie-Pascale Grimon, Steve Ross, Rhema Vaithianathan and seminar participants at APPAM, PPCW, NBER Children's 2025 Spring Meeting, University of Connecticut, UT Austin, the University of Missouri, University of Massachusetts Amherst Department of Resource Economics, the Gray Symposium at Clark University, Texas Tech University and the 2025 Workshop on the Economics of Risky Behaviors. Allegheny County Department of Human Resources additionally provided access to the data. They had the opportunity to review these research findings before they were released. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Child Maltreatment Investigations and Family Well-being
Katherine Rittenhouse, David Simon, and Lindsey Lacey
NBER Working Paper No. 34334
October 2025
JEL No. I1, I30, J13

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 this interaction 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. Finally, investigations increase the likelihood that a parent is jailed, suggesting that investigations also detect criminal activity within the home. An exploration of mechanisms suggests that it is the investigation, rather than home removal or parental criminal justice system involvement, which improves child health.

Katherine Rittenhouse
University of Texas at Austin
katherine.rittenhouse@austin.utexas.edu

David Simon
University of Connecticut
Department of Economics
and NBER
david.simon@uconn.edu

Lindsey Lacey
Allegheny County Department of
Human Services
lindsey.lacey@alleghenycounty.us

1 Introduction

One in three children growing up in the U.S. experiences a child maltreatment investigation by age 18 (Kim et al. 2017). Investigations, in which a social worker visits a child's home to ensure their safety, are the first point of contact families have with child protection services (CPS). Designed to protect children from harm, such oversight may be essential to preventing abuse, injury, or even death. However, detractors argue that investigations impinge on parental rights, disrupt family autonomy, and disproportionately impact racial minorities with limited benefits to children.[1] 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 interaction with CPS, there is limited evidence on the effects of investigations for child maltreatment. This paper contributes novel causal evidence to this debate. Our paper is among the first to estimate the effects of early-stage interactions with CPS on child and family outcomes. Unlike prior work that focuses on foster care placements, we examine the consequences of the more common choice to investigate a family.[2] This first point of contact with CPS has been understudied despite its prevalence and potential to shape longer-term family interactions and outcomes within child protection systems. We implement a regression discontinuity design in which certain high-risk referrals are defaulted into an investigation, allowing us to isolate the impact of this interaction. We link CPS records to Medicaid claims and parental criminal justice records to directly measure impacts on child health and safety.

[1] 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 2018, 2020). Large racial disparities in family CPS involvement have led scholars and activists to critique CPS as a "family policing system" with particularly negative impacts on Black communities (Roberts 2022).

[2] While 6% of children are placed in foster care by age 18, 38% of children are investigated for maltreatment (Kim et al. 2017; Wildeman and Emanuel 2014).

The primary goal of investigations is to protect children from harm, and yet effects of interactions with CPS on child health and safety are largely unknown. We find that these early CPS interactions can yield large declines in injuries and increases in preventative care. However, these gains appear to be short lived, underscoring the need for supplementary or alternative interventions that build on these initial benefits.

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, the Allegheny Family Screening Tool (AFST), to help workers decide which families to screen in for investigation. The 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 a family 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. Balance tests on predetermined characteristics and McCrary tests in the running variable confirm the validity of this design. We then estimate reduced form effects of being over this threshold on a range of outcomes. As such our results are applicable to children on the margin of being screened in, and will not necessarily generalize to those children who would always be screened in or out.[3]

Children above the threshold for a default investigation are 1.2 percentage points (33%) less likely to have an injury-related Medicaid claim in the 60 days following a referral.[4] Effects are concentrated among injuries occurring at home, suggesting that investigations may impact the home environment in ways which improve child safety. Default investigations also increase the likelihood of wellness visits and immunizations by 3.3 percentage points (28%) and 2.0 percentage points (30%), respectively. Combining these and other measures into a "poor health index," we find that a default investigation reduces poor health by 0.07 standard deviations.

We perform a number of supplementary analyses to support these results. Subgroup analysis shows that effects are similar across child age, race and gender, suggesting that investigations are protective across demographic groups. We show that our results are driven by improvements in health rather than declines in healthcare utilization. Specifically, we find similar reductions in "deferrable" and "non-deferrable" injuries, suggesting our results are not driven by injuries that can be treated without medical attention. Our results are also robust to dropping those referrals most likely to discourage families from seeking health care.[5] Finally, our results are robust to a wide set of specifications such as alternative bandwidths,

[3]Our empirical approach does not allow us to speak to the effects of investigations for all children. Our results apply to those children around the high-risk threshold who are only investigated due to the high-risk protocol.

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

[5]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.

alternative functional forms of the running variable, and the inclusion of different sets of controls.

We explore several possible mechanisms for improvements in child health, including child foster care placement and parental criminal justice system involvement.[6] The effect of an investigation on the likelihood of home removal is small in magnitude and statistically insignificant during the time period where we see impacts on child health, suggesting our results are not driven by foster care placements. Turning to parental outcomes, we find that 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.[7] A mediation analysis shows that declines in child injuries are similar across households where parents are more vs. less likely to be jailed. This suggests that removal of a parent involved in criminal activity is not a primary mechanism for our main results. All together, our analysis of potential mechanisms implies that it is the investigations themselves that are protective. Surveillance with the implicit threat of home removal may provide a strong incentive for parents to change their behavior during the period of investigation.

We make three major contributions to the literature on the causal effects of child protection systems. First, we study a previously-unexplored intervention. 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; Helénsdotter, Forthcoming).[8] We instead study the effects of investigations,

[6]Parental incarceration has been shown to have both positive and negative effects on children in various contexts (Norris, Pecenco, and Weaver 2021; Dobbie et al. 2018; Bhuller et al. 2018; Arteaga 2023).

[7]This analysis presents the first causal evidence of the effects of child maltreatment investigations on parental outcomes.

[8]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. Few papers that estimate effects of CPS interactions on child well-being have looked at margins other than home removal. 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.

an intervention experienced by one third of children in the U.S. Second, we study effects of CPS interactions on child health outcomes, which directly relates to the stated goal of child welfare agencies to protect children from harm.[9] Third, we identify impacts for a novel group of children - those at the margin of investigation.[10] These children are likely to be better off and at lower risk than those at the margin of home removal. Our results show that early stage child welfare interventions can improve outcomes for this group of children.

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 do not study the direct effects of algorithms, but instead exploit a feature of the AFST as exogenous variation to study the effects of investigations on child and family well-being. That said, identification in our quasi-experiment comes from an interaction between human and algorithm decision making. Our results imply that, at least for some children, information provided by an algorithm can help identify at-risk children who would be passed over by human judgment alone.

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.

[9] Prior work on the effects of foster care has tended to focus on outcomes related to education and criminal justice system involvement (Doyle 2007, 2008; Bald et al. 2022; Gross and Baron 2022; Baron and Gross 2022; Roberts 2019). Exceptions include Doyle (2013), who finds increases in emergency room visits and Helénsdotter (Forthcoming), who shows increases in mortality driven by suicides.

[10] To the best of our knowledge, we are also the first to use a regression discontinuity design to study effects of interactions with the child welfare system.

2 Background

2.1 Allegheny County Child Welfare Investigations

In Allegheny County, 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. After a screener makes their recommendation, a supervisor will make the final decision on whether to open an investigation.[11] 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 or less severe forms of child maltreatment. 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 majority of reports in Allegheny county, just as allegations of neglect make up the majority of child protection referrals nationwide.[12]

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.[13] To achieve these aims, the caseworker assigned to the investigation conducts one

[11] 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.

[12] 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. For more details on the screening decision see details at https://www.alleghenycounty.us/Human-Services/Programs-Services/Children-Families/When-a-Report-is-Made.aspx.

[13] 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.

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 at least every one to four weeks throughout the period of investigation. All home visits over the course of an investigation may be unannounced, such that parents cannot predict when monitoring for child safety will occur. 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, 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 of 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 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. A CYF case involves periodic check-ins with the family, suggested or mandated services (including, e.g., mental health or substance abuse services), or, typically as a last resort, removal of the child from the home. Both mandated services and 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. The AFST is designed 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. From these variables, the model calculates a continuous latent risk score for each child on a referral.[14]

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.[15] This referral-level latent score is mapped to discrete values between 1 and 20 based on predetermined cutoffs.[16] 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

[14]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 AFST has been updated several times since its introduction and has used different input variables and estimators to make its predictions depending on the AFST version. For a complete list of the variables used in a recent version of the AFST, see Vaithianathan et al. (2019). 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. Race was not included as a predictor. Initially, the AFST used a probit model to predict latent risk scores, while later versions use a LASSO model. See Appendix B for more details on the how the AFST score is constructed and its various versions.

[15]The original version of the AFST introduced in August was substantially revised in December 2016 to reduce the large number of missing scores that were generated for many children. Given these changes in how the score was implemented, we follow Goldhaber-Fiebert and Prince (2019) and drop referrals before December 2016. See Appendix B for more details.

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

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." Figure A1 shows what the screener would see when the score falls below and above the cutoff for a default screen-in.

3 Data

We obtained administrative data from the Allegheny County Department of Human Services on the universe of child maltreatment referrals to CYF from June 2016 through December 2023.[17] 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 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 then create four main categories of Medicaid claims to measure child health around the time of any given 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; Hooft et al. 2013; McConnell et al. 2020). The final two categories measure well-child visits and

[17] In our main analyses we limit our sample to referrals through December 2022, ensuring that we keep a one-year follow up period where we observe all outcomes for each referral.

immunizations, respectively.[18] 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.[19] Figure A2 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.

Finally, we use several variables from Allegheny County criminal justice records. Namely, we connect parents to monthly indicators of their criminal justice system involvement, including jail stays and 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. In our analysis sample we exclude three categories of referrals. First, as discussed above, calls including allegations of physical or sexual abuse are mandated by the state to be investigated, and as such are excluded from our sample. We also exclude referrals involving families with an active CYF case or ongoing investigation. 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. We therefore drop referrals (across all AFST versions for consistency) where there is not a child under the age of 16 in the household.[20] Our main

[18]Table A1 lists the ICD-10 codes associated with each category.

[19]To create the index we first multiply the indicators for wellness visits and immunizations by negative one, so that having these services decreases poor health. We then 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.

[20]Note, while these three categories of referrals are excluded from our analysis sample, they are included

bandwidth (referrals with scores of 17 or 18) includes 18,854 child-referrals, including 12,726 unique children and 7,809 unique referrals.

All outcomes are observed 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).[21] Approximately half of child-referrals are female and 39% are Black. Note, this indicates a significant over-representation of Black children relative to their share of the population.[22] On average about three children are associated with a referral. In the six months prior to the given referral, 28% of children were included on a different referral, 14% 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.[23] Referrals are most likely to come from educational professionals (16%), social workers (15%), and parents or relatives (13%). 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 relative to the full sample. They are more likely to have been on a previous referral (43%) and had an investigation (24.5%) within the prior six months. The vast majority are enrolled in Medicaid (87%). Child-referrals within the bandwidth are also more likely to be investigated (63.7%), and removed from the home (7.5%) within twelve months.

