The Impact of Maltreatment Investigations on Child Wellbeing*

Jason Baron† Brian A. Jacob‡ Richard Lombardo§

March 2026

Abstract

Child Protective Services (CPS) investigations are among the most frequent state interventions in family life: by age 18, roughly 37 percent of U.S. children experience an investigation. Despite their prevalence, there is little causal evidence on the effects of investigations themselves. We study the impact of CPS investigations using linked administrative data from Michigan that include the universe of screened-in and screened-out hotline reports from 2017–2022 merged to statewide public school records. Our research design exploits the quasi-random assignment of hotline screeners who differ in their propensity to screen in reports for investigation. We examine subsequent CPS contact as a proxy for child safety, as well as educational outcomes including chronic absenteeism, standardized test scores, and high school graduation. While naïve comparisons suggest that investigated children fare worse, instrumental variables estimates indicate that marginal investigations have small positive effects. Point estimates suggest reductions in subsequent CPS involvement, though these effects are modest and not statistically significant. For educational outcomes, we find precise null effects on chronic absenteeism and test scores and a marginally significant increase in high school graduation. Marginal treatment effect estimates reveal substantial heterogeneity: investigations appear beneficial for higher-risk children but may be harmful for lower-risk children. These findings suggest caution in reducing investigations at the margin and highlight the importance of targeting investigations toward children at greatest risk.

Keywords: Child protective services; maltreatment investigations; examiner assignment research design; marginal treatment effects

*We thank seminar participants at the University of Connecticut (Economics) and the University of Michigan (CIERS) for helpful comments and suggestions. This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI).

†Duke University and NBER. Email: jason.baron@duke.edu.

‡University of Michigan and NBER. Email: bjacob@umich.edu.

§Harvard University. Email: richardlombardo@fas.harvard.edu.

1 Introduction

Child Protective Services (CPS) investigations are among the most widespread forms of government intervention in American family life. By age 18, approximately 37 percent of all U.S. children experience a CPS investigation, with rates exceeding 50 percent for Black children (Kim et al., 2017). Among children in the lowest income decile, roughly 50 percent are reported to CPS at some point during childhood (Baron et al., 2026). These rates far exceed the approximately 5 percent of children who enter foster care by age 18 (Doyle, Emanuel and Eyzaguirre, 2025), underscoring that CPS investigations—not foster care placements—represent the predominant form of contact between families and the child welfare system.

Despite the extraordinary prevalence of CPS investigations, there is remarkably little causal evidence on their consequences. A large literature, following the seminal work of Doyle (2007, 2008), examines the effects of foster care placement on child outcomes (e.g., Bald et al., 2022; Baron and Gross, 2024). Yet foster care placement represents only a small share of system involvement, accounting for roughly 3 percent of investigations in our setting. Investigations themselves affect far more families and serve as the gateway to a wide range of potential interventions, including monitoring, service provision, and possible child removal. Understanding whether investigations help or harm children is therefore central to evaluating the role of the child welfare system.

The effects of investigations are theoretically ambiguous. Investigations may benefit children by identifying and interrupting abuse or neglect, increasing monitoring of families, and connecting caregivers to services. The presence of caseworkers in the home may also create a deterrence effect: when families know that authorities are observing the household and may return unannounced, caregivers may be less likely to engage in behaviors that place children at risk (Rittenhouse, Simon and Lacey, 2025). At the same time, investigations can impose significant costs on families, including intrusive surveillance, stigma, stress, and fear of child removal. Qualitative evidence suggests that CPS contact may lead some families to withdraw from institutions such as schools and healthcare providers that could trigger future reports (Fong, 2023). These concerns have motivated policy reforms in many states aimed at reducing the number of families subject to investigation, including diversion programs, alternative response pathways, and "warmlines" that connect families to services without triggering a formal CPS report.

Studying CPS investigations presents two key empirical challenges. The first is data availability. Most administrative data systems observe only cases that are screened in for investigation, making it difficult to observe outcomes for children whose reports were screened

out. We address this challenge using administrative data from Michigan's centralized CPS intake hotline, which records the universe of screened-in and screened-out reports between 2017 and 2022. We link these records to statewide public school data, allowing us to track educational outcomes for children regardless of whether their report was investigated.

The second challenge is identification. Children are not randomly selected for investigation, and naïve comparisons between investigated and non-investigated children are likely to reflect differences in underlying risk rather than the causal effects of investigation itself. Our research design exploits institutional features of Michigan's centralized CPS hotline. Incoming calls are routed to the next available screener in a queue, generating quasi-random assignment of reports to hotline workers. Because screeners differ in their propensity to screen in reports for investigation, assignment to a stricter or more lenient screener creates plausibly exogenous variation in whether a report is investigated. We use this variation in screener leniency as an instrument for investigation in an examiner-assignment research design, following a growing literature that exploits quasi-random variation in decision-maker discretion to estimate causal effects (e.g., Kling, 2006; Aizer and Doyle, 2015; Dobbie, Goldin and Yang, 2018; Norris, Pecenco and Weaver, 2021; Collinson et al., 2024, 2025). This research design allows us to identify a local average treatment effect (LATE), which represents the impact of investigation for children at the margin—those for whom screeners may disagree about whether an investigation is warranted. We argue that this LATE is policy relevant: it captures the effects of marginal changes in screening thresholds that shift which children are investigated at the margin.

We examine outcomes in two domains. First, we study subsequent CPS involvement within 36 months of the focal report, including additional investigations, substantiated allegations, case openings, and foster care placement. Because the primary goal of CPS investigations is to ensure child safety, these outcomes are informative about both underlying child wellbeing and the degree of subsequent concern from mandated reporters, family members, and others in the child's environment. At the same time, subsequent CPS contact is an imperfect proxy for child wellbeing because it depends not only on underlying safety, but also on whether someone reports concerns to the system. An investigation could therefore affect future CPS involvement both by changing child safety and by changing reporting behavior. For this reason, we also examine educational outcomes—chronic absenteeism, standardized test scores, and high school graduation—which provide a more direct, welfare-relevant measure of how children are doing. Together, these outcomes allow us to assess whether investigations affect child wellbeing more broadly.

Our main findings are as follows. Naïve OLS estimates suggest that children who are in-

vestigated experience worse outcomes on average, consistent with the fact that higher-risk children are more likely to be investigated. In contrast, our instrumental variables estimates indicate that marginal investigations have small and generally positive effects. For child welfare outcomes, point estimates suggest reductions in subsequent CPS involvement—including future investigations, substantiated allegations, and case openings—though these estimates are modest and statistically insignificant. For educational outcomes, we find precise null effects on chronic absenteeism and test scores and a marginally significant increase in high school graduation.

While average effects are modest, we find substantial heterogeneity in the impacts of investigation. We find little evidence of systematic differences across observable characteristics such as gender, race, age, or poverty status. However, marginal treatment effect estimates reveal pronounced heterogeneity along unobserved dimensions of risk. Investigations are beneficial for children with low resistance to investigation—those most likely to be screened in regardless of screener assignment—but harmful for children with high resistance to investigation, who are screened in only by screeners with the highest screening propensities. For children with low resistance to investigation, we estimate that investigations reduce chronic absenteeism by approximately 18 percentage points and subsequent case openings by about 13 percentage points. In contrast, for children with high resistance, investigations increase chronic absenteeism by roughly 8 percentage points and subsequent case openings by about 14 percentage points. These results suggest that the consequences of investigations depend critically on where the screening margin lies.

To our knowledge, only one other study provides causal estimates of the impacts of CPS investigations. Rittenhouse, Simon and Lacey (2025) study investigations in Allegheny County, Pennsylvania, where referrals above a threshold in a predictive risk model are defaulted into investigation. Their regression discontinuity design identifies the effects of investigation for children near a high-risk cutoff and focuses primarily on child health outcomes. Their setting is limited to "general protective services" (GPS) referrals, which involve only allegations of neglect; allegations of physical or sexual abuse are automatically screened in by law and thus do not involve discretionary screening decisions. They find that investigations reduce injury-related medical claims and increase preventive healthcare utilization during the roughly 60-day period when investigations are typically ongoing, suggesting that investigations improve child safety while they are active. However, these effects fade after the investigation period ends, with little evidence of longer-term impacts.

Our study complements this important work by examining a different margin of investigation. While Rittenhouse, Simon and Lacey (2025) focus on children near a high-risk cutoff

within the subset of referrals that primarily capture neglect, we exploit quasi-random variation in the leniency of hotline screeners to identify the effects of investigations for children at the margin of discretionary screening decisions across a broader set of allegation types. This distinction is important because the consequences of investigation may differ across types of maltreatment; for example, investigations of neglect may operate through different mechanisms than investigations involving physical or sexual abuse. In addition, around the high-risk threshold in their setting, many children who are initially screened out are subsequently re-referred and investigated, leading to relatively limited persistence in differences in investigation exposure over time. In contrast, our design generates persistent differences in investigation exposure over multiple years following the focal report, allowing us to study medium- and longer-run outcomes. Finally, our continuous instrument enables us to estimate marginal treatment effects that characterize how the impacts of investigation vary across children with different propensities for investigation, providing insight into heterogeneity in treatment effects and the consequences of policy changes that shift screening thresholds.

These findings have direct implications for current policy debates about the scope of CPS involvement in the United States. Many states have sought to reduce the number of investigations through diversion programs and alternative response systems. Our results suggest caution in reducing investigations at the margin. We find little evidence that marginal investigations harm children—and some evidence of modest benefits, including an increase in high school graduation—indicating that marginal investigations may be beneficial. At the same time, our heterogeneity results show that the effects of investigations vary substantially across children: investigations benefit higher-risk children but may harm lower-risk children. Taken together, these findings suggest that both broad reductions and broad expansions of investigations should be approached with caution, underscoring the importance of targeting investigations toward children at greatest risk. In ongoing work, we use our marginal treatment effect estimates to evaluate counterfactual policy changes—such as reductions in screening rates—by estimating a range of policy-relevant treatment effects (PRTEs).

