# Child Maltreatment Investigations and Family Well-being*

Lindsey Lacey†, Katherine Rittenhouse‡and David Simon§

November 1, 2024

PRELIMINARY AND INCOMPLETE – PLEASE DO NOT CITE

**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 are largely unknown. Investigations aim to protect children from harm and direct services to families in need. However, investigations also bring the implicit threat of child removal, and as such may be stressful and disruptive. Using a regression discontinuity design, we study the impacts of these interactions on child health, and parent service use and mental 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. Parents who are defaulted into an investigation have higher rates of mental and behavioral healthcare utilization in the following year.

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

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

†Allegheny County Analytics

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

§University of Connecticut and NBER

# 1 Introduction

Every year over three and a half million families are investigated for child abuse and neglect, with 37% of all children experiencing an investigation before age 18 (Kim et al. 2017). Investigations aim to protect children from harm and direct services to high-risk families. However, these interactions may also be burdensome and stressful, particularly for parents. Despite the widespread and controversial nature of this interaction, empirical evidence on the effects of child protection system (CPS) investigations is severely limited.

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

Our setting is Allegheny County, Pennsylvania (Pittsburgh). Allegations of maltreatment are made to the county's office of Children, Youth and Families (CYF). Call screeners at CYF are charged with evaluating the allegations and determining whether to screen in the family for an investigation. An investigation involves a social worker visiting the child's home to make determinations about their safety and well-being, and may result in mandated services or removal of the child from the home (i.e., foster care placement). Allegheny County has a rich Data Warehouse of linked administrative data across multiple public agencies, including from CYF, county-provided public services, and state medicaid claims. For every referral to CYF, we are able to link each individual (children and parents) to the Data Warehouse and follow them over time.

In 2016, Allegheny County implemented a predictive risk model to help workers decide which families to screen in for investigation. The Allegheny Family Screening Tool (AFST)

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

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

We focus on outcomes related to child health, and parent service use and mental health. Much of the debate on child welfare investigations centers around these impacts. Defenders of the child welfare system argue that investigations are necessary to protect children from harm and safeguard basic child rights to health and safety. There are many examples of tragic cases in which a child was referred to CPS, screened out and subsequently was a victim of severe abuse (Heimpel 2019). Additionally, investigations are seen as a way to reach high-need families, and link them with beneficial services. Detractors argue that investigations impinge on parental rights and family autonomy, with limited benefits to child safety. They may note that 80% of investigated allegations go unsubstantiated, suggesting that they may have been unnecessary to begin with. Qualitative evidence shows that investigations are disruptive and stressful to families, and may alienate parents from community support and

services (Fong 2018). Severe racial disparities throughout the child welfare system have led to critiques of the child welfare system by scholars and activists as a "family policing system" weaponized against Black communities (Roberts 2022). As such, there is a growing movement to reduce investigations, limit CPS involvement in families' lives and, at the extreme, to abolish CPS entirely. We bring empirical evidence to this debate by observing potential benefits and costs to families in terms of mental and physical health in the context of "as good as random" variation in which high risk children get investigated.

Children above the threshold for a default investigation are 35% less likely to have an injury-related Medicaid claim in the 60 days following a referral.[1] Default investigations increase the likelihood of wellness visits and immunizations by 28% and 30%, respectively. Combining these and other measures into a "poor health index," we find that a default investigation reduces poor health by 0.07 standard deviations. We do not find evidence of longer-term impacts on child health.

For parents, we study reduced form effects of a default investigation on service use and mental healthcare utilization. We find that while investigations have no impact on parents' enrollment in Medicaid, public benefits, substance abuse or housing services, they do increase involvement in mental health services by 25% in the short run, and 19% in the long run. Taking a deeper look into effects on Medicaid-funded mental health services, we find that parents above the threshold for a default investigation are 49% more likely to receive mental health crisis services in the two months following a referral, 47% more likely to receive outpatient mental health services such as therapy in the two months following a referral, and 21% more likely to receive outpatient mental health services in the 3-12 months following a referral.

We contribute to a growing literature on the economics of child protection systems. Several papers study the effects of foster care, relying on the random assignment of cases

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

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; Roberts 2019). This literature has 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. However, this approach can only assess the impact of removal, often a last resort for child welfare agencies. We identify the impact of investigations, an early-stage interaction which affects one third of U.S. children. Additionally, we identify impacts for a different group of children - those at the margin of investigation, as opposed to at the margin of removal.

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

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

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, and Section 6 concludes.

# 2 Background

## 2.1 Allegheny County Child Welfare Investigations

In Allegheny county, the Office of Children, Youth, and Families (CYF) is charged with protecting children from abuse and neglect. Alleged maltreatment is first brought to the attention of CYF by third-party reporters, including mandated reporters (e.g., educators or doctors) and non-mandated reporters (e.g. neighbors or family members). Reports can be made by phone or through an online service known as "Childline." A call screen worker then reviews and recommends whether to "screen in" each referral for an investigation.[2] After a screener makes their recommendation, a supervisor will make the final decision on whether to open an investigation.[3] The state of Pennsylvania distinguishes between "child protective services" referrals which involve allegations of physical or sexual abuse, and "general protective services" (GPS) which tend to involve allegations of neglect. Child protective service referrals are screened in 100 percent of the time by law and do not allow for screener discretion. As such, our analysis is limited to GPS referrals, which make up the majority of reports in Allegheny county.[4]

If a report is screened in, an investigation is opened. Investigations aim to determine whether the allegations of maltreatment are credible, whether children are safe in their home, and whether the family requires various services or court interventions to ensure child safety.[5] To achieve these aims, the caseworker assigned to the investigation conducts one or more home visits, and conducts individual interviews with parents and children as well

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

[3]Anecdotally, call 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.

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

[5]Information on investigation procedures in Allegheny County is derived from review of internal documents detailing investigative practice standards.

as the reporter and other individuals connected to the family. Home visits must occur within a maximum of three days from the time of referral, and every one to four weeks throughout the period of investigation. The initial home visit is unannounced. At this visit, the caseworker informs the parents they are being investigated for maltreatment, discloses the allegations, walks through the home, and conducts one-on-one interviews with each household member. During home visits, caseworkers look for food, clothing, and appropriate toys, any environmental dangers, evidence suggesting criminal activity or drug use, weapons, and overall cleanliness.

Investigations are supposed to be closed within a maximum 60 days, after which one of several outcomes is possible. If there is not enough evidence to support the allegations, they are deemed unfounded and no further action is taken. The social worker may still connect the family with or provide information about optional services, but CYF will not have any continued interaction with the family. If the investigator instead finds reasons to be concerned about the child's well-being, they may open a CYF case for the family. A CYF case may involve periodic check-ins with the family, suggested or mandated services (including, e.g., mental health or substance abuse services), or, often as a last resort, removal of the child from the home. Note, 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 calls. The Allegheny Family Screening Tool (AFST) draws on data from the county's Data Warehouse, which integrates human services data across multiple govern-

ment agencies including the Department of Human Services, the County Housing Authority, Jail, Probation Office, and Health Department.[6] The AFST is meant to inform, rather than replace, human decision-making.