Next, Table 2 presents summary statistics at the parent-referral level, separately for all

in our calculation of past and future CYF involvement at the individual-referral level.

[21]Our calculation of complier means is discussed in greater detail in Section 4.1 below.

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

[23]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.

parents associated with referrals to CYF (Column 1), parent-referrals which fall within our main bandwidth (Column 2), and compliers (Column 3). Our main bandwidth includes 19,001 parent-referrals, including 12,581 unique parents and 7,794 unique referrals. We define an individual as a mother if their stated role is "parent" and their gender is female. We define fathers similarly. Note, this definition means 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 model:

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

Where Y_{irt} is the outcome for individual i who is associated with referral r that occurs in month-year t. 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 for a default investigation. 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 specification clusters at the referral level to account for auto-correlation within referrals. We use a triangular kernel and a linear functional form, although results are robust to alternative decisions.

Figure 1 presents the distribution of algorithm scores separately by AFST version.[24] 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).

Each version's respective MSE-optimal bandwidth is also 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. Thus the bandwidth choice for our main analyses is conservative, intuitive for our setting, and allows for consistent estimation across models. That being said, for each of our main results, we test robustness to alternative bandwidths. We additionally test robustness of our main specification by adding controls for demographics and referral case characteristics and using a quadratic form of the running variable.

[24] We present formal manipulation tests in Figure A3, discussed in Section 4.2.

4.1 Compliers

Our regression discontinuity design identifies the Intent-to-Treat (ITT) of a default investigation on child and family outcomes. The decision to investigate in this context comes from an interaction between human and algorithm decision making. Treatment effects are identified off of families which the AFST flags as high risk, but who call screeners would have chosen not to investigate. By investigating the characteristics of these compliers, we can learn about the types of high-risk individuals who are passed over by human judgment and may benefit from additional information provided by the AFST. To better understand the characteristics of this group, we use the method in Abadie (2002), which involves modifying Equation 1 to describe characteristics of compliers directly below the threshold.[25] These control complier means are included in Column (3) of Tables 1 and 2. Compliers are largely similar to children within the wider analysis bandwidth in terms of age and gender, lagged system involvement, and allegation categories. They are more likely to be Black, and less likely to be reported by Anonymous or Other sources, which are typically considered less reliable.[26] 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 similar to parents within the wider analysis bandwidth, though they are also more likely to be Black, and are less likely to have been jailed 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 might arise if call screeners pay more attention to cases which are identified as high-risk by

[25]Specifically, we estimate a fuzzy regression discontinuity design, where a lack of investigation opened is instrumented for by a risk score below the threshold, and Y_{irt} is set equal to the value of the relevant observable characteristic if an investigation is not opened, and zero if an investigation is opened.

[26]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 were subject to the high risk protocol.

the algorithm, 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. Discussions with call screeners at CYF revealed that they run the AFST only after examining the details of the referral and making a preliminary decision. Further, screening out a high-risk protocol requires a supervisor to override the default investigation, making it administratively burdensome 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 calculated using pre-determined administrative histories of the children and parents on the referral, and call screeners and families do not have the ability to change these inputs. Running the algorithm multiple times does not change the score. Informal discussions with call screeners suggest that they do not attempt to change or manipulate the score in any way.[27]

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 A3 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.

[27]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.

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 variables of interest. Out of the 34 regressions, three 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 being over the threshold for a default screen-in results in a strong first stage effect on investigations. The high-risk protocol may have little to no effect if most referrals with a score below the cutoff are also screened in for an investigation. Likewise, there would be little first stage power if many referrals above the cutoff are screened out through a supervisor override.

The first stage regression discontinuity graph is shown in Figure 2. 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 risk scores. Again, we see a clear and discontinuous increase around the threshold of a risk score of 18.

We next turn to regression estimates of the first stage. We estimate Equation 1, where Y_{irt} 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 A4 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_{irt} equal to an indicator for having at least one investigation in the twelve months following the referral. 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.

We explore this future treatment of the control group more in Figure A5. In Panel (a) we report coefficients from seven separate regressions, setting Y_{irt} equal to an indicator for any investigation within different time periods from the date of reference referral (i.e., the extensive margin over the year). 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 effect of being over the threshold on the likelihood of any investigation attenuates steadily over the six months following the reference referral before stabilizing at around 7 percentage points for months 6-12. Control children are re-referred and investigated relatively quickly after the reference referral. Panel (b) performs a similar exercise for the intensive margin. Here, Y_{irt} is set equal to the number of investigations in each time period over the year following the reference referral. We find that the coefficient is largely stable over time, attenuating slightly after 6 months from 0.19 to 0.15. Together, these analyses imply that while many control children are eventually screened in, treated children are also re-referred and re-investigated in the year following the reference referral.

Having established the effect of being over the high-risk threshold on investigations, we next turn to the impact of a default investigation on child health outcomes.

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 may direct families to health services, provide counseling to improve safety of the home, and monitor the physical safety and medical care of children, all of which may lead to improved health outcomes. However, investigations are disruptive and may increase stress within the family or cause fractures between families and the medical system, potentially 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 being over the high risk threshold on outcomes in the 60 days after a call, when an investigation is likely ongoing. During this period, families are subject to visits from a caseworker, which may or may not be announced in advance. Investigations are typically mandated to last a maximum of 60 days. Figure A6 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 Medicaid 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 1.17 percentage points less likely to have an injury-related Medicaid claim (statistically significant at the 5% level). This represents a 33% decline in injuries relative to the control group mean of 3.5%. Being defaulted into an investigation increases the probability of having a wellness visit by 28% (3.24 percentage points), and getting an immunization by 30% (2.02 percentage points), statistically significant at the 1% and 5% level, respectively. There are no statistically significant effects on ICD 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).[28] Being over the high risk threshold decreases the poor health index by 0.07 standard deviations in the first two months after a call, statistically significant at the 1% level.

In order to translate these reduced form estimates into the effect of an investigation, we divide by the first stage estimate of 0.192 (as estimated in Table 4). This implies that investigations cause a 6 percentage point reduction in the likelihood of injury (0.0117/0.192), a 16.8 percentage point increase in the likelihood of a wellness visit, and a 10 percentage point increase in the likelihood an immunization in the two months following a referral. While these effects are large, they occur during the period of an intensive treatment: surveillance by a professional social worker with the power to remove the child from the home, and who can visit the house unannounced.

We use additional information from the Medicaid claims to further explore the observed reduction in injuries. Specifically, injury diagnoses are in some cases paired with an ICD-10 code describing the place the injury occurred. Using this data, we separately estimate effects for injuries which occur at home, not at home, and injuries not associated with a

[28]In unreported results, we limit abuse/neglect codes to the relevant age groups as defined in Schnitzer et al. (2011). Effects are close to zero and statistically insignificant.

place of occurrence.[29] Table A2 reports results. Children subject to a default investigation are 0.3 percentage points (134%) less likely to have an injury occurring at home in the two months following a referral. This effect is statistically significant at the 5% level. We do not find statistically significant effects on injuries which do not occur at home, and the point estimate corresponds to a 12% decline relative to the control group mean. Finally, we find a 1.3 percentage point (47%) decline in injuries which are not associated with any place of occurrence code. These results suggest that investigations have a stronger effect on injuries occurring in the home, consistent with changes in the home environment improving child safety.

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 A7. Recall that the first stage is largely similar across subgroups. We do not find statistically distinguishable effects across subgroups, though we see suggestive evidence that reductions in injuries are somewhat 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 the child health index, suggesting that investigations are beneficial across all groups of complier children.

5.1 Dynamics

Next we consider how child health effects evolve over the year following a referral. For a given child health outcome, we estimate Eq. 1 separately for Y_{irt} measured in each 60-day window up to 360 days after the referral, for a total of six regressions per outcome. We exclude Medicaid claims made on the day of the referral, as we cannot determine whether the referral occurred prior to or after the Medicaid claim on that day.

[29] Table A1 lists the codes associated with each category. We define a visit as an injury which occurs in a given place if both an injury code and place of occurrence code is associated with a given child on a given date.

Figure 4 presents results for each health outcome. For injuries, wellness visits, immunizations, and the poor health index we see improvements in health during the 60 days following the referral, corresponding to the results in Table 5. However, in all these cases the estimates fall close to zero and become statistically insignificant after 60 days. One interpretation of this pattern is that health benefits primarily accrue during the period of investigation, and the large effects we document during the investigation do not result in lasting safety improvements to the child's home environment.

Alternatively, this pattern may reflect the fact that the control group "catches up" to the treated group, attenuating estimates of longer-term effects.[30] We argue that this explanation is less consistent with the body of evidence. First, if such "catch-up" were driving the dynamics, we would expect a gradual attenuation of the estimates as more of the control group incrementally gets investigated over the six months following the reference referral (see Figure A5 (a)). Instead, we see a near immediate decline in the magnitude of the health effects following the investigation period across all health outcomes. Further, recall that the control group only catches up on the extensive margin. As shown in Figure A5, treated children are investigated at similar rates to control children in the year following the reference referral. If each investigation has similar effects on child health outcomes, then control group "catch up" on the extensive margin would not impact our ability to detect longer-term effects.

To further investigate if eventual control group treatment on the extensive margin is attenuating our effects, we implement a dynamic regression discontinuity design (DRDD). This approach isolates the effect of a default investigation for a focal child-referral, holding constant past and future investigation trajectories.[31] This insulates our estimates from

[30]As discussed in Section 4.3, many control compliers are re-referred and investigated within a year from the reference referral.

[31]The dynamic RDD was introduced in Cellini, Ferreira, and Rothstein (2010) in the context of bond authorizations for school facility investments, and has been developed in more recent work including Baron (2022) and Biasi, Lafortune, and Schönholzer (2025).

control group contamination that occurs when screened-out children under the high risk threshold are later investigated.

We present the estimating equation and results for the DRDD in Appendix C. Overall, the results are similar to the simple dynamics estimates shown in Figure 4, though the estimated standard errors are larger. Effects remain concentrated in the 60 days following a referral, attenuating once the period of investigation ends. This suggests that control group catch up is not driving our results. Thus results are most consistent with the hypothesis that child health improvements last only in the 60 days following a referral, when an investigation is likely ongoing. Active surveillance may be a key mechanism for the short-term improvements in health, and alternative interventions may be necessary in order to induce longer-term change. We explore potential mechanisms in more detail in Section 6.

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. Column 2 includes a quadratic rather than linear form for the running variable. This more flexible functional form ensures we are not mistaking a quadratic change in outcomes for a discontinuity around the cutoff. Column 3 shows results using a linear running variable but including additional controls.[32] If assignment to treatment is truly random at the cutoff then additional controls should not materially change the estimated coefficient. The fourth column additionally controls for a six month lag of the value of the outcome variable. The

[32]We include as controls demographic variables (Black, White, 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).

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 specifications and outcomes, the estimated coefficients remain similar to the baseline. Effects are statistically indistinguishable across specifications and, for the most part, retain similar levels of statistical significance.[33] Across outcomes, including a quadratic form of the running variable or cutting the bandwidth in half tends to increase the estimated discontinuity.

