The Causal Impact of Removing Children from Abusive and Neglectful Homes

Anthony Bald, Eric Chyn,
Justine Hastings, and Margarita Machelett*

November 16, 2021

Abstract

This paper measures impacts of removing children from families investigated for abuse or neglect. We use removal tendencies of child protection investigators as an instrument. We focus on young children investigated before age six and find that removal significantly increases test scores and reduces grade repetition for girls. There are no detectable impacts for boys. This pattern of results does not appear to be driven by heterogeneity in pre-removal characteristics, foster placements, or the type of schools attended after removal. The results are consistent with the hypothesis that development of abused and neglected girls is more responsive to home removal.

JEL Classification: H75, I21, I24, I28, I38, J12, J13, J24
Keywords: foster care, educational attainment, early intervention

*Bald: Harvard University and RIPL, Email: abald@g.harvard.edu; Chyn: Dartmouth College, RIPL, and NBER, Email: eric.t.chyn@dartmouth.edu; Hastings: Amazon, RIPL, and NBER, Email: jhastings@ripl.org; Machelett: Bank of Spain and RIPL, Email: margarita.machelett@bde.es. The opinions and conclusions expressed are solely those of the authors. We thank Mark Howison, the team at Research Improving People's Lives (RIPL), the state of Rhode Island, and the Rhode Island Department of Children, Youth, and Families for making this project possible. We are also thankful for helpful comments and feedback from the editor, three anonymous referees, Martin Andresen, Isaiah Andrews, Dmitry Arkhangelsky, Joe Doyle, Sebastian Fanelli, Florian Gunsilius, Sara Heller, Brian Jacob, Erzo Luttmer, John Pepper, Jonah Rockoff, Bruce Sacerdote, Emilia Simeonova, Maya Rossin-Slater, Doug Staiger, Sarah Turner, Ernesto Villanueva, Christina Weiland, and seminar participants at the University of Michigan, the 2018 Annual Meeting of the Southern Economic Association, and the NBER Children's Program Meeting. Finally, we are grateful for financial support from the Smith Richardson Foundation and the Laura and John Arnold Foundation.

1 Introduction

Each year, child protective service agencies in the U.S. investigate more than four million allegations of abuse or neglect (U.S. Department of Health and Human Services, 2016). As a result of these investigations, authorities annually remove nearly 200,000 children from their homes and place them into foster care (U.S. Department of Health and Human Services, 2016). The goal of removal is to protect children by reducing exposure to abuse and neglect.

There is relatively little evidence on the causal impact of child protective service removal on children. Abused children have lower academic performance and are more likely to have social or emotional conditions such as aggressive behavior or depression (Fantuzzo and Mohr, 1999; Wolfe et al., 2003; Holt et al., 2008; Doyle and Aizer, 2018).[1] Because removal is more likely in more severe cases, the relationship between removal and outcomes may not be causal. Doyle (2007; 2008) addressed the endogeneity of removal from home by using the removal tendencies of quasi-randomly assigned child protective service investigators as an instrument for removal. He studied later-life outcomes of children who were subject to an investigation between the ages of five and fifteen using data from Illinois and found that removal increased delinquency and arrests while decreasing labor market activity.

This paper focuses on young children and provides new evidence on the impact of removal based on comprehensive administrative data from Rhode Island. The data contain approximately two decades of child protective services case records joined to administrative records on academic outcomes in public schools. We study the impacts of removal in early childhood (before age six) for two reasons. First, nearly half of removed children are under the age of six (U.S. Department of Health and Human Services, 2016). Second, the literature on child development suggests that early life events and interventions can have particularly strong influences on outcomes (Cunha et al., 2006; Heckman, 2006; Cunha and Heckman, 2007; Almond and Currie, 2011; Heckman et al., 2013; Heckman and Mosso, 2014; Elango

[1] Currie and Tekin (2012) study long-term outcomes of children, finding that maltreatment is associated with increases in the likelihood of committing crime.

et al., 2015; Almond et al., 2018). Our analysis is the first to estimate causal impacts of home removal for this important group of children.[2]

We use the removal tendency of child protective service investigators as an instrument for removal.[3] This approach provides estimates of causal impacts of removal for the marginally removed child, that is, those whose removal outcomes would have been different had they being assigned to investigators with different removal propensities. Our main specification uses a standard leave-out mean removal rate as the measure of the tendency for each investigator. Prior literature has used this type of measure for judges and other authorities (Kling, 2006; Doyle, 2007, 2008; Aizer and Doyle, 2015; Eren and Mocan, 2017; Sampat and Williams, 2019; Dobbie, Goldin and Yang, 2018; Dobbie, Grönqvist, Niknami, Palme and Priks, 2018; Bhuller et al., 2018, 2020). In our sample, the leave-out removal rate is a statistically significant, positive predictor of removal and is uncorrelated with child and case characteristics. We present all regression results separately for girls and boys. Our analysis of effects by gender is motivated by prior research, which shows that girls and boys may respond differently to social programs and family conditions (Heckman et al., 2010; Bertrand and Pan, 2013; Heckman et al., 2013; Elango et al., 2015; Conti et al., 2016; Heckman et al., 2017; Garcia et al., 2018; Autor et al., 2019).

Our main finding is that there are significant and positive effects of removal on achievement outcomes for young girls and no corresponding significant effects for young boys. For young girls, the point estimate for the impact of removal indicates a 1.367 student-level standard deviation increase in average standardized test scores (math and reading) in the years after removal. These large effects are similar to findings from the Perry Preschool program, where girls randomly assigned to receive high-quality early education had 0.806 higher standardized

[2]Note that age six is the compulsory school starting age in Rhode Island during our sample period (Rhode Island, 2016). Fitzpatrick et al. (2020) provide evidence that reports of child maltreatment increase when children enroll in school. Their findings suggest that the composition of investigated children may change at age six because educators may be an important source of information for instances of neglect and abuse.

[3]Our research design relies on the fact that cases are assigned using a rotation list, which effectively randomizes investigators to cases. Note that we exclude sex abuse reports from all analysis because these cases may be assigned non-randomly. These cases make up only five percent of all investigations.

test scores (Heckman et al., 2013). For young boys, the estimates are imprecisely estimated, and we cannot rule out substantively large positive or negative impacts on their outcomes. We can statistically reject the hypothesis that the effects on test scores are equal for young girls and young boys (p-value < 0.10).

In line with the results for test scores, we find that removal has beneficial impacts on additional measures of schooling achievement. For young girls, we find that removal reduces the likelihood of repeating a grade by 42.6 percentage points. Removed young girls are also significantly less likely to participate in special education.[4] In contrast, we find no detectable impacts of removal on grade repetition or special education participation for young boys. But, as with the test score results, the lack of precision in the results for young boys suggests caution in the interpretation, as we cannot rule out large negative or positive impacts on these measures of schooling progress.

We examine whether these results are due to multiple hypothesis testing or attrition in the form of changes in public school enrollment. Following Anderson (2008), we calculate adjusted "q-values" that control for the false discovery rate (FDR). Using the set of results for gender subgroups, we find that the impacts on test scores and retention for young girls are significant at the 10 percent level using the FDR-adjusted q-values. For attrition, we find no statistically significant impacts of removal on enrollment for young girls or young boys. A caveat for this analysis is that the point estimates are large for boys. These point estimates for girls and boys are not statistically different.

Next, we investigate potential explanations for the differences in the impacts of removal on test scores by gender. Our analysis provides some limited but suggestive evidence that the pattern of results stems from differences in how girls and boys respond to removal. We conduct a subsample analysis that focuses on siblings and find estimates that are imprecise but in line with our main analysis. The point estimates suggest that sisters and brothers

[4]We measure participation based on whether the child has a written Individualized Education Program (IEP). An IEP can be given as early as pre-school, and children are assessed each year until they are deemed to no longer be in need. Note that having an IEP does not generally exempt a student from testing in Rhode Island. In the academic year 2013, 89 percent of Rhode Island students with an IEP took standardized exams.

from the same household respond differently to removal. Further, we find that young girls and boys generally have similar placement outcomes after removal (i.e., type of foster care and days spent in the foster care system) and attend schools with broadly similar types of characteristics (in terms of school value-added and student body composition). We also find little evidence that suggests the heterogeneous impacts on achievement are due to differences between girls and boys in terms of complier characteristics or parental responses to removal.[5] An important caveat is that many of the estimates that test potential mechanisms are relatively imprecise. This implies that we have limited ability to rule out economically meaningful gender differences for some mediators.

As a final analysis, we study the impact of removal for older children (investigated at age six or later). We study schooling outcomes and later-life outcomes such as juvenile delinquency, high school graduation, the likelihood of having a teen birth, and post-secondary school enrollment. This analysis of later-life outcomes focuses on older children because a child removed at a young age will not be old enough for us to observe outcomes by the end of our sample period. For older children of either gender, we find no statistically significant effects on any outcome. The point estimates tend to suggest that removed boys have worse outcomes in terms of the school index and high school graduation. For older girls, the non-significant point estimates are imprecise and have no consistent pattern.[6] That is, the signs of some point estimates indicate there are beneficial effects (e.g., increased college enrollment) while at other times the estimates indicate adverse impacts (e.g., lower average test scores). The estimates are sufficiently imprecise that we cannot reject that the effects of removal are equal for older girls and older boys.

Overall, these findings contribute to a broad literature on the impact of interventions for

[5]For example, the share of compliers that have a married parent is similar among young girls and young boys. To analyze parental behavior, we study parent perpetrators of abuse and neglect. Approximately 95 percent of the perpetrators in our sample are parents. Using samples of parent perpetrators for young girls and young boys, we find no statistically significant impacts of removal on criminal charges and incarceration.

[6]The imprecision in the results for older girls is a potential concern and stands in contrast to our analysis for younger girls where we consistently find beneficial impacts of removal on test scores and other schooling outcomes.

children from disadvantaged backgrounds that shows early-life interventions can have large causal impacts on children's outcomes (Garces et al., 2002; Ludwig and Miller, 2007; Almond et al., 2010; Heckman et al., 2013; Campbell et al., 2014; Elango et al., 2015; Aizer et al., 2016; Chetty et al., 2016; Hoynes et al., 2016; Isen et al., 2017; Chyn, 2018; Currie et al., 2020; Garcia et al., 2018; Doyle, 2019; Heckman and Karapakula, 2019*a*,*b*; Heckman et al., 2020).[7] Our results extend this literature by focusing on interventions for young children at risk of abuse and neglect and suggest that the impacts of removal for young girls may be particular to age. In addition, our findings complement the results from a growing literature showing heterogeneous program impacts by gender. As in our results, a number of studies find that schooling and social program interventions can have larger positive impacts for girls (Hastings et al., 2006; Kling et al., 2007; Angrist and Lavy, 2009; Heckman et al., 2013; Deming et al., 2014; Hoynes et al., 2016; Garcia et al., 2018).

2 Background: Child Protective Services and Case Assignment in Rhode Island

Figure 1 illustrates the process for child abuse and neglect investigations and home removal decisions in Rhode Island. An investigation of child abuse or neglect begins when an allegation is reported to the DCYF Child Protective Services (CPS) hotline.[8] The CPS hotline workers record details of the allegation, identify previous or pending investigations, and determine whether the report meets the criteria to initiate an investigation. If the criteria are not met, DCYF expunges the records of the allegation after a specified period. If the allegations meet the criteria for an investigation, a CPS report is created and forwarded to the central Investigative Unit (IU). A supervisor from this unit then assigns the case to a field Child Protective Investigator (CPI).[9]

[7] See Almond and Currie (2011) and Heckman and Mosso (2014) for a review of the literature on child development and the impact of interventions for children.

[8] Details on DCYF policies and procedures come from conversations with DCYF staff and documentation from the 2018 DCYF Policy Manual (Rhode Island Department of Children, Youth, and Families, 2018).

[9] In Rhode Island, there is one central Investigative Unit, which assigns cases to CPIs regardless of geography.

The supervisor assigns the authorized reports using an internal "rotation list," which effectively randomizes cases to available field CPIs. This rotation list is an ordered spreadsheet of CPIs that does not depend on investigator characteristics such as age, ethnicity, or geographic consideration. Each day, the supervisor assigns cases as they arrive based on this ordered list, and CPIs with non-assigned cases move to the top of the list for the next day's rotation.[10] The supervisor uses the rotation list to assign cases even when the child has had previous investigations. In interviews, the supervisor who assigns cases stated that the goal of the list is to provide "fairness" so that each field CPI will receive a similar mix of cases. The only exception for assigning cases through the list is when there is an allegation of sex abuse. For these allegations, the supervisor may assign the case to a CPI of the same gender as the victim.[11] Every case assigned outside of the rotation list is flagged in the case management system. As discussed further in Section 3, we use this flag to exclude cases from our analysis.

The CPI investigating the case decides whether there is sufficient evidence of abuse or neglect to warrant removal (i.e., out-of-home placement).[12] If there is sufficient evidence, the CPI petitions the Rhode Island Family Court (RIFC) for placement of the child into DCYF custody. According to conversations with DCYF staff, the RIFC typically follows the recommendation made by investigators. The average investigation (including those that do not end in removal) lasts less than one month.[13]

CPIs have limited ability to impact investigated children and their families other than

[10]Cases left unassigned on a day can be voluntarily picked by CPIs outside of this rotation list. These cases are flagged and excluded from the analysis.

[11]Note that sex abuse cases comprise five percent of all investigations, and we exclude these from our analysis.