The AFST uses a wide range of variables to predict risk, including information on previous investigations/allegations, child or parent involvement in the court system (such as time spent in jail), medical history (such as behavioral health issues), and demographics such as age, gender, and presence of biological mother and father.[7] From these variables, the model calculates a continuous latent "risk score" for each child on a referral. The AFST then takes the maximum latent value across all children on the referral, such that there is one latent risk score per referral.[8] This referral-level latent score is mapped to discrete values between 1 and 20 based on predetermined cutoffs. Call screeners ultimately see this discrete score, and are instructed to consider it in their ultimate recommendation. In most cases, call screeners are shown the score along with a general category of low risk (AFST scores 1-9), medium risk (10-14), or high risk (15-17). While scores below 18 are meant simply to advise the screener's decision, referrals with scores of 18 and above are defaulted to be screened in, and require supervisor approval to be screened out. In such cases, the call screener does not observe a numerical score, but rather a notification that the referral falls under the "high risk protocol." See Figure A1 for a visual example of what the screener would see.[9] The AFST has been updated several times since its introduction. In December

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

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

[8]When the AFST was first introduced in August 2016, there was a separate latent risk score assigned to each child who was an alleged victim of abuse or neglect. However, this generated many missing scores, particularly for younger children who had no administrative records stored in the data warehouse. Therefore, in December 2016, the risk scoring was changed to use information from everyone in the household (including parents), to assign a risk score to each individual. The maximum risk score across all children on the referral is then used as the latent risk score (and running variable in our discontinuity design). Because this was a substantial change in how the score was implemented, we drop referrals before December 2016.

[9]A "low risk" protocol was also introduced in November 2018 such that all children who got a score below

2018 the algorithm was revised to use a LASSO model and was re-trained on all available data. In July 2019 the lasso model was improved to refine how the model counted incoming criminal justice outcomes. For our main analyses, we pool and normalize algorithm scores, as discussed further in the next section.

# 3 Data

We obtained administrative data from the Allegheny County Department of Human Services on the universe of child maltreatment referrals to CYF from December 2016 through December 2023.[10] For each referral, we observe the raw (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, these individuals are also linked to the Allegheny County Data Warehouse. This allows us to observe monthly indicators of involvement with various public institutions that have integrated their data into the data warehouse. For each individual in our sample, we observe use of public benefits, enrollment in Medicaid, involvement in the criminal justice system (including a distinct indicator for juvenile probation), and use of various county- and Medicaid-funded services including housing support, mental and behavioral health services, and substance abuse counseling. Moreover, we observe detailed Medicaid claims data for both physical and mental health services, described further below.

11 (or 12 in later versions) were defaulted to be screened-out for an investigation. However, we did not find any strong "first stage" effects of being below the threshold on having an investigation opened.

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

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 we observe outcomes for individuals. Table 1 presents summary statistics at the child-referral level, separately for all children referred to CYF (Column 1) and child-referrals which fall within our main bandwidth (Column 2).

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

Of all child-referrals in our sample, approximately half are female and 39% are Black. Note, this indicates a significant over-representation of Black children, relative to their share of the population.[11] Less than half of child-referrals have two parents listed on the referral, and on average about three children are associated with the referral. 75% of children have received public benefits in the prior six months, 72% have been enrolled in medicaid, 3% were involved in juvenile probation and 13% have had a recent investigation. The most common allegations categories are caregiver substance abuse, inadequate care or home, neglect, exposure to risk, and physical maltreatment. Referrals are most likely to come from educational professionals (15%), social workers (15%), parents or relatives (13%), therapists (12%) and legal professionals (12%). 46% of child-referrals are screened in for investigation, and 5% are followed by the child being removed from their home within twelve months.

Children within our main bandwidth (with risk scores of 17 or 18) are more likely to be Black, less likely to be in a two-parent household, and come from households with a

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

larger number of children. They are also somewhat more likely to be referred by legal professionals and social workers, and somewhat less likely to be reported by friends and neighbors or educational professionals. These child-referrals are 39% more likely to result in investigation, and 53% more likely to be removed from their home within twelve months.[12]

Next, Table 2 presents summary statistics at the parent-referral level, separately for all parents associated with referrals to CYF (Column 1) and parent-referrals which fall within our main bandwidth (Column 2). We define an individual as a mother if their stated role is "parent" and their gender is female. We define fathers similarly. Note, there may be multiple mothers or fathers included in any given referral if, for example, there are multiple father figures associated with the referred children. As noted in Table 1, 37% of child-referrals in our main bandwidth are associated with a two-parent household. In 43% of parent-referrals in our analysis bandwidth, the parent is listed as the alleged perpetrator. Within the bandwidth, 75% of parents have received public benefits in six months prior to the referral, 61% were enrolled in Medicaid, 30% had an interaction with the criminal court system, and 26% were involved in a previous investigation.

## 3.1 Medicaid Claims

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

We create four main categories of Medicaid claims to measure child health in the months after a referral. The first category includes ICD codes related to injuries. The second includes codes which have been identified as suggestive of child abuse or neglect in prior work

[12]Children in with risk scores of 17 or 18 are also more likely to have received public benefits, be enrolled in Medicaid, be involved in juvenile probation and are more likely to have had a previous investigation within the prior six months. Child-referrals within the bandwidth are more likely to have an allegation of caregiver substance abuse, and less likely to have an allegation of physical maltreatment

(Schnitzer et al. 2011; McConnell et al. 2020). The final two categories measure well-child visits and immunizations, respectively. Finally, we normalize and combine these categories in 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. Figure 2 presents the average value of the poor health index in the twelve months after a referral for each risk score between 10 and 20. Note that poor health is increasing (i.e., health is worsening) with higher risk scores.

In our analysis of the longer-term health impacts of investigations, we include two additional categories in our index. The first includes codes related to ambulatory care sensitive conditions (ACSC). This category follows Grimon and Mills (2023) in identifying hospital visits for conditions which may be avoidable if proper preventative care is received. The fourth category includes infectious sicknesses such as the flu and lower respiratory tract infections (pneumonia and bronchitis). We also observe detailed claims data for Medicaid-funded mental and behavioral health services. These include visits with providers who are exclusively licensed to treat mental health conditions. Again, for each individual in our sample we observe claims from 2015 through 2023, including exact date, the category and name of the service provided, and the associated ICD-10 diagnosis codes. We focus our analysis on two categories of behavioral health claims: outpatient mental health services, and crisis services. Outpatient services are primarily made up of various forms of individual, group and family therapy. Crisis services include walk-in crises, mobile crises, telephone crises, and other similar services.

# 4 Empirical Strategy

Our empirical strategy compares individuals with latent risk score directly above relative to below the threshold for being "defaulted" into an investigation in a fuzzy regression

discontinuity design. We estimate the reduced form effects using the following regression model:

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

where $Y_{ir}$ is the outcome for individual $i$ who is associated with referral $r$. Recall that each referral $r$ is assigned a single risk score, equal to the maximum risk score of all children on the associated referral. $D_r$ is equal to one if referral $r$ has a score higher than the default screen-in threshold, and $z_r$ is the normalized algorithm score relative to the cutoff. $\beta_1$ is the coefficient of interest, that captures the effect of crossing the threshold. We allow for different slopes on either side of the threshold, and include month-year fixed effects $\gamma_t$. Since treatment is assigned based on the AFST score, which is identical for all children on the referral, our main specifications cluster at the referral level to account for auto-correlation within referrals of score assignment. We use a triangular kernel and a linear functional form.

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

For each of our main results, we test robustness to alternative bandwidth choices and specifications. In particular, we show results using the MSE-optimal bandwidth both pooling across all AFST versions and separately by AFST version. Each version's respective MSE-optimal bandwidth is shown in Figure 1. The MSE-optimal bandwidth selection pro-

cedure recommended by Cattaneo, Idrobo, and Titiunik (2019) results in bandwidths that are uniformly larger than our preferred choice of 17-18. Optimal bandwidth selections reflect a bias-variance trade off, allowing larger bandwidths to reduce variance at the cost of increased bias. We thus argue the bandwidth choice for our main analyses is conservative, intuitive for our setting, and allows for consistent estimation across models.