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 CPS. 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 in the 60 days after the referral, when an investigation is likely ongoing. During an investigation parents may be explicitly told to get necessary medical care for their children, and the parents know that the social worker could come to

[33] For injuries, specifying a quadratic running variable or cutting the bandwidth in half causes the standard errors to increase resulting in a loss of statistical significance. However, the estimated coefficient actually becomes slightly larger than in the baseline specification.

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.

Next, we separately estimate effects on "deferrable" vs. "non-deferrable" injuries. If results are driven by avoidance behavior, we would expect null effects on more serious, non-deferrable injuries which are thought to require immediate medical attention. We follow prior work in defining non-deferrable injuries as those observed at similar rates on weekends vs. weekdays (Dobkin 2003; Card, Dobkin, and Maestas 2009; Doyle Jr et al. 2015). The intuition behind this approach is that diagnoses which are closer to uniformly distributed are less likely deferred to the weekends. To implement this approach, we first calculate the share of weekend visits associated with each 3-digit injury code for children in our sample.[34] We then create four distinct cutoffs, defining non-deferrable injuries as those 3-digit ICD codes with weekend rates within each of 2, 3, 4 and 5 percentage points of 2/7.[35] ICD-10 codes included in each definition of non-deferrable injuries are listed in Table A1. For example, the three codes with weekend rates closest to 2/7 are: traumatic amputation of the wrist, hand and fingers; foreign body in alimentary tract; and open wound of the head.

We estimate Equation 1 separately for each definition of deferrable and non-deferrable injuries. Coefficients and standard errors are reported as percent of the control group mean in Figure A9. While the means are smaller for non-deferrable injuries, making estimates somewhat noisier, we find similar declines in percent terms for both categories. This pattern is inconsistent with avoidance behavior, given that non-deferrable injuries also see a decline.

Finally, declines in healthcare utilization may be more likely if parents suspect the mal-

[34] Here the sample is defined as all children receiving a referral in our time period, regardless of whether they are in our primary analysis bandwidth.

[35] In doing this, we also attempted to define non-deferrable injuries as those 3-digit ICD codes with weekend rates within 1 percentage point of 2/7, but this definition includes only two ICD codes, making estimation infeasible for this group.

treatment 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 Figure A8. Estimated coefficients for the full sample are shown as dark gray squares, estimates for a subsample excluding reports by Medical professionals are shown as light gray diamonds, and estimates for a subsample excluding reports by all mandated reporters are shown as hollow circles. While we lose some precision when excluding all mandated professionals, point estimates are similar and statistically indistinguishable across subsamples for each outcome considered. Together, these analyses suggest that reductions in injuries are driven by improvements in health, rather than changes in healthcare utilization.

6 Mechanisms and Other Outcomes of Interest

Above we have documented that investigations lead to short-term improvements in child health and increases in preventative care. Investigations may directly lead to these improvements by, for example, surveillance during the investigation period, providing families with counseling that enhances home safety, or promoting the use of preventative health care. Alternatively, investigations may be the starting point for future interventions which ultimately promote child safety such as home removals. In this section we explore three additional outcomes which may act as mechanisms. We study the effect of being over the high-risk threshold on removing the child from the home, removing a parent from the home through jail, or bringing parents to court for criminal activity. These analyses are intrinsically interesting beyond their use in uncovering potential mechanisms. To the best of our knowledge, no prior work has studied the effect of investigations on these outcomes.

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.[36] 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 in home removals, or if they are more likely the direct effects of investigations themselves.

We estimate 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_{irt} equal to an indicator for having any home removal over different time periods from the date of referral. Figure 5 plots coefficients and 95% confidence intervals for six such regressions. 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. In summary, it seems that short-run improvements in child health are not driven by the removal of the child from their home.

[36] Earlier work in Illinois finds that placing the marginal child in foster care leads to 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 but increases in juvenile delinquency for children in South Carolina (Roberts 2019). Finally, Helénsdotter (Forthcoming) studies foster care placements in Sweden, finding that marginally-removed children have higher mortality rates, largely driven by suicide.

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. Substantiation of alleged maltreatment could 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 potentially lead to improvements in child health and safety.[37]

We estimate Equation 1, where Y_{irt} is an indicator variable for either parental jail stay or court appearance. 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.

Panels (a) and (b) of Figure 6 shows the regression discontinuity plot for each outcome using the full sample of 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). Effects on court appearances are close to zero and statistically insignificant in the full sample.[38]

[37] Several recent studies examine the effects of parental incarceration on childrens' long-run outcomes. The conclusions of this literature are mixed, with studies finding positive, negative and null effects on child outcomes in various contexts (Norris, Pecenco, and Weaver 2021; Dobbie et al. 2018; Bhuller et al. 2018; Arteaga 2023). To the best of our knowledge, no study has causally identified the effects of parental incarceration on child health and safety in the short run.

[38] Finding impacts on jail but not court appearances could be due to parents being initially jailed, followed by the charges being dropped before a court hearing was held. We confirmed with Allegheny County that it is common in PA for an alleged criminal to be first put in jail. If held in jail, a parent must be brought before the judge or have charges dropped. Alternatively, court appearances might be more likely to happen during the month of the referral, while jail time happens over the following months. Since we don't count occurrences during the of the referral due to the coarseness of our data, this may explain the absence of an

Panels (c) and (d) of Figure 6 plot the coefficients and standard errors from estimating Equation 1 on the full sample, and separately by parental race and gender subgroups. The positive effect of a default investigation on jail time 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). We also find a 6 percentage point (17%) increase in court appearances for White fathers, statistically significant at the 10% level.[39]

It is possible that the removal of bad actors from the home could be beneficial to child health. If parents who are abusive or otherwise involved in criminal activity are removed from the household, children may end up living in a safer environment. To investigate, we predict for each parent in our bandwidth the effect of a default investigation on likelihood of jail stays given their baseline characteristics. Specifically, we estimate the following regression discontinuity equation:

$$Y_{irt} = \alpha + \beta_1 D_r + \beta_2 z_r + \beta_3 (D_r \times z_r) + \mathbf{X} D_r + \mathbf{X} z_r + \mathbf{X}(D_r \times z_r) + \gamma_t + \epsilon_{irt} \quad (2)$$

Here, all notation is analogous to notation in Eq. 1. Y_{irt} is an indicator for parent i being placed in jail in the two months following referral r. $\mathbf{X}$ is a vector of baseline coefficients at the parent and referral level.[40] We estimate the predicted treatment effect for each parent ($\hat{TE_i}$) by summing the estimated coefficients on D_r and $\mathbf{X}D_r$ interacted with $\mathbf{X}_i$. For each referral, we take the maximum $\hat{TE_i}$ among all parents associated with the referral. We then split referrals into those with above-median and below-median predicted treatment effects.

effect on court appearances.

[39] In unreported results, we do not find evidence that investigations affect childrens' criminal justice system involvement.

[40] We include allegation and reporter categories as defined in Table 1, indicators for each of Black mothers, Black fathers, White mothers and White fathers; parent's age; indicators for whether they are a member of the household and whether they are an alleged perpetrator; and all lagged system involvement variables described in Panel B of Table 2.

We estimate Eq. 1 separately for each of these groups, setting Y_{irt} equal to indicators for each of parental jail, open investigation, and our five child health outcomes of interest. Results are reported in Table A3. Column (1) shows that parents on referrals with high *predicted* treatment effects are indeed more likely to be induced to jail time by a default investigation. Column (2) shows that the first stage effect on investigations is similar across referrals with high vs. low predicted treatment effects. Columns (3) through (7) reveal that child health effects are broadly similar in magnitude across these two groups, in particular when evaluating effects relative to the mean. The group of referrals with low predicted jail effects sees a 0.06 standard deviation reduction in the poor health index, while the group with high predicted jail effects sees a 0.09 standard deviation reduction, both significant at the five percent level. Likewise, the magnitude of the decline in injuries is similar for both groups, though slightly smaller and not statistically significant for children of parents with high predicted jail effects.

This analysis suggests that parental jail time is not the main mechanism for the improvements in child health, as both groups see similar sized effects. Overall, our results are most consistent with the hypothesis that the surveillance and counseling received by the family during the period of the investigation is responsible for improvements in child health and safety. Further intervention may be necessary in order to extend these improvements beyond the time period of the investigation.

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 often unnecessary.

We provide the first causal 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 to identify the causal effects of investigations on child and family well-being. Children 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. For marginal children, investigations reduce injuries and increase preventative healthcare in the two months following a referral. Scaling by the first stage, we find that investigations lead to a 6 percentage point reduction in injuries, a 16.8 percentage point increase in having a wellness visit, and a 10 percentage point increase in the likelihood of an immunization in the two months following a referral. We do not find longer-term impacts on child health, suggesting that investigations may only be protective while they are ongoing. Turning to parents, we find that investigations increase the likelihood of jail time in the two months following a referral.

We explore several possible mechanisms for our main results. We do not see short-term impacts of investigations on home removals, suggesting that health improvements are not driven by foster care placements. While parental jail may in theory improve child outcomes, we find in our setting that the magnitude of the health effects are similar across children whose parents are more vs. less likely to be jailed. As such, this is unlikely to be the main mechanism behind our child health results. Instead, investigations themselves appear to be protective, at least while they are ongoing. Active surveillance and the threat of home removal during the period of investigation may lead parents to invest resources in child health and safety in the short run.

Our findings contribute new evidence to the debate around the role of the child welfare system in families' lives. We identify positive effects of investigations for complier children. Critics of CPS may argue that other interventions could achieve similar results without the associated costs to parents. Calls for defunding or abolishment of CPS often include a suggestion that funds be reallocated to cash assistance or other material supports for families (Roberts 2022). Indeed, a large body of evidence shows that both cash and in-kind transfers may have long-term benefits to children.[41] While cash or in-kind transfers may address underlying economic needs, it is not clear that providing these supports to the marginal families in our sample would be as effective at promoting short term health improvements. Our results show that the more targeted intervention of home investigations leads to large reductions in injuries, potentially addressing more immediate family crises which may not be solely due to unmet economic needs. Additional research is needed to better understand which interventions work for which children. Moreover, 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.

[41] For an overview, see Aizer, Hoynes, and Lleras-Muney (2022).

References

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

Aizer, Anna, Hilary Hoynes, and Adriana Lleras-Muney. 2022. “Children and the US social safety net: Balancing disincentives for adults and benefits for children.” *Journal of Economic Perspectives* 36 (2): 149–174.

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. 2022. “School spending and student outcomes: Evidence from revenue limit elections in Wisconsin.” *American Economic Journal: Economic Policy* 14 (1): 1–39.

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.

Biasi, Barbara, Julien Lafortune, and David Schönholzer. 2025. “What works and for whom? effectiveness and efficiency of school capital investments across the us.” *The Quarterly Journal of Economics,* qjaf013.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. “Does Medicare save lives?” *The quarterly journal of economics* 124 (2): 597–636.

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.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. 2010. “The value of school facility investments: Evidence from a dynamic regression discontinuity design.” *The Quarterly Journal of Economics* 125 (1): 215–261.

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.

Dobkin, Carlos. 2003. "Hospital staffing and inpatient mortality." *Unpublished Working Paper.*

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.