Our study also complements a broader literature in economics that examines the efficiency of child protection decisions by exploiting quasi-random variation in the discretion of judges or caseworkers. Seminal work by Doyle (2007, 2008) uses the quasi-random assignment of investigators to study the effects of foster care placement in Illinois, finding that removal can worsen outcomes for marginal children. Subsequent research has examined other settings and margins of removal decisions, with mixed results depending on the context. For example, Bald et al. (2022) find that removal improves educational outcomes for young girls investigated for maltreatment in Rhode Island, while Gross and Baron (2022) and Baron and

Gross (2024) document positive effects of foster care placement on educational outcomes and later-life criminal outcomes in Michigan. In contrast, Helénsdotter (2025) introduces a novel instrument based on judges' propensities to order removal and finds that foster care placement increases mortality and other adverse health outcomes in Sweden. These studies focus on the share of cases in which children are removed from their homes. In contrast, we study an earlier and common margin of the child welfare system—the decision to investigate a report of maltreatment—which serves as the gateway to subsequent interventions.

Finally, our study follows a growing literature that uses marginal treatment effects (MTEs) to study heterogeneity in treatment effects in settings with quasi-random assignment of decision makers. Building on the framework developed by Heckman and Vytlacil (1999, 2005, 2007) and subsequent work on identification and extrapolation with instrumental variables (Mogstad and Torgovitsky, 2018), recent applied research has combined leniency designs with MTE estimation to characterize how treatment effects vary across individuals with different propensities for treatment. For example, Bhuller et al. (2020) study incarceration decisions using quasi-random judge assignment and show that the effects of imprisonment vary systematically with defendants' unobserved likelihood of incarceration. Similarly, Agan, Doleac and Harvey (2023) examine prosecutorial discretion in misdemeanor cases, and Humlum, Munch and Rasmussen (2025) study caseworker assignment in unemployment programs. Related work studies heterogeneous treatment effects using structural selection models and bounding approaches rather than MTE estimation. For example, Rose and Shem-Tov (2021) combine regression discontinuities in sentencing guidelines with a Roy-style selection model to estimate bounds on the heterogeneous effects of incarceration. We apply this framework to the child welfare system, estimating marginal treatment effects of CPS investigations to characterize how the impacts of investigation vary across the distribution of children's unobserved resistance to investigation. This approach allows us to evaluate how the consequences of CPS investigations may change as policy shifts the margin of screening decisions.

The remainder of the paper proceeds as follows. Section 2 provides background on the child welfare system. Section 3 describes the data sources and analysis sample, while Section 4 discusses the research design and probes the identification assumptions. Section 5 presents the main results, heterogeneity analyses, and marginal treatment effects. Section 6 concludes and outlines three key extensions that we plan to incorporate in the next draft: (i) estimating PRTEs to evaluate counterfactual changes in screening rates, (ii) exploiting a policy-driven change in screening behavior in 2019 to study treatment effects at an alternative margin, and (iii) analyzing mechanisms using detailed information on investigation processes and outcomes.

2 Background

The child welfare system in the United States is designed to protect children from abuse and neglect. The process typically begins when a concerned individual—a mandated reporter such as a teacher, doctor, or police officer, or a non-mandated reporter such as a neighbor or family member—contacts a CPS hotline to report concerns about a child's safety. In Michigan, the vast majority of reports (97% percent) are made by phone, with the remaining share submitted through an online reporting system. For ease of exposition, we refer to all reports as "calls" throughout the paper.

Figure 1 illustrates the CPS process in Michigan, which functions similarly to many other states. The state operates two centralized intake centers, located in Detroit and Grand Rapids, that receive and process reports from across the state (and share the same phone number). When a call is received, a hotline screener gathers information about the child and the alleged maltreatment and determines whether the report meets the statutory criteria for investigation. Calls last approximately 15 minutes on average, at the end of which the screener makes a binary decision: whether to "screen in" the report for a formal investigation or "screen out" the report. Reports that are screened out receive no further CPS action. Roughly 50 percent of calls in our sample are screened in.

At the time of the call, screeners gather detailed information provided by the reporter, including the nature of the allegation, the identity and type of reporter (e.g., teacher, medical professional, family member), and information about the child and household. In addition, screeners have access to administrative records through the state's case management system, including prior CPS contact for the child and family (e.g., prior screened-in and screened-out reports), as well as information on the findings of past investigations or foster care involvement. Screeners may also access other administrative systems to verify addresses, identify household members, and observe participation in public assistance programs. Together, this information provides a snapshot of both the current allegation and the family's prior system involvement. At the same time, screeners must make decisions quickly—typically within a single call—and not all available information is necessarily consulted in every case. Screening decisions are guided by relatively broad statutory criteria and agency guidelines that emphasize assessing both the validity of the allegation and the risk of harm to the child. In practice, this framework leaves substantial room for discretion, as screeners must weigh incomplete information and make judgment calls about whether a report warrants investigation.

Importantly for our analysis, incoming reports—whether submitted by phone or through the online portal—are assigned to screeners through a centralized, queue-based system. Reports

enter a common queue and are routed to the available screener who has been waiting the longest since their previous assignment. This assignment process generates quasi-random variation in which screener handles a given report, conditional on the exact day and shift in which the report occurs.

If a report is screened in, it is transferred from the centralized intake center to a county child welfare office, where an investigator is assigned to the case. Investigations typically begin within 24 hours for higher-risk cases and within several days for lower-risk cases, and must generally be completed within 30 days. The investigation process involves direct interaction with the family and is designed to assess both the credibility of the allegation and the child's safety. In most cases, investigators conduct one or more home visits, during which they observe the home environment, assess the child's immediate safety, and interview caregivers and children. Investigators typically speak separately with children, parents, and other household members, and may also contact individuals outside the household such as teachers or medical providers. In addition, investigators may review administrative records, including prior CPS reports, school records, or medical information, to inform their assessment. Throughout the investigation, caseworkers evaluate both immediate safety concerns and longer-run risk, and in some cases may implement temporary safety plans or take steps to mitigate perceived risks.

In most offices, screened-in reports are assigned to investigators on a rotational basis. Cases cycle through investigators according to who is next in the rotation, rather than based on the investigator's specific characteristics or expertise. At the conclusion of the investigation, the caseworker must determine whether there is sufficient evidence to substantiate the allegation of maltreatment. Approximately one quarter of investigations result in substantiated findings. Substantiation decisions are based on the evidence available in the case rather than on assessed risk, meaning that serious allegations may remain unsubstantiated while less severe allegations may be substantiated.

If the investigator does not substantiate the allegation, the case is closed with no further action. If an allegation is substantiated, several outcomes are possible depending on the investigator's assessment of the child's safety. In lower-risk situations, the case may be closed after referral to community-based services, such as parenting programs or food pantries. In other cases, the child welfare agency may open a case and provide ongoing monitoring and services to the family, which can include regular home visits, service coordination, and participation in support programs (e.g., mental health counseling or substance abuse treatment). In rare situations where the investigator determines that the child faces severe or imminent risk of harm (approximately 3% of screened-in calls in our sample), the agency may petition

the court to remove the child from the home and place the child in foster care.

The screening decision is therefore the gateway to a cascade of potential interventions. Understanding the causal effects of this initial decision is important because it determines which families are subject to the investigative process and its downstream consequences.

3 Data

3.1 Data Sources and Matching

Our analysis draws on two primary administrative data sources. The first is child welfare records from the Michigan Department of Health and Human Services (MDHHS), which include the universe of reports made to Michigan's centralized CPS intake hotline between 2017 and 2022. These records capture both reports that are screened in for investigation and those that are screened out. For each call, we observe the identity of the assigned screener, the screening decision, and information about the children and allegations involved. For screened-in cases, we additionally observe investigation outcomes including substantiation decisions, case openings, and foster care placements. In addition, we observe detailed information for all reports that are screened in for investigation from 2008 to 2025, which allows us to create measures of child welfare involvement prior to the *focal call*.

The second data source is K–12 public school records from the Michigan Department of Education (MDE), which include enrollment, attendance, standardized test scores, and graduation outcomes for all students enrolled in Michigan public schools—including both traditional public and charter schools—from the 2002–03 school year onward.

We link records across these two systems through probabilistic matching conducted by the Michigan Education Data Center (MEDC). Because no common unique identifier exists across the child welfare and education systems, records are matched based on first and last name, date of birth, and gender. The matching algorithm assigns a posterior probability to each potential match, reflecting the likelihood that two records correspond to the same individual. MEDC applies additional post-processing rules and manual review to remove likely false positives, particularly in cases with similar names but inconsistent birth dates.

Overall, 75.6 percent of all CPS records are successfully matched to a Michigan public school record. This match rate is consistent with expectations, as a substantial share of children in CPS records are too young to have enrolled in school during our sample period. Among children with hotline calls during our sample period who would have reached school age by the end of the sample, the match rate is 88%. Even among school-age children, we do not expect a complete match rate because some children move out of state, attend private

schools or are homeschooled. The resulting linked dataset provides comprehensive coverage of CPS involvement for children observed in Michigan public schools and forms the basis for our analysis.

3.2 Analysis Sample

Our analysis sample is constructed at the child-by-call level, reflecting the fact that a single call may involve multiple children and that a given child may be the subject of multiple reports over time. The data include reports made to the centralized hotline between January 2017 and December 2022 and impose several sample restrictions.

We begin with 1,283,791 child-by-call observations between January 2017 and December 2022. We first restrict the sample to observations that successfully match to the education data, dropping 221,105 child-by-call observations. We next exclude observations that are likely false positive matches: children born after September 1, 2018, who are unlikely to have reached school age by the end of our sample period ($N = 43,983$), and children who appear in the education records more than 15 years prior to the CPS report ($N = 18,703$). We then exclude calls assigned to screeners with fewer than 200 total calls ($N = 2,854$), ensuring that each screener's propensity to screen in reports is estimated with sufficient precision to serve as a reliable instrument. Our final analysis sample includes 997,146 child-by-call observations corresponding to 382,827 unique children. 48% of children have a single call during this period, 20% have two calls, and 32% have more than two calls.

To facilitate linking child welfare and educational records, we assign each report to an academic year. Specifically, reports from July 1 of year t through June 30 of year $t + 1$ are assigned to academic year $t + 1$. For children enrolled in public schools prior to the report, we use earlier educational records to construct baseline covariates. Similarly, we use prior child welfare records to construct measures of CPS involvement before the focal report, which are also included as baseline covariates.