## 4.1 Validity

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

We run three empirical tests to validate the research design. First, we test for manipulation in the running variable using local polynomial density estimators (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.

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 demographic

and referral characteristics (Panel A), as well as six-month lags of each of our main outcome variables of interest (Panel B). Out of the 23 regressions, two 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 Medicaid enrollment. We directly test for this by estimating Equation 1, setting the outcome variable to an indicator for Medicaid enrollment in the 1-2 months, and 3-12 months after the referral. Results are reported in Panel C 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 the long term.

## 4.2 Relevance

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

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

We estimate Equation 1, where $Y_{ir}$ is a binary variable equal to one if an investigation is opened on referral $r$, and zero otherwise. Results are presented in Column (1) of Table A2.

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 A5 plots the first stage estimate separately for several subgroups. The first stage is similar across age and race. Finally, Figure A6 plots the first stage estimate separately by AFST version, when the MSE-optimal bandwidth for each version is used. Estimates for AFST versions 2 through 4 are almost exactly the same in magnitude to our main estimate, while the first stage for version 1 is approximately 10 percentage points, or half as large.

Next, we estimate Equation 1, setting $Y_{ir}$ equal to an indicator for *any* investigation in the next twelve months. Families are often referred more than once, meaning it is possible that a child who is initially screened out is re-referred and investigated at a later date. Column (2) of Table A2 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 A2 implies that many families below the threshold who are initially screened out are re-referred and investigated in the following year.[13] As such, our first stage can be interpreted in part as a shift forward in the timing of investigation. Downstream effects may thus be concentrated in the short term.

Children categorized as high risk by the algorithm on average already have a high level of exposure to protective services, and the results in Table A2 should be interpreted in this context. Figure A4 descriptively illustrates how the likelihood of exposure to CYF evolves over the months around the referral. We graph these raw means separately for those who fall directly below (control) and above (treatment) the cutoff. Panel (a) of A4 shows that roughly 40% of the control group and 50% of the treatment group were exposed

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

to at least one investigation in the year before the referral. Exposure increases for both treatment and control children in response to the referral (though the increase is larger for the treated group). The control group continues to see an increase in their cumulative likelihood of investigation increasing to up to 80% a year after the referral. Panel (b) shows cumulative exposure to ongoing services, and panel (c) documents cumulative likelihood of home removals. About 6% of the control group had services opened in the year prior to the referral, increasing to around 18% at 12 months post-referral. Home removals are less common but follow a similar pattern; cumulative likelihood of home removal increases to approximately 8% in the control group by 12 months post-referral. [14]

The patterns in Figure A4 show that a substantial portion of the control group has previously had an investigation with a number of those turning into services and at times home removals.

# 5 Results

In this section we show effects of being over the high-risk threshold on a variety of downstream outcomes for children and families. We present results in their reduced form, before scaling by the estimated first stage to estimate the local average treatment effect of investigations. In subsections below we study effects on child health, as well as effects on parental mental health and service use, and finally we investigate potential mechanisms.

## 5.1 Child Physical Health

Protecting children from harm is the primary goal of investigations for maltreatment. However, the effects of such investigations on child health are *a priori* ambiguous. Investigations

[14]Note that these graphs are distinct from 3 as they look at the cumulative average likelihood of having had any investigation (or a related outcome) in the time leading up to and following the referral, rather than estimating the discontinuity in the immediate likelihood of a new investigation being opened.

serve as a path to direct families to health services, prevent maltreatment, and surveil the medical care of children, possibly leading to improved health outcomes. However, investigations could be disruptive, increasing stress within the family, or cause fractures between families and the medical system, leading to poorer health outcomes (Fong 2020). For example, investigations may lead parents to be suspicious of mandated reporters (such as nurses and doctors) making them less likely to take their child in to receive preventative health care.

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

We separately estimate effects on each of the four main categories of Medicaid claims identified in section 3 above: injuries, abuse/neglect, wellness visits and immunizations. Results are reported in Table 5. Investigations have positive effects on health outcomes. Children above the threshold are 35% (1.2 percentage points) less likely to have an injury-related Medicaid claim (statistically significant at the 5% level). Investigations increase the probability of a wellness visit by 28% (3.3 percentage points), and immunizations by 30% (2.0 percentage points), each statistically significant at the 1% and 5% level, respectively. There are no statistically significant effects on codes linked to abuse and neglect. Our pooled index of poor health indicates that investigations decrease poor health by 0.07 standard deviations in the first two months after a call (statistically significant at the 1% level).

We next estimate effects in the 3-12 months after an investigation. Here we study impacts

on two additional measures of health – ACSC, and infectious illness. Results are reported in Table A1.[15] In the long run, we do not see a statistically significant impact on the poor health index. Only one of the six components shows a statistically significant impact. Visits for ambulatory care sensitive conditions decline by 0.025 percentage points (10%), statistically significant at the 10% level. Although suggestive, this is consistent with an improvement in preventative healthcare reducing long-term problems.

## 5.2 Parent Service Use and Mental Health

We next turn to the effects of investigation on parents. Parents may be affected by investigations in a variety of ways. Qualitative work documents the stress associated with being investigated for child maltreatment, even when allegations are ultimately unsubstantiated (Fong 2020). Investigated parents report a loss of trust in mandated reporters, which may lead them to disengage from professionals such as health or service providers Fong 2018. However, investigations also aim to connect struggling parents to programs such as substance abuse or mental health services. We study the impact of investigations on parents' enrollment in various public benefits and services, including an in-depth analysis of medicaid-funded mental and behavioral health services.

Table 6 presents results from estimating Eq. 1, where the sample is parents associated with referrals in our main analysis bandwidth. $Y_{ir}$ is set to each of the following outcomes: Medicaid enrollment, enrollment in other public benefits, housing services, receipt of county- or medicaid-funded mental health services, and receipt of substance abuse services. Each outcome is measured both during the two months after the relevant referral, when the investigation is likely ongoing (Panel A), and in the 3-12 months after the referral, after the investigation has concluded (Panel B). We do not find any evidence of effects on enrollment

[15]For completeness, we also show effects on these additional health measures in the short run, and present effects on each poor health index in both the short and the long run.

in Medicaid or other public benefits in either the short or long term. Similarly, effects on housing services and substance abuse services are small in magnitude and statistically insignificant. Finally, default investigation increases the likelihood that parents receive mental health services by 2.4 percentage points (21%) in the short run, and 3.7 percentage points (19%) in the long run. These results are statistically significant at the five percent and one percent level, respectively.[16]

An increase in mental health services may be welfare improving, if parents are getting the help they need, or could indicate a reduction in mental health which leads more parents to take up services. While we do not have detailed information on county-funded mental health services, we explore effects on different categories of Medicaid-funded mental and behavioral health services, including outpatient services and crisis services. A long-run increase in outpatient visits such as therapy may indicate that parents are connected to services they find helpful, while an increase in crisis services may indicate that investigations cause significant mental distress. Table 7 reports coefficients and standard errors from estimating Eq. 1, where $Y_{ir}$ is an indicator variable for each of the following outcomes: any Medicaid-funded visit to a mental or behavioral health provider, any outpatient mental health service, and any crisis service. Again, estimates are reported separately for short- and long-run outcomes in Panels A and B, respectively. Associated regression discontinuity plots for each outcome variable considered are presented in Figure 5. Consistent with the increase in mental and behavioral health services observed in Table 6 above, parents with an AFST score above the threshold for a default investigation are more likely to have a Medicaid-funded mental health visit in both the short and long run. A default investigation increases the likelihood of any medicaid-funded mental or behavioral health service by 2.4 percentage points (30%) in the short run, and 2.5 percentage points (18%) in the long run (statistically significant