Doyle Jr, Joseph J, John A Graves, Jonathan Gruber, and Samuel A Kleiner. 2015. "Measuring returns to hospital care: Evidence from ambulance referral patterns." *Journal of Political Economy* 123 (1): 170–214.

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

———. 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.

Helénsdotter, Ronja. Forthcoming. "Surviving Childhood: Health Effects of Removing a Child From Home." *Review of Economic Studies.*

Hooft, Anneka, Jocelyn Ronda, Paula Schaeffer, Andrea G Asnes, and John M Leventhal. 2013. "Identification of physical abuse cases in hospitalized children: accuracy of International Classification of Diseases codes." *The Journal of pediatrics* 162 (1): 80–85.

Hornby Zeller Associates, Inc. 2018. *Allegheny County Predictive Risk Modeling Tool Implementation: Process Evaluation.* Technical report. Allegheny County Department of Human Services. https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Process-Evaluation-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-4.pdf.

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.

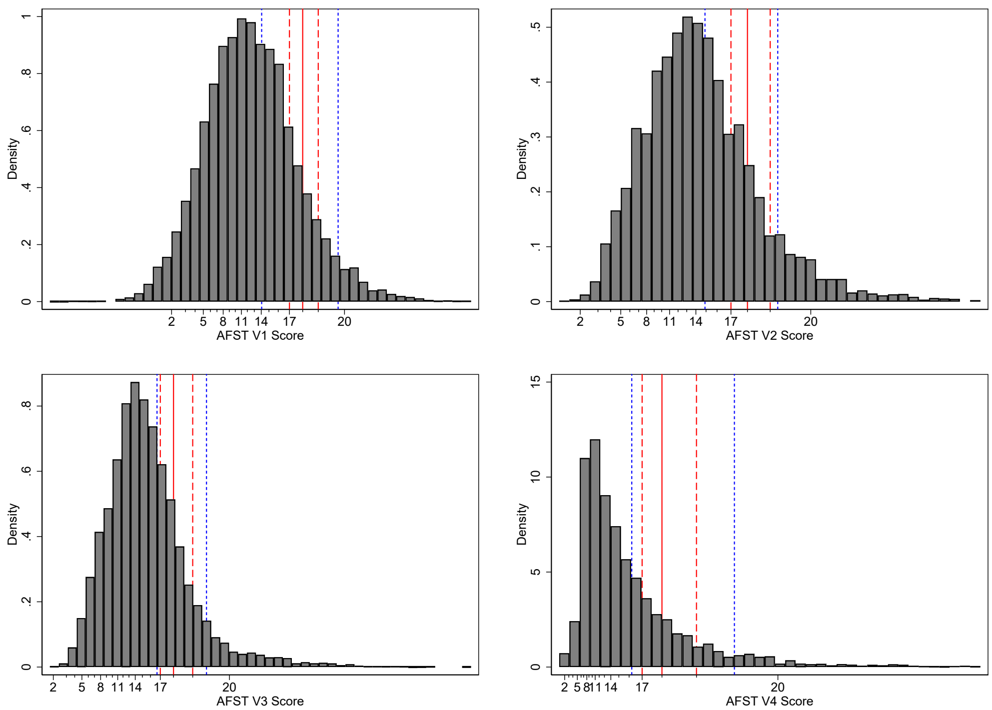
Vaithianathan, Rhema, Emily Kulick, Emily Putnam-Hornstein, and D Benavides-Prado. 2019. "Allegheny family screening tool: Methodology, version 2." *Center for Social Data Analytics,* 1–22.

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.*

Wildeman, Christopher, and Natalia Emanuel. 2014. "Cumulative risks of foster care placement by age 18 for US children, 2000–2011." *PloS one* 9 (3): e92785.

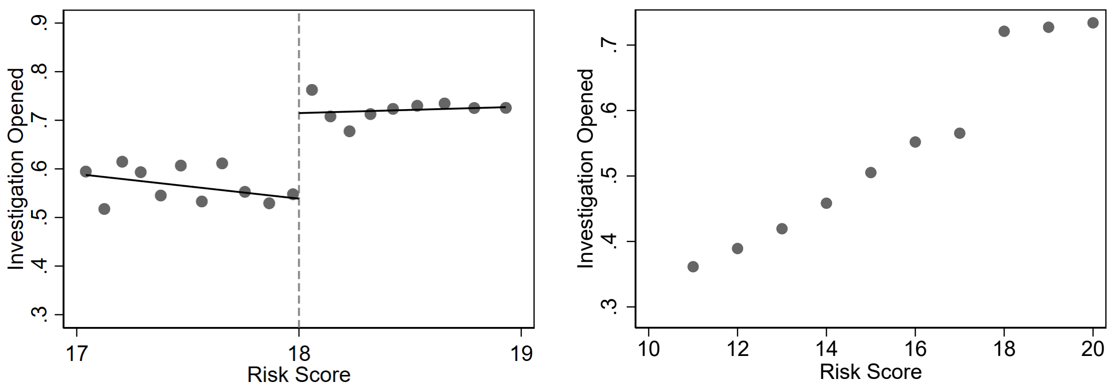
Figures

Figure 1: Density and Bandwidths by AFST Version



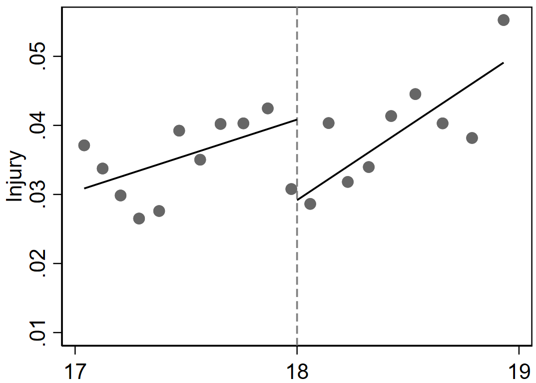
This figure plots the density of the latent risk score for each of the four AFST versions. Several discrete risk scores are marked with ticks on the X-axis. The solid vertical red line shows the cutoff for a high-risk protocol at the discrete risk score of 18. The long-dashed vertical red lines show the cut offs for discrete scores 17 and 19 (the two endpoints of our main bandwidth). The short-dashed blue lines show the endpoints of the MSE-optimal bandwidth for each AFST version.

Figure 2: First Stage

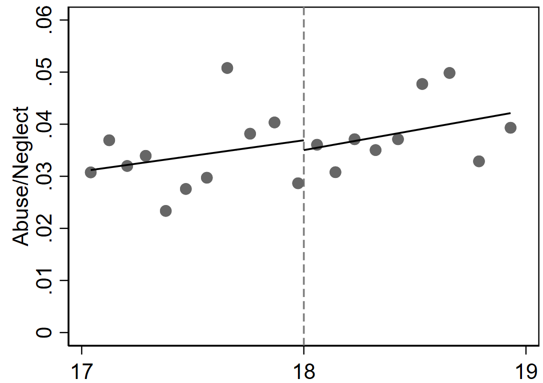


Each of these figures plots the first stage of being over the threshold for a "default" investigation on the likelihood of having an investigation opened. In panel (a), the x-axis is the pooled and normalized risk score, ranging from 17 to 19 (our main RD bandwidth). Markers plot the probability of investigation in twenty quantile-spaced risk bins. 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.

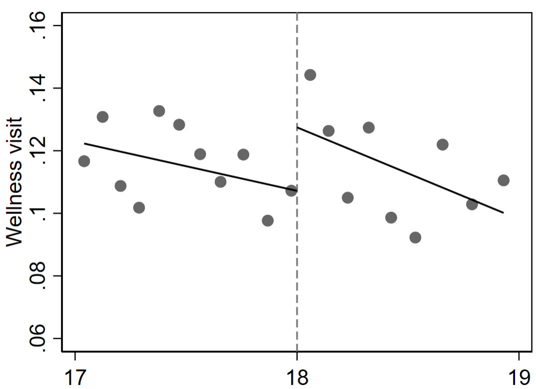
Figure 3: Regression Discontinuity Plots: Child Health



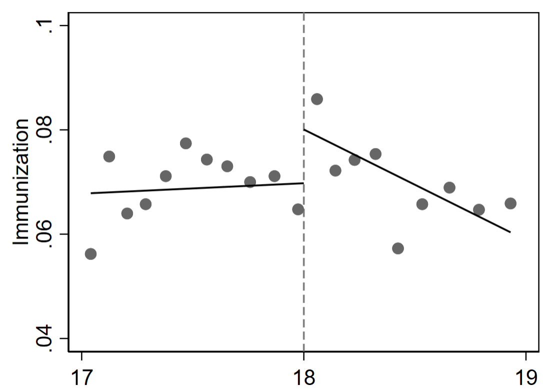
(a) Injury



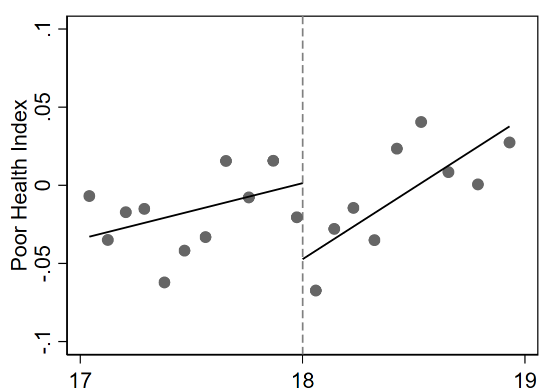
(b) Abuse/Neglect



(c) Wellness Visit



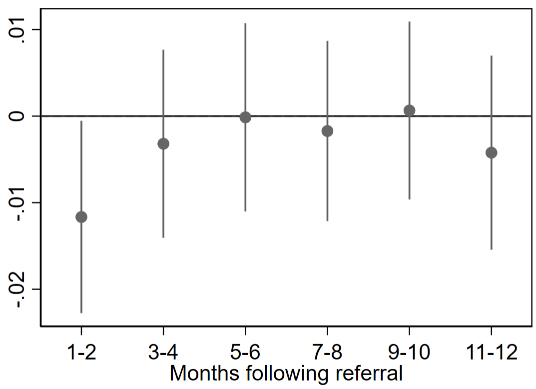
(d) Immunization



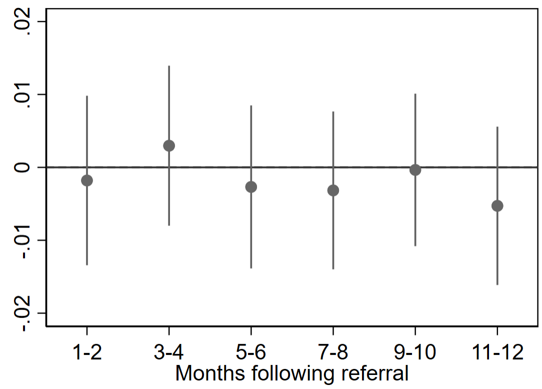
(e) Index

For each outcome variable reported in Table 5, this figure plots the average value in each of twenty quantile-spaced risk score bins within the bandwdith. See Table A1 for details on the ICD codes used to construct each outcome. Risk scores to the right of the vertical line represent those which are subject to a default investigation under the high-risk protocol.