For outcomes measured over a fixed horizon, such as high school graduation and subsequent CPS contact within three years, we use a child-by-call dataset. To examine educational outcomes that vary over time, such as standardized test scores and chronic absenteeism, we use a child-by-call-by-academic-year panel, beginning in the academic year of the report. In all cases, we restrict attention to outcomes measured after the focal call.

Table 1 presents summary statistics. Column (1) describes the population of Michigan public school students as a point of reference, while Column (2) presents the full analysis sample. Columns (3) and (4) report statistics separately for reports that were screened in and screened out, respectively.

Relative to the population of Michigan public school students, children subject to CPS reports are more likely to be Black and economically disadvantaged, more likely to receive special education services, and perform substantially worse on standardized achievement tests. These patterns are consistent with prior evidence on the concentration of CPS involvement among disadvantaged populations (Ryan et al., 2018).

In contrast, observable differences between children in screened-in and screened-out reports are relatively modest. Rates of economic disadvantage and baseline test scores are similar across groups, though screened-in reports are somewhat more likely to originate from law enforcement and less likely to originate from family members or medical providers. These patterns suggest that screening decisions are not strongly driven by observable child characteristics alone, consistent with the role of discretion in hotline screening decisions. Finally, children in screened-in reports are, on average, younger than those in screened-out reports, which likely explains why they are less likely to have prior CPS involvement.

4 Research Design

Our research design exploits quasi-random variation in the assignment of hotline screeners at Michigan's centralized CPS intake centers. As described in Section 2, incoming reports enter a queue and are routed to the screener who has been waiting the longest since their previous call. Conditional on the exact day and shift, this queue-based system generates quasi-random assignment of reports to hotline workers.

As we show below, screeners differ in their propensity to screen in otherwise similar reports for investigation. As a result, children whose reports are assigned to relatively strict screeners are more likely to be investigated than similar children assigned to more lenient screeners. We exploit this variation in screener stringency to instrument for whether a report is screened in for investigation. This examiner-assignment research design has been widely used to study the causal effects of discretionary institutional decisions (e.g., Kling, 2006; Aizer and Doyle, 2015; Dobbie, Goldin and Yang, 2018; Norris, Pecenco and Weaver, 2021; Collinson et al., 2024, 2025)

4.1 Screener Leniency Instrument

Following the standard approach in the examiner-assignment literature, we measure screener stringency using a leave-out measure of the assigned worker's screening tendency (Chyn, Frandsen and Leslie, 2025; Goldsmith-Pinkham, Hull and Kolesár, 2025). We first residualize the screening decision with respect to the rotation structure of the hotline by regressing the screening decision on day-by-shift fixed effects. These fixed effects capture the exact queue

conditions under which reports are assigned to screeners. We then construct the screener leniency instrument as the leave-out mean of these residualized screening decisions among all other reports handled by the same screener. Formally, for child i assigned to screener w, the instrument is

$$Z_{iw} = \frac{1}{n_w - 1} \sum_{k \neq i}^{n_w} \tilde{SI}_{kw}, \tag{1}$$

where $\tilde{SI}_{kw}$ denotes the residualized screening decision for report k handled by screener w, and n_w is the total number of reports handled by screener w. The leave-out construction ensures that child i's own screening decision does not mechanically enter the instrument.

Our sample includes 284 screeners assigned to an average of 2,268 reports each, providing substantial variation in screening tendencies. Figure 2 displays the distribution of the screener leniency instrument. A screener at the 10th percentile screens in reports at a rate approximately 4 percentage points lower than the average screener working in the same day-by-shift rotation group, while a screener at the 90th percentile screens in reports at a rate about 4 percentage points higher. Relative to an average screen-in rate of roughly 50 percent, this represents meaningful variation in screening behavior across workers.

4.2 Estimation

We estimate the causal effects of CPS investigations using two-stage least squares (2SLS). The first-stage equation relates the investigation decision to screener leniency:

$$INV_{iw} = \gamma_1 Z_{iw} + \gamma_2 X_{iw} + \Theta_{rt} + \eta_{iw}, \tag{2}$$

where INV_{iw} is an indicator for whether report i assigned to screener w is screened in for investigation, Z_{iw} is the screener leniency instrument, and X_{iw} is a vector of child and report characteristics. The term Θ_{rt} represents day-by-shift fixed effects capturing the rotation structure of the hotline.

The second-stage equation relates investigation status to outcomes of interest:

$$Y_{iw} = \beta_1 \widehat{INV}_{iw} + \beta_2 X_{iw} + \theta_{rt} + \epsilon_{iw}. \tag{3}$$

The coefficient β_1 identifies the local average treatment effect (LATE) of investigation for compliers—children whose investigation status is influenced by the leniency of their assigned screener. Our baseline specifications include controls for child demographics, prior educational outcomes, and prior CPS involvement. We do not control for the type or number of allegations reported to the hotline because these are coded by the screener and could therefore be endogenous to the screening decision. We cluster standard errors at the call level, since calls are the unit of random assignment to screeners and a single call may involve multiple children.

4.3 Identification Assumptions

Interpreting β_1 as the causal effect of investigation requires four standard assumptions: relevance, exogeneity, monotonicity, and the exclusion restriction.

Relevance. This assumption requires that screener stringency predicts whether a report is screened in for investigation ($\gamma_1 \neq 0$). Table 2 reports first-stage estimates of the effect of screener leniency on the screening decision. The results show a strong positive relationship between screener tendencies and the probability that a report is investigated. In our preferred specification (Column 5), a 10 percentage point increase in a screener's propensity to screen in reports increases the probability that a given report is screened in by roughly 9.6 percentage points. The corresponding F-statistic of 1,913 indicates that the instrument is strong.

Exogeneity. This assumption requires that the unobserved determinants of children's outcomes are independent of the leniency of the assigned screener ($\text{Cov}[Z, \epsilon] = 0$). A key implication is that observable characteristics should not differ systematically across children induced into investigation by screener assignment.

Table 3 evaluates this implication. Column (1) reports naïve differences in observable characteristics between children whose reports were screened in versus screened out. Screened-in children are younger, more likely to be Black, and have worse prior educational outcomes.[1] While these differences are modest, they are consistent with higher-risk children being more likely to be investigated.

Column (2) reports the corresponding differences among treated and control compliers, which assess whether children induced into investigation by screener assignment differ in observable characteristics. For a given covariate X, control complier means are estimated by regressing $(1 - INV) \cdot X$ on $(1 - INV)$ (instrumented by Z) and including exact day-by-shift fixed effects. Similarly, complier means in the treated state are estimated by regressing $INV \cdot X$

[1] Screened-in children are less likely to have prior CPS contact, which may reflect their younger age and therefore shorter exposure to the system.

on INV (instrumented by Z) with the same fixed effects (Abadie, 2003). In contrast to the naïve comparisons, the complier differences are small and generally statistically insignificant. Notably, this is also true for predicted high school graduation—a summary measure that aggregates baseline information into a single index of child risk—where the estimated difference is close to zero and precisely estimated. This pattern is consistent with the quasi-random assignment of screeners and supports the exogeneity of the instrument.

Average Monotonicity. Interpreting our estimates as a weighted average of complier treatment effects requires an average monotonicity assumption (Frandsen, Lefgren and Leslie, 2019). For average monotonicity to hold, the covariance between each child's screener-specific treatment status and screener stringency must be weakly positive. A testable implication of average monotonicity is that the first-stage relationship between screener leniency and investigation status should be positive across all subgroups of children. Table 4 reports subgroup-specific first-stage estimates by gender, race, and age. In all cases the estimated relationship between screener leniency and the probability of investigation is positive and statistically significant, consistent with the average monotonicity assumption.[2]

Exclusion Restriction. The instrument must affect outcomes only through the investigation decision. This assumption is plausible in our setting for two reasons. First, hotline screeners play a narrow and short-lived role in the CPS process: calls last approximately 15 minutes and the screener's sole responsibility is to determine whether a report meets the criteria for investigation. Second, once a report is screened in, it is transferred to a local child welfare office and assigned to a different investigator. Screeners have no role in the subsequent investigation, case management, or service provision. As a result, screener assignment is unlikely to affect child outcomes except through the decision to initiate an investigation.

4.4 Characterizing Treatment

Before turning to the main results, we characterize the nature of treatment in our setting. In particular, we document that CPS investigations generate persistent differences in subsequent system contact and describe the population of compliers for whom our estimates apply.

Persistence of Treatment

Figure 3 plots instrumental variables estimates of the effect of being screened in for investigation on the probability of a CPS investigation in each month following the hotline call. By

[2]Previous studies implementing examiner assignment research designs have made the stronger assumption of pairwise monotonicity, yet recent advances indicate that the pairwise monotonicity assumption is not necessary to estimate LATEs (Frandsen, Lefgren and Leslie, 2019). Unlike average monotonicity, pairwise monotonicity requires that children who are investigated by a particularly lenient screener must also have been investigated by a stricter screener, and vice versa.

construction, the effect is close to one in the initial month, since children whose reports are screened in are investigated during the focal period. The estimate falls below one because some control compliers are screened in shortly after the initial call. One month after the focal call, the estimated difference is 0.88, implying that approximately 12 percent of control compliers are subsequently screened in for investigation within the first month.

A natural concern is that children whose reports are initially screened out may be quickly re-referred and investigated, limiting the scope for persistent differences in exposure to investigation. However, the figure shows that this is not the case. Although the treatment effect declines over time, it remains large and statistically significant for several years. The gap between treatment and control children narrows gradually over time, as expected, but remains substantial: even 72 months after the focal call, being screened in during the focal call increases the probability of subsequent investigation by approximately 35 percentage points.

This pattern indicates that the initial screening decision generates long-lasting differences in the likelihood that a child is investigated by CPS. As a result, our estimates capture the effects of a persistent shift in exposure to investigations rather than a short-lived intervention.

Characteristics of Compliers

Our instrumental variables estimates identify the effects of CPS investigations for compliers: children whose investigation status is influenced by the leniency of the assigned screener. These are children at the margin of investigation—cases for which screeners may disagree about whether to screen in the report.