[16]This result is consistent with Grimon (2021), who finds in the same setting that case openings (as opposed to investigations) increase involvement in substance abuse and mental health services for mothers.

at the one and five percent level, respectively). When considering only outpatient mental health services (Column 2), a default investigation increases the likelihood of services by 2.1 percentage points (46%) in the short run, and 1.8 percentage points (20%) in the long run. Effects are statistically significant at the one percent and five percent level, respectively. To the extent that outpatient mental health services such as therapy are helpful in addressing underlying struggles or conditions, this result is consistent with investigations connecting parents to beneficial services. Column (3) estimates effects of a default investigation on mental health crisis services. In the two months after a referral, a default investigation increases the use of crisis services by 0.34 percentage points, or 55% relative to the mean (statistically significant at the 5% level). In the long run, we do not find any statistically significant effects on crisis services. This is consistent with investigations causing short-run disruptions, leading to mental health crises. Alternatively, referrals may be prompted by crises which are more likely to be addressed with mental health services if the family is screened in for an investigation.

Overall, these result suggest that investigations increase the takeup of mental healthcare services. To the extent that these services are helpful in treating underlying mental health conditions, parents may benefit from ongoing interactions. However, investigations do not appear to connect parents to other useful benefits and services, such as Medicaid, public benefits, housing services or substance abuse services.

## 5.3 Mechanisms and Child Welfare System Outcomes

After an investigation, CYF can open an ongoing case with the family, and at any point during the investigation or ongoing case may remove a child from their home. Each of these downstream actions likely have their own impacts on child and family well-being. Indeed, Grimon (2020) studies the effects of case openings on parents in Allegheny County, finding that when a case is opened mothers are more likely to enroll in mental health and substance

abuse treatment services. Other work has studied the impacts of child removals on child well-being, finding different effects in different time periods and settings (Doyle 2007, 2008, 2013; Bald et al. 2022; Gross and Baron 2022; Roberts 2019). In this section, we explore whether our results can be explained by changes in downstream case openings and removals, or if they are more likely the effects of investigations themselves. Note, our estimated effects on child health occur primarily in the short term, while an investigation is ongoing. As such, these effects are unlikely to be driven by case openings or removals, which tend to occur closer to the end of the 60-day period given for investigations. Longer-term impacts on parental mental health and healthcare utilization are more likely to work through case openings and child removals.

We estimate the reduced form effect of a default investigation on the probability that a child or parent has a case opened in the twelve months following a referral, and the probability that a child is removed from their home, or a parent has a child removed from their home, in the twelve months following a referral. Results are reported in Table A2. Effects for both children and parents are statistically insignificant at conventional levels, although not small in magnitude relative to the mean.

# 6 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 how they impact children and families. While investigations aim to protect children from harm and direct services to families in need, they are also highly criticised for being stressful, burdensome, and in many cases unnecessary. We provide the first empirical evidence on the impacts of CPS investigations on child health and parent mental health.

We leverage an algorithm-generated risk score in a regression discontinuity design. Chil-

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

We find that investigations are indeed protective of children's health, reducing injuries and increasing preventative healthcare in the two months following a referral. For parents, investigations increase utilization of mental health services, including crisis services in the short term and outpatient mental health in both the short and long term.

# References


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

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.

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.

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

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

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

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

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

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.

Heimpel, Daniel. 2019. "In Aftermath of Latest Child Death, L.A. Contends with Potential Foster Care Panic." *The Imprint.*

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.

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.

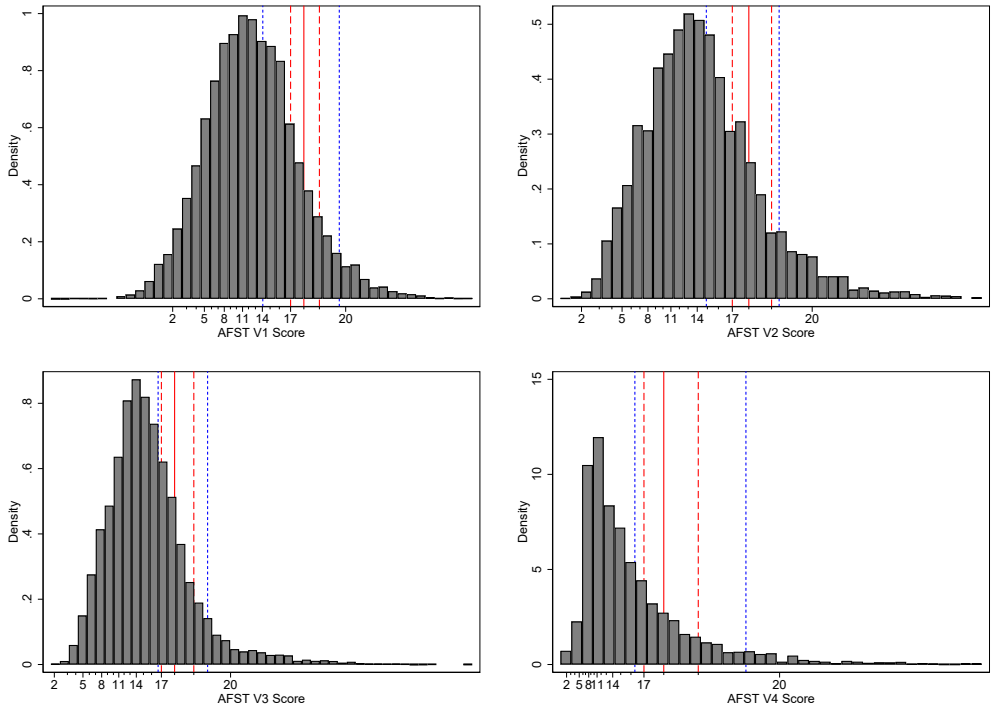
Roberts, Kelsey V. 2019. "Foster care and child welfare." *PhD dissertation, Clemson University.*

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.