[12]The assigned CPI also decides whether an allegation of abuse or neglect is founded or unfounded (see Figure 1). DCYF dismisses unfounded allegations, and children are not removed in those cases. An important concern is that a CPI with a high removal rate might also have a high or low rate of determining that allegations are unfounded, thereby generating sample selection bias. Appendix B discusses this concern and conducts analysis using data on a limited sample of unfounded records (the only data accessible because DCYF deletes older records after a specified period). As summarized in Appendix B, the results suggest there is no significant correlation between the CPI's unfoundedness and removal rates. The CPI's unfoundedness rate is also not significantly correlated with observable case characteristics. An important caveat is we can only examine the relationship between the instrument and unfounded investigations for a limited sample of years that do not cover our main analysis sample.

[13]In the sample of first investigations (described in Section 3), the average investigation lasts about 22 days. In cases where the CPI recommends removal, the average duration is 18 days.

through the removal decision.[14] The circumstances of the case largely determine the type of placement and the duration of time in the foster care system. DCYF places children in a family setting (relatives or a licensed foster family) or a supervised environment (a group home or shelter). The field CPI is not involved in a case once the investigation is closed following the removal decision. After removal, a social worker handles case management.

When a child is in DCYF custody, parents can work with the social worker to arrange visits, although visitation frequency varies depending on case-specific factors. DCYF releases children from custody upon reunification with parents, adoption, or aging out of the child welfare system. Reunification with parents occurs only after a parent has completed conditions stipulated by DCYF (e.g., parents may be required to follow a visitation plan or complete mental health counseling with a DCYF service provider). DCYF staff monitor whether a parent complies with conditions for reunification.

3 Data

We use data from anonymized administrative records housed in a secure enclave. All personally identifiable information has been removed from the data and replaced with anonymous identifiers. These identifiers allow researchers with approved access to join records associated with an individual across a range of social programs and government services (Hastings, 2019; Hastings et al., 2019). This section describes the samples and key measures that we construct. Appendix C provides further details and statistics on the approach for joining records.

3.1 Sample of Children Investigated at Young Ages

We construct a sample of investigated children from records of founded abuse and neglect investigations conducted by DCYF during January 1, 2000 to December 31, 2015. The sample is based on four main restrictions. First, we exclude investigations where the Investigative

[14] In Section 4.4, we provide a detailed discussion of the exclusion restriction necessary for our empirical analysis.

Unit supervisor may have assigned the case without using the rotation list (e.g., sex abuse investigations). Second, we exclude investigations that occur after the first investigation associated with each child (ages 0-18). Third, we drop investigations assigned to CPIs with outlier removal tendencies and exclude investigations assigned to CPIs who received less than ten cases.[15] Fourth, we focus on children *under* the age of six at the time of the first investigation (hereafter, referred to as "young" children).[16] The investigations sample contains 6,287 young girls and 7,387 young boys.

3.2 Samples and Outcomes for Main Analysis

We join the above sample of investigated young children with records from the Rhode Island Department of Education (RIDE) to create a schooling outcomes sample. The schooling outcomes sample is defined as the set of all investigated children who are observed in records for public (and charter) school enrollment and standardized exam performance. The enrollment records cover the academic years 2003-2016 and include information on the school attended, the grade enrolled, special education participation as indicated by receipt of a written Individualized Education Program (IEP), and attendance during the school year. The standardized exam file contains math and reading test scores for exams taken in grades 3-8 during the academic years 2005-2016.

There are 2,614 young girls and 3,142 young boys in the schooling outcomes sample. Note that investigated children who were born after 2008 are *not* in the sample because they are too young to have taken standardized exams during the period 2005-2016. In addition, investigated children who moved from Rhode Island or enrolled in a private school are not included.[17,18]

[15] We define outliers as values of CPI removal tendency that fall below the first percentile and above the ninety-ninth percentile.

[16] In Section 5.5, we provide results using alternative age ranges to define a sample of young children. In Section 7, we report results studying children who were ages six to 18 at the time of the first investigation.

[17] Official reports indicate that fewer than 10 percent of children living in Rhode Island attend a private or parochial school (Rhode Island Kids Count, 2016).

[18] Out-of-state migration is a possible reason why some investigated children are not observed in a Rhode Island public school. In the investigations sample, we observe 6,350 young children who enrolled in kindergarten before 2015. Among these children, 308 (4.9 percent) do not enroll in a Rhode Island public

Our main schooling outcomes are standardized test scores for exams taken in grades 3-8. We construct a school-year panel with the average of math and reading scores (standardized by grade and year) for children in the schooling outcomes sample.[19] This panel contains 9,980 student-year observations with non-missing math and reading scores for 2,614 young girls and 12,344 student-year observations with non-missing math and reading scores for 3,142 young boys.

For children in the schooling outcomes sample, we also study grade repetition, special education participation, and average yearly absences during elementary and middle school (grades K-8).[20] We measure special education participation based on whether a child ever has a written IEP. A child who has an IEP has at least one of 13 disability categories (e.g., developmental delay and emotional disturbance) defined by the Individuals with Disabilities Education Act (IDEA) (Individuals with Disabilities Education Act, 2004). The determination of an IEP can start as early as pre-school, when the child is three to four years old. More than half of students with an IEP in Rhode Island are identified with special needs prior to entering first grade.[21] For absences, we compute the average annual absences across grades.

Finally, we construct an enrollment outcome sample that includes all investigated children who were born between 1995 and 2008. These are the cohorts that would have been age-eligible to attend grades 3-8 during the period in which we can observe test scores (i.e., 2005-2016). This sample of investigated children includes 3,971 young girls and 4,770 young boys. We create a yearly panel for these children and use RIDE enrollment records to measure

(or charter) school in first grade. Based on an analysis of exit codes for this sample, 70 percent of these non-continuing students moved out of state.

[19] All test scores are post-investigation since young children are removed before age six and tested starting around age eight.

[20] Note that the grade repetition outcome is only defined when children are enrolled for two consecutive years. In our schooling outcomes sample, there are 10 young girls and 19 young boys for whom we cannot measure grade repetition because they are enrolled for only one academic year during our sample period. In addition, note that a given child may not have a complete set of academic years for which we can measure grade retention or IEP enrollment. For example, if a child transfers (permanently) from a public to a private school in fifth grade, we would observe IEP enrollment only from third to fourth grade. We keep these children in our analysis and compute grade retention and IEP participation for the grades available.

[21] About 28 percent of children receive their IEP for the first time in kindergarten. An additional 25 percent of children receive an IEP before starting kindergarten and enroll in an Early Childhood Special Education program for young children with developmental delays and disabilities, as mandated by IDEA.

whether they enrolled in school during the years that we would expect attendance based on their date of birth. In addition to enrollment, we study whether the investigated child had a non-missing standardized exam score in each year of the panel. The panel that we study for the enrollment sample has 17,164 and 21,420 observations for young girls and young boys, respectively.

3.3 Child Outcomes for Mediating Factor Analysis

In Section 6, we study several factors that may mediate the impact of removal from home on child schooling outcomes. For our main schooling outcomes sample, we study two types of mediating factors using data from DCYF and RIDE. First, we study foster care outcomes associated with the first investigation. The DCYF data contain information on the number of days that a removed child was placed in any type of foster care, with relatives, with a foster family, in a group home, or in other less common care settings (e.g., in an emergency shelter). The data also contain information on whether a removed child was adopted. Second, we study school mobility and characteristics of public schools attended by children in our sample. For school mobility, we construct a measure of switching public schools. For attended school characteristics, we construct school-level measures of test score value-added, average test scores, the fraction of enrolled students who are minorities, and the fraction of students who receive a special education services (IEP). Value-added for each school is estimated using all years available for the school and excluding the students in our DCYF investigation sample. We regress average standardized test scores (the average of math and reading scores) on lagged test scores (including their square and cube), as well as indicators for a student's race, gender, special education status, and free or reduced-price lunch status. We use the mean residuals within a school as a single measure of value-added.[22] The fraction of minority and IEP students at a school are calculated in each school year. We join all school characteristics to a child-level panel covering grades 3-8 to measure the impact

[22]Our approach is similar to the methods applied in prior studies such as Kane et al. (2008) and Chetty et al. (2014). See Appendix C for details on the estimation of school value-added and how we join this measure to the student-level data.

of removal on the characteristics of the schools attended post-investigation. We observe school characteristics in at least one academic year for every child in our schooling outcomes sample.[23]

3.4 Parent Perpetrators and Crime Outcomes

For our mediator factors analysis, we also study the impacts of removal on outcomes of parents of children in our schooling outcomes sample. We obtain information on parents from DCYF records on perpetrators associated with an investigation.[24] For young children in our sample, 95 percent of children have at least one parent listed as a perpetrator. We use this information to create a sample of parent perpetrators. Specifically, we join this sample to criminal charge and incarceration records (1995-2017) from the Rhode Island Department of Corrections (RIDOC). The outcome of interest is whether a parent perpetrator is charged or incarcerated at any point in the four years following the conclusion of an investigation. We also study this outcome by type of committed offense, such as property charges, drug charges, public offenses (e.g., disorderly conduct), or sex-related offenses. Because the criminal justice data source ends in 2017, these measures will be partially censored depending on the associated investigation's end date. For the children in our schooling outcomes sample, there are $2,333$ parent perpetrators associated with young girls and $2,777$ parent perpetrators associated with young boys.

3.5 Descriptive Statistics

Table 1 presents summary statistics for the investigations sample of young children in Rhode Island. Column 1 shows that 59 percent and 16 percent of investigated children are white and Hispanic, respectively.[25] Race in the sample differs notably from Doyle (2007;

[23]Sample sizes vary slightly across these school-related outcomes due to missing data. For example, there are 10 young girls in our sample for whom we cannot measure school mobility because they are enrolled only in one academic year during our sample period. See Appendix C for further details.

[24]The DCYF investigation records have information on household characteristics, but there is no information on parent identity aside from the information contained in perpetrator records.

[25]Nationally, 45 percent of child abuse victims were white, and 22 percent were Hispanic (U.S. Department of Health and Human Services, 2016).

2008), which studied the impact of removal for a sample from Illinois where 76 percent of investigated children were African American. This contrast partly reflects differences in the demographics between the two states. That is, nine percent of children in Rhode Island and 15.8 percent of children Illinois are African-American (U.S. Census, 2018). In terms of family background, 12 percent of the investigated children in our sample are from married households.

The DCYF data report all allegations associated with an investigation. An allegation of neglect occurs in about 80 percent of investigations. Allegations of physical abuse or physical neglect (e.g., neglect that results in a physical injury) occur much less frequently, in about 14 and seven percent of investigations, respectively. These statistics are broadly in line with national statistics, where allegations of neglect and physical abuse occur in 75 and 18 percent of investigations, respectively (U.S. Department of Health and Human Services, 2016).

We also observe who reported the allegation associated with the investigations. For 82 percent of children, the reporter in the case is a professional such as a teacher, physician, social worker, or police officer. The remaining fraction of reports are made by family, friends, or other individuals such as neighbors or anonymous reporters.

Column 1 of Table 1 shows that removal from home occurs in 20 percent of the sample of first investigations. This rate is lower than the removal rate observed in Doyle (2007; 2008), which studied older children from Illinois during the 1990s when the state's placement rate (27 percent) was one of the nation's highest. Columns 2 and 3 provide separate summary statistics for children subject to investigations that do not and do result in home removal, respectively. Column 4 reports the p-values from tests of differences in means for each summary statistic. Investigations that do not end in removal have significantly different child and case characteristics from investigations where removal does occur. Children who are not removed are slightly older than those who are removed (2.0 years old versus 1.1), live in households with seven percentage point higher marriage rates (p-value < 0.01), and are about four percentage points less likely to be African-American (p-value < 0.01). The final

row of Table 1 shows that children who are removed spend roughly 450 days in foster care, which is less than the average four-year stay in Doyle's (2007; 2008) study of Illinois.

4 Empirical Strategy

Consider the following regression model of the relationship between child outcomes and removal:

$$Y_i = \beta_0 + \beta_1 R_i + \beta_2 X_i + \epsilon_i, \tag{1}$$

where Y_i is a post-investigation outcome for child i, R_i is an indicator for whether the child was removed during the first investigation, X_i is a vector of child and case characteristics (including fixed effects for the investigation year), and ϵ_i is an error term. The child and case characteristics are those listed in Table 1. Standard OLS estimates of Equation 1 will be biased if home removal (R_i) is correlated with unobserved determinants of child outcomes (ϵ_i). The descriptive statistics in Table 1, as well as prior research, suggest that observed and unobserved family and home conditions affect both the likelihood of removal and child outcomes (Berger et al., 2009, 2015; Wildeman and Waldfogel, 2014).

To address the endogeneity concern in Equation 1, we rely on an instrumental variable (IV) strategy that uses a measure of the removal tendency of the investigator j who handles case c associated with child i. We denote the removal tendency as Z_{ijc}, and the first-stage equation is:

$$R_i = \alpha_0 + \alpha_1 Z_{ijc} + \alpha_2 X_i + \nu_i, \tag{2}$$

where Z_{ijc} is a leave-out removal tendency measure that is similar to measures calculated in the literature using judge decision tendencies as instruments for individual case decisions (Kling, 2006; Doyle, 2007, 2008, 2013; Dahl et al., 2014; Aizer and Doyle, 2015; Mueller-Smith, 2015; Eren and Mocan, 2017; Sampat and Williams, 2019; Dobbie, Grönqvist, Niknami,

Palme and Priks, 2018; Dobbie, Goldin and Yang, 2018; Bhuller et al., 2018, 2020). In our setting, we construct this measure to account for the fact that 30 percent of the cases in the DCYF sample include siblings.[26] We exclude the focal child i and siblings on case c by defining the leave-out removal tendency for each case as:

$$Z_{ijc} = \frac{1}{N_j - n_c} \left(\sum_{k \neq c}^{N_c} R_k' \right), \tag{3}$$

where N_j is the total number of children assigned to the investigator j, n_c is the number of children on case c, and N_c is the number of cases assigned to the investigator. We define k to index the cases handled by investigator j, and R_k' is the number of children removed on case k. The leave-out measure is an average of home removals that excludes children from the same case. We calculate the measure of removal tendency using all cases for investigator j within an eight-year window.[27] When we estimate Equation 1 using this leave-out measure, we report two-way clustered standard errors at the investigator (CPI) and case (i.e., family) level.