Table 5 describes the characteristics of compliers. Panel A compares baseline characteristics of the full sample to those of compliers. Column (1) reports average characteristics for the full sample of calls. Following Chyn, Frandsen and Leslie (2025), Column (2) reports (unweighted) average complier characteristics across the treated and untreated states. Overall, compliers are similar to the full sample across most observable dimensions. They are slightly more likely to be low-income, somewhat more likely to be reported by family members, and less likely to be reported by schools or law enforcement. Prior CPS involvement is nearly identical across groups. These patterns suggest that compliers are broadly representative of the population of reported children.

Panel B describes the experiences of treated compliers within the child welfare system. Among compliers who are screened in, approximately 21 percent have substantiated allegations, 12 percent have cases opened, and about 3 percent experience foster care placement. These rates are similar to those in the full sample, though substantiation rates are somewhat

lower among compliers, consistent with these cases being closer to the margin of investigation.

Taken together, these results suggest that our estimates apply to children at the margin of investigation who are broadly similar to the overall population of reported children. This contrasts with Rittenhouse, Simon and Lacey (2025), where compliers are drawn from a higher-risk population near the investigation threshold, and is consistent with the fact that roughly half of reports in our sample are screened in.

5 Results

This section presents our main findings on the effects of CPS investigations. We begin by examining impacts on subsequent child welfare involvement, including later investigations, substantiations, case openings, and removals. For these outcomes, we measure CPS contact beginning 30 days after the initial hotline call and extending through three years. This choice is motivated by the dynamic effects of investigation shown in Figure 4, which plots monthly instrumental variables estimates of the effect of being screened in on subsequent CPS involvement (excluding the outcome of the focal call). The figure shows a sharp decline in CPS contact in the first month following the focal report, with smaller effects in the second month and estimates close to zero thereafter. This pattern is consistent with a mechanical *incapacitation* effect: because investigations typically last approximately 30 days, children who are screened in are actively under investigation during this period, reducing the likelihood that additional reports are screened in. Including this initial month would therefore mechanically bias estimates toward finding reductions in subsequent CPS contact. We therefore exclude the first 30 days from our main outcome measures to focus on longer-run effects. We then study educational outcomes, including chronic absenteeism, standardized test scores, and high school graduation. Finally, we explore heterogeneity in treatment effects by observable characteristics and estimate marginal treatment to characterize the distribution of treatment effects along unobservable determinants of treatment likelihood.

5.1 Child Welfare Outcomes

Panel A of Table 6 reports naïve OLS estimates comparing children whose reports were screened in for investigation to those whose reports were screened out. These estimates suggest that investigated children experience modestly higher rates of future CPS contact. For example, screened-in children are 0.7 percentage points more likely to experience a subsequent investigation within 36 months relative to screened-out children, representing a 1.3 percent increase relative to the control mean of 0.53. Similar positive associations appear for substantiations and case openings. However, these OLS estimates likely reflect

differences in underlying risk rather than the causal effects of investigations. Children whose reports are screened in are systematically higher risk along both observable and unobservable dimensions. As a result, naïve comparisons confound the effects of investigation with the factors that lead cases to be screened in.

Panel B of Table 6 presents instrumental variables estimates that use screener leniency to isolate quasi-random variation in investigation status. In contrast to the OLS estimates, the IV estimates suggest that marginal investigations reduce subsequent CPS involvement, though the estimates are modest in magnitude and not statistically significant. For example, investigations decrease the probability of a subsequent investigation by 2.4 percentage points relative to a control complier mean of 52.6 percent. We observe similarly negative and statistically insignificant point estimates for substantiations, case openings, and removals.

5.2 Educational Outcomes

We next examine whether CPS investigations affect children's educational outcomes. Panel A of Table 7 reports OLS estimates. These naïve comparisons suggest that investigated children experience modestly worse educational outcomes: they are more likely to be chronically absent and have lower standardized test scores. However, as with the CPS outcomes, these associations likely reflect underlying differences in risk rather than the causal effects of investigations.

Panel B of Table 7 presents instrumental variables estimates using screener leniency as an instrument for investigation. In contrast to the OLS estimates, the IV estimates show little evidence that investigations harm educational outcomes. The estimated effects on chronic absenteeism and test scores are close to zero and statistically insignificant. At the same time, we find evidence that investigations may increase the probability of high school graduation. The point estimate implies a 6.2 percentage point increase relative to a control complier mean of 58.7 percent, though it is statistically significant only at the 10 percent level.

Taken together, the results across outcomes suggest modest declines in subsequent CPS contact—consistent with improvements in child safety or reductions in community concerns about the child—and a marginally significant increase in high school graduation. These findings provide little evidence that marginal investigations harm children and instead suggest that marginal investigations may modestly improve child wellbeing.

5.3 Heterogeneity by Child Observable Characteristics

We next examine whether the effects of CPS investigations vary across observable child characteristics. Figure 5 reports instrumental variables estimates separately by gender, age,

race, poverty status, and prior child welfare involvement for both educational and CPS outcomes.

Overall, we find little evidence of meaningful heterogeneity across observable characteristics. For both educational and CPS outcomes, the estimated effects are broadly similar across subgroups and are generally not statistically distinguishable from one another. For example, the estimated effects on subsequent CPS investigations are negative across most subgroups, and the effects on high school graduation are positive across groups, though imprecisely estimated.

There is some suggestive variation in point estimates across subgroups. For instance, the reduction in subsequent CPS involvement and the increase in high school graduation is larger for female relative to male children. However, these differences are not statistically significant and should be interpreted with caution.

Taken together, these results suggest that the impacts of CPS investigations are relatively homogeneous across observable characteristics. In the next section, we explore heterogeneity along unobserved dimensions using a marginal treatment effects framework.

5.4 Marginal Treatment Effects

Our instrumental variables estimates identify the impact of CPS investigations for children at the margin of investigation—those whose investigation status is influenced by the leniency of their assigned screener. This LATE parameter is policy relevant in the sense that it captures the effects of marginal changes in screening behavior. However, while our estimates of LATE are informative about the consequences of marginal screening decisions, they do not by themselves characterize how outcomes would change under alternative screening policies.

To examine heterogeneity beyond the LATE, we estimate MTEs, which characterize how the impact of investigation varies across children with different levels of unobserved resistance to investigation. MTEs allow us to trace out treatment effects along the full distribution of children and provide insight into how the consequences of investigation vary as the screening margin shifts.

Following the framework of Heckman and Vytlacil (1999, 2005) and recent applications such as Bhuller et al. (2020), we model observed outcomes as

$$Y = I \cdot Y(1) + (1 - I) \cdot Y(0),$$

where I indicates whether a child's report is screened in for investigation, and $Y(1)$ and $Y(0)$ denote potential outcomes with and without investigation. Screening decisions follow

a latent index model:

$$I = \mathbb{1}[v(X, Z) - U > 0],$$

where v is an unknown function, X represents observed characteristics, Z is screener leniency, and U is an unobserved scalar capturing a child's resistance to investigation. The unobserved component U summarizes factors known to screeners but not observed in our data that influence screening decisions (e.g., tone of the caller or nuanced information about the allegation).

After normalizing $U \mid X = x$ to be uniformly distributed over $[0, 1]$, the index $v(X, Z)$ corresponds to the propensity score, defined as $p(X, Z) \equiv P(I = 1 \mid X = x, Z = z)$. The marginal treatment effect is defined as

$$\text{MTE}(x, u) = E[Y(1) - Y(0) \mid X = x, U = u],$$

which captures the causal effect of investigation for children at different points in the distribution of unobserved resistance to treatment.

Intuitively, MTEs trace out how treatment effects vary across children who are induced into investigation as screener leniency changes. At low values of U, the MTE reflects the effect of investigation for children who are likely to be investigated regardless of screener assignment, including when assigned to relatively lenient screeners; these children can be interpreted as higher-risk cases. At higher values of U, the MTE reflects the effect for children who are only investigated when assigned to relatively strict screeners and can be interpreted as lower-risk children at the margin of investigation.

We estimate MTEs using a local instrumental variables (local IV) approach, which identifies the MTE as the derivative of the conditional expectation of outcomes with respect to the propensity score (Heckman and Vytlacil, 1999, 2000). The local IV framework relies on a latent index model of treatment choice, which is equivalent to the strict monotonicity condition underlying the LATE framework: changes in the instrument shift individuals' treatment decisions in a common direction through a threshold-crossing rule (Vytlacil, 2002). This assumption is stronger than the average monotonicity condition sufficient for identifying LATEs in our 2SLS framework. Moreover, point identification of the MTE without further restrictions generally requires broad support of the propensity score conditional on observables, which is often difficult to achieve in practice. Following the applied literature (e.g., Bhuller et al., 2020), we therefore assume separability between observed and unobserved heterogeneity in treatment effects. This assumption implies that the shape of the MTE does not vary with X, while the level of treatment effects may shift with observables. Under these

assumptions, the MTE is identified over the region of (unconditional) common support of the propensity score.

We focus on two outcomes: chronic absenteeism and subsequent case opening. Chronic absenteeism is observed for the full sample, while test scores and high school graduation are only available for selected subsamples. Focusing on absenteeism avoids conflating heterogeneity in treatment effects with variation in outcome availability. We also examine case opening as a relatively stringent CPS outcome that is more informative than subsequent investigations but more common than removal.

Panel (a) of Figure 6 shows the distribution of estimated propensity scores for treated and untreated children. The region of common support is indicated by the vertical dashed lines, after trimming the top and bottom 1 percent of the overlapping distribution. The MTE is identified only over this region of common support. Within this region, we estimate MTEs using a local instrumental variables approach and a global quadratic polynomial specification. Panels (b) and (c) of Figure 6 plot the estimated MTEs for chronic absenteeism and case opening, respectively. The horizontal axis represents unobserved resistance to investigation, with lower values corresponding to children who are more likely to be investigated.

For both outcomes, the estimated MTEs are upward sloping. For children with low unobserved resistance to investigation (i.e., higher-risk children), the effects of investigation are negative, indicating reductions in chronic absenteeism and subsequent case opening. In contrast, for children with high resistance to investigation (i.e., lower-risk children), the estimated effects are positive. In terms of magnitude, for the highest-risk children, chronic absenteeism decreases by approximately 18 percentage points and subsequent case openings decrease by about 13 percentage points. For the lowest-risk children, chronic absenteeism increases by roughly 8 percentage points and subsequent case openings increase by about 14 percentage points.