# 7 Figures

Figure 1: Density and Bandwidths by AFST Version




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

Figure 2: Poor Health Index by Risk Score

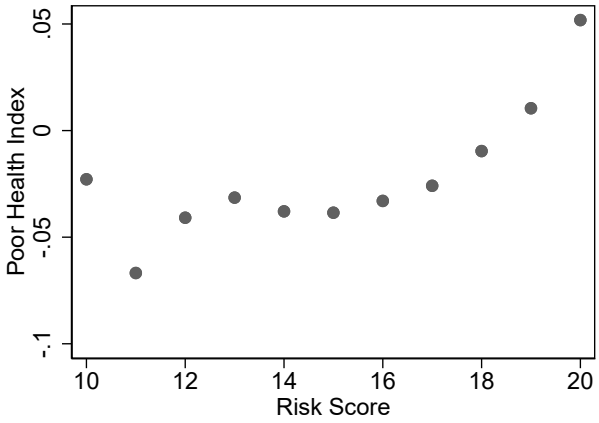
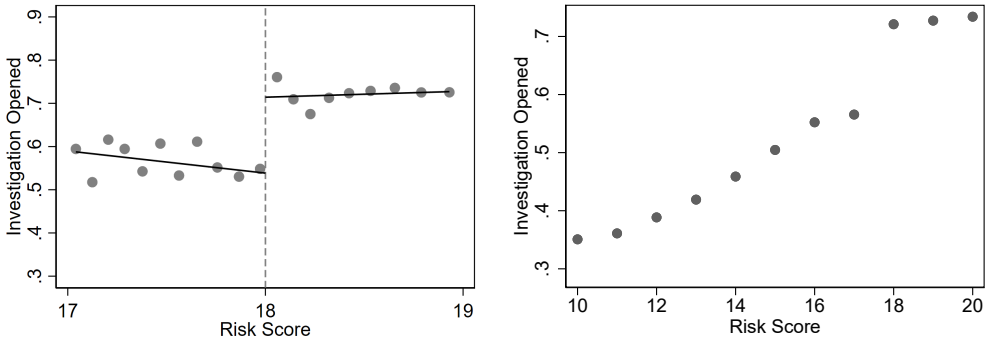
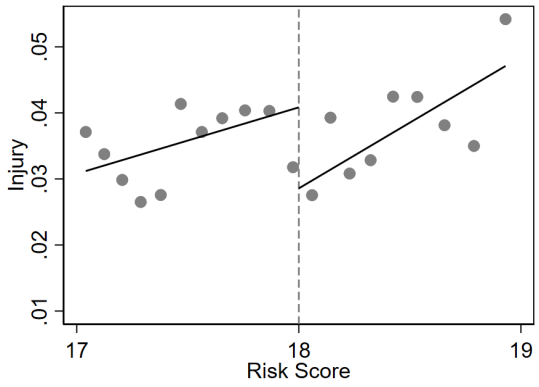



Figure 3: First Stage

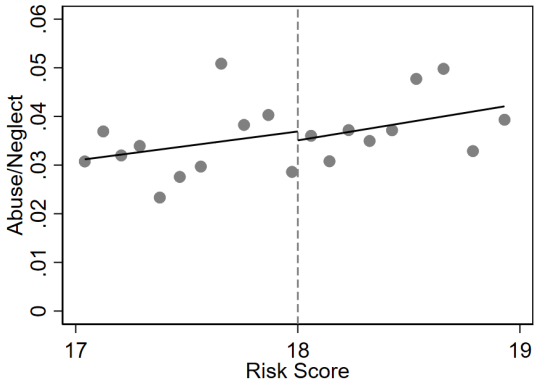



Each of these figures plots the first stage of being over the threshold for a "default" investigation on the likelihood of having an investigation opened. In panel (a), the x-axis is pooled and normalized risk score, ranging from 17 to 19 (our main RD bandwidth) The dashed vertical line shows the threshold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

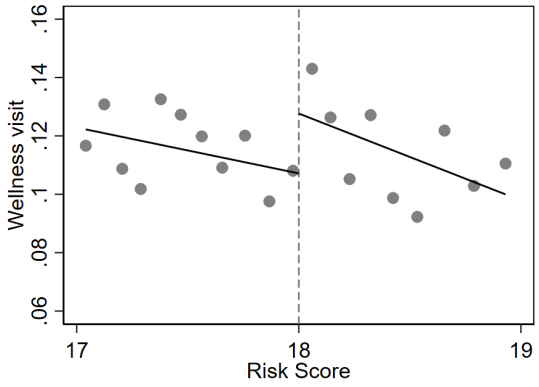
Figure 4: Regression Discontinuity Plots: Child Health



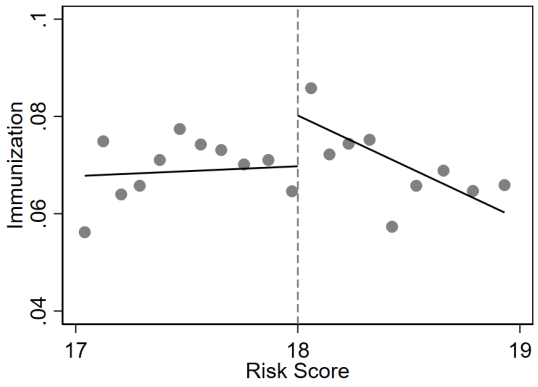

(a) Injury



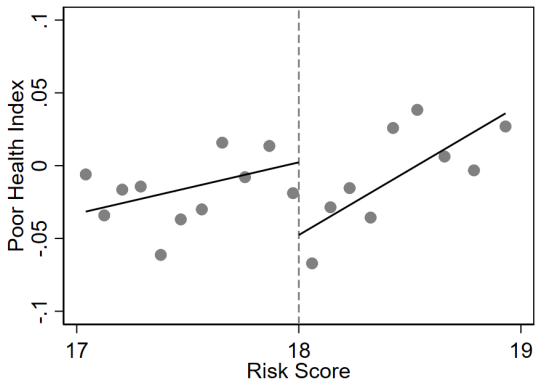

(b) Abuse/Neglect




(c) Wellness Visit



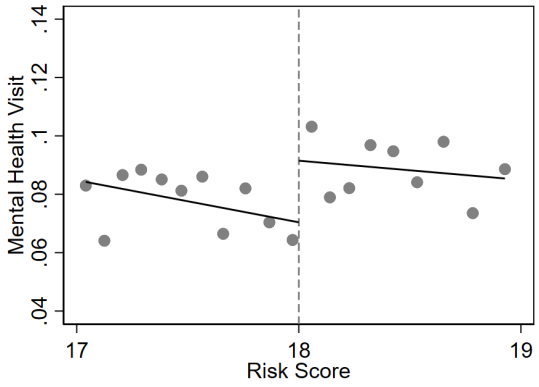

(d) Immunization



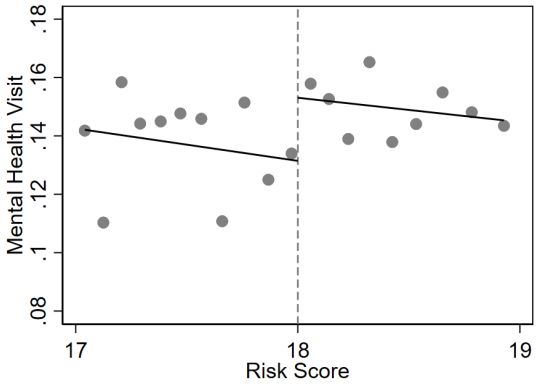

(e) Index

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

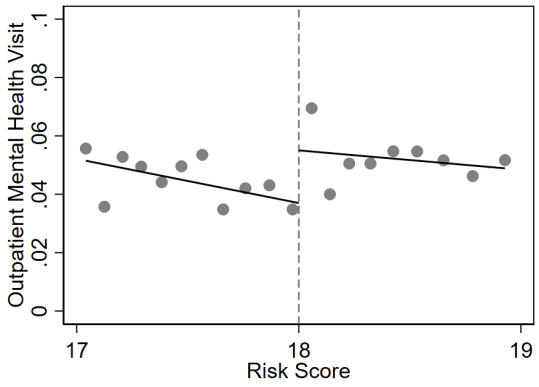
Figure 5: Regression Discontinuity Plots: Parent Mental Health



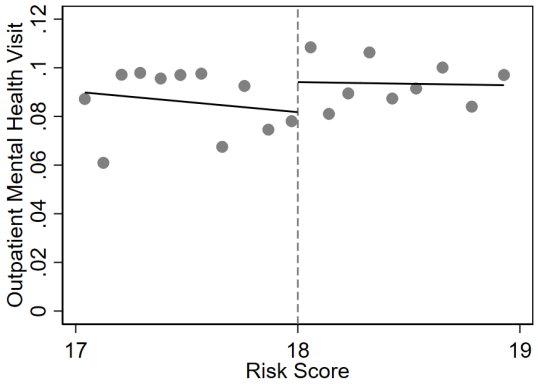

(a) Mental Health Visit (1-2 months)