If there are heterogeneous impacts of removal, our approach can identify a local average treatment effect (LATE) of removal for children with marginal cases (i.e., those where investigators may disagree about the decision to remove a child from their home). We must make two assumptions to interpret IV estimates of the parameter β_1 from Equation 1 as a LATE of removal (Imbens and Angrist, 1994). First, the measure of CPI removal tendency defined in Equation 3 must affect child outcomes by changing only the probability of removal.

[26] See Appendix Section C.5 for further details on siblings.

[27] Following prior studies, we allow CPI tendency to evolve over time (Doyle, 2007, 2008). This approach accounts for potential changes in child protection policies or shifts in local social and economic factors that may impact child protection agency work (Hegar and Scannapieco, 1995; Doyle and Peters, 2007; Schneider et al., 2017). In particular, our data cover a period of reform and budgetary cuts for DCYF. The DCYF budget for fiscal year 2008 was cut by $60.4 million from the prior year (from $293.1 to $232.7 million), and Tom Dwyer, longtime head of child welfare for Rhode Island, left office in 2007. In addition, the Great Recession began in late 2007. We allow our measure of CPI removal tendency to vary separately for the first (2000-2007) and second halves (2008-2015) of the period covered by our data. As demonstrated in this section, the first-stage coefficient on this version of removal tendency is substantively large and has a statistically significant impact. Note that Section 5.5 shows that results for young girls are consistently positive and significant using alternative definitions for the leave-out measure of CPI removal tendency.

Sections 4.3 and 4.4 provide evidence suggesting this assumption is plausible in our setting by examining random assignment of investigators and analyzing whether CPI removal tendency is correlated with other post-removal decisions such as the type of placement or whether police were notified during an investigation. Second, we assume that there is a monotonic impact of CPI assignment on removal across children. That is, a child removed by a lenient investigator would also be removed by a more strict investigator. A violation of this assumption may occur if CPI removal tendencies vary with case characteristics. For example, a given CPI may be relatively strict when it comes to removing non-white children, but lenient when it comes to removing white children. If there is a non-monotone impact of removal tendency, the IV estimate will not identify a well-defined LATE.[28]

As tests of monotonicity, Section 4.5 follows prior work to examine this assumption in our setting (Bhuller et al., 2020; Dobbie, Goldin and Yang, 2018). We show that the first-stage coefficient for the tendency measure defined in Equation 3 is positive in various sub-samples. Similarly, Section 4.5 also shows that the first-stage coefficient for a reverse-sample tendency measure is positive in various subgroups.[29] Finally, to further address monotonicity concerns, we carry out robustness tests in Section 5.5. Specifically, we follow Aizer and Doyle (2015) by dividing our sample into mutually exclusive subgroups and constructing group-specific versions of the leave-out measure of CPI removal tendency. For example, one version of our approach calculates the instrument for each allegation type (i.e., neglect versus non-neglect). We use this version of the leave-out measure as the instrument in our robustness test. This approach relaxes the monotonicity assumption by allowing each CPI to have different tendencies depending on the allegation type. In addition to constructing a measure based on subgroups defined by the allegation type, we also focus on subgroups defined by gender, minority status, parent marital status, and reporter type.

[28]Under non-monotonicity, the IV estimate would be a weighted average of marginal treatment effects where the weights do not sum to one (Angrist et al., 1996; Heckman and Vytlacil, 2005).

[29]We calculate the reverse sample tendency by dividing the sample into subgroups (e.g., by race) and constructing instruments using the complement for each subgroup. For example, we recalculate the removal tendency for white children using all observations outside this subgroup (all non-white children).

4.1 Variation in Child Protective Investigator (CPI) Removal Tendency

Figure 2 plots the distribution of the leave-out CPI removal tendency from Equation 3 in our sample of investigated young children. We observe 102 CPIs during 2000-2015, and these CPIs see an average of 387 children across all years. Figure 2 shows that CPIs differ in their propensities to recommend home-removals. The mean of the removal tendency measure is 0.178, while the 25th and 75th percentiles of the distribution are 0.140 and 0.216, respectively. The standard deviation is 0.055.[30] Appendix C.4 provides further statistics and information on the CPIs in our analysis sample.

4.2 First-Stage Impact

Figure 2 also plots the predicted probability of home removal from a local linear regression. The probability of being removed from home increases with CPI removal tendency. Consistent with this, Table 2 reports results from Equation 2, measuring the impact of our instrument on removal of the child from the home due to the investigation. This table shows the results separately by gender for the investigations sample (Columns 1 and 2) and the schooling outcomes sample (Columns 3 and 4). Overall, the leave-out measure of mean CPI removal tendency is significant and highly predictive of removal across samples. For example, the estimate in Column 1 implies that moving from a CPI in the lowest quartile of removal tendency to one in the highest quartile would increase the likelihood of removal by about 4.5 percentage points ($= 0.594 \times 0.076$), relative to a mean removal rate of 20.8 percentage points.[31] The point estimates suggest that removal tendency has a smaller impact for boys, but we cannot reject the hypothesis of equal first-stage impacts between girls and boys. Finally, note that the F-statistics for the excluded instrument are 37.70 and 69.16 in the

[30] In Doyle (2007), the standard deviation is nine percent in the delinquency sample, 10 percent in the teen motherhood sample, and seven percent in the labor market outcomes sample.

[31] Doyle (2007; 2008) discusses the possibility that the coefficient on the impact of CPI removal tendency may be less than one due to measurement error.

investigation samples of young girls and young boys, respectively.[32,33] These are above the thresholds for weak instruments discussed in Staiger and Stock (1997), Stock et al. (2002), and Stock and Yogo (2005).[34]

4.3 Instrument Validity: Testing Random Assignment

According to the assignment process described in Section 2, investigations in our samples should be quasi-randomly assigned to CPIs. To test this implication, we regress the removal tendency on baseline child and case characteristics for various samples of young investigated children. Table 3 reports the point estimates and results from a test of the joint significance of baseline characteristics. Specifically, Columns 1-4 report these results for young children by gender in the investigation and schooling outcomes samples described in Section 3. The point estimates are generally small (i.e., less than one percentage point) and not statistically significant. We consistently fail to reject the null hypothesis that the coefficients for child and case characteristics are jointly zero. For example, Column 1 shows that the chi-squared test statistic is 10.249 with a p-value of 0.673 for young girls in the investigations sample.[35]

[32]The F-statistics in the test score and additional schooling outcome samples are lower. For example, Table 4 shows that the F-statistics on the excluded instrument are 17.70 and 14.00 when we analyze the test score samples of young girls and young boys, respectively. Similarly, Table 5 shows that the F-statistics for young girls and young boys are 15.04 and 12.52 in the school index outcomes sample, respectively.

[33]As an additional robustness test, we follow the recommendation from Andrews et al. (2019) and compute Anderson-Rubin (AR) confidence intervals for the effects on test score and non-test score outcomes. The AR confidence intervals are robust to weak identification and are efficient in the just-identified case. Appendix Table A17 reports the AR confidence interval results and reproduces our main estimates for convenience. The conclusions based on the AR confidence intervals match what we observe for the main results. For example, we still observe positive impacts of removal for removed young girl, as none of their AR confidence intervals for all test score outcomes, grade repetition, likelihood of having an IEP, and the school index contain zero.

[34]An additional consideration is that the leave-out instrument is estimated. To address this, we use an approach that bootstraps the first-stage F-statistic. For each case worker, the bootstrap procedure samples (with replacement) their investigations and calculates leave-out removal rates within the sampled data. We create 250 bootstrap samples which we use to estimate first stage models. The mean and median of the F-statistics for our samples are slightly smaller than the ones associated with the first stage for the main analysis. For example, when we analyze test scores, the F-statistics for the first stage are 17.70 and 14.00 in the young girl and young boy samples, respectively (see Column 2 of Table 4). In the bootstrap results, we obtain a mean F-statistic of 14.60 for young girls and a mean of 11.67 for young boys. See Appendix Table A18 for further details. The medians are 14.22 and 11.29, which are close to the mean values.

[35]Appendix G shows that we obtain similar results in the sample that we analyze for studying enrollment as an outcome.

4.4 Instrument Validity: Exclusion Restriction

The random assignment of cases to investigators is sufficient for a causal interpretation of the reduced form impact of being assigned to a stricter investigator. However, interpreting IV estimates as measuring the causal impact of removal in Equation 1 further requires that the removal tendency of an investigator should affect children only through the decision to remove a child from home and not through any other channel. For example, the exclusion restriction would be violated if CPIs also determined the duration of foster care, the number of placements, or the type of foster placement. In Appendix Table A1, we test whether CPI removal tendency is correlated with these foster care outcomes for the subgroup of children who have been removed. We also test if there is a correlation between whether police are notified during an investigation, which might also affect child well-being. We find no significant correlations between removal tendencies and foster care outcomes in these samples of removed children, although the estimates are imprecise. The point estimate for the length of stay in foster care is larger in magnitude for removed young girls relative to the estimate for their male counterparts. However, this is not a statistically significant difference (p-value $= 0.39$).[36] For the remaining foster care outcomes of removed children, the point estimates have more similar magnitudes by gender. This pattern of results is consistent with the idea that CPIs have limited ability to influence a child's outcomes once a child is placed into DCYF custody (as discussed in Section 2).

4.5 Monotonicity

To interpret IV estimates from Equation 1 as a LATE of removal for marginal investigations, we must assume monotonicity in the impact of the CPI removal tendency on the likelihood of removal across children in our sample. As noted in Bhuller et al. (2020) and Dobbie, Goldin

[36]Note that CPIs do not directly determine the length of stay in foster care. However, it is possible that the removed children assigned to very strict investigators have shorter foster care stays because their cases may have lower average unobserved severity. The results in Appendix Table A1 show there is no statistically significant correlation between removal tendency and length of stay. An important caveat for this result is that the standard errors are large.

and Yang (2018), one testable implication of monotonicity is that the first-stage estimates should be non-negative for any subgroup of the investigations sample. The results in Table 2 provide an initial indication that there is no evidence of a violation of monotonicity across all cases by showing that the first-stage is non-negative for the subgroups defined by gender. Appendix Table A2 expands on these results by providing additional results for narrower subgroups based on various case characteristics. The first-stage impacts of removal tendency are consistently positive.[37] An additional implication of monotonicity is that CPIs should be stricter for a specific type of investigation if they are stricter in other investigation types. To test this implication, we estimate first-stage models where we recalculate the leave-out instrument for each subgroup using all investigations outside of the subgroup. For example, we estimate a first-stage model for Hispanic children using the CPIs' removal tendency calculated for all non-Hispanic investigations. Appendix Table A3 shows that these estimates are positive and almost always statistically different from zero.

4.6 Interpreting the LATE in Our Analysis

Assuming the exclusion restriction and monotonicity assumptions hold, the IV estimates of the parameter β_1 from Equation 1 are a local average treatment effect (LATE) of removal for children who would have received a different removal decision had their case been assigned to a different investigator. To better understand this treatment effect parameter, we examine characteristics of compliers in our sample of first investigations for girls and boys separately. We calculate these characteristics following the approach from Abadie (2003), Dahl et al. (2014), and Dobbie, Goldin and Yang (2018).[38]

Each row of Appendix Table A4 provides information on the overall sample mean for a case characteristic and the complier-specific mean. We provide these statistics separately

[37]The magnitudes of the first-stage estimates for subgroups defined by each case characteristic (shown in the rows of Appendix Table A2) are generally similar to the impact in the sample of all investigations.

[38]Similar to Dahl et al. (2014) and Dobbie, Goldin and Yang (2018), we define compliers in our setting as children whose removal decision would have been different had their case been assigned to the most lenient versus the strictest investigator. We consider investigators in the top percentile of removal tendency as "strict" and investigators in the bottom percentile of removal tendency as "lenient." See Appendix Section D and the notes to Appendix Table A4 for further details on our calculation of complier characteristics.

for girls and boys investigated at young ages. For each gender, we see that compliers are generally similar to the average child in our investigation sample. The main exception is that compliers in the sample of young girls are less likely to be white. Comparing Columns 2 and 4, we also see that complier girls and boys have similar characteristics except in terms of race. For example, complier young girls are about 17 percentage points less likely to be white than complier young boys.

5 Main Results

5.1 Standardized Test Scores for Young Children

Table 4 presents estimates of the impact of removal on standardized test scores for young girls (Panel A) and young boys (Panel B).[39] Columns 1 and 2 provide IV estimates on our main outcome, the average of math and reading scores with and without controls for case characteristics. Similarly, Columns 3-6 provide estimates separately for math and reading scores. Robust standard errors that are two-way clustered at the case (family) and investigator level are reported throughout. As detailed in Section 3, the sample for this analysis contains $2,614$ young girls and $3,142$ young boys, which differs from the sample of all investigated young children in Table 1. The sample is smaller because some investigated children are not old enough to attend grades 3-8 during the school years with available test-score information (2005-2016) and others do not attend a public (or charter) school in Rhode Island.

The results in Panel A show that the marginal removal has a significant and positive impact on the average standardized test scores for young girls. Column 1 shows that the point estimate for removal is 1.349 standard deviations. We obtain nearly identical results when we include controls for case characteristics in Column 2.[40] Results for standardized math and reading scores are similarly large in magnitude and statistically significant. Evaluations of

[39] We report results for test scores using the pooled sample of young girls and boys in Appendix Table A5. In the pooled sample, we find that removal increases test scores by 0.739 standard deviations. This result is statistically significant at the 10 percent level.