These heterogeneous effects are not necessarily obvious ex ante. While higher-risk children may benefit from increased monitoring or service provision during an investigation, it is less clear why investigations would improve outcomes such as school attendance or reduce subsequent CPS contact. Indeed, many recent policy reforms have been motivated by the view that investigations are intrusive and yield limited benefits for families. Conversely, for lower-risk children, it is not immediately apparent why an investigation would worsen outcomes, rather than simply impose short-lived costs. Our findings therefore provide new empirical evidence on how the consequences of investigations vary across the risk distribution.

Finally, as shown by Heckman and Vytlacil (1999, 2005, 2007), conventional treatment effect

parameters such as the average treatment effect (ATE) can be expressed as weighted averages of the MTE. Recovering these parameters for the full population requires support of the propensity score over the unit interval. Because our propensity score has limited support, we follow Bhuller et al. (2020) and construct rescaled versions of these parameters by normalizing the relevant weights to integrate to one over the region of common support. Within this region, the ATEs for both outcomes are relatively small and close to zero.

6 Conclusion and Next Steps

CPS investigations are among the most common forms of state intervention in family life, yet there is little causal evidence on their consequences. This paper studies the effects of CPS investigations using quasi-random assignment of hotline screeners who differ in their propensity to screen in reports for investigation. Leveraging linked administrative data that include both screened-in and screened-out reports, we estimate the causal impact of investigations on subsequent child welfare involvement and educational outcomes.

Our findings suggest that marginal investigations have small and generally positive effects. We find no evidence that marginal investigations increase subsequent CPS contact, and point estimates indicate modest reductions in later investigations, substantiations, and case openings, though these effects are not statistically significant. For educational outcomes, we find precise null effects on chronic absenteeism and test scores and a marginally significant increase in high school graduation. At the same time, we document substantial heterogeneity in treatment effects. Marginal treatment effect estimates indicate that investigations are beneficial for higher-risk children—those most likely to be investigated regardless of screener assignment—but may be harmful for lower-risk children.

Taken together, our findings provide new evidence on the effects of CPS investigations at the margin of screening decisions. Our baseline estimates are local to children whose investigation status is influenced by screener assignment, and therefore reflect the consequences of marginal changes in screening behavior. For the next version of this draft, we plan to pursue three approaches to move beyond this local average treatment effect and better inform policy-relevant counterfactuals. First, we plan to use our MTE estimates to recover PRTEs that map directly to proposed policy changes, such as reductions in screening rates through diversion programs or "warmline" alternatives. This approach allows us to evaluate how outcomes would change under counterfactual policies that shift the screening margin.

Second, we exploit a policy-driven change in screening behavior that occurred in Michigan beginning in late 2019, which led to a sharp and sustained decline in screen-in rates (Figure 7). This policy change allows us to study the effects of a large, system-wide shift in screening in-

tensity and to estimate treatment effects at a different margin than our examiner-assignment design. We plan to combine variation in screener tendencies with the timing of the policy change, and assess robustness to potential confounding from the COVID-19 pandemic by excluding the most affected periods.

Third, we plan to explore the mechanisms underlying our results by leveraging detailed information on investigation processes and outcomes. In particular, we examine whether the benefits of investigation for higher-risk children operate through increased monitoring and deterrence, service provision, or more intensive interventions such as case openings or foster care placement. To do so, we use detailed data from investigations, including case dispositions, safety assessments, and risk assessments. We also explore why investigations may adversely affect lower-risk children. These analyses will help shed light on how and for whom CPS investigations are most effective.

References

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics*, 113(2): 231–263.

Agan, Amanda, Jennifer L Doleac, and Anna Harvey. 2023. "Misdemeanor prosecution." *The Quarterly Journal of Economics*, 138(3): 1453–1505.

Aizer, Anna, and Joseph J. Doyle, Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly-Assigned Judges." *Quarterly Journal of Economics*, 130(2): 759–803.

Bald, Anthony, Eric Chyn, Justine S. Hastings, and Margarita Machelett. 2022. "The causal impact of removing children from abusive and neglectful homes." *Journal of Political Economy*.

Baron, E Jason, and Max Gross. 2024. "Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators." *Review of Economics and Statistics*.

Baron, E. Jason, Ezra G. Goldstein, Christopher Mills, and Emily Putnam-Hornstein. 2026. "The Intergenerational Persistence of Child Maltreatment." Working Paper.

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad. 2020. "Incarceration, recidivism, and employment." *Journal of political economy*, 128(4): 1269–1324.

Chyn, Eric, Brigham Frandsen, and Emily Leslie. 2025. "Examiner and judge designs in economics: A practitioner's guide." *Journal of Economic Literature*, 63(2): 401–439.

Collinson, Robert, Deniz Dutz, John Eric Humphries, Nicholas S Mader, Daniel Tannenbaum, and Winnie van Dijk. 2025. "The effects of eviction on children." National Bureau of Economic Research.

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie Van Dijk. 2024. "Eviction and poverty in American cities." *The Quarterly Journal of Economics*, 139(1): 57–120.

Dobbie, Will, Jacob Goldin, and Crystal S Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108(2): 201–40.

Doyle, Joseph J., Natalia Emanuel, and Raimundo Eyzaguirre. 2025. "Cumulative Risks of Foster Care Placement for U.S. Children: A Research Note Comparing Birth Cohort and Synthetic Cohort Approaches." *Working paper*.

Doyle, Jr., Joseph J. 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review*, 97(5): 1583–1610.

Doyle, Jr., Joseph J. 2008. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." *Journal of Political Economy*, 116(4): 746–770.

Fong, Kelley. 2023. *Investigating Families: Motherhood in the Shadow of Child Protective Services*. Princeton, NJ:Princeton University Press.

Frandsen, Brigham R, Lars J Lefgren, and Emily C Leslie. 2019. "Judging Judge Fixed Effects." National Bureau of Economic Research Working Paper 25528.

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár. 2025. "Contamination Bias in Linear Regressions." *American Economic Review*. Forthcoming.

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): 170–199.

Heckman, James J, and Edward J Vytlacil. 1999. "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the national Academy of Sciences*, 96(8): 4730–4734.

Heckman, James J, and Edward J Vytlacil. 2000. "Local Instrumental Variables." National Bureau of Economic Research Working Paper 252.

Heckman, James J., and Edward J. Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica*, 73(3): 669–738.

Heckman, James J., and Edward J. Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*. Vol. 6, , ed. James J. Heckman and Edward Leamer, 4779–4874. Elsevier.

Helénsdotter, Ronja. 2025. "Surviving childhood: Effects of removing a child from home." *Review of Economic Studies*.

Humlum, Anders, Jakob Munch, and Mette Rasmussen. 2025. "What works for the unemployed? Evidence from quasi-random caseworker assignments." National Bureau of Economic Research.

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.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review*, 96(3): 863–876.

Mogstad, Magne, and Alexander Torgovitsky. 2018. "Identification and Extrapolation of Causal Effects with Instrumental Variables." *Annual Review of Economics*, 10: 577–604.

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.

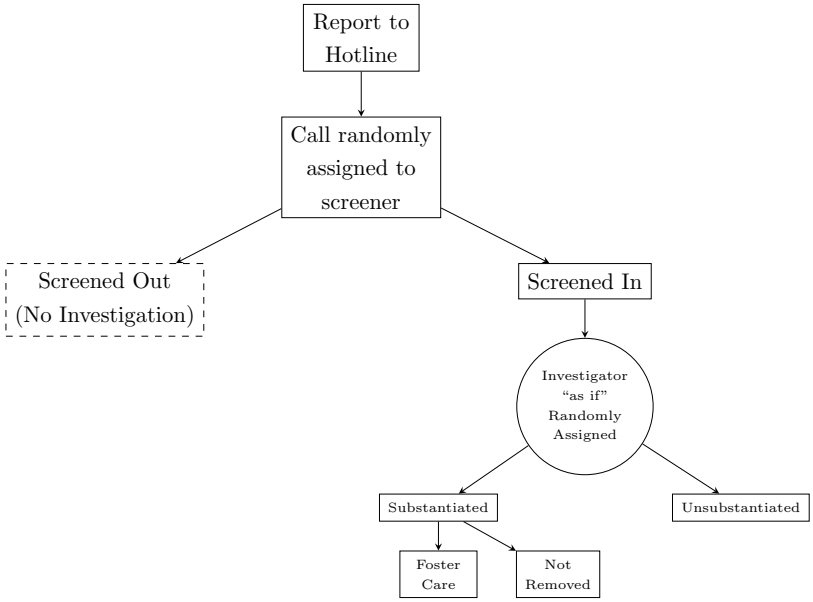
Rittenhouse, Katherine, David Simon, and Lindsey Lacey. 2025. "Child Maltreatment Investigations and Family Well-being." National Bureau of Economic Research Working Paper 34334.

Rose, Evan K, and Yotam Shem-Tov. 2021. "How does incarceration affect reoffending? Estimating the dose-response function." *Journal of Political Economy*, 129(12): 3302–3356.

Ryan, Joseph P, Brian A Jacob, Max Gross, Brian E Perron, Andrew Moore, and Sharlyn Ferguson. 2018. "Early exposure to child maltreatment and academic outcomes." *Child maltreatment*, 23(4): 365–375.