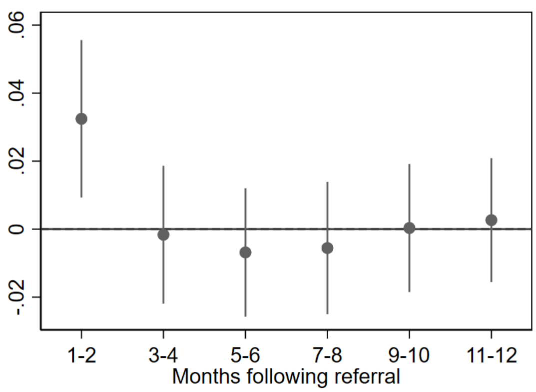
Figure 4: Dynamic Effects on Child Health Outcomes Relative to Time of Referral



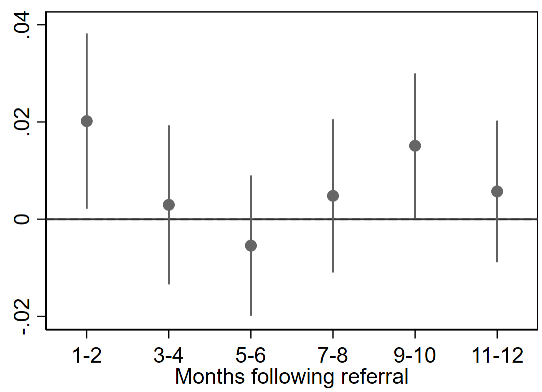
(a) Injuries



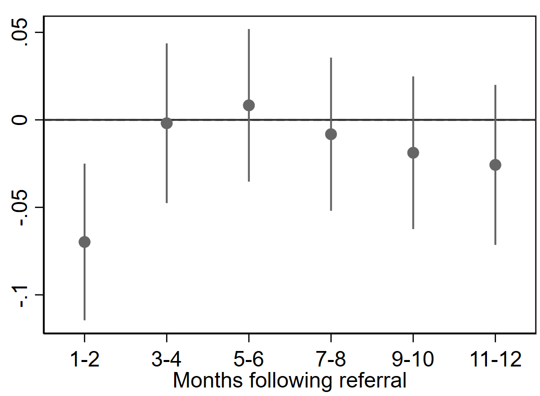
(b) Abuse / Neglect



(c) Wellness Visit



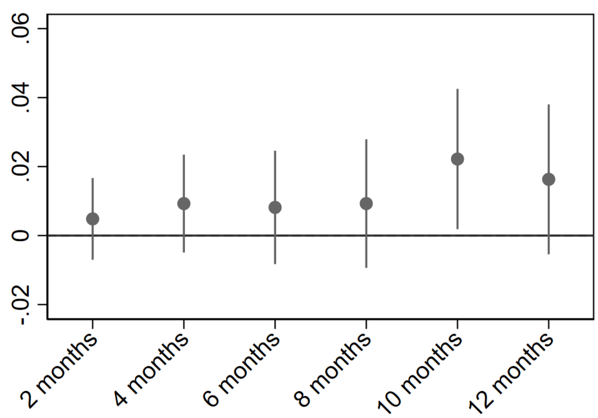
(d) Immunization



(e) Poor Health Index

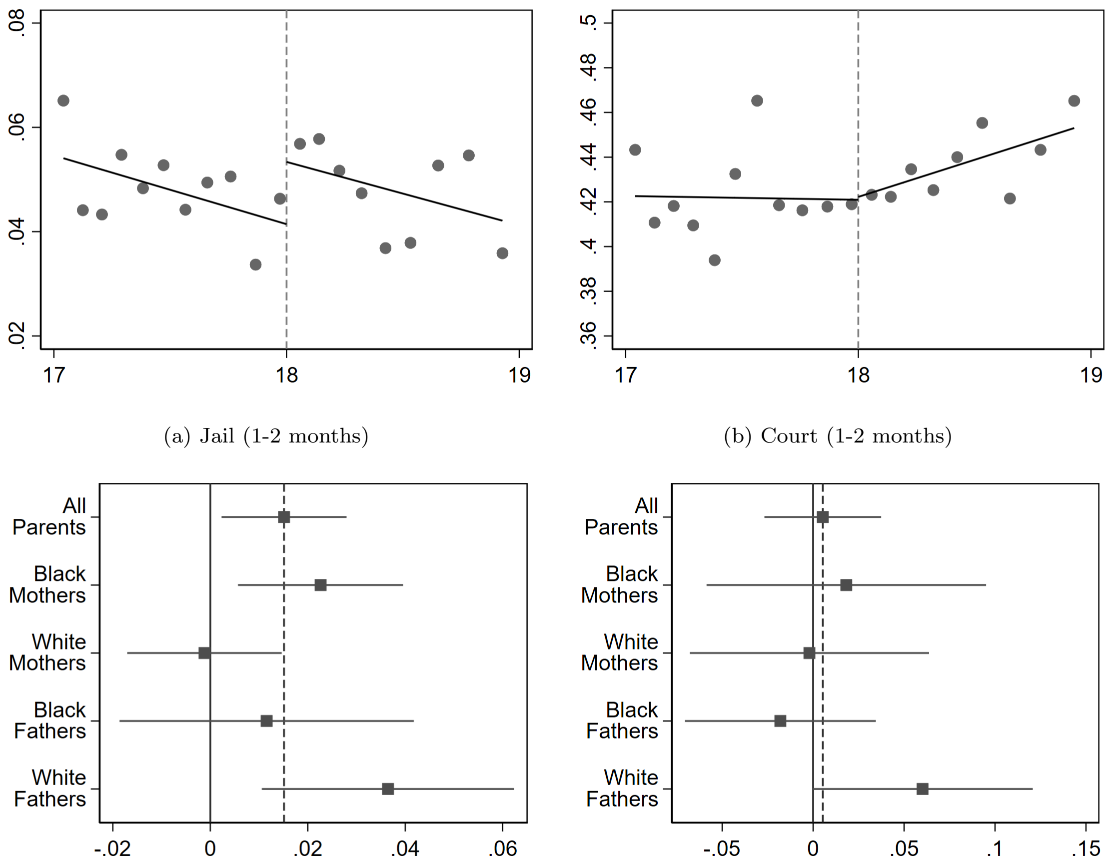
Each subfigure plots coefficients and 95% confidence intervals for 6 separate regressions estimating Equation 1, where Y_{irt} is equal to a child health outcome in six different time periods. The child health outcome is described in the subplot captions. See Table A1 for details on the ICD codes used to construct each outcome. The x-axis shows the 60-day time period in which Y_{irt} is measured relative to the referral. The day of the referral is excluded, so that month 1-2 includes days 1-60 after the referral. The sample is all child-referrals in our main bandwidth (risk scores of 17-18).

Figure 5: 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 confidence bands from seven different regressions estimating Equation 1, are plotted on the Y-axis. In each regression Y_{irt} indicates, starting from the time after referral, home removal within the given cumulative time frame as reported on the x-axis.

Figure 6: Parent Criminal Justice System Involvement



(a) Jail (1-2 months)

(b) Court (1-2 months)

(c) Jail (subgroup effects)

(d) Court (subgroup effects)

Panels (a) and (b) respectively show the regression discontinuity plots on parental jail time and court appearances in the 1-2 months following the referral. The x-axis is the pooled and normalized risk score, ranging from 17 to 19 (our main RD bandwidth). Markers plot the probability of the outcome of interest occurring in twenty quantile-spaced risk bins. The dashed vertical line shows the threshold value for a default investigation. Panels (c) and (d) plot coefficients and 95% confidence intervals from estimating Equation 1 for different subgroups of parents within our main analysis sample. Y_{irt} 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	0.591
White	0.429	0.354	0.244
Age	8.290	8.765	9.094
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	0.227
Medicaid (Prev. 6 months)	0.722	0.870	0.992
Panel C: Allegation Categories			
Caregiver Substance Abuse	0.235	0.283	0.279
Inadequate Care or Home	0.253	0.257	0.298
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	0.069
Parent or Relative	0.131	0.151	0.206
Educational Professional	0.155	0.135	0.270
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	18854	

This table presents summary statistics for our main analysis sample, at the child-referral level (one child may be included multiple times if she appears on multiple referrals.) Column (1) presents means for the full sample, across the full distribution of risk scores. 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	19001	

This table presents summary statistics for our main analysis sample, at the parent-referral level (parents may be included multiple times if they appear on multiple referrals.) Column (1) presents average values for the sample, across the full distribution of risk scores. 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: Balance Tests

	RD Estimate	Outcome mean	Obs.
Panel A: Demographics			
Female	0.00232	0.505	18854
Black	-0.0212	0.466	18854
White	0.0218	0.353	18854
Age	-0.192	8.559	18837
Panel B: Allegation Categories			
Caregiver Substance Abuse	-0.00573	0.297	18854
Inadequate Care/Home	-0.0219	0.258	18854
Neglect	0.00708	0.105	18854
Exposure to Risk	0.0302	0.181	18854
Physical Maltreatment	-0.00787	0.167	18854
Panel C: Reporter Categories			
Anonymous	0.0262*	0.0806	18854
Legal Professional	-0.00507	0.120	18854
Medical Professional	-0.000857	0.0669	18854
Parent or Relative	-0.0172	0.144	18854
Educational Professional	-0.0165	0.131	18854
Social Worker	0.0161	0.189	18854
Therapist	-0.00354	0.111	18854
Other Reporter	0.00259	0.156	18854
Panel D: Lagged Outcomes			
Investigation (Prev. 6 months)	-0.0107	0.219	18854
Removal (Prev. 6 months)	0.00164	0.00247	18854
Medicaid (Prev. 6 months)	-0.0116	0.852	18854
Injuries (Prev. 6 months)	-0.00564	0.0795	18854
Abuse/Neglect (Prev. 6 months)	-0.0106	0.0651	18854
Wellness Visit (Prev. 6 months)	-0.0262*	0.216	18854
Immunization	-0.0107	0.153	18854
Poor Health Index (Prev. 6 months)	0.00735	-0.0139	18854
Panel E: Parent Demographics and Lagged Outcomes			
Mother	0.0130*	0.415	19001
Black	-0.0285	0.460	19001
White	0.0288	0.408	19001
Age	0.200	36.31	18793
Household Member	0.00679	0.750	19001
Alleged Perpetrator	-0.00209	0.438	19001
Jail (Prev. 6 months)	-0.000495	0.0791	19001
Court (Prev. 6 months)	0.00207	0.441	19001
Investigation (Prev. 6 months)	-0.00325	0.235	19001
Panel F: Medicaid Enrollment			
Medicaid (1-2 months)	-0.00936	0.875	18854
Medicaid (1-12 months)	-0.00633	0.898	18854

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

Column (1) of this table presents coefficients and standard errors from estimating Equation 1, setting Y_{irt} 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 (our main RD bandwidth). The sample for Panel E is all parents on referrals with a risk score of 17-18. Column (2) presents control group means (means for observations to the left of the cutoff).