[40] Appendix Table A6 shows that the estimates from a model that includes birth cohort fixed effects are similar to our main results.

high-quality early education programs targeting disadvantaged children serve as an important point of comparison for these impacts. Heckman et al. (2013), for example, found that the Perry Preschool program increased female standardized test scores by 0.806 standard deviations. As another benchmark, Bharadwaj et al. (2013) and Chyn et al. (2021) find that neo-natal investments for babies born at very low birth weight increase standardized test scores by $0.15 - 0.34$ standard deviations in elementary and middle school.

Our estimated impact is large in magnitude, but note that complier young girls in our sample would have had very low standardized test scores if they had not been removed. Following the approach from Dahl et al. (2014) and Bhuller et al. (2020), we estimate outcomes for compliers if they had not been removed, finding that the mean complier among young girls would have had an average standardized test score of -1.753.[41] This implies that young girls at the margin benefit from removal, but they are still likely to have below-average test scores.

In contrast to the results for young girls, Panel B shows that there are no detectable impacts on any measure of test scores of young boys at the margin of removal. The point estimates for boys are generally an order of magnitude smaller than what we obtain for girls, although the standard errors in these estimates are large and the confidence intervals contain effect sizes that are substantively large. We can reject the hypothesis of equal impacts of removal by gender for our preferred estimate in Column 2 (p-value $= 0.054$).[42]

In Appendix Figure A1, we report estimates and confidence intervals for impacts on average standardized test scores separately for each grade (3-8) in the panel of test scores. For girls, we find positive point estimates that are similar in magnitude across grades. This pattern suggests that the benefits of removal are persistent and may be due to permanent changes in child ability prior to third grade. We also find that the contrast between the

[41] For a detailed discussion of our calculation of the complier average outcome when not removed, see Appendix D.

[42] The p-value from the test of equality of coefficients is calculated from a pooled regression that includes the samples of young girls and young boys. The regression includes all variables fully interacted with each respective gender. We test for differences in the gender-specific effects of removal using the results from this specification.

impacts on test scores for girls and boys is constant across grades. For young boys, the estimates are never significant and are generally smaller in magnitude than the results for girls.

5.2 Grade Retention, Special Education, and Attendance for Young Children

Table 5 tests for impacts on additional schooling outcomes for the same sample of young children with test scores. As discussed in Section 3, we measure impacts on ever repeating a grade, ever participating in special education services (i.e., having a written IEP), and the average number of absences during elementary and middle school (grades K-8). Due to the number of outcomes, Table 5 reports the estimates from a specification with case controls. We report the estimates from a specification without case controls in Appendix Table A7.

The results in Panel A from Table 5 show that marginally removed young girls are significantly less likely to ever be retained at school. The point estimate shows that removal decreases the likelihood of any grade repetition by 42.6 percentage points. As with test scores, this impact is large, but the complier mean rate of repeating a grade when not removed is also high (48.7 percent). Panel A also shows that removal has a significant and large (51.1 percentage points) reduction in special education needs as measured by ever having a written IEP during grades K-8.[43,44] There is suggestive evidence of a decrease in the mean number of absences for young girls, although the point estimate is not statistically significant.

In Panel B, the results show that there are no statistically significant effects of removal on any non-test-score schooling outcomes for young boys at the margin of removal. In Columns 1 and 2, the point estimates suggest that removal decreases the likelihood of grade retention or special education needs by four and 19.5 percentage points, respectively. For absences, the estimate suggests that removal has relatively small benefits. Although the signs of these impacts are consistent with school improvements for young boys, it is worth highlighting that

[43]Appendix Table A9 shows the results when re-defining these additional non-test-score schooling outcomes for grades 3-8. We find similar results for the measures defined over this grade range.

[44]Note that an IEP does not imply that a student is exempt from testing. In the academic year 2013, 89 percent of students with an IEP took a standardized exam.

the estimates are sufficiently imprecise that we cannot rule out substantively large adverse effects of removal for these outcomes.

Overall, the results for retention, special education participation, and absences are consistent with the pattern of heterogeneous impacts by gender observed for test scores. To summarize these schooling results, we construct a school index measure, which is the equally-weighted average of the standardized (z-score) measures for the three outcomes.[45] One interpretation of this index is that lower values indicate that children have more schooling ability or less difficult experiences in school. This is summarized in Column 4 of Table 5, which shows that removal leads to a large and significant improvement in the school index for young girls at the margin of removal (i.e., a 0.998 standard deviation *decrease* in the average standardized measures of grade retention, special education and absences). The corresponding estimate for boys is much smaller in magnitude and not significant. Yet, it is worth noting that we cannot reject the hypothesis that the effects on this index are equal for boys and girls (i.e., a test of the hypothesis of equal impacts has a p-value equal to 0.18).

5.3 Attrition Due to Changes in Public School Enrollment or Test-Taking

A concern for interpreting these schooling results is that removal may affect whether a child attends a Rhode Island public school and takes a standardized exam. This would generate selection into the analysis of test scores and additional schooling outcomes. To address this possibility, we construct a balanced panel with indicators for enrollment and exam-taking during the years in which each student is expected to take the exams based on their date of birth. Note that the sample for this analysis is larger than what appears in Tables 4 and 5 because investigated children that never appear in the school enrollment records are included.[46]

[45] To standardize each component, we calculate the mean and standard deviation of each outcome using investigated children by gender. Next, we compute the standard score by taking each outcome and subtracting the mean for all investigated children of the same gender and dividing by the standard deviation.

[46] Appendix G reports first-stage and tests of randomization results for the sample of children included in this enrollment analysis. The first-stage and randomization results are similar to what we observe for the schooling outcomes sample.

Table 6 shows that there are no significant impacts of removal on enrollment or exam-taking for young girls and young boys. The insignificant point estimates suggest that, if anything, marginally removed young girls are more likely to be observed in the test score panel. For young boys, the point estimates are larger in magnitude and negative for both outcomes. These estimates range from -0.35 to -0.26 and are imprecise. We cannot reject the hypothesis of equal impacts on enrollment or exam-taking for girls and boys (p-values range from 0.17 to 0.33 depending on whether case controls are used). Although we lack precision in this analysis, the overall results do not suggest that attrition from public school or selective test-taking drive our results for young girls.

5.4 Multiple Hypothesis Testing

Given that our analysis tests for impacts for multiple outcomes, one concern might be that the findings for young girls are an artifact of multiple hypothesis testing. To manage the risk of false positives, we follow the recommended practice of adjusting per comparison p-values (Anderson, 2008). We use the two-step procedure from Benjamini et al. (2006) to calculate "q-values" that control for the false discovery rate (FDR), which is the proportion of rejections that are false positives (Type I errors). Appendix Table A10 shows that the IV estimates for test scores and grade retention of young girls are significant at the 10 percent level after adjusting for the fact that we analyzed multiple outcomes (i.e., impacts for average test scores, retention, participation in special education, and average absences) for boys and girls.

5.5 Robustness Tests

In this section, we report results from several robustness tests for standardized test scores.[47] We begin with checks related to changes in the sample definition. For comparison, Column 1 of Appendix Table A11 reproduces the estimate for the impact of removal on the

[47]We also show results for all robustness exercises for our school index measure (i.e., the measure based on retention, special education participation, and average absences) in Appendix Tables A12, A14, and A16.

average of standardized tests from our preferred specification. Recall that this specification includes children investigated before age six whose assigned CPI handled at least 10 children from other cases. Columns 2-4 provide results for samples of children who were assigned to CPIs with at least 100, 200 or 300 other cases. Our main conclusions generally do not change based on the results for these alternative samples. The main differences are that we lose statistical significance when we impose the 300-case restriction for young girls (a change that reduces the sample by half), and the sign and magnitude of the estimates for young boys change notably when we impose the 200- or 300-case restriction. Column 5 shows that the results are similar when we include children involved in sex abuse allegations. The estimate for girls remains positive and significant, while the estimate for boys continues to be imprecise.[48] Columns 6 and 7 test whether the results are sensitive to the definition of the age range used to define the sample of young investigated children. The estimates are similar to our main results when we define the sample of young children as those investigated during ages 0-4 or during ages 0-6.

Next, we check whether the results are robust to using an approach that allows CPI removal tendency to vary with case characteristics. As discussed in Section 4, this allows us to relax the assumption of monotonicity necessary to interpret our main results as the LATE of removal for marginal investigations (Imbens and Angrist, 1994). The case characteristics that we consider are sex, race (non-minority and minority), parent marital status, allegation type, and reporter type. Based on each of these characteristics, we define mutually exclusive groups of children and calculate CPI removal tendency for the group. For example, each CPI will have a leave-out removal tendency that is calculated separately for non-minority (white) and minority (non-white) children.

Appendix Table A13 shows the results for test scores when we allow CPI removal tendency to vary with each of these case characteristics.[49] For comparison, Column 1 repeats the results

[48] Recall that sex abuse investigations are excluded from the main analysis since these case assignments may take into consideration the gender of the CPI.

[49] Appendix E reports robustness tests results where we follow Mueller-Smith (2015) and use LASSO to select the instruments with greatest predictive power for removal in the first stage equation (Belloni et al.,

from our preferred estimate. Columns 2-6 report impacts for each of the characteristic-specific versions of the instrument. The results are broadly similar to our preferred estimates in that we consistently find evidence of beneficial effects of removal for young girls. For young girls, Panel A shows that the effects are consistently positive, and the point estimates are larger than one in four out of the five instrument versions. The estimates are also statistically significant except in the specifications that use versions of the instrument that vary either by gender or parent marital status. Panel B shows that the IV estimates for young boys remain imprecise, and the point estimate changes signs based on the version of the instrument specified.

Finally, we test whether our results are robust to alternative definitions of the instrument in terms of the set of investigations or time period used to calculate CPI removal tendencies. Similar to our previous tables, Column 1 of Appendix Table A15 reproduces our preferred estimates for test scores. Recall that this estimate is based on an instrument that is calculated using removal tendencies measured with an eight-year period for each CPI and using all children (i.e., those with first and subsequent investigations).[50] By calculating the removal rate within an eight-year period, we allow the removal tendency of a given CPI to change over time.[51] Columns 2-4 of Panel A show consistently positive and statistically significant results for young girls when we use a measure of removal tendency that is based only on first investigations or based on pooling investigation decisions for all years (2000-2015). The point estimates are always greater than one standard deviation. Additionally, Columns 5-8 show similar results when we use residualized removal measures of these alternative instruments.[52]

2014).

[50]For the analysis of impacts on outcomes, we only use the first investigation associated with a child. In the construction of the instrument, we use first and subsequent investigations. This provides a larger sample to calculate removal tendencies, allowing us to increase statistical power.

[51]As discussed in Section 4, we allow the CPI removal tendency to vary over time by calculating the measure separately for the 2000-2007 and 2008-2015 periods, respectively. As in prior studies such as Doyle (2007), an alternative approach is to calculate the leave-out instrument for each year to allow CPI tendency to evolve. A concern is that some CPIs in our sample see relatively few children within a year, thereby making it difficult to infer their tendency.

[52]Specifically, we follow Dahl et al. (2014) and Dobbie, Grönqvist, Niknami, Palme and Priks (2018) by constructing residualized measures of the instrument. We re-calculate these measures as follows. First, we regress removal on investigation year fixed effects. Second, we use the residuals from this regression to

Panel B shows that the results for young boys are never statistically significant, and the point estimates are sensitive to the instrument definition.

5.6 Marginal Treatment Effects

To further examine the impacts of removal on test scores of young children, we explore heterogeneity by examining marginal treatment effects (MTEs). MTEs are treatment effects for individuals with a particular "resistance" to treatment (Cornelissen et al., 2016). These effects are defined under a generalized Roy model. In our context, let Y_1 and Y_0 denote the potential outcomes if a child is removed or not removed, respectively. We assume that each of these is a linear functions of both observable (X) and unobservable factors. The choice to remove a child by a CPI is given by the indicator function $I = 1(v(X, Z) - U)$, where v is any function, Z is the leave-out removal tendency instrument, and U is an unobserved continuous random variable. Since U enters the removal equation with a negative sign, it is interpreted as resistance to treatment (removal). We can re-write the CPI choice equation as $P(X, Z) > U_d$, where $P(X, Z)$ is the propensity score and U_d represents quantiles of the unobserved resistance to removal (U).

The MTE is defined as $\mathbb{E}(Y_{1i} - Y_{0i}|X = x, U_d = u)$, and the dependence of the MTE on U_d reflects unobserved heterogeneity in treatment effects (Heckman and Vytlacil, 1999; Heckman et al., 2001; Heckman and Vytlacil, 2005, 2007). As in prior studies, we assume separability between observed and unobserved heterogeneity in treatment effects (Carneiro et al., 2011; Bhuller et al., 2020; Brinch et al., 2017; Cornelissen et al., 2018). Given this assumption and the exogenous instrument condition from Section 4, the MTE is identified over the common support of the propensity score $P(X, Z)$ (Carneiro et al., 2011; Bhuller et al., 2020; Brinch et al., 2017). Panels A and B of Appendix Figure A2 show the propensity score distribution for the removed and non-removed children in the young girl and young boy samples, respectively. The dashed red lines indicate the upper and lower points of the propensity score with common support (after trimming five percent of the sample).

construct a removal tendency measure analogue to our standard instrument in Equation 3.

Panels A and B of Appendix Figure A3 show the MTEs for young girls and young boys, respectively. We use a local IV approach using a global quadratic polynomial specification and construct confidence intervals using 100 bootstrap replications.[53] In addition to reporting the MTE, each panel also reports an estimate of the average treatment effect (ATE). As demonstrated by Heckman and Vytlacil (1999; 2005; 2007) and Heckman et al. (2006), the ATE can be expressed as an average of MTEs. Estimating the ATE for the full population requires full support for the propensity score over the unit interval. In line with Carneiro et al. (2011) and Bhuller et al. (2020), we report estimates of the ATE for the region of common support.