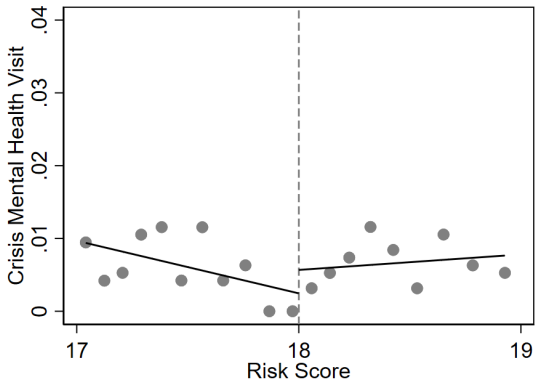

(b) Mental Health Visit (3-12 months)



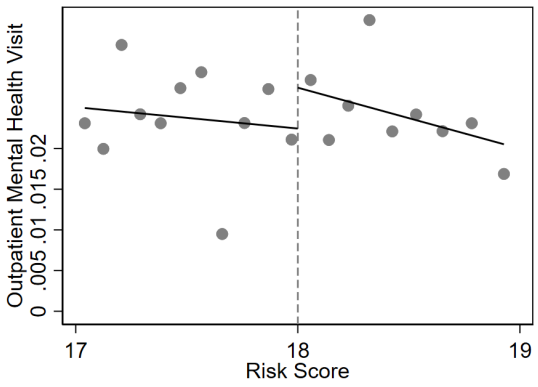

(c) Outpatient Mental Health Visit (1-2 months)




(d) Outpatient Mental Health Visit (3-12 months)




(e) Crisis Mental Health Visit (1-2 months)




(f) Crisis Mental Health Visit (3-12 months)

For each outcome variable reported in Table 7, this figure plots the average value in each of twenty risk score bins within the bandwdith. Risk scores to the right of the vertical line represent those which are subject to a default investigation under the high-risk protocol.

# 8 Tables

Table 1: Summary Statistics (Child-Referral)

| | (1) Full sample | (2) Within Bandwidth |
|---|---|---|
| **Panel A: Child Demographics** | | |
| Female | 0.494 | 0.506 |
| Black | 0.393 | 0.467 |
| White | 0.429 | 0.354 |
| Age | 8.29 | 8.76 |
| **Panel B: Household Composition** | | |
| Two-Parent HH | 0.441 | 0.376 |
| Number of Children | 2.88 | 3.18 |
| **Panel C: Lagged System Involvement** | | |
| Public Benefits (Prev. 6 months) | 0.750 | 0.890 |
| Medicaid (Prev. 6 months) | 0.722 | 0.870 |
| Juvenile Probation (Prev. 6 months) | 0.027 | 0.040 |
| Investigation (Prev. 6 months) | 0.135 | 0.245 |
| **Panel D: Allegation Categories** | | |
| Caregiver Substance Abuse | 0.235 | 0.283 |
| Inadequate Care or Home | 0.253 | 0.257 |
| Neglect | 0.112 | 0.108 |
| Exposure to Risk | 0.184 | 0.181 |
| Physical Maltreatment | 0.202 | 0.164 |
| **Panel E: Reporter Categories** | | |
| Anonymous | 0.082 | 0.083 |
| Friend or Neighbor | 0.032 | 0.027 |
| Legal | 0.116 | 0.123 |
| Medical | 0.071 | 0.067 |
| Parent or Relative | 0.131 | 0.151 |
| Educational Professional | 0.155 | 0.135 |
| Social Worker | 0.154 | 0.176 |
| Therapist | 0.120 | 0.109 |
| **Panel F: Referral Outcomes** | | |
| Investigation Opened | 0.459 | 0.637 |
| Home Removal (12 months) | 0.049 | 0.075 |
| Observations | 128423 | 18857 |

This table presents summary statistics for our main analysis sample, at the child-referral level. (I.e., one child may be included multiple times if she appears on multiple referrals.) Column (1) presents average values for the full sample, across the distribution of risk score. Column (2) presents average values for child-referrals within our main RD bandwidth (risk score of 17-18).

Table 2: Summary Statistics (Parent-Referral)

| | (1) Full sample | (2) Within Bandwidth |
|---|---|---|
| **Panel A: Parent Demographics** | | |
| Mother | 0.451 | 0.412 |
| Father | 0.545 | 0.587 |
| Black | 0.378 | 0.461 |
| White | 0.472 | 0.411 |
| Age | 36.24 | 36.72 |
| Alleged Perpetrator | 0.471 | 0.433 |
| **Panel B: Lagged System Involvement** | | |
| Public Benefits (Prev. 6 months) | 0.615 | 0.745 |
| Medicaid (Prev. 6 months) | 0.483 | 0.611 |
| Criminal Court (Prev. 6 months) | 0.207 | 0.304 |
| Investigation (Prev. 6 months) | 0.146 | 0.262 |
| Observations | 130864 | 19003 |

This table presents summary statistics for our main analysis sample, at the parent-referral level. (I.e., one parent may be included multiple times if she appears on multiple referrals.) Column (1) presents average values for the full sample, across the distribution of risk score. Column (2) presents average values for parent-referrals within our main RD bandwidth (risk score of 17-18).

Table 3: Validity Tests

| | Default Investigation | Outcome mean | Obs. |
|---|---|---|---|
| **Panel A: Demographic and Referral Characteristics** | | | |
| Female | 0.00211 | 0.505 | 18857 |
| | (0.0155) | | |
| Black | -0.0207 | 0.466 | 18857 |
| | (0.0262) | | |
| White | 0.0215 | 0.353 | 18857 |
| | (0.0242) | | |
| Age | -0.196 | 8.559 | 18840 |
| | (0.203) | | |
| Two-Parent HH | 0.0206 | 0.378 | 18857 |
| | (0.0263) | | |
| Number of Children | 0.0885 | 3.131 | 18857 |
| | (0.111) | | |
| Caregiver Substance Abuse | -0.00584 | 0.297 | 18857 |
| | (0.0236) | | |
| Inadequate Care/Home | -0.0221 | 0.258 | 18857 |
| | (0.0240) | | |
| Neglect | 0.00795 | 0.105 | 18857 |
| | (0.0176) | | |
| Exposure to Risk | 0.0299 | 0.181 | 18857 |
| | (0.0196) | | |
| Physical Maltreatment | -0.00704 | 0.167 | 18857 |
| | (0.0200) | | |
| **Panel C: Lagged Outcomes (Previous 6 months)** | | | |
| Investigation | -0.0110 | 0.219 | 18857 |
| | (0.0232) | | |
| Removal | 0.00164 | 0.00247 | 18857 |
| | (0.00158) | | |
| Medicaid | -0.0116 | 0.852 | 18857 |
| | (0.0144) | | |
| Injury | 0.00215 | 0.0785 | 18857 |
| | (0.00891) | | |
| Abuse/Neglect | -0.00691 | 0.0638 | 18857 |
| | (0.00827) | | |
| Sickness | -0.000131 | 0.0114 | 18857 |
| | (0.00328) | | |
| ACSC | -0.0237* | 0.182 | 18857 |
| | (0.0135) | | |
| Wellness visit | -0.0282* | 0.221 | 18857 |
| | (0.0153) | | |
| Immunization | -0.0142 | 0.156 | 18857 |
| | (0.0129) | | |
| Poor Health Index | 0.00431 | 0.00478 | 18857 |
| | (0.0177) | | |
| JPO | 0.00794 | 0.0319 | 18857 |
| | (0.00608) | | |
| Magisterial Court | 0.0108 | 0.0437 | 18857 |
| | (0.00684) | | |
| **Panel D: Medicaid enrollment** | | | |
| Medicaid (1-2 months) | -0.00936 | 0.875 | 18857 |
| | (0.0146) | | |
| Medicaid (3-12 months) | -0.00657 | 0.892 | 18857 |
| | (0.0135) | | |