Table 4: First Stage

	(1)	(2)
	Investigation	Any Investigation 12 months
Default Investigation	0.192***	0.0713***
	(0.0252)	(0.0200)
Mean	0.566	0.771
RightBW	8691	8724
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_{irt} 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: Child Physical Health

	(1) Injury	(2) Abuse/Neglect	(3) Wellness visit	(4) Immunization	(5) Index
Default Investigation	-0.0117**	-0.00180	0.0324***	0.0202**	-0.0698***
	(0.00567)	(0.00593)	(0.0118)	(0.00921)	(0.0228)
Mean	0.0353	0.0338	0.116	0.0687	-0.0175
RightBW	8724	8724	8724	8724	8724
LeftBW	10128	10128	10128	10128	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_{irt} equal to each of the outcome variables described in the column titles, measured in the 60 days after a referral. See Table A1 for details on ICD codes used to construct each outcome. 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: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Investigation						
Default Investigation	0.192***	0.231***	0.194***	0.194***	0.229***	0.221***
	(0.0252)	(0.0365)	(0.0245)	(0.0244)	(0.0357)	(0.0336)
Observations	18805	18805	18788	18788	18788	18788
Panel B: Injury						
Default Investigation	-0.0117**	-0.0128	-0.0113**	-0.0107*	-0.0121	-0.0112
	(0.00567)	(0.00792)	(0.00566)	(0.00556)	(0.00778)	(0.00738)
Observations	18854	18854	18837	18837	18837	18837
Panel C: Abuse/Neglect						
Default Investigation	-0.00180	0.00159	-0.00122	-0.000176	0.00344	0.00823
	(0.00593)	(0.00828)	(0.00591)	(0.00582)	(0.00818)	(0.00775)
Observations	18854	18854	18837	18837	18837	18837
Panel D: Wellness Visit						
Default Investigation	0.0324***	0.0538***	0.0300***	0.0301***	0.0484***	0.0394**
	(0.0118)	(0.0174)	(0.0116)	(0.0116)	(0.0171)	(0.0163)
Observations	18854	18854	18837	18837	18837	18837
Panel E: Immunization						
Default Investigation	0.0202**	0.0362***	0.0190**	0.0198**	0.0341**	0.0297**
	(0.00921)	(0.0136)	(0.00908)	(0.00906)	(0.0134)	(0.0128)
Observations	18854	18854	18837	18837	18837	18837
Panel F: Index						
Default Investigation	-0.0698***	-0.102***	-0.0649***	-0.0657***	-0.0945***	-0.0737**
	(0.0228)	(0.0334)	(0.0226)	(0.0225)	(0.0328)	(0.0312)
Observations	18854	18854	18837	18837	18837	18837
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

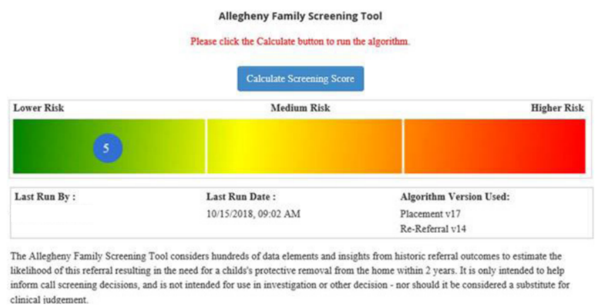
Standard errors in parentheses

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

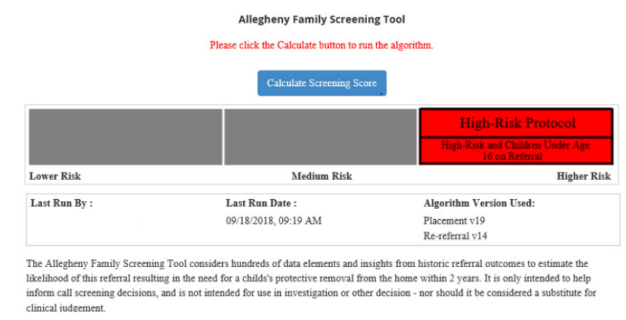
This table reports reduced-form results from estimating Equation 1, setting Y_{irt} equal to each of the outcome variables described in the panel titles. Modifications to the main analysis specification are noted in the table footer. The first column is identical to our baseline result. The second column includes a quadratic polynomial form for the running variable. The third column, using a linear form of the running variable, adds additional demographic and reporter controls. The fourth column adds controls for a 6-month lag of the outcome variable, the fifth column includes all of the previous controls and a quadratic form of the running variable. Finally, the fifth column cuts the baseline 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



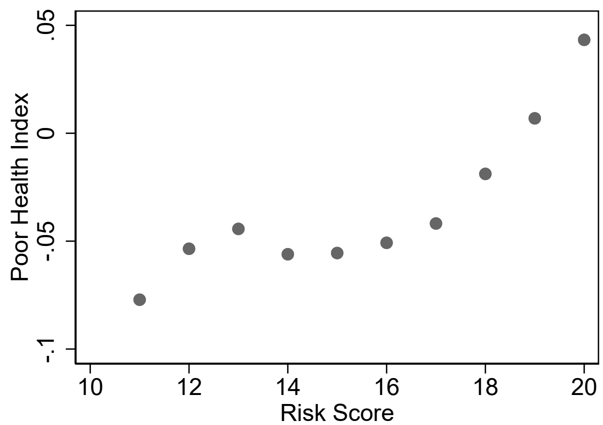
(a) No protocol



(b) High-risk protocol

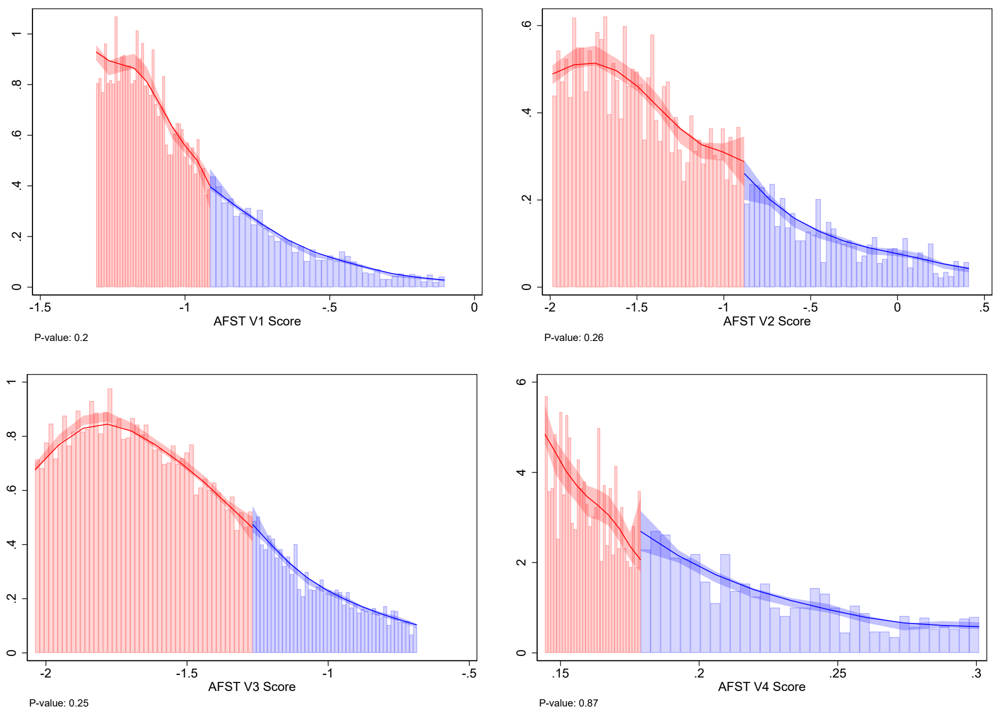
This figure shows two versions of what a screener might see after running the AFST. Panel (a) shows the AFST result for a risk score of 5, and panel (b) shows the AFST result for a high-risk protocol referral (risk score above 17).

Figure A2: Poor Health Index by Risk Score



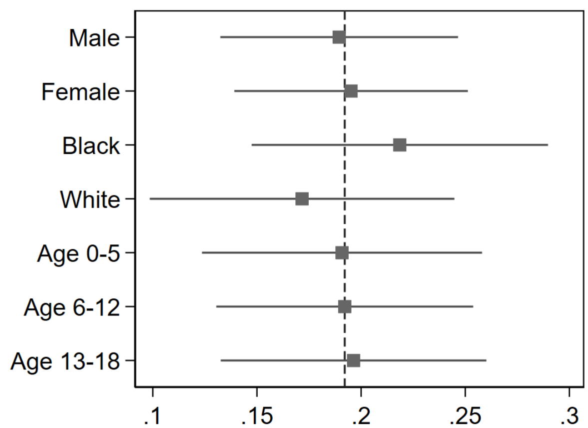
In this figure, the x-axis shows the discrete risk score for a given child-referral and the y-axis shows the average of our calculated 12-month poor health index in that risk score bin. That is, the index captures health in the year following the referral. The sample is all child-referrals with a discrete 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 A3: Manipulation Testing



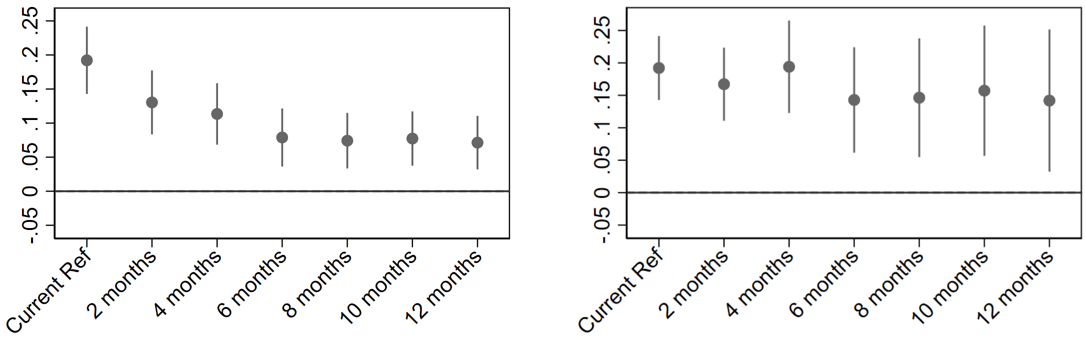
This figure reports results from a test for manipulation of the running variable around the high-risk protocol cutoff, following Cattaneo, Jansson, and Ma (2020). The test allows for separate bandwidths on each side of the cutoff, and is conducted for each AFST version. P-values from the test are noted below each subfigure.