The results in Panel A show that the MTE estimates for test scores are most positive for young girls with low unobserved resistance to treatment. The estimates decrease as the unobserved resistance increases and become negative at the highest quantiles. The decline at the upper levels of resistance suggests that young girls on the margin of placement with the highest removal rate CPIs (who likely have less severe unobserved abuse or neglect cases) benefit less from removal.[54] For young boys, the results in Panel B show that the MTE estimates are usually negative over the region of common support, and the estimates decline with increases in the resistance to treatment. As a robustness exercise for the MTE analysis, we examined MTE estimates using models that exclude covariates. We find that the MTE results for young girls are positive and overall similar regardless of whether covariates are included. In contrast, the MTE estimates for young boys are sensitive to this specification choice, a finding that may not be surprising given the large confidence intervals that we observe in the analysis for young boys.[55] We caution against strong interpretation of the

[53]To estimate the MTEs, the predicted probability of removal is estimated using a probit specification. Note that we conducted robustness checks on the MTE estimates and found similar results when we used linear and cubic specifications.

[54]Table 1A of Heckman and Vytlacil (2005) reports how various treatment effect parameters (e.g., the ATE or the LATE) are different weighted averages of the MTEs. Appendix Figure A4 reports the IV weights for the samples of young girls and young boys. These results show that both young girls and young boys with relatively larger (i.e., more positive) MTEs have larger IV weights.

[55]Appendix Figure A5 reports the MTE results without covariates for young girls and young boys. For young girls, the estimates are consistently positive, and the confidence intervals rule out zero across the unobserved resistance to treatment. For young boys, we find imprecisely estimated MTEs. Relative to the

MTE analysis for young boys.

6 Understanding Gender Differences in the Impact of Removal for Young Children

What explains the pattern of gender differences in the impacts of removal for young children? This section considers three categories of explanations. First, it is possible there are gender differences in the pre-investigation characteristics of compliers that could help determine the effects of removal. Second, removal may have heterogeneous effects on mediating factors such as the type of foster care placement, school mobility or characteristics, or parental behavior. Third, girls and boys may respond differently to the same treatment of removal in early life.

6.1 Complier Characteristics

One possibility is that the compliers among young girls are different in terms of pre-investigation background characteristics relative to their male counterparts. If effects vary by these complier characteristics, this could explain why we observe gendered effects on test scores and the other schooling outcomes such as grade repetition. As noted in Section 4.6, the average characteristics for compliers are generally similar for young girls and young boys except in terms of racial composition. Specifically, the average young girl complier is much more likely to be a minority relative to her male counterparts. Appendix Table A4 shows that the fraction of compliers who are white is only 42.4 percent for young girls compared with 59.8 percent for young boys.

To understand the importance of race in our analysis, Appendix Tables A19 and A20 reports impacts of removal by gender and minority status subgroups (and other case characteristic subgroups). These results do not provide strong evidence that the difference in

MTE results with covariates, the results for boys in Panel B reveal MTE estimates that are more positive but are imprecise across the distribution of unobserved resistance to treatment. Note that the more positive (but not significant) MTE estimates without covariates for young boys are in line with the 2SLS point estimates in Table 4.

the minority share among girl and boy compliers explains the pattern of effects. Although the estimates are not always statistically significant, the results suggest that removal has substantively large and beneficial estimated impacts on test scores and the school index for both non-minority (white) and minority young girls. The results also show that there are no significant impacts of removal for either non-minority (white) and minority young boys. In this case, the sign of the point estimates for young boys differs by subgroup, but these results are imprecise and should be interpreted cautiously.

6.2 Differences in Mediating Factors

As detailed in Section 3, we have extensive measures of mediating factors that could help determine the impact of removal. Specifically, we focus on factors such as types of foster care outcomes associated with the first investigation, school mobility and characteristics, and parent behavior. Our focus is on testing whether there are gender differences in any of these potential mediators.

Table 7 reports impacts of removal on foster care outcomes associated with the first investigation such as the number of days spent in each type of foster care and the likelihood of adoption. The results show little evidence of differences in these post-removal outcomes for young girls (Panel A) and young boys (Panel B). For example, removal has statistically significant and large positive impacts on the days spent in any foster care for both genders. The point estimates are larger for young girls but we fail to reject the hypothesis that these estimated effects equal for young girls and young boys (p-value $= 0.717$).[56,57]

Next, we test for gender differences in the impact of removal on school mobility or the types of schools that children attend. Table 8 provides estimates for impacts on school

[56]The foster care outcomes in Table 7 are based on placement records for the first investigation. Alternatively, we can measure foster care outcomes associated with any subsequent investigation. When we analyze the total days spent in foster care including time from the first and subsequent investigations, we also find statistically significant and large impacts of removal on the first investigation. The point estimates are larger for young girls, but we fail to reject the hypothesis that these estimated effects on total days are equal for young girls and young boys.

[57]By definition, the number of days spent in each type of foster care after the first investigation is zero for children who are not removed from their home.

mobility and several characteristics of schools attended for grades 3-8. The results provide no strong evidence of gendered treatment effects. For young girls and young boys, there are no statistically significant impacts, and we cannot reject the hypothesis that the point estimates are equal for young girls and young boys. That said, an important caveat is that the standard errors are generally large.

As a last test of mediating factors, we examine two outcomes that proxy for changes in behavior for adult household members. First, we study a sample of parent perpetrators of child abuse or neglect and estimate the impact of child removal on their criminal charges and incarceration in the four-year period after an investigation concludes.[58,59] Second, we explore whether removal affects the likelihood of having a *subsequent* CPS investigation or removal in our schooling outcomes sample of children. This outcome could proxy for changes in parental behavior given that more than 95 percent of perpetrators in our main sample are parents. To parallel our analysis of parental criminal outcomes, we measure subsequent CPS investigations within the four-year period after the conclusion of an investigation.

We do not find strong evidence for the hypothesis that differential impacts on parent behavior could mediate the heterogeneous pattern of effect of removal. Appendix Table A21 reports impacts separately for the parent perpetrators of young girls (Panel A) and young boys (Panel B).[60] Column 1 shows that there are no statistically significant impacts of removal on the likelihood that a parent perpetrator is charged or incarcerated for any crime after an investigation concludes. Columns 2-5 also show there are no significant impacts on different types of crime such as property offenses, drug-related offenses, public offenses (e.g., disorderly

[58]Note that all perpetrators in the sample are associated with an investigation where DCYF has substantiated the report of abuse or neglect. The data contain no information on the residence of a perpetrator.

[59]There are at least three reasons why charges and incarceration of parent perpetrators might increase following removal. First, during the hearing and removal decision process, evidence may be uncovered which would trigger an adult criminal charge that results in post-investigation incarceration. Second, the DCYF system could affect reporting behavior because parents must regularly check-in with case management staff (who are not CPIs) as part of a child reunification plan. Third, removal may adversely affect the mental health of perpetrators resulting in changes in criminal behavior.

[60]The unit of analysis is a parent who is listed as a perpetrator. We split the analysis by gender of the investigated child. If a parent is associated with siblings of both genders, they are included in the results for both young girls (Panel A) and young boys (Panel B).

conduct), or sex offenses.[61] Appendix Table A22 reports impacts of removal on subsequent CPS investigations. There are no statistically significant impacts of removal on subsequent investigations or removals, although the point estimates indicate that removal reduces the likelihood of a future investigation for both young girls and young boys.[62]

6.3 Analysis of Siblings

A final explanation that we consider is that girls and boys could respond differently to the same treatment of home removal during early life. This hypothesis is motivated by prior research that finds that biology and social processes drive development advantages for young girls in terms of language, temperament, and socioemotional development (Else-Quest et al., 2006; Zahn-Waxler et al., 2008; Schore, 2017; Magnuson et al., 2016). To test for gender differences in the impact of removal per se, we compare the impacts of removal between brothers and sisters from the same household.

Table 9 reports estimated impacts of removal for the subsample of young children who have opposite sex siblings. In this analysis, the specification is a modified version of Equation 1 that interacts removal with indicators for gender. In this approach, the IV model has two endogenous variables, which are interaction terms for removal and an indicator for being a girl and for removal and an indicator for being a boy. The first stage has two instruments, which are the leave-out measures interacted with gender. The second and first stages both control for the main effects for gender.

The main specifications in Columns 1 and 2 show that, while the results are not precisely estimated, the point estimates for young girls who have siblings are nearly identical to the effects in Table 4 for the main sample. In contrast to these large and positive estimates, the effects for boy siblings are negative. As an additional robustness test, we show that there is a similar pattern of results in Columns 3 and 4 when we restrict the sample to two children

[61] We also examined whether removal had impacts on criminal behavior of fathers in the sample of parent perpetrators. We find no significant effects in this analysis. The results are relatively imprecise due to the fact that fathers are a minority of the parent perpetrators in our sample.

[62] We also estimated the impact of removal on subsequent investigations related to sex abuse allegations and found no statistically significant impacts.

per household by only studying the oldest siblings of the opposite gender in each family. Due to the large standard errors in our estimates, we cannot statistically reject the hypothesis of equal effects for siblings. Although the large size of the estimated coefficients and their lack of statistical significance suggest caution in the interpretation, these results provide limited but suggestive evidence that young girls are more positively affected by home removal than their brothers.

7 Impacts of Removal on Outcomes for Older Children

Finally, we study the effects of removal on the older children who are investigated at ages 6-18. We study post-investigation schooling outcomes and the following (post-investigation) later-life outcomes: having any juvenile court conviction by age 18, graduation from high school by age 19, teen birth, and enrollment in any post-secondary institution by age 22.[63] In contrast to the analysis in Section 5, we study these later-life outcomes only for older children since a child investigated before age six will generally not be old enough to be at risk for a given later-life outcome by the end of the period covered by the data sources.

Appendix Table A23 reports tests of randomization for the sample of older children. These results provide an important caveat for the analysis of impacts of removal for older children. Column 2 shows that we reject the null hypothesis at the one percent level in a joint test of the statistical significance of case characteristics in the sample of older investigated girls.[64] To help assess whether this imbalance threatens the validity of IV estimates for older children, we conduct two tests, which we discuss in detail in Appendix F. First, we find that estimates of the impact of removal are not sensitive to the inclusion of case characteristic controls. This

[63]Details on the sample construction and outcomes are provided in Appendix C. Note that we construct schooling outcomes of older children (i.e., the measures of grade retention, special education participation (IEP), and average absences) using only school year observations that occur after the year that an investigation concludes. Most test score results for older children are based on a sample of children investigated at ages 6-12 because children investigated at later ages will not be enrolled in the testing grades (3-8). Results for teenage parenthood for boys should be interpreted cautiously since information on fathers is available only in 82 percent of the birth records that we use to construct this outcome.

[64]The regression estimates show that older girls who have physical neglect or emergency cases see CPIs who have 1.5 and 1.6 percentage points higher removal tendencies.

provides some reassurance to the extent that observed case characteristics are correlated with unobserved explanatory variables (Altonji et al., 2005). Second, unlike the analysis for young children, we can analyze test scores in the periods *before* an investigation begins for older children. This placebo analysis finds that there are no statistically significant impacts on pre-investigation test scores.

Appendix Table A24 reports estimates for the impact of removal for older girls (Panel A) and older boys (Panel B). Across outcomes, we find no statistically significant impacts of removal for either gender. The estimates are imprecise, and we cannot rule out substantively large positive or negative impacts. For older girls, the point estimates do not consistently point to beneficial impacts. For example, the results suggest removal increases the likelihood of having a teenage birth but improves enrollment in a post-secondary institution. Note that the mixed pattern of point estimates for older girls contrasts with our analysis for young girls where we reassuringly find consistently beneficial impacts of removal. The results for older boys provide some weak but suggestive evidence that removal has detrimental effects in terms of decreases in test scores, increases in adverse school experience, and lower likelihoods of both high-school graduation and post-secondary attendance.

Comparing the estimates for older and younger girls allows us to examine whether the effects of removal are specific to age. For girls investigated at older ages, the estimated impact on average test scores is -0.230 standard deviations. Despite the large standard error associated with this estimate, we can reject the hypothesis that the effects for older and younger girls are equal at the five percent significance level. This pattern is consistent with the literature on the importance of early-life interventions (Cunha et al., 2006; Heckman, 2006; Cunha and Heckman, 2007; Almond and Currie, 2011; Heckman et al., 2013; Bharadwaj et al., 2013; Heckman and Mosso, 2014; Elango et al., 2015; Almond et al., 2018; Chyn et al., 2021).

As a final discussion point, we benchmark our results relative to prior studies of home removal. Using a similar IV approach, Doyle (2007) studied older children investigated at

ages five to 15 in Illinois. He found statistically significant and large positive impacts on teenage pregnancy (29 percentage points) and juvenile delinquency (47 percentage points) for older girls.[65] In our sample, the positive point estimate for teenage pregnancy for older girls is much smaller in magnitude, but the standard error is sufficiently large that we cannot rule out the effect size observed by Doyle (2007).

8 Conclusion

Child protection authorities in the U.S. remove about 200,000 children from their homes annually (U.S. Department of Health and Human Services, 2016). Despite this fact, there is relatively little research on the causal impacts of this policy. This paper provides new evidence on the effects of home removal by using comprehensive administrative data on educational outcomes. We focus on children removed before the age of six and examine heterogeneous effects by gender. Our analysis is motivated by the growing literature showing the importance of early-life interventions (Cunha et al., 2006; Heckman, 2006; Cunha and Heckman, 2007; Almond and Currie, 2011; Heckman et al., 2013; Heckman and Mosso, 2014; Elango et al., 2015; Almond et al., 2018) and differential responses by gender (Heckman et al., 2010; Bertrand and Pan, 2013; Heckman et al., 2013; Elango et al., 2015; Conti et al., 2016; Heckman et al., 2017; Garcia et al., 2018; Autor et al., 2019).