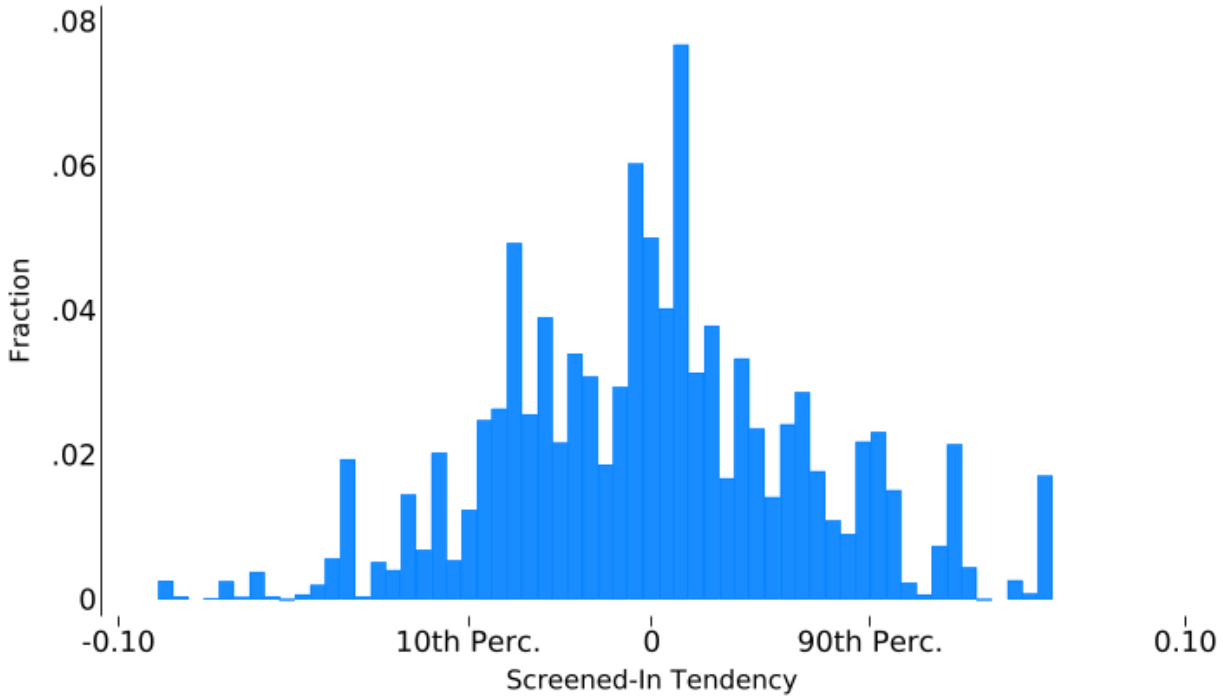
Vytlacil, Edward. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica*, 70(1): 331–341.

Figure 1: Child Welfare Investigation Process in Michigan



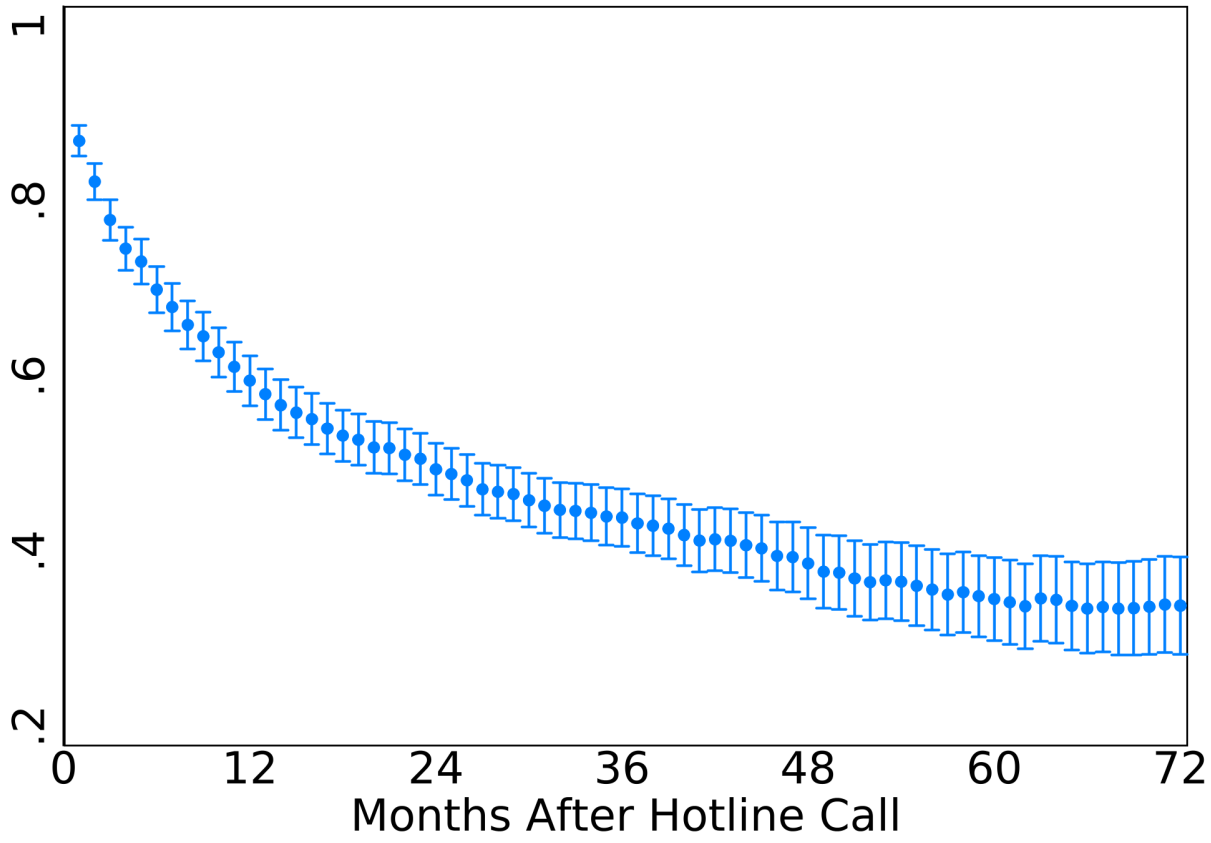
Notes. This figure illustrates the sequence of decisions in Michigan's child welfare system following a report of suspected maltreatment. Reports are first received by the centralized CPS hotline, where a screener decides whether to screen the report in for investigation or screen it out with no further action. Screened-in reports are referred to a local child welfare office, where an investigator conducts an investigation and determines whether the allegation is substantiated. Substantiated cases may result in case closure, referral to services, ongoing monitoring through an open case, or in rare circumstances, foster care placement. The figure is a simplified representation of the institutional process described in Section 2.

Figure 2: Distribution of Screener Leniency Instrument



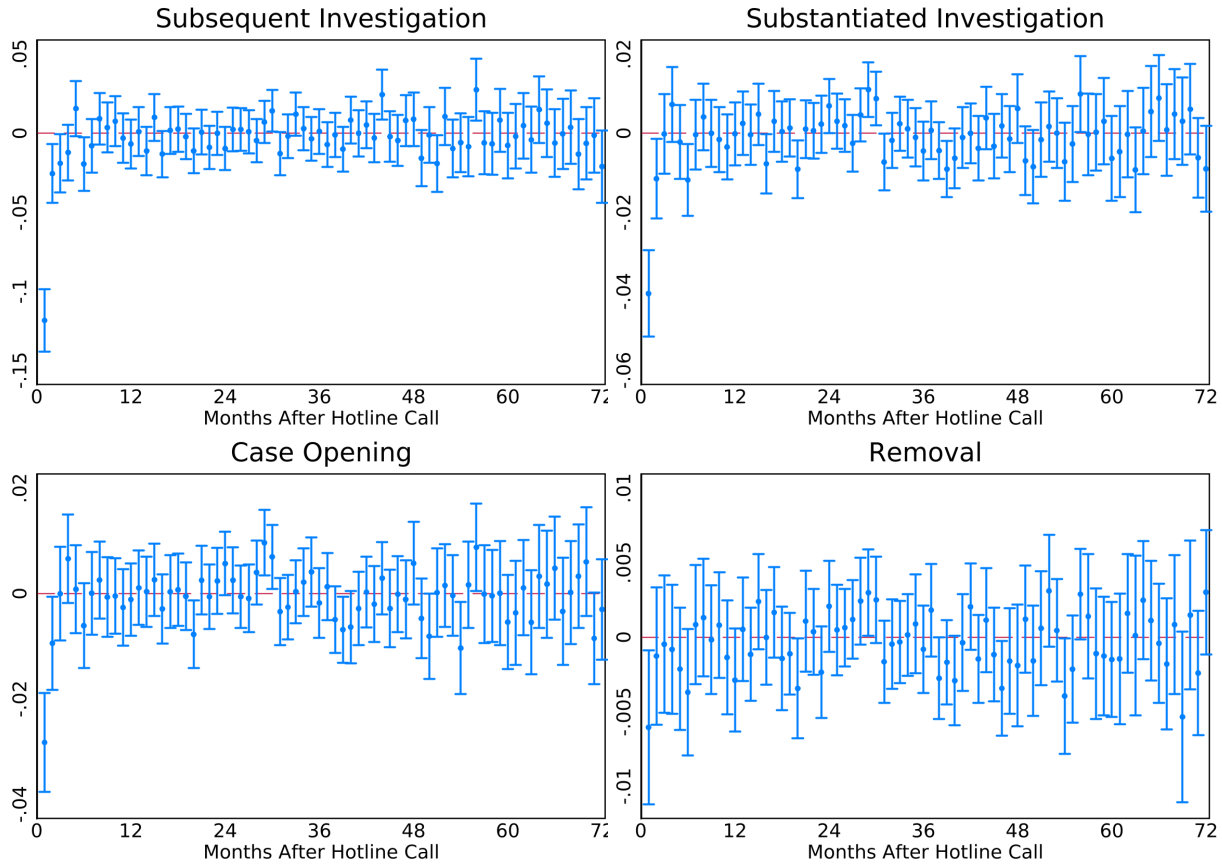
Notes. This figure displays the distribution of the screener leniency instrument. The instrument is constructed as the leave-out mean screening rate of the assigned hotline worker, calculated using all other reports handled by the same screener after residualizing the screening decision with respect to complaint day-by-shift fixed effects. Each observation represents a screener, and the distribution reflects variation in screening tendencies across workers within rotation groups. The horizontal axis measures screener leniency relative to the average screener working in the same shift-by-day rotation group.