Figure A4: 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_{irt} 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 A5: Cumulative Effect on Investigation

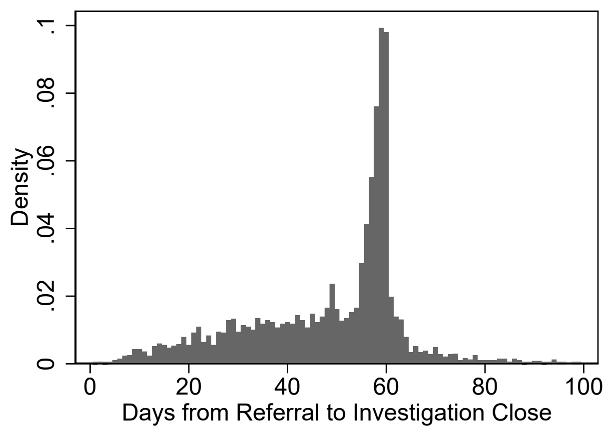


(a) Any investigation

(b) Number of investigations

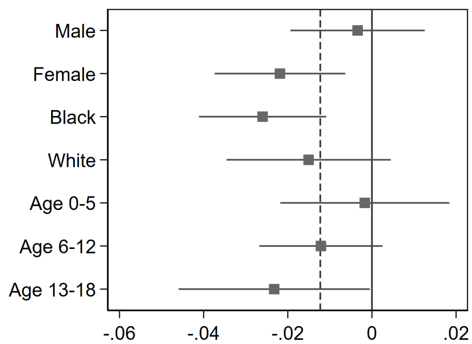
This figure shows the cumulative likelihood of having at least one investigation for our treatment and control groups in the year following a referral. In Panel (a), coefficients and standard errors for seven different regressions estimating Equation 1 are reported, where Y_{irt} is an indicator for an investigation occurring within the cumulative time frame post 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 coefficients and standard errors for seven different regressions estimating Equation 1, where Y_{irt} is the number of investigations occurring within the cumulative time frame post referral reported on the x-axis.

Figure A6: Investigation Length

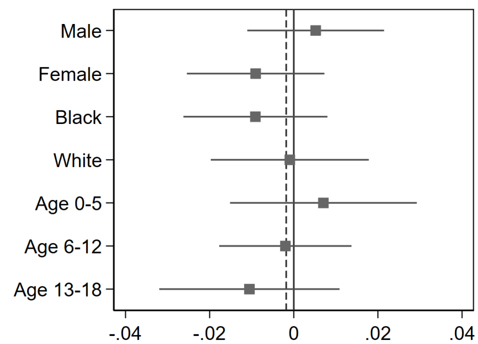


This figure shows the distribution of investigation length, defined as the days between a referral date and investigation close date, for the sample of screened-in referrals with a risk score of 17-18. One observation is included per referral. Referrals with investigation length above 100 days (2% of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

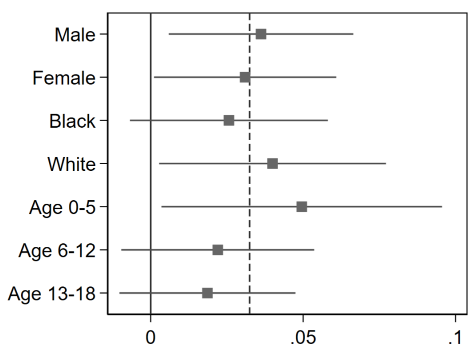
Figure A7: Child Health Results by Subgroup



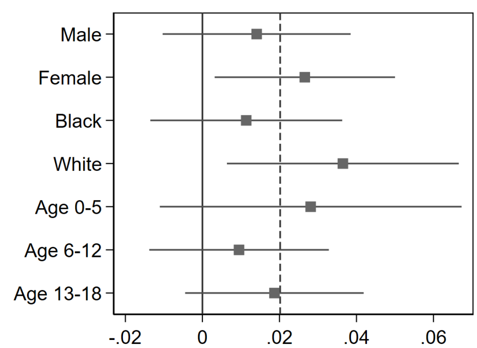
(a) Injuries



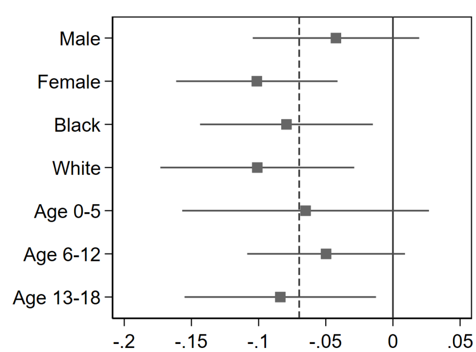
(b) Abuse / Neglect



(c) Wellness Visit



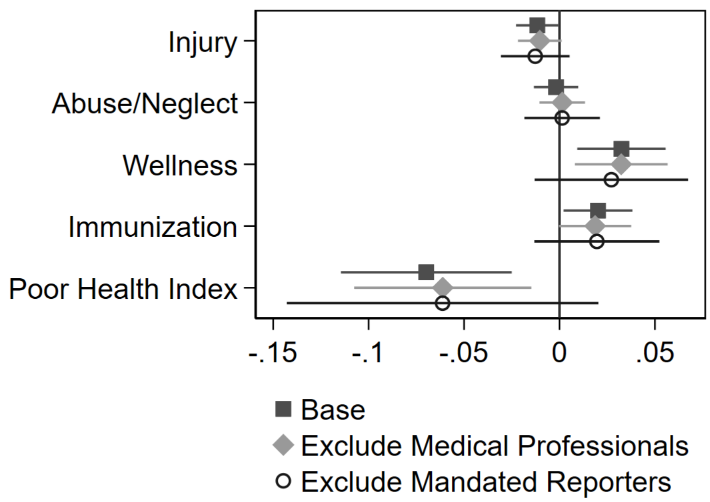
(d) Immunization



(e) Poor Health Index

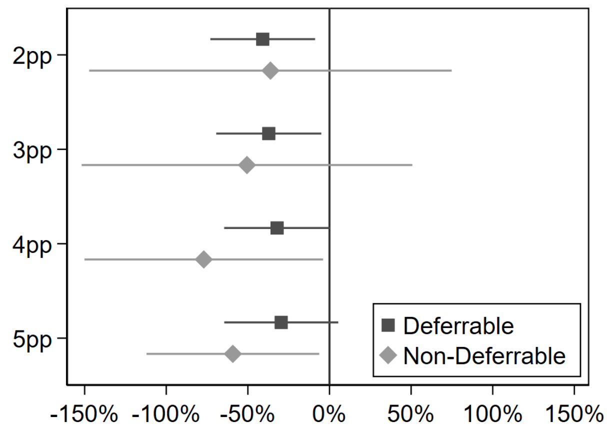
This figure shows coefficients and 95% confidence intervals from estimating Equation 1 for different subgroups of our main analysis sample, where the subgroup of interest is listed on the Y-axis. Y_{irt} is equal to the outcome variable, measured in the 60 days following the referral, described in the subplot label. The vertical dashed line shows the estimated coefficient for the full sample, as reported in Table 5.

Figure A8: Exclude Medical and Mandated Reporters



This figure plots coefficients and 95% confidence intervals from estimating Equation 1, setting Y_{irt} equal to each of our main short-run child health outcomes (labeled on the Y-axis). Grey squares indicate our baseline estimates, corresponding to Table 5. Grey diamonds show estimates when the sample is limited to reports not made by a medical professional, and hollow circles show estimates when the sample is limited to reports made by non-mandated reporters.

Figure A9: Injury - Deferrable vs. non-Deferrable



This figure plots estimates and 95% confidence intervals for eight separate regressions estimating Equation 1, where Y_{irt} represents different subsets of injuries defined as deferrable or non-deferrable. Here we define non-deferrable injuries as those with 3-digit ICD codes whose weekend rates fall within either 2, 3, 4 and 5 percentage points of 2/7. Each estimate matches to the corresponding definition labeled on the Y-axis. The injury codes included in each definition of deferrable / non-deferrable are reported in Table A1. The coefficients and confidence intervals are normalized as percentages, relative to the control group mean.

Table A1: ICD-10 Codes

Category	**Codes**	**Description**
Injury (S00-T32)	S00-S09	Injuries to the head
	S10-S19	Injuries to the neck
	S20-S29	Injuries to the thorax
	S30-S39	Injuries to the abdomen, lower back, lumbar spine, pelvis and external genitals
	S40-S49	Injuries to the shoulder and upper arm
	S50-S59	Injuries to the elbow and forearm
	S60-S69	Injuries to the wrist, hand and fingers
	S70-S79	Injuries to the hip and thigh
	S80-S89	Injuries to the knee and lower leg
	S90-S99	Injuries to the ankle and foot
	T07	Injuries involving multiple body regions
	T14	Injury of unspecified body region
	T15-T19	Effects of foreign body entering through natural orifice
	T20-T25	Burns and corrosions of external body surface, specified by site
	T26-T28	Burns and corrosions confined to eye and internal organs
	T30-T32	Burns and corrosions of multiple and unspecified body regions
Place - home	Y92.0	Non-institutional (private) residence as the place of occurrence of the external cause
Place - non-home	Y92.1	Institutional (nonprivate) residence as the place of occurrence of the external cause
	Y92.2	School, other institution and public administrative area as the place of occurrence of the external cause
	Y92.3	Sports and athletics area as the place of occurrence of the external cause
	Y92.4	Street, highway and other paved roadways as the place of occurrence of the external cause
	Y92.5	Trade and service area as the place of occurrence of the external cause
	Y92.6	Industrial and construction area as the place of occurrence of the external cause
	Y92.7	Farm as the place of occurrence of the external cause
	Y92.8	Other places as the place of occurrence of the external cause
Place - no info.	No Y.92 code	
Non-def. (1pp)	S68	Traumatic amputation of wrist, hand and fingers
	T18	Foreign body in alimentary tract
Non-def. (2pp)	S01	Open wound of head
	S27	Injury of other and unspecified intrathoracic organs
	S97	Crushing injury of ankle and foot
Non-def. (3pp)	S30	Superficial injury of abdomen, lower back, pelvis and external genitals
Non-def. (4pp)	S03	Dislocation and sprain of joints and ligaments of head
	S10	Superficial injury of neck
	S11	Open wound of neck

Continued on next page

Table A1 – continued from previous page

Category	Codes	Description
	S38	Crushing injury and traumatic amputation of abdomen, lower back, pelvis and external genitals
	S40	Superficial injury of shoulder and upper arm
	S85	Injury of blood vessels at lower leg level
	S90	Superficial injury of ankle, foot and toes
	S91	Open wound of ankle, foot and toes
	T27	Burn and corrosion of respiratory tract
Non-def. (5pp)	S02	Fracture of skull and facial bones
	S20	Superficial injury of thorax
	S41	Open wound of shoulder and upper arm
	S51	Open wound of elbow and forearm
	S57	Crushing injury of elbow and forearm
	S60	Superficial injury of wrist, hand and fingers
	S61	Open wound of wrist, hand and fingers
	S81	Open wound of knee and lower leg
Immunization	Z23	Encounter for immunization
Wellness visit	Z00	Encounter for general examination without complaint, suspected or reported diagnosis
	Z01	Encounter for other special examination without complaint, suspected or reported diagnosis
Abuse and Neglect	We follow McConnell et al. (2020) in defining ICD codes related to maltreatment. We use detailed ICD-10 codes shared by Margaret A. McConnell. See replication code for details.	Summary of included diagnoses: anogenital herpes, viral infection, gonococcal infection, other severe malnutrition, retinal hemorrhage, dental caries, pelvic inflammatory disease (unspecified), solar radiation dermatitis, fracture of skull, fracture of spine and trunk, fracture of upper limb, intracranial injury excluding those with skull fractures, internal injury of chest – abdomen and pelvis, contusion with intact skin surface, some specific burns, injury to nerves and spinal cord, poisoning by drugs – medicinal and biological substances, drowning, non-fatal submersion, other and unspecified effects of external causes.