We use the removal tendency of quasi-experimentally assigned CPIs as an instrument for removal and estimate causal effects for children on the margin of home removal. For young girls, we find that removal causes statistically significant and substantial improvements in performance on standardized exams, as well as decreases in grade retention and special education needs. Estimates show similar impacts on test scores starting from the first testing grade and onward, which suggests a permanent change in ability prior to when a young removed girl begins taking exams. We do not find significant positive impacts of removal for

[65] Warburton et al. (2014) also study crime for older investigated children and use an IV strategy based on caseworkers. They find imprecise IV estimates of the impact of foster care placement. Lindquist and Santavirta (2014) provide evidence showing that, among children placed at ages 13-18, foster care is associated with higher crime.

young boys. However, we suggest caution in the interpretation of the effects of removal for this group of children. A key caveat is that the point estimates for young boys are imprecisely estimated, and we cannot rule out large positive or negative impacts. For all results, we show that our main conclusions are robust to several checks, including allowing for heterogeneity in investigator removal tendency by case and child characteristics.

We investigate several potential explanations for the gendered pattern of treatment effects. An analysis of siblings provides limited but suggestive evidence that young girls have larger benefits from removal relative to the effects for their brothers. We find no evidence of notable differences in the complier characteristics of girls and boys, and we find that young children of both genders have similar foster care and school experiences subsequent to removal. Overall, this suggests that the impact of home removal per se varies based on the gender of young children.

Our findings echo prior studies of schooling and social program interventions that find girls respond positively and significantly to interventions aimed at improving educational opportunity or community environment (Hastings et al., 2006; Kling et al., 2007; Angrist and Lavy, 2009; Heckman et al., 2013; Deming et al., 2014; Hoynes et al., 2016). In addition, our finding that increases in academic performance accrue to girls removed before age six contributes to the literature on the importance of early-life conditions (Cunha et al., 2006; Heckman, 2006; Cunha and Heckman, 2007; Almond and Currie, 2011; Heckman et al., 2013; Heckman and Mosso, 2014; Elango et al., 2015; Almond et al., 2018).

Given the prevalence of home removal, we conclude by emphasizing the need for additional research on the impacts of home removal. To the best of our knowledge, we provide the first estimates of the causal impacts of home removal at early ages. Prior work by Doyle (2007; 2008) provides compelling evidence on the causal effects for children removed at older ages. One caveat for our analysis is that many of the estimates of potential mediators are relatively imprecisely estimated. This limits the conclusions we can draw on the mechanisms that drive the effects of removal that we detect. Future research that uses larger samples and

administrative data from other states can help facilitate a more complete understanding of the effects of removal on neglected and abused children.

References

Abadie, Alberto. (2003). 'Semiparametric instrumental variable estimation of treatment response models', *Journal of Econometrics* 113(2), 231–263.

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

Aizer, Anna, Eli, Shari, Ferrie, Joseph and Lleras-Muney, Adriana. (2016). 'The Long-Run Impact of Cash Transfers to Poor Families', *American Economic Review* 106(4), 935–971.

Almond, Douglas and Currie, Janet. (2011), Chapter 15 - Human capital development before age five, *in* **David Card and Orley Ashenfelter and David Card**., eds, 'Handbook of Labor Economics', Vol. 4, Part B, Elsevier, pp. 1315–1486.

Almond, Douglas, Currie, Janet and Duque, Valentina. (2018). 'Childhood Circumstances and Adult Outcomes: Act II', *Journal of Economic Literature* 56(4), 1360–1446.

Almond, Douglas, Doyle, Joseph J., Kowalski, Amanda E. and Williams, Heidi. (2010). 'Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns', *The Quarterly Journal of Economics* 125(2), 591–634.

Altonji, Joseph G., Elder, Todd E. and Taber, Christopher R. (2005). 'Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools', *Journal of Political Economy* 113(1), 151–184.

Anderson, Michael L. (2008). 'Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects', *Journal of the American Statistical Association* 103(484), 1481–1495.

Anderson, T. W. and Rubin, Herman. (1949). 'Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations', *The Annals of Mathematical Statistics* 20(1), 46–63. Publisher: Institute of Mathematical Statistics.

Andresen, Martin Eckhoff. (2018). 'Exploring marginal treatment effects: Flexible estimation using Stata', *The Stata Journal* 18(1), 118–158.

Andrews, Isaiah, Stock, James H. and Sun, Liyang. (2019). 'Weak Instruments in Instrumental Variables Regression: Theory and Practice', *Annual Review of Economics* 11(1), 727–753.

Angrist, Joshua D., Imbens, Guido W. and Rubin, Donald B. (1996). 'Identification of Causal Effects Using Instrumental Variables', *Journal of the American Statistical Association* 91(434), 444–455.

Angrist, Joshua and Lavy, Victor. (2009). 'The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial', *American Economic Review* 99(4), 1384–1414.

Autor, David, Figlio, David, Karbownik, Krzysztof, Roth, Jeffrey and Wasserman, Melanie. (2019). 'Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes', *American Economic Journal: Applied Economics* 11(3), 338–381.

Belloni, Alexandre, Chernozhukov, Victor and Hansen, Christian. (2014). 'High-Dimensional Methods and Inference on Structural and Treatment Effects', *The Journal of Economic Perspectives* 28(2), 29–50.

Benjamini, Yoav, Krieger, Abba M. and Yekutieli, Daniel. (2006). 'Adaptive Linear Step-up Procedures that Control the False Discovery Rate', *Biometrika* 93(3), 491–507.

Berger, Lawrence M., Bruch, Sarah K., Johnson, Elizabeth I., James, Sigrid and Rubin, David. (2009). 'Estimating the "Impact" of Out-of-Home Placement on Child Well-Being: Approaching the Problem of Selection Bias', *Child Development* 80(6), 1856–1876.

Berger, Lawrence M., Cancian, Maria, Han, Eunhee, Noyes, Jennifer and Rios-Salas, Vanessa. (2015). 'Children's Academic Achievement and Foster Care', *Pediatrics* 135(1), 109–116.

Bertrand, Marianne and Pan, Jessica. (2013). 'The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior', *American Economic Journal: Applied Economics* 5(1), 32–64.

Bharadwaj, Prashant, Loken, Katrine and Neilson, Christopher. (2013). 'Early Life Health Interventions and Academic Achievement', *American Economic Review* 103(5), 1862–1891.

Bhuller, Manudeep, Dahl, Gordon B., Loken, Katrine V. and Mogstad, Magne. (2018). 'Intergenerational Effects of Incarceration', *AEA Papers and Proceedings* 108, 234–240.

Bhuller, Manudeep, Dahl, Gordon B., Løken, Katrine V. and Mogstad, Magne. (2020). 'Incarceration, Recidivism, and Employment', *Journal of Political Economy* 128(4), 1269–1324.

Brinch, Christian N., Mogstad, Magne and Wiswall, Matthew. (2017). 'Beyond LATE with a Discrete Instrument', *Journal of Political Economy* 125(4), 985–1039.

Campbell, Frances, Conti, Gabriella, Heckman, James J., Moon, Seong Hyeok, Pinto, Rodrigo, Pungello, Elizabeth and Pan, Yi. (2014). 'Early Childhood Investments Substantially Boost Adult Health', *Science* 343(6178), 1478–1485.

Carneiro, Pedro, Heckman, James J and Vytlacil, Edward J. (2011). 'Estimating Marginal Returns to Education', *American Economic Review* 101(6), 2754–2781.

Chetty, Raj, Friedman, John N. and Rockoff, Jonah E. (2014). 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates', *American Economic Review* 104(9), 2593–2632.

Chetty, Raj, Hendren, Nathaniel and Katz, Lawrence F. (2016). 'The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment', *American Economic Review* 106(4), 855–902.

Chyn, Eric. (2018). 'Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children', *American Economic Review* 108(10), 3028–3056.

Chyn, Eric, Gold, Samantha and Hastings, Justine S. (2021). 'The returns to early-life interventions for very low birth weight children', *Journal of Health Economics* 75(102400).

Conti, Gabriella, Heckman, James J. and Pinto, Rodrigo. (2016). 'The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour', *The Economic Journal* 126(596), F28–F65.

Cornelissen, Thomas, Dustmann, Christian, Raute, Anna and Schönberg, Uta. (2016). 'From LATE to MTE: Alternative methods for the evaluation of policy interventions', *Labour Economics* 41, 47–60.

Cornelissen, Thomas, Dustmann, Christian, Raute, Anna and Schönberg, Uta. (2018). 'Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance', *Journal of Political Economy* 126(6), 2356–2409. Publisher: The University of Chicago Press.

Cunha, Flavio and Heckman, James. (2007). 'The Technology of Skill Formation', *American Economic Review* 97(2), 31–47.

Cunha, Flavio, Heckman, James J., Lochner, Lance and Masterov, Dimitriy V. (2006), Interpreting the Evidence on Life Cycle Skill Formation, *in* **E. Hanushek and F. Welch**., eds, 'Handbook of the Economics of Education', Vol. 1, Elsevier, pp. 697–812.

Currie, Janet, Mueller-Smith, Michael and Rossin-Slater, Maya. (2020). 'Violence while in Utero: The Impact of Assaults During Pregnancy on Birth Outcomes', *The Review of Economics and Statistics* . Forthcoming.

Currie, Janet and Tekin, Erdal. (2012). 'Understanding the Cycle: Childhood Maltreatment and Future Crime', *The Journal of Human Resources* 47(2), 509–549.

Dahl, Gordon B., Kostøl, Andreas Ravndal and Mogstad, Magne. (2014). 'Family Welfare Cultures', *The Quarterly Journal of Economics* 129(4), 1711–1752.

Deming, David J., Hastings, Justine S., Kane, Thomas J. and Staiger, Douglas O. (2014). 'School Choice, School Quality, and Postsecondary Attainment', *American Economic Review* 104(3), 991–1013.

Dobbie, Will, Goldin, Jacob and Yang, Crystal S. (2018). 'The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges', *American Economic Review* 108(2), 201–240.

Dobbie, Will, Grönqvist, Hans, Niknami, Susan, Palme, Mårten and Priks, Mikael. (2018), The Intergenerational Effects of Parental Incarceration, Technical Report w24186, National Bureau of Economic Research.

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

Doyle, 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.

Doyle, Joseph J. (2013). 'Causal effects of foster care: An instrumental-variables approach', *Children and Youth Services Review* 35(7), 1143–1151.

Doyle, Joseph J and Aizer, Anna. (2018). 'Economics of Child Protection: Maltreatment, Foster Care, and Intimate Partner Violence', *Annual Review of Economics* 10, 87–108.

Doyle, Joseph J. and Peters, H. Elizabeth. (2007). 'The Market for Foster Care: An Empirical Study of the Impact of Foster Care Subsidies', *Review of Economics of the Household* 5(4), 329–351.

Doyle, Orla. (2019). 'The First 2,000 Days and Child Skills', *Journal of Political Economy* 128(6), 2067–2122.

Durose, Matthew R, Cooper, Alexia D and Snyder, Howard N. (2014), *Recidivism of prisoners released in 30 states in 2005: Patterns from 2005 to 2010*, Vol. 28, US Department of Justice, Office of Justice Programs, Bureau of Justice.

Elango, Sneha, García, Jorge Luis, Heckman, James J and Hojman, Andrés. (2015), Early Childhood Education, Working Paper 21766, National Bureau of Economic Research.

Else-Quest, Nicole M., Hyde, Janet Shibley, Goldsmith, H. Hill and Hulle, Carol Ann Van. (2006). 'Gender differences in temperament: a meta-analysis.', *Psychological bulletin* 132(1), 33–72.

Eren, Ozkan and Mocan, Naci. (2017), Juvenile Punishment, High School Graduation and Adult Crime: Evidence from Idiosyncratic Judge Harshness, NBER Working Paper 23573, National Bureau of Economic Research, Cambridge, MA.

Fantuzzo, John W. and Mohr, Wanda K. (1999). 'Prevalence and Effects of Child Exposure to Domestic Violence', *The Future of Children* 9(3), 21–32.

Fitzpatrick, Maria D., Benson, Cassandra and Bondurant, Samuel R. (2020), Beyond Reading, Writing, and Arithmetic: The Role of Teachers and Schools in Reporting Child Maltreatment, Technical Report w27033, National Bureau of Economic Research.

Garces, Eliana, Thomas, Duncan and Currie, Janet. (2002). 'Longer-Term Effects of Head Start', *The American Economic Review* 92(4), 999–1012.

Garcia, Jorge Luis, Heckman, James J. and Ziff, Anna L. (2018). 'Gender differences in the benefits of an influential early childhood program', *European Economic Review* 109, 9–22.

Hastings, Justine S. (2019), Fact-Based Policy: How Do State and Local Governments Accomplish It?, Technical report, The Hamilton Project (Brookings Institution).

Hastings, Justine S., Howison, Mark, Lawless, Ted, Ucles, John and White, Preston. (2019), Unlocking Data to Improve Public Policy, preprint, Open Science Framework.

Hastings, Justine S, Kane, Thomas J and Staiger, Douglas O. (2006). 'Gender and Performance: Evidence from School Assignment by Randomized Lottery', *American Economic Review* 96(2), 232–236.

Heckman, James J, Holland, Margaret L, Makino, Kevin K, Pinto, Rodrigo and Rosales-Rueda, Maria. (2017), An Analysis of the Memphis Nurse-Family Partnership Program, NBER Working Paper 23610, National Bureau of Economic Research.

Heckman, James J and Karapakula, Ganesh. (2019*a*), Intergenerational and Intragenerational Externalities of the Perry Preschool Project, Working Paper 25889, National Bureau of Economic Research.