Figure 3: Effect of Initial Screen-In Decision on Subsequent Investigations



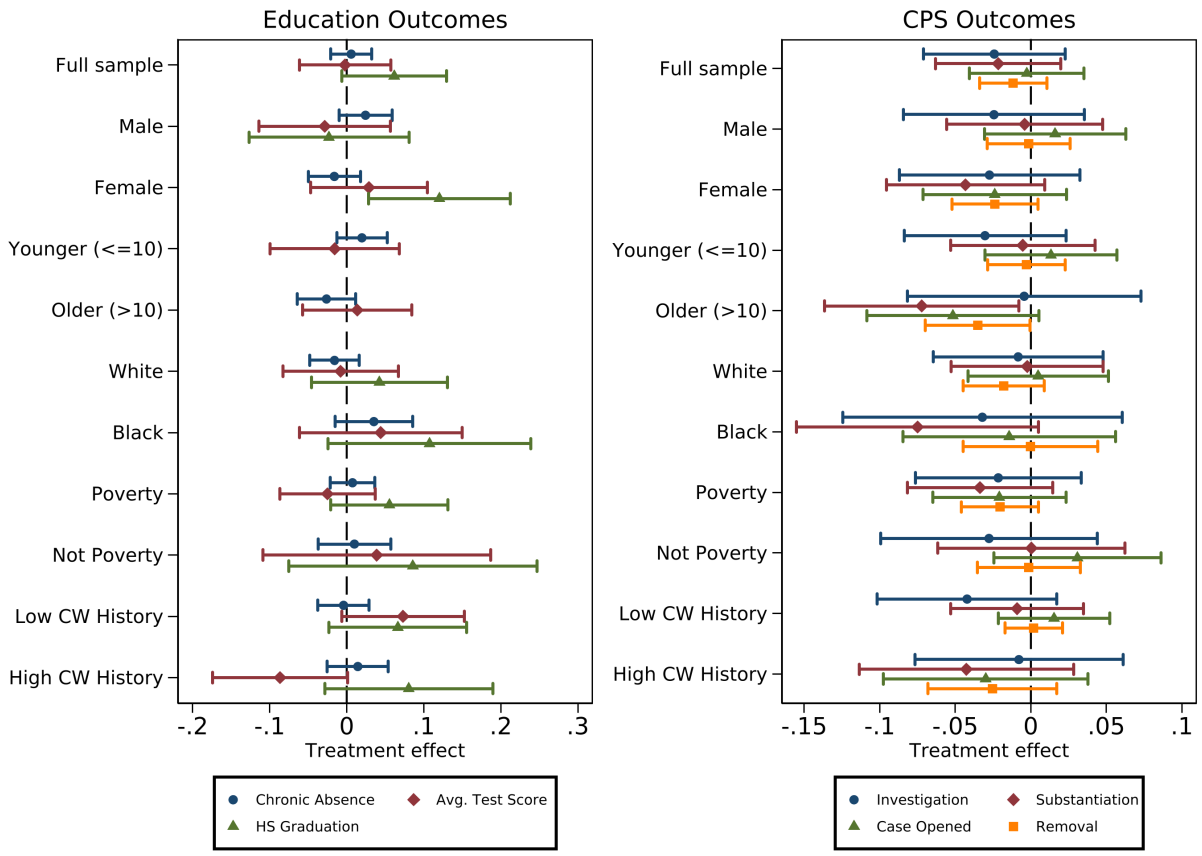
Notes. This figure plots instrumental variables estimates of the effect of being screened in for investigation on the probability of a subsequent CPS investigation at each month t following the focal hotline call. Each point corresponds to a separate 2SLS regression in which the outcome is an indicator for whether the child is screened in for any investigation in month t, and the treatment is an indicator for being screened in at the initial report. Screened-in status is instrumented using screener leniency. The figure reports point estimates and 95 percent confidence intervals. All specifications include baseline covariates and day-by-shift fixed effects. Standard errors are clustered at the call level.

Figure 4: Effect of Initial Screen-In on Subsequent CPS Outcomes by Month



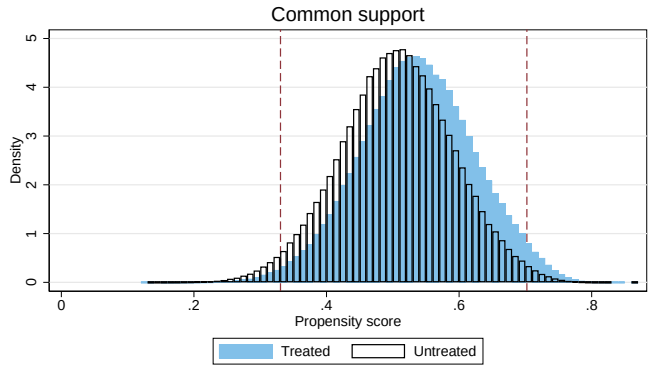
Notes. This figure plots instrumental variables estimates of the effect of being screened in for investigation at the focal hotline call on subsequent CPS outcomes in each month following the call. Outcomes include any subsequent investigation, substantiated investigation, case opening, and foster care placement, and exclude outcomes associated with the focal report. Each point corresponds to a separate 2SLS regression in which the outcome is an indicator for whether the child experiences the given CPS event in month t. Screened-in status is instrumented using screener leniency. Vertical bars represent 95 percent confidence intervals, and standard errors are clustered at the call level. The figure shows a sharp decline in CPS contact in the first month following the focal report, with smaller effects thereafter, consistent with a temporary mechanical effect during the investigation period.

Figure 5: Treatment Effect Heterogeneity by Observable Features

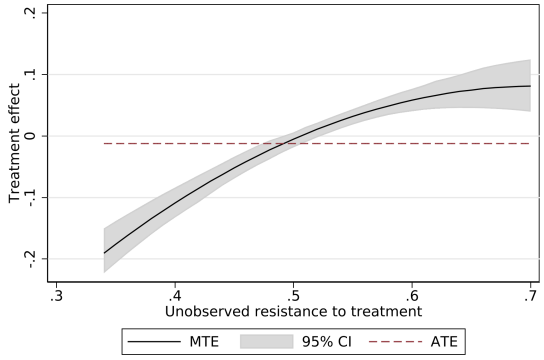


Notes. This figure reports subgroup-specific instrumental variables estimates of the effect of being screened in for investigation on CPS and educational outcomes. Each marker corresponds to a separate 2SLS estimate within the indicated subgroup, instrumenting screened-in status with screener leniency. Horizontal bars represent 95 percent confidence intervals. The left panel reports effects on educational outcomes (chronic absenteeism, average test scores, and high school graduation), and the right panel reports effects on CPS outcomes (subsequent investigation, substantiation, case opening, and removal). Subgroups are defined by gender, age, race, poverty status, and prior child welfare involvement. "Low" and "high" prior CPS involvement are defined using the age-specific median number of prior investigations: low involvement is at or below the median for children of the same age, and high involvement is above that median. High school graduation and CPS outcomes are estimated on the time-invariant call-child sample; the other educational outcomes are estimated using the call-child-year sample. Age-specific estimates are not shown for high school graduation because graduation is only observed for the older children sample. All specifications include baseline covariates and day-by-shift fixed effects. When a subgroup-defining covariate is constant within the estimation sample, it is dropped from the regression. Standard errors are clustered at the call level. Estimates are obtained by splitting the sample by subgroup and re-estimating the same specification within each subgroup. CPS outcomes are measured between 30 days to 36 months after the initial hotline call.

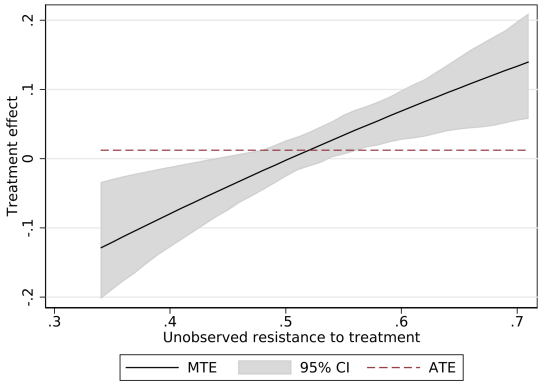
Figure 6: Marginal Treatment Effects



(a) Common Support



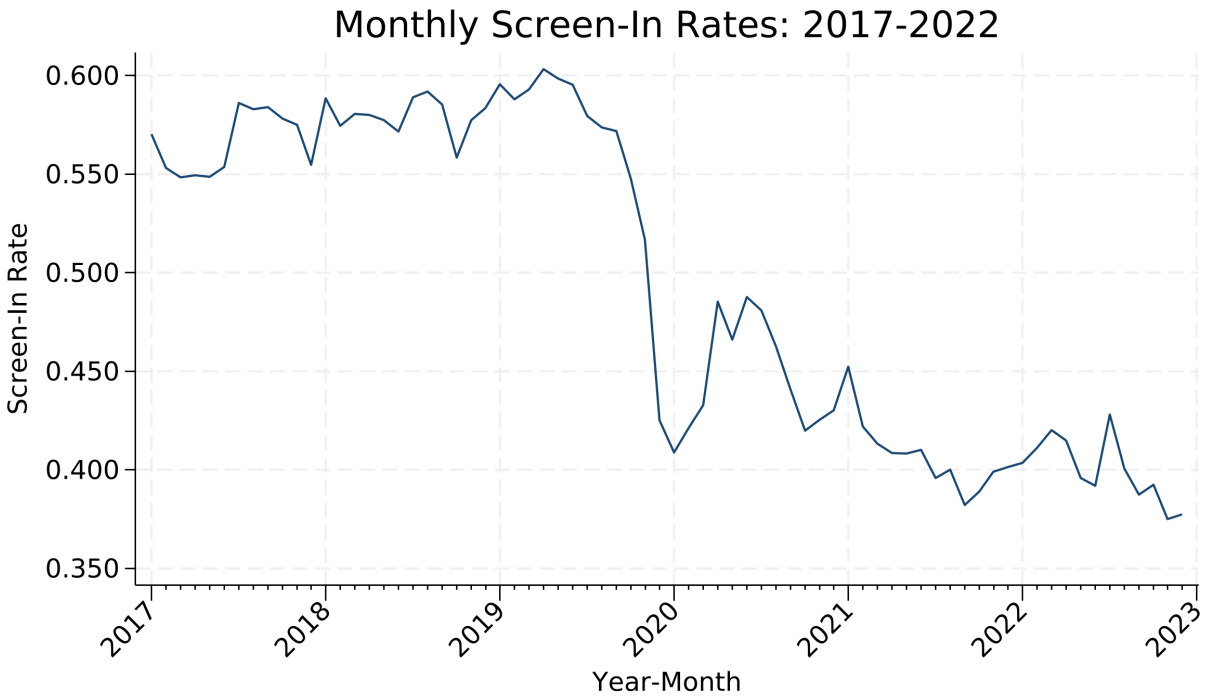
(b) Education: Chronic Absenteeism



(c) CPS: Case Opened

Notes. Panel (a) plots the distribution of estimated propensity scores for being screened in for investigation separately for treated and untreated children. The dashed vertical lines indicate the region of common support after trimming the top and bottom 1 percent of the overlapping distribution. Panels (b) and (c) report marginal treatment effect (MTE) estimates for chronic absenteeism and subsequent case opening, respectively. The horizontal axis represents unobserved resistance to investigation, where lower values correspond to children more likely to be investigated. The solid line plots the estimated MTE, the shaded region shows 95 percent confidence intervals, and the dashed horizontal line indicates the average treatment effect (ATE). MTE estimates are obtained using a local instrumental variables approach with a global quadratic polynomial specification, estimated over the trimmed region of common support. Standard errors are constructed using 100 bootstrap replications and are clustered at the call level. CPS outcomes exclude events occurring within the first 30 days after the hotline call.