This table reports the ICD-10 diagnostic codes used in the constructing our health outcomes from medicaid records. We use the descriptions associated with each code as reported by the American Academy of Professional Coders (AAPC) at: https://www.aapc.com/codes/. Each category corresponds to a different outcome or indicator variable used in our study. “Non-def” refers to different definitions of non-deferrable injuries, defined as those 3-digit ICD codes whose weekend rates fall within either 1, 2, 3, 4 and 5 percentage points of 2/7. We would like to thank Margaret McConnell for providing us with the ICD-10 codes for constructing our measure of child abuse and neglect, which matches similar measures used in the literature. These codes came from ICD-9 codes indicative of abuse provided in McConnell et al. 2020, which were in turn adapted from Schnitzer et al. 2011, and Hooft et al. 2013, and then cross-walked to ICD-10 codes using the NBER general equivalence mapping (see: https://www.nber.org/research/data/icd-9-cm-and-icd-10-cm-and-icd-10-pcs-crosswalk-or-general-equivalence-mappings).

Table A2: Injuries: Place of Occurrence

	(1) Base	(2) At home	(3) Not at home	(4) No information
Default Investigation	-0.0117**	-0.00290**	-0.00135	-0.0125**
	(0.00567)	(0.00146)	(0.00312)	(0.00508)
Mean	0.0353	0.00217	0.0114	0.0267
RightBW	8724	8724	8724	8724
LeftBW	10128	10128	10128	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_{irt} equal to each of the outcome variables described in the column titles. The injury codes included in each definition of an injury being "at-home", "not at home", or "no information" are reported in Table A1. 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 A3: Mechanisms: Parent Jail and Child Health

	(1) Jail (parents)	(2) Investigation	(3) Injury	(4) Abuse/Neglect	(5) Wellness	(6) Immunization	(7) Index
Panel A: Low Predicted Jail Effects							
Default Investigation	-0.00378	0.194***	-0.0151**	-0.00209	0.0275*	0.0108	-0.0608**
	(0.00763)	(0.0338)	(0.00700)	(0.00721)	(0.0149)	(0.0120)	(0.0288)
Mean	0.0380	0.550	0.0354	0.0361	0.0875	0.0572	0.0229
Panel B: High Predicted Jail Effects							
Default Investigation	0.0351***	0.198***	-0.00971	-0.00341	0.0407**	0.0325**	-0.0899**
	(0.0107)	(0.0366)	(0.00932)	(0.00970)	(0.0187)	(0.0142)	(0.0364)
Mean	0.0594	0.587	0.0354	0.0310	0.150	0.0829	-0.0674

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_{irt} equal to each of the outcome variables described in the column titles, in the period of 1-2 months following the referral. The sample for the regressions in Column (1) is all parent-referrals with a risk score of 17-18. The sample for the regressions in Columns (2)-(7) is all child-referrals with a risk score of 17-18. See Table A1 for details on ICD codes used to construct each child health outcome. Panel A further restricts the sample such that the maximal predicted likelihood of jail for any parent on the referral is below the median (Columns 2-7). Panel B is similarly defined but for parents whose predicted likelihood of jail is above the median. See the text in section 6 for more details.

Appendix B: Allegheny Family Screening Tool

In this appendix we discuss additional details about the implementation and use of the AFST.

Allegheny County Analytics contracted with external academics to create the AFST, which was first launched in 2016. Coefficients were originally estimated using a probit model on all available referrals in Allegheny county from April 2010 and April 2014 (76,964 observations) (V1). Coefficients are only estimated a single time, and are updated (using available new data) only when each new version of the model comes out.

The AFST assigns a separate latent risk score to each child who is an alleged victim of abuse or neglect. While adult histories have always been 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 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 (Hornby Zeller Associates, Inc. 2018).

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.

The High Risk Protocol has also evolved over time. Version 1 applies a default screen in to investigation for anyone over the risk score threshold of 18. Later versions 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. For simplicity, we exclude referrals where no child is under the age of 16 from our entire analysis sample.

Appendix C: Dynamic Regression Discontinuity Design

Here, we present the details of our dynamic regression discontinuity design (DRDD), introduced in Section 5.1. This approach estimates the effect of a default investigation, holding constant past and future investigation trajectories. Specifically, we estimate the following equation:

$$Y_{ird} = \alpha + \beta_1 D_r + \beta_2 f_g(z_r) + \beta_3 (D_r \times f_g(z_r)) + \sum_{\tau=-6}^{6} [m_{i,d-\tau} + n_{i,d-\tau} + f_g(z_{i,d-\tau})] + \gamma_{t(d)} + \epsilon_{ird} \quad \text{(C1)}$$

Where Y_{ird} is an outcome for child i associated with referral r occurring on date d. For each child health outcome we run 12 different regressions, for Y_{ird} measured in 60 day windows within the 360 days before and after the reference referral r.[42]

D_r and z_r are defined as in Eq. 1. We again include fixed effects for the month-year of the referral, $\gamma_{t(d)}$. For the purposes of the DRDD the functional form of the running variable is a flexible polynomial of degree g (f_g).[43] We further modify our main equation by including various controls related to each of twelve 60-day periods in the 360 days before and after the reference referral. Each 60-day window is referenced by τ, for $\tau \in [-6, -1] \cup [1, 6]$. $m_{i,d-\tau}$ represents a vector of indicators equal to one if child i is associated with any referral in period $d - \tau$, and zero otherwise. $n_{i,d-\tau}$ is a vector of indicators equal to 1 if the referral measured by $m_{i,d-\tau}$ led to an investigation, and zero otherwise.[44] β_1 now captures the effect of a default investigation on a child health outcome Y_{ird} measured in a given 60 day period relative to the referral, controlling for the previous and subsequent trajectory of referrals and investigations.

Following the DRDD literature and our main specification, we limit the reference referral to be within the high-risk bandwidth, while the subsequent and previous trajectory of referrals are by necessity allowed to be at any point in the risk distribution (Baron 2022). To account for unobserved heterogeneity introduced in including these referrals outside our "high-risk" bandwidth, we follow the literature and flexibly control for risk score with a cubic functional form (Baron (2022)), though the results are robust to quadratic and linear functional forms.

Figure C1 shows results from this exercise for each of our child health outcomes. Across all outcomes, effects on child health cluster around zero and are not statistically significant in the year leading up to the referral, offering additional evidence of the validity of our identification strategy. For child health outcomes in the year following the referral the

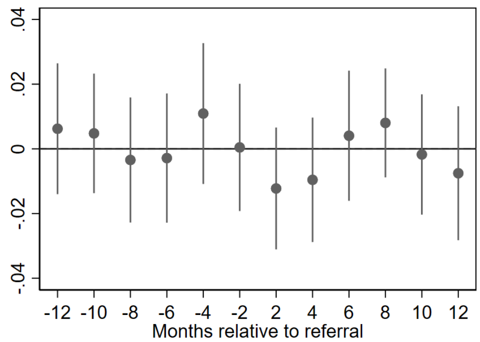
[42] Outcomes measured on the exact day of the referral are excluded, as we cannot determine whether they occurred before or after the referral.

[43] In this case we don't normalize our risk scores but instead interact our risk scores with AFST version fixed effects, and assign an AFST value of zero to those windows in which a referral does not occur.

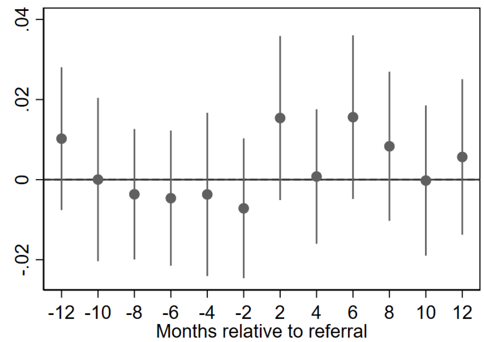
[44] In cases where there are multiple referrals for a child within the same 60 day window, we choose the first referral which leads to an investigation or, if there is no investigation that month, the first referral within the window.

patterns are overall similar to conventional RDD results shown in figure 4. For injuries, wellness visits, and immunizations, effects remain concentrated in the 60 days after referral, and in all cases the effects attenuate after 60 days. Though the standard errors are somewhat larger than in our conventional RDD results, the poor health index retains significance at the 5% level. This suggests that control group "catch-up" was not attenuating our previous results, and is consistent with the hypothesis that the protective effects of investigations last primarily while investigations are ongoing. Active surveillance of these families may be a key mechanism for the documented short-term improvements in health, and alternative interventions may be necessary in order to induce longer-term change. We explore potential mechanisms in more detail in Section 6 of the paper.

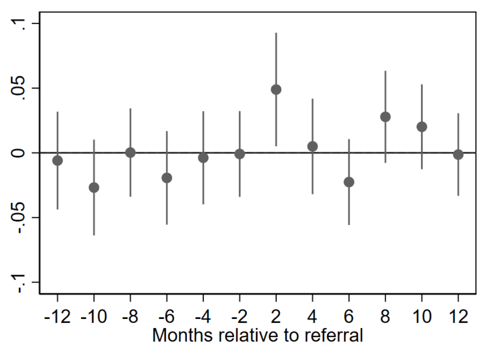
Figure C1: Dynamic RDD



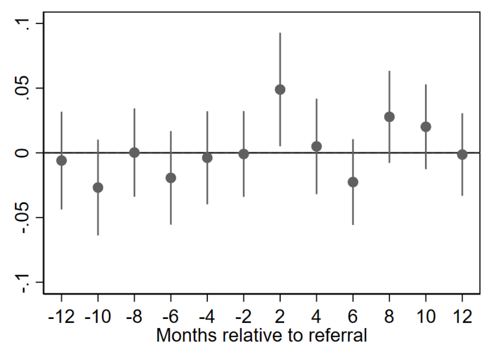
(a) Injuries



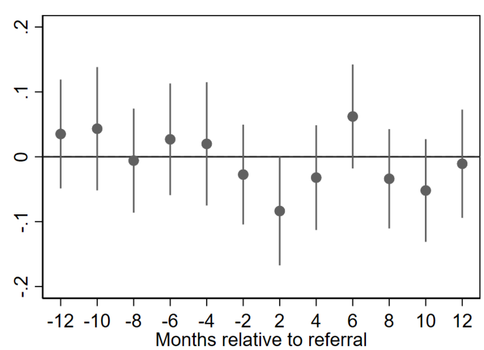
(b) Abuse / Neglect



(c) Wellness Visit



(d) Immunization



(e) Poor Health Index

Each subfigure plots coefficients and 95% confidence intervals for 12 separate regressions estimating Equation C1, where Y_{ird} is equal to a child health outcome in twelve different time periods. The child health outcome is described in the subplot captions. The x-axis shows the 60-day time period in which Y_{ird} is measured relative to the referral. The day of the referral is excluded, so that month 1-2 includes days 1-60 after the referral. The sample is all child-referrals in our main bandwidth (risk score of 17-18).