Standard errors in parentheses

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

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

Table 4: First Stage

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

Standard errors in parentheses

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

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

Table 5: Results: Child Physical Health

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

Standard errors in parentheses

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

This table reports reduced-form results from estimating equation 1, setting $Y_{ir}$ equal to each of the outcome variables described in the column titles. 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: Parents - Services

| | (1) Medicaid | (2) Public Benefits | (3) Housing | (4) Mental Health | (5) Drug/Alcohol |
|---|---|---|---|---|---|
| ***Panel A: 1-2 months after referral*** | | | | | |
| Default Investigation | -0.000469 | 0.00186 | 0.00144 | 0.0272*** | 0.00350 |
| | (0.0149) | (0.0135) | (0.00762) | (0.0102) | (0.00960) |
| Mean | 0.590 | 0.732 | 0.0582 | 0.111 | 0.0999 |
| ***Panel B: 3-12 months after referral*** | | | | | |
| Default Investigation | -0.00916 | 0.00447 | -0.00538 | 0.0359*** | -0.00509 |
| | (0.0148) | (0.0132) | (0.00896) | (0.0127) | (0.0115) |
| Mean | 0.631 | 0.758 | 0.0864 | 0.193 | 0.143 |
| RightBW | 8767 | 8767 | 8767 | 8767 | 8767 |
| LeftBW | 10232 | 10232 | 10232 | 10232 | 10232 |

Standard errors in parentheses

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

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

Table 7: Parents - Medicaid-funded Mental Health Services

| | (1) MH Claim | (2) Outpatient MH | (3) MH Crisis |
|---|---|---|---|
| ***Panel A: 1-60 days after referral*** | | | |
| Default Investigation | 0.0167* | 0.0195*** | 0.00462** |
| | (0.00867) | (0.00678) | (0.00215) |
| Mean | 0.0811 | 0.0457 | 0.00782 |
| ***Panel B: 61-360 days after referral*** | | | |
| Default Investigation | 0.0285*** | 0.0186** | 0.00368 |
| | (0.0109) | (0.00876) | (0.00462) |
| Mean | 0.136 | 0.0861 | 0.0239 |
| RightBW | 8766 | 8766 | 8766 |
| LeftBW | 10235 | 10235 | 10235 |

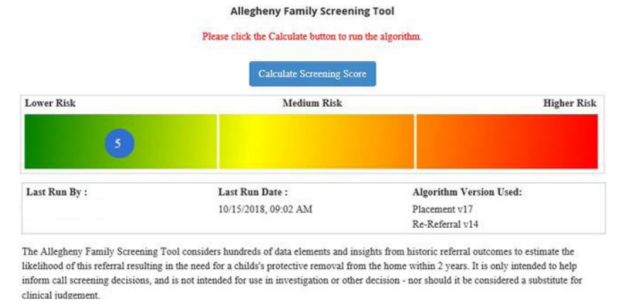
Standard errors in parentheses

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

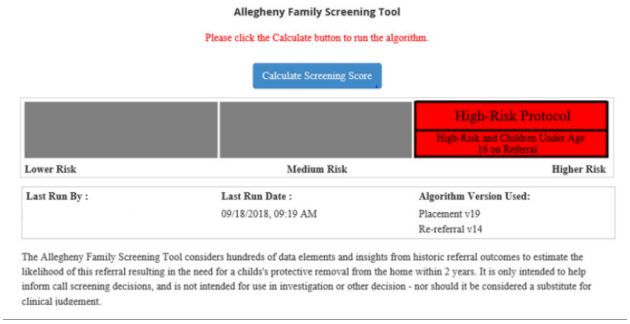
This table reports reduced-form results from estimating equation 1, setting $Y_{ir}$ equal to each of the outcome variables described in the column titles. Average values of the outcome variable below the threshold, as well as the number of observations in each of the left and right bandwidth, are reported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