Figure 7: Monthly Screen-In Rates, 2017–2023



Notes. This figure plots the monthly share of reports that are screened in for investigation in Michigan between January 2017 and December 2023. Each point represents the fraction of all reports in a given month that are screened in by the centralized CPS intake system. Screen-in rates are relatively stable at approximately 55–60 percent prior to late 2019, followed by a sharp and sustained decline to roughly 40 percent. This change corresponds to a policy shift that reduced screening intensity.

Table 1: Summary Statistics

	(1) MI Student	(2) All Sample	(3) Screened In	(4) Screened Out
Panel A: Child Characteristics				
Female	0.486	0.513	0.506	0.520
White	0.730	0.715	0.702	0.727
Black	0.214	0.293	0.307	0.280
Hispanic	0.081	0.075	0.074	0.075
Asian/American Indian	0.044	0.011	0.011	0.011
Age at Report	11.3	10.8	10.7	10.9
Economically Disadvantaged	0.528	0.870	0.877	0.864
Special Education	0.142	0.276	0.270	0.282
Attendance Rate	0.927	0.882	0.883	0.881
Standardized Math Score	-0.001	-0.585	-0.592	-0.578
Panel B: Reporter Category				
Education		0.201	0.208	0.195
Medical		0.139	0.124	0.154
Family		0.227	0.211	0.242
Law Enforcement		0.211	0.240	0.183
Other		0.222	0.217	0.226
Panel C: Prior CW Involvement				
Any prior investigation		0.782	0.768	0.795
Any prior substantiation		0.465	0.446	0.482
Any prior removal		0.130	0.118	0.141
Screen-in rate		0.488	1.000	0.000
Sample Counts				
Calls		652,237		
Children		382,827		
Child-by-call pairs		997,146		
Child-by-call pairs with data at $t = -1$		686,075		

Notes. This table reports means for characteristics measured in the school year before the hotline call (t = -1). Column (1) reports Michigan student benchmarks (2018–2019) where available. Columns (2)–(4) report means for the full analysis sample, screened-in calls, and screened-out calls. The Sample Counts panel reports the number of unique calls, unique children, child-by-call observations in the analysis panel, and child-by-call observations with t = -1 data.

Table 2: First-Stage Estimates: Instrument Relevance

	(1)	(2)	(3)	(4)	(5)
Screener Tendencies	1.199***	0.992***	0.991***	0.966***	0.963***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Complaint Day x Shift FE	No	Yes	Yes	Yes	Yes
Demog Controls	No	No	Yes	Yes	Yes
Report Controls	No	No	No	Yes	Yes
Prior CW Controls	No	No	No	No	Yes
F-Stat	2897	2004	2003	1917	1913
N	982,534	982,534	982,534	982,534	982,534

Notes. The outcome is an indicator for whether a hotline call is screened in. Entries report first-stage coefficients on screener leniency. Column (1) is bivariate; Column (2) adds complaint-day-by-shift fixed effects; Columns (3)–(5) sequentially add child demographics, report characteristics, and prior child-welfare controls. Standard errors clustered at the call level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Balance Test: Instrument Exogeneity

	(1) Full Sample	(2) Compliers
Child Characteristics		
Female	-0.012***	0.022
	(0.001)	(0.016)
White	-0.028***	-0.059***
	(0.001)	(0.022)
Black	0.032***	0.027
	(0.001)	(0.021)
Age at time of complaint	-0.467***	-0.137
	(0.011)	(0.167)
Prior Child Welfare Risk		
Any prior investigation	-0.049***	0.013
	(0.001)	(0.017)
Any prior substantiation	-0.047***	-0.005
	(0.001)	(0.021)
Prior Educational Outcomes		
Attendance rate (prior year)	-0.002***	-0.003
	(0.000)	(0.007)
Chronic absenteeism (prior year)	0.011***	0.011
	(0.001)	(0.022)
Special education (prior year)	-0.012***	-0.005
	(0.001)	(0.018)
Homeless (prior year)	-0.004***	-0.007
	(0.001)	(0.013)
Std. math score (prior year)	-0.016***	-0.043
	(0.004)	(0.076)
Std. ELA score (prior year)	-0.017***	-0.065
	(0.004)	(0.074)
Predicted HS Grad	0.013***	0.001
	(0.000)	(0.007)
School Context (Prior Year)		
School in city	0.019***	0.014
	(0.001)	(0.021)
Charter school	0.006***	-0.001
	(0.001)	(0.014)
Share White students	-4.090***	-2.708*
	(0.103)	(1.574)
Share FRL students	-0.012***	-0.015
	(0.001)	(0.012)

Notes. Each row reports the estimated difference in the listed baseline characteristic between screened-in and screened-out children. Column (1) reports the difference between screened-in and screened-out children from OLS, controlling for shift-by-day fixed effects. Column (2) reports the analogous local average treatment effect for compliers, estimated by 2SLS instrumenting screened-in status with screener leniency and including the same fixed effects. Predicted HS graduation is the estimated probability of high school graduation from a logit model estimated using the full set of predetermined baseline controls. Regressions are estimated at the call-by-child level. Standard errors are clustered at the call level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Test of Average Monotonicity

	(1) Female	(2) Male	(3) White	(4) Black	(5) Hispanic	(6) 0–3	(7) 4–6	(8) 7–11	(9) 12+
Screener Tendencies	0.894***	0.907***	0.925***	0.860***	0.771***	0.814***	0.986***	0.677***	1.271***
	(0.043)	(0.043)	(0.039)	(0.055)	(0.130)	(0.044)	(0.044)	(0.205)	(0.273)
F-statistic	429	440	565	242	35	344	494	11	22
Observations	140,379	137,575	188,290	100,054	20,561	142,820	120,027	7,852	4,553

Notes. This table assesses average monotonicity across child subgroups. Each column reports the first-stage coefficient on screener leniency from a subgroup-specific regression of screened-in status on the instrument, with shift-by-day fixed effects and baseline controls. F-statistics are for the excluded instrument. Standard errors clustered at the call level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Complier Characteristics and Experiences

	(1) Full Sample	(2) Compliers
Panel A: Complier Characteristics		
Socio-Demographics		
Female	0.510	0.513
White	0.590	0.599
Black	0.302	0.297
Older (above median age)	0.500	0.507
Low Income (prior year)	0.869	0.885
Reporter & Risk Characteristics		
Reported by School	0.173	0.158
Reported by Law Enforcement	0.220	0.156
Reported by Family Member	0.236	0.332
Prior CPS Investigation	0.742	0.736
Composite Score		
Predicted HS Grad	0.722	0.719
Panel B: Complier Experiences in the System		
Investigation Outcomes		
Foster Care Placement	0.027	0.029
Case Opened	0.129	0.121
Substantiated	0.249	0.209

Notes. Panel A reports baseline characteristics. Column (1) shows full-sample means. Column (2) shows complier means identified using variation in screener leniency with shift-by-day fixed effects. Panel B reports investigation outcomes: Column (1) is the mean among screened-in calls, and Column (2) is the corresponding treated-complier mean from the same instrumental-variables framework.

Table 6: Treatment Effects on CPS Outcomes

	(1) Investigation	(2) Substantiation	(3) Case Opened	(4) Removal
Panel A: OLS Estimates				
Screened In	0.007***	0.010***	0.009***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
	{0.530}	{0.217}	{0.166}	{0.051}
Panel B: IV Estimates				
Screened In	-0.024	-0.022	-0.003	-0.012
	(0.024)	(0.021)	(0.019)	(0.011)
	{0.526}	{0.211}	{0.151}	{0.046}
Shift-by-day FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	763,730	763,345	763,345	762,984

Notes. Panel A reports OLS estimates and Panel B reports IV estimates, where screened-in status is instrumented by screener leniency. Outcomes are 36-month CPS measures that exclude events in the first 30 days after the hotline call. Columns (1)–(4) report the component outcomes shown in the column headings. Standard errors clustered at the call level are in parentheses. Control means (Panel A) and control complier means (Panel B) are reported in braces. All specifications include shift-by-day fixed effects and baseline controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Treatment Effects on Education Outcomes

	(1) Chronic Absence	(2) Avg. Test Score	(3) HS Graduation
Panel A: OLS Estimates			
Screened In	0.013***	-0.019***	-0.001
	(0.001)	(0.002)	(0.002)
	{0.453}	{-0.534}	{0.598}
Panel B: IV Estimates			
Screened In	0.006	-0.002	0.062*
	(0.013)	(0.030)	(0.035)
	{0.473}	{-0.530}	{0.587}
Shift-by-day FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N	4,588,528	1,824,575	264,325

Notes. This table reports treatment effects on education outcomes. Panel A reports OLS estimates and Panel B reports IV estimates, where screened-in status is instrumented by screener leniency. Time-varying outcomes (Columns 1 and 2) are estimated using student-year observations from the report year onward; high school graduation (Column 3) is a time-invariant call-child outcome estimated on post-report observations. Avg. test score is the row mean of standardized math and ELA scores. Standard errors clustered at the call level are in parentheses. Control means (Panel A) and control complier means (Panel B) are reported in braces. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.