Heckman, James J. and Karapakula, Ganesh. (2019*b*), The Perry Preschoolers at Late Midlife: A Study in Design-Specific Inference, Technical Report w25888, National Bureau of Economic Research.

Heckman, James J., Liu, Bei, Lu, Mai and Zhou, Jin. (2020), Treatment Effects and the Measurement of Skills in a Prototypical Home Visiting Program, Technical Report w27356, National Bureau of Economic Research.

Heckman, James J. and Mosso, Stefano. (2014). 'The Economics of Human Development and Social Mobility', *Annual Review of Economics* 6(1), 689–733.

Heckman, James J, Urzua, Sergio and Vytlacil, Edward. (2006). 'Understanding Instrumental Variables in Models with Essential Heterogeneity', *Review of Economics and Statistics* 88(3), 389–432.

Heckman, James J. and Vytlacil, Edward. (2005). 'Structural Equations, Treatment Effects, and Econometric Policy Evaluation1', *Econometrica* 73(3), 669–738.

Heckman, James J. and Vytlacil, Edward J. (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 Vytlacil, Edward J. (2007), Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation, *in* **James J. Heckman and Edward E. Leamer**., ed., 'Handbook of Econometrics', Vol. 6, Part B, Elsevier, pp. 4779–4874.

Heckman, James, Moon, Seong Hyeok, Pinto, Rodrigo, Savelyev, Peter and Yavitz, Adam. (2010). 'Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program', *Quantitative Economics* 1(1), 1–46.

Heckman, James, Pinto, Rodrigo and Savelyev, Peter. (2013). 'Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes', *American Economic Review* 103(6), 2052–2086.

Heckman, James, Tobias, Justin L. and Vytlacil, Edward. (2001). 'Four Parameters of Interest in the Evaluation of Social Programs', *Southern Economic Journal* 68(2), 211–223.

Heckman, J. J. (2006). 'Skill Formation and the Economics of Investing in Disadvantaged Children', *Science* 312(5782), 1898–1900.

Hegar, Rebecca and Scannapieco, Maria. (1995). 'From Family Duty to Family Policy: The Evolution of Kinship Care', *Child Welfare* 74(1), 201–215.

Holt, Stephanie, Buckley, Helen and Whelan, Sadhbh. (2008). 'The impact of exposure to domestic violence on children and young people: A review of the literature', *Child Abuse & Neglect* 32(8), 797–810.

Hoynes, Hilary, Schanzenbach, Diane Whitmore and Almond, Douglas. (2016). 'Long-Run Impacts of Childhood Access to the Safety Net', *American Economic Review* 106(4), 903–934.

Imbens, Guido W. and Angrist, Joshua D. (1994). 'Identification and Estimation of Local Average Treatment Effects', *Econometrica* 62(2), 467.

Individuals with Disabilities Education Act. (2004), '20 U.S.C. Section 1400'.

Isen, Adam, Rossin-Slater, Maya and Walker, W. Reed. (2017). 'Every Breath You Take—Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970', *Journal of Political Economy* 125(3), 848–902.

Kane, Thomas J., Rockoff, Jonah E. and Staiger, Douglas O. (2008). 'What does certification tell us about teacher effectiveness? Evidence from New York City', *Economics of Education Review* 27(6), 615–631.

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

Kling, Jeffrey R., Liebman, Jeffrey B. and Katz, Lawrence F. (2007). 'Experimental Analysis of Neighborhood Effects', *Econometrica* 75(1), 83–119.

Lindquist, Matthew J. and Santavirta, Torsten. (2014). 'Does placing children in foster care increase their adult criminality?', *Labour Economics* 31, 72–83.

Ludwig, Jens and Miller, Douglas L. (2007). 'Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design', *Quarterly Journal of Economics* .

Magnuson, Katherine A., Kelchen, Robert, Duncan, Greg J., Schindler, Holly S., Shager, Hilary and Yoshikawa, Hirokazu. (2016). 'Do the effects of early childhood education programs differ by gender? A meta-analysis', *Early Childhood Research Quarterly* 36, 521–536.

Mogstad, Magne, Torgovitsky, Alexander and Walters, Christopher. (2019), The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables, Technical Report w25691, National Bureau of Economic Research, Cambridge, MA.

Mueller-Smith, Mike. (2015), The Criminal and Labor Market Impacts of Incarceration. Unpublished.

Rhode Island. (2016), 'Rhode Island General Law. The Rhode Island Board of Education Act, Stat. 16-19-1'.

Rhode Island Department of Children, Youth, and Families. (2018), 'Policy Manual'.

Rhode Island Kids Count. (2016), '2016 Rhode Island Kids Count Factbook'.

Sampat, Bhaven and Williams, Heidi L. (2019). 'How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome', *American Economic Review* 109(1), 203–236.

Schneider, William, Waldfogel, Jane and Brooks-Gunn, Jeanne. (2017). 'The Great Recession and risk for child abuse and neglect', *Children and Youth Services Review* 72, 71–81.

Schore, Allan N. (2017). 'All Our Sons: The Developmental Neurobiology and Neuroendocrinology of Boys at Risk', *Infant Mental Health Journal* 38(1), 15–52.

Staiger, Douglas and Stock, James H. (1997). 'Instrumental Variables Regression with Weak Instruments', *Econometrica* 65(3), 557–586. Publisher: [Wiley, Econometric Society].

Stock, James H, Wright, Jonathan H and Yogo, Motohiro. (2002). 'A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments', *Journal of Business & Economic Statistics* 20(4), 518–529.

Stock, James H and Yogo, Motohiro. (2005), Testing for Weak Instruments in Linear IV Regression, *in* 'Identification and Inference for Eonometric Models: Essays in Honor of Thomas Rothenberg', p. 29.

U.S. Census. (2018), 'S0901: Children Characteristics, 2012-16 American Community Survey 5-year Estimates'.

U.S. Department of Health and Human Services. (2016), 'Child Maltreatment'.

Warburton, William P., Warburton, Rebecca N., Sweetman, Arthur and Hertzman, Clyde. (2014). 'The Impact of Placing Adolescent Males into Foster Care on Education, Income Assistance, and Convictions', *Canadian Journal of Economics* 47(1), 35–69.

Wildeman, Christopher and Waldfogel, Jane. (2014). 'Somebody's Children or Nobody's Children? How the Sociological Perspective Could Enliven Research on Foster Care', *Annual Review of Sociology* 40(1), 599–618.

Wolfe, David A., Crooks, Claire V., Lee, Vivien, McIntyre-Smith, Alexandra and Jaffe, Peter G. (2003). 'The Effects of Children's Exposure to Domestic Violence: A Meta-Analysis and Critique', *Clinical Child and Family Psychology Review* 6(3), 171–187.

Zahn-Waxler, Carolyn, Shirtcliff, Elizabeth A. and Marceau, Kristine. (2008). 'Disorders of Childhood and Adolescence: Gender and Psychopathology', *Annual Review of Clinical Psychology* 4(1), 275–303.

9 Tables and Figures

Table 1: Descriptive Statistics for the DCYF Investigations Sample

		(1)	(2)	(3)	(4)
		Sample: Investigated Young Children (Age < 6)			
		All	Non-removed	Removed	p-value
Demographics	Female	0.460	0.457	0.471	0.178
		(0.498)	(0.498)	(0.499)	
	White	0.589	0.590	0.583	0.450
		(0.492)	(0.492)	(0.493)	
	Black	0.167	0.160	0.196	0.000
		(0.373)	(0.367)	(0.397)	
	Hispanic	0.162	0.170	0.130	0.000
		(0.368)	(0.376)	(0.337)	
	Other race	0.082	0.080	0.092	0.039
		(0.274)	(0.271)	(0.288)	
	Age	1.805	1.978	1.122	0.000
		(1.763)	(1.762)	(1.593)	
Family	Married couple	0.122	0.136	0.068	0.000
		(0.327)	(0.341)	(0.251)	
	Unmarried couple	0.292	0.302	0.251	0.000
		(0.455)	(0.459)	(0.433)	
	Single/other	0.586	0.562	0.682	0.000
		(0.493)	(0.496)	(0.466)	
	English language	0.972	0.970	0.978	0.020
		(0.165)	(0.170)	(0.145)	
	Other language	0.028	0.030	0.022	0.020
		(0.165)	(0.170)	(0.145)	
Allegation	Neglect	0.794	0.811	0.727	0.000
		(0.404)	(0.391)	(0.445)	
	Physical neglect	0.065	0.059	0.088	0.000
		(0.246)	(0.235)	(0.284)	
	Physical abuse	0.141	0.130	0.184	0.000
		(0.348)	(0.336)	(0.388)	
Reporter	Professional	0.824	0.825	0.817	0.272
		(0.381)	(0.380)	(0.387)	
	Family/friend	0.128	0.126	0.136	0.129
		(0.334)	(0.331)	(0.343)	
	Other reporter	0.049	0.049	0.047	0.685
		(0.215)	(0.216)	(0.212)	
Invest. Type	Emergency	0.104	0.054	0.298	0.000
		(0.305)	(0.227)	(0.458)	
	Immediate	0.571	0.606	0.433	0.000
		(0.495)	(0.489)	(0.496)	
	Routine	0.326	0.340	0.269	0.000
		(0.469)	(0.474)	(0.444)	
Post Invest.	Removed	0.203	0.000	1.000	0.000
		(0.402)	(0.000)	(0.000)	
	Days, Foster Care	92.623	0.000	456.570	0.000
		(267.337)	(0.000)	(431.566)	
N		13,674	10,900	2,774	

Notes: This table reports descriptive statistics for young children (investigated before age six). This sample of investigated children is described in Section 3.1. Columns 2-3 report statistics for non-removed and removed children, respectively. Column 4 reports the p-value from a t-test of difference in means for Columns 2-3.

Table 2: First-Stage Results

	(1)	(2)	(3)	(4)
Sample:	Investigations Sample		Schooling Outcomes Sample	
Dependent variable:	Removed (=1)			
CPI removal tendency	0.594***	0.582***	0.649***	0.403***
	(0.096)	(0.069)	(0.166)	(0.113)
Age/gender group	Young Girls	Young Boys	Young Girls	Young Boys
Mean of dependent variable	0.208	0.199	0.181	0.174
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
N (Individuals)	6,287	7,387	2,614	3,142

Notes: This table summarizes the first-stage impact of CPI removal tendency. Columns 1-2 report results for all young children (investigated before age six) included in the investigations sample described in Section 3.1. Columns 3-4 report results for the investigated children who matched to the school test score and enrollment records. Note that the schooling outcomes sample does not include children in the investigations sample who are not be age-eligible to appear in testing grades (3-8) during the school years covered by the data sources. The first-stage results are from a regression of removal on CPI removal tendency, controls for case characteristics, and investigation year fixed effects (FE). Removed is an indicator for home removal at the child's first investigation. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table 3: Tests of Randomization

	(1)	(2)	(3)	(4)
Sample:	Investigations Sample		Schooling Outcomes Sample	
Dependent variable:	CPI Removal Tendency			
Black	-0.001	0.000	-0.004	-0.000
	(0.002)	(0.003)	(0.003)	(0.004)
Hispanic	0.003	0.001	0.002	-0.001
	(0.002)	(0.002)	(0.005)	(0.006)
Other race	0.001	0.004	-0.003	-0.001
	(0.003)	(0.003)	(0.005)	(0.006)
Age	-0.000	0.000	0.000	0.000
	(0.001)	(0.000)	(0.001)	(0.001)
Married couple	-0.002	0.002	-0.005	-0.003
	(0.002)	(0.002)	(0.004)	(0.003)
Unmarried couple	-0.001	-0.001	-0.002	0.000
	(0.002)	(0.001)	(0.003)	(0.003)
English language	-0.002	0.003	0.000	0.001
	(0.004)	(0.005)	(0.007)	(0.007)
Neglect	0.001	-0.002	-0.001	-0.004
	(0.002)	(0.002)	(0.004)	(0.004)
Physical neglect	0.002	0.005*	0.000	0.005
	(0.004)	(0.003)	(0.006)	(0.005)
Professional reporter	-0.004	-0.004	0.002	-0.002
	(0.004)	(0.004)	(0.006)	(0.005)
Family/friend reporter	-0.006	-0.002	-0.004	-0.002
	(0.004)	(0.004)	(0.006)	(0.005)
Emergency investigation	-0.001	0.001	0.001	-0.002
	(0.003)	(0.003)	(0.005)	(0.004)
Immediate investigation	0.001	0.002	0.005*	0.003
	(0.002)	(0.002)	(0.003)	(0.002)
Age/gender group	Young Girls	Young Boys	Young Girls	Young Boys
Chi-squared statistic	10.249	17.679	18.696	16.974
p-value of joint significance	0.673	0.170	0.133	0.201
Mean of CPI removal tendency	0.176	0.180	0.178	0.183
Investigation year FE	Yes	Yes	Yes	Yes
N (Individuals)	6,287	7,387	2,614	3,142

Notes: This table summarizes tests of random case assignment. Columns 1-2 report results for the young children (investigated before age six) included in the investigations sample described in Section 3.1. Columns 3-4 report results for the investigated children who matched to the school test score and enrollment records. The test statistics are from a regression of CPI removal tendency on the set of case characteristics and investigation year fixed effects. The chi-square test-statistic and p-value reported are from a test for joint significance of all variables except investigation year fixed effects. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table 4: Impact of Removal on Test Scores of Young Children

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Young Girls (Age < 6)						
Dependent variable:	Average z-score		Math z-score		Reading z-score	
Removed (= 1)	1.349**	1.367**	1.472***	1.471***	1.232*	1.271**
	(0.600)	(0.567)	(0.596)	(0.561)	(0.647)	(0.615)
Mean of dependent variable	-0.394	-0.394	-0.462	-0.462	-0.328	-0.328
Complier mean if not removed	-1.753	-1.753	-1.854	-1.854	-1.638	-1.638
Case controls	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	14.959	17.696	14.293	16.901	14.984	17.697
N	9,980	9,980	10,006	10,006	10,014	10,014
Individuals	2,614	2,614	2,614	2,614	2,614	2,614
Panel B. Young Boys (Age < 6)						
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score		Math z-score		Reading z-score	
Removed (= 1)	0.158	0.044	0.135	-0.003	0.193	0.102
	(0.594)	(0.562)	(0.584)	(0.574)	(0.644)	(0.601)
Mean of dependent variable	-0.571	-0.571	-0.517	-0.517	-0.630	-0.630
Complier mean if not removed	-0.981	-0.981	-0.931	-0.931	-1.057	-1.057
Case controls	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	10.523	13.999	10.563	13.896	10.761	14.397
N	12,344	12,344	12,385	12,385	12,406	12,406
Individuals	3,142	3,142	3,142	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on test scores for young girls (Panel A) and young boys (Panel B). As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to the school test score and enrollment records. We standardize scores at the grade-year level and construct a yearly panel of tests taken in grades 3-8 during school years 2005-2016. All results are from two-stage least squares models with a leave-out measure of CPI removal tendency as an instrument for removal. Columns 1-2 report impacts for the average of standardized math and reading scores. Columns 3-4 and 5-6 report results for math scores and reading scores, respectively. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table 5: Impact of Removal on Additional Schooling Outcomes of Young Children

	Panel A. Young Girls (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	-0.426**	-0.511*	-4.576	-0.998**
	(0.170)	(0.295)	(5.368)	(0.448)
Mean of dependent variable	0.129	0.280	12.562	0.000
Complier mean if not removed	0.487	0.856	10.610	0.711
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.594	15.039	15.039	15.039
N	2,604	2,614	2,614	2,614
Individuals	2,604	2,614	2,614	2,614
	Panel B. Young Boys (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	-0.040	-0.195	-0.496	-0.152
	(0.273)	(0.290)	(6.610)	(0.513)
Mean of dependent variable	0.173	0.473	12.667	0.000
Complier mean if not removed	0.357	0.946	14.055	0.495
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	12.760	12.518	12.518	12.518
N	3,123	3,142	3,142	3,142
Individuals	3,123	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on schooling outcomes for young girls (Panel A) and young boys (Panel B). As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to the school test score and enrollment records. Columns 1-3 report impacts on measures of whether an investigated child was ever retained, ever participated in special education (i.e., had an IEP), and the average number of days absent during grades K-8, respectively. Column 4 reports results for an index that is constructed from standardized measures of retention, IEP, and absence measures. Note that there are 10 young girls and 19 young boys for whom we cannot measure grade repetition because they are enrolled only in one academic year during our sample period. For these children, we compute the School Index measure using only the IEP and average attendance outcomes. All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table 6: Impact of Removal on School Enrollment and Test-taking

	Panel A. Young Girls (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Enrolled (=1)		Tested (=1)	
Removed (= 1)	0.110	0.087	0.089	0.073
	(0.231)	(0.198)	(0.224)	(0.195)
Mean of dependent variable	0.633	0.633	0.570	0.570
Complier mean if not removed	0.583	0.583	0.537	0.537
Case controls	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	19.653	27.951	19.653	27.951
N	17,164	17,164	17,164	17,164
Individuals	3,971	3,971	3,971	3,971
	Panel B. Young Boys (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Enrolled (=1)		Tested (=1)	
Removed (= 1)	-0.352	-0.337	-0.283	-0.264
	(0.253)	(0.243)	(0.259)	(0.248)
Mean of dependent variable	0.639	0.639	0.564	0.564
Complier mean if not removed	0.932	0.932	0.715	0.715
Case controls	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	26.194	32.476	26.194	32.476
N	21,420	21,420	21,420	21,420
Individuals	4,770	4,770	4,770	4,770

Notes: This table reports results for the impact of removal on public school enrollment and test-taking outcomes for young girls (Panel A) and young boys (Panel B). As described in Section 3.2, the sample for this analysis is the set of investigated children whose birth cohorts made them age-eligible to attend grades 3-8 during the period in which we observe test scores (i.e., the academic years 2005-2016). For this sample, the table reports results from a yearly panel with measures of annual enrollment and test-taking. All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table 7: Impact of Removal on Foster Care Outcomes

	Panel A. Young Girls (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Days in Any Foster Care	Days w/ Relative	Days w/ Foster Family	Days in Group Home	Days in Other Care	Adopted (=1)
Removed (= 1)	342.502***	203.278**	142.962	9.492	-13.230	0.070
	(126.152)	(83.778)	(116.768)	(5.974)	(19.569)	(0.072)
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.039	15.039	15.039	15.039	15.039	15.039
N	2,614	2,614	2,614	2,614	2,614	2,614
Individuals	2,614	2,614	2,614	2,614	2,614	2,614
	Panel B. Young Boys (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Days in Any Foster Care	Days w/ Relative	Days w/ Foster Family	Days in Group Home	Days in Other Care	Adopted (=1)
Removed (= 1)	413.748**	180.357*	128.337	46.612	58.443	0.048
	(207.274)	(99.360)	(150.693)	(34.108)	(38.092)	(0.109)
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	12.518	12.518	12.518	12.518	12.518	12.518
N	3,142	3,142	3,142	3,142	3,142	3,142
Individuals	3,142	3,142	3,142	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on foster care placement outcomes for young girls (Panel A) and young boys (Panel B). As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to test score and enrollment records. All foster care outcomes are associated with the child's first investigation, which implies the means of placement outcomes are zero for non-removed children. Column 1 reports impacts on days in foster care. Column 2-5 split days in foster care into four categories: days spent with relatives, days spent with a foster family (non-relatives), days spent in a group home, and other days spent in foster care, respectively. Column 6 reports impacts on adoption, an indicator for whether the child is adopted upon discharge from foster care. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table 8: Impact of Removal on School Mobility and School-level Characteristics

	Panel A. Young Girls (Age < 6)				
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Moved Schools (=1)	Value-Added	Avg. Test Scores	% Black	% IEP
Removed (= 1)	-0.153	0.053	0.275	0.003	-0.052
	(0.126)	(0.040)	(0.208)	(0.064)	(0.038)
Mean of dependent variable	0.351	-0.046	-0.115	0.112	0.165
Complier mean if not removed	0.427	-0.102	-0.483	0.146	0.240
Case controls	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	16.177	16.016	16.016	16.061	16.061
N	10,519	10,574	10,574	10,602	10,602
Individuals	2,604	2,614	2,614	2,614	2,614
	Panel B. Young Boys (Age < 6)				
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Moved Schools (=1)	Value-Added	Avg. Test Scores	% Black	% IEP
Removed (= 1)	-0.118	-0.039	-0.377	0.053	-0.039
	(0.141)	(0.038)	(0.275)	(0.064)	(0.082)
Mean of dependent variable	0.365	-0.050	-0.152	0.124	0.182
Complier mean if not removed	0.439	-0.024	-0.072	0.153	0.267
Case controls	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.464	16.123	16.123	16.179	16.179
N	13,112	13,207	13,207	13,230	13,230
Individuals	3,125	3,141	3,141	3,142	3,142

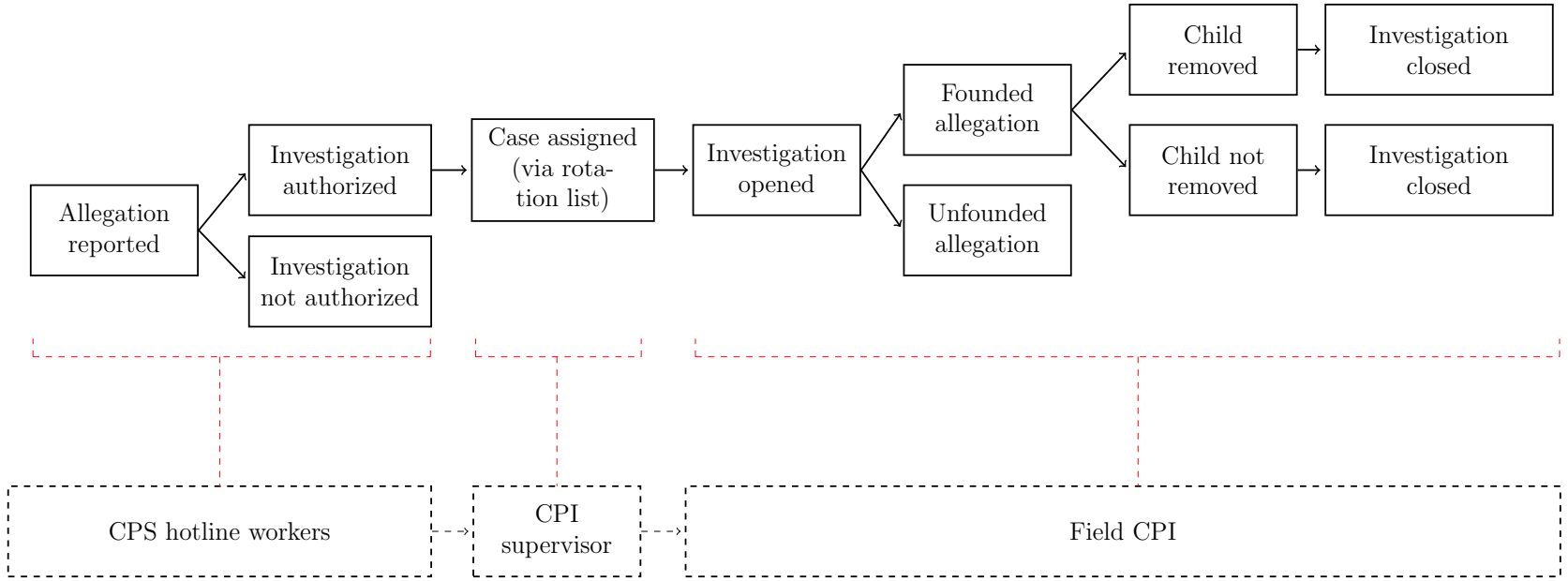
Notes: This table reports results for the impact of removal on school mobility and school-level characteristics for young girls (Panel A) and young boys (Panel B). As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to test score and enrollment records. All measures are based on a panel of observations covering grades 3-8. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table 9: Impact of Removal on Test Scores of Young Siblings

	(1)	(2)	(3)	(4)
Dependent variable:	Average z-score			
Removed (= 1) × Female	1.316	1.193	1.070	0.861
	(1.092)	(0.901)	(1.069)	(0.849)
Removed (= 1) × Male	-0.351	-0.471	-0.516	-0.669
	(0.844)	(0.778)	(0.915)	(0.857)
Sample	All	All	Oldest	Oldest
Mean of dependent variable				
Female	-0.498	-0.500	-0.507	-0.507
Male	-0.675	-0.675	-0.665	-0.665
Complier mean if not removed				
Female	-1.400	-1.400	-1.253	-1.253
Male	0.135	0.135	-0.231	-0.231
Case controls	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	3.333	4.168	3.615	4.531
N	5,546	5,546	4,764	4,764
Individuals	1,342	1,342	1,155	1,155

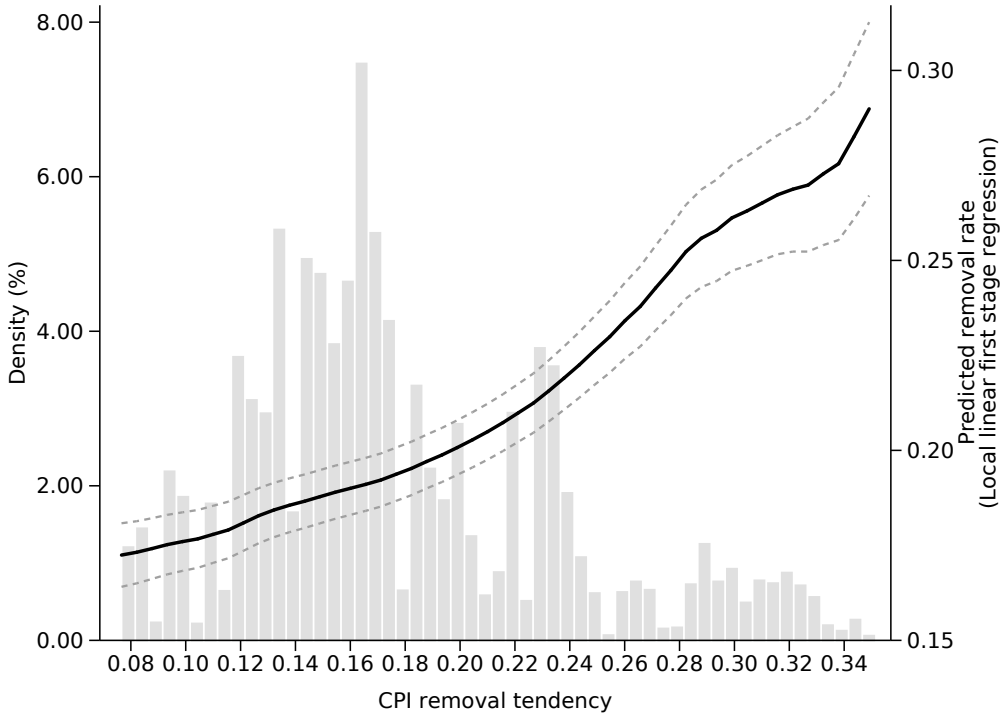
Notes: This table reports results for the impact of removal on test score outcomes for the subset of children in the schooling outcomes sample who are opposite sex siblings. Results are based on estimating IV models where there are two endogenous variables which are interactions between removal status and gender dummy variables. The first-stage has two instruments which are the leave-out measures interacted with the same gender dummy variables. Columns 1-2 report impacts using all young siblings. Columns 3-4 report impacts using a