# A Appendix A: Additional Figures and Tables

Figure A1: Screener View of AFST Output



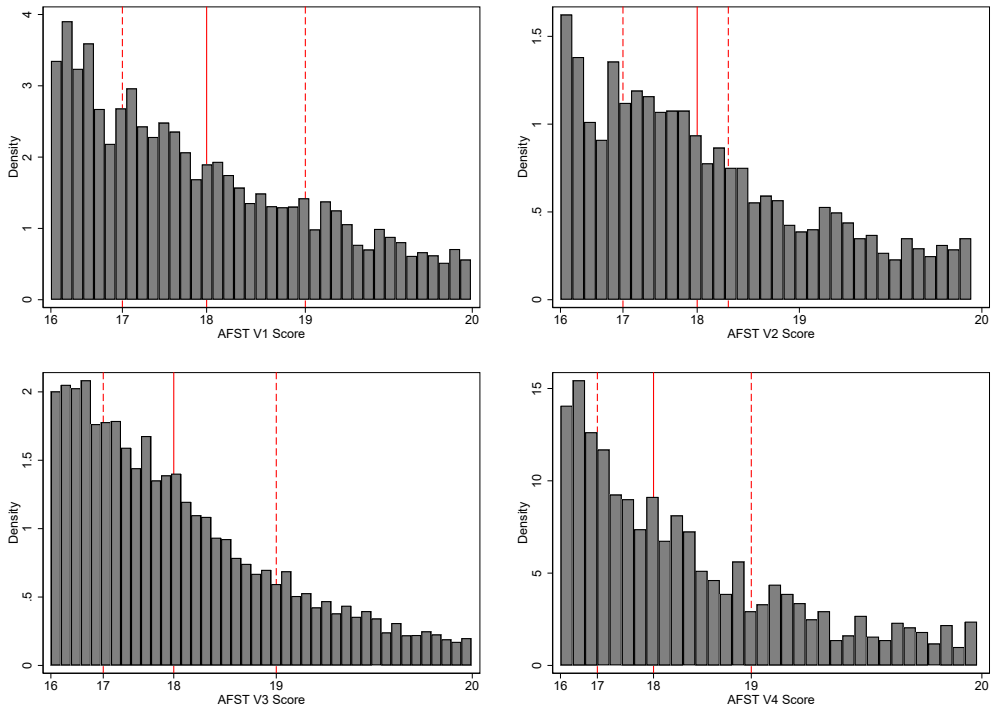

(a) No protocol




(b) High-risk protocol

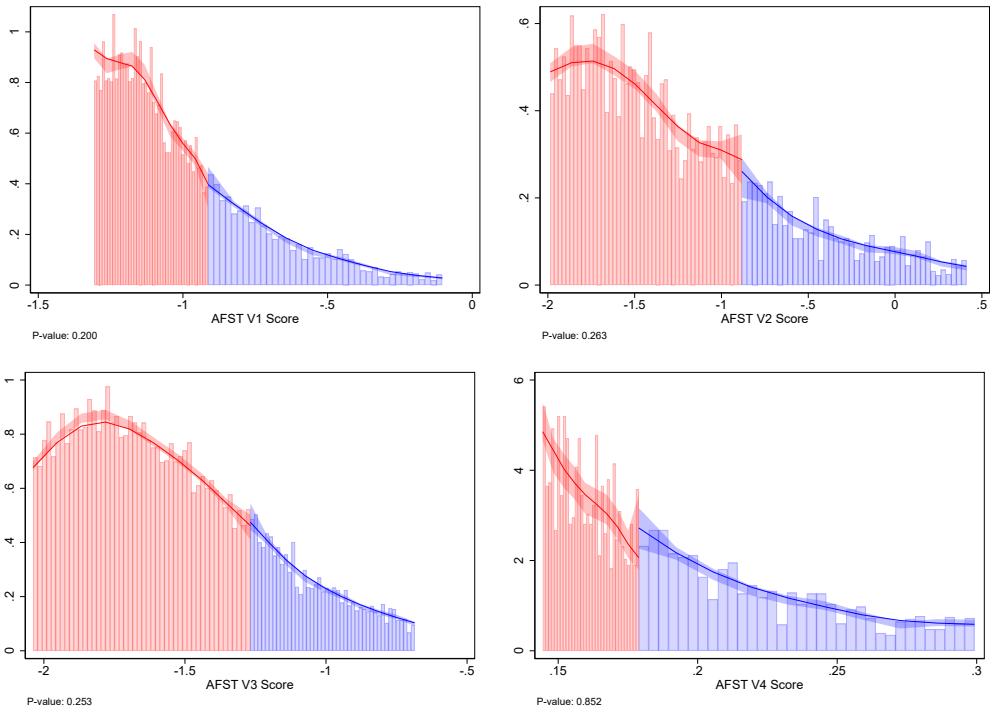
This figure shows two versions of what a screener might see after running the AFST. Panel (a) shows the AFST result for a risk score of 5, and panel (b) shows the AFST result for a high-risk protocol referral (risk score above 17).

Figure A2: Density by AFST Version



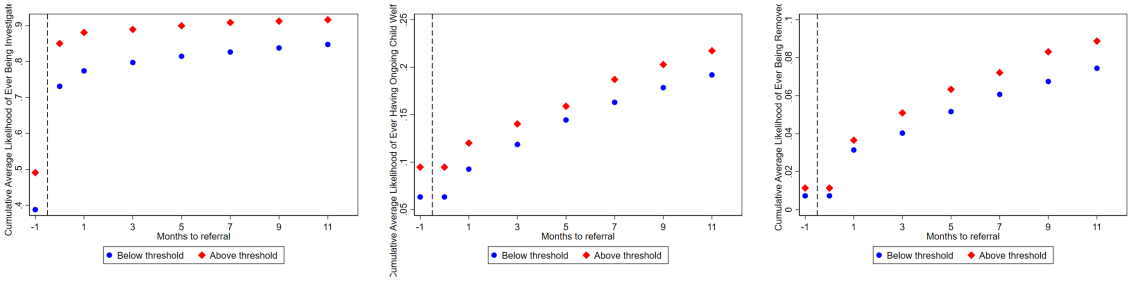

This Figure replicates Figure 1 for a smaller range of risk scores around the threshold. For each of the four AFST versions, this figure plots the density of latent risk score. Latent risk score is on the x-axis, with discrete scores marked by ticks. The solid vertical red line shows the cutoff for a high-risk protocol at discrete risk score of 18. The dashed vertical red lines at discrete scores of 17 and 19 show the endpoints for our main bandwidths.

Figure A3: Manipulation Testing




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

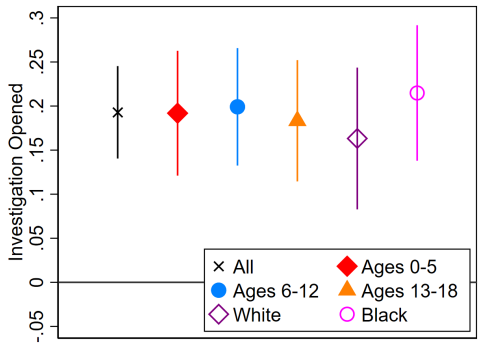
Figure A4: Control Group Exposure to Child Protection Systems +/- 12 Months to Referral




(a) Ever Investigated (b) Ever Had Services (c) Ever Removed from Home

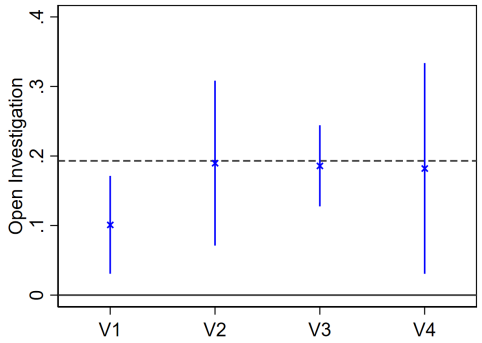
This cumulative likelihood of having a given exposure to CYF relative to the time of the referral. Each outcome retrospectively looks over the 12 months before the referral up until the time relative to the referral given on the X-axis. So the point at time -1: is the likelihood of having had the given intervention within the 12 months before the referral, and then the graph continues to show how that likelihood changes at each month up to a year after the referral.

Figure A5: First Stage by Age and Race



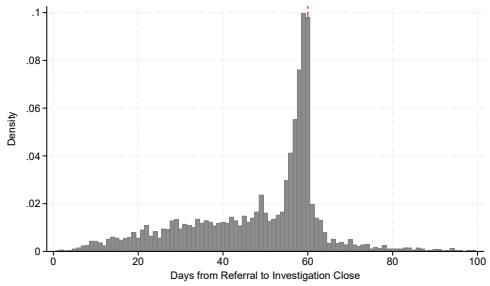

This figure shows coefficients and 95% confidence intervals from estimating equation 1 for different subgroups of our main analysis sample, where $Y_{ir}$ is equal to one if a referral is investigated, and zero otherwise.

Figure A6: First Stage by AFST Version




This figure shows coefficients and 95% confidence intervals from estimating equation 1 separately by AFST version, where $Y_{ir}$ is equal to one if a referral is investigated, and zero otherwise. The MSE-optimal bandwidth is calculated separately for each AFST version.

Figure A7: First Stage by AFST Version




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. Referrals with investigation length above 100 days (2% of referrals) are excluded from the figure for readability.

Table A1: Child Physical Health

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

Standard errors in parentheses

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

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

Table A2: Downstream CYF

| | (1) Open Case (Children) | (2) Open Case (Parents) | (3) Removal (Children) | (4) Removal (Parents) |
|---|---|---|---|---|
| Default Investigation | 0.0401 | 0.0199 | 0.0177 | 0.0175 |
| | (0.0257) | (0.0229) | (0.0111) | (0.0170) |
| Mean | 0.309 | 0.310 | 0.0691 | 0.122 |
| RightBW | 8723 | 8766 | 8723 | 8766 |
| LeftBW | 10132 | 10235 | 10132 | 10235 |

Standard errors in parentheses

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

This table reports reduced-form results from estimating equation 1, setting $Y_{ir}$ equal to each of the outcome variables described in the column titles, in the year following the referral. Average values of the outcome variable below the threshold, as well as the number of observations in each of the left and right bandwidth, are reported below the coefficients and standard errors. The sample is all child-referrals (in Columns 1 and 3) or parent-referrals (in Columns 2 and 4) with a referral risk score of 17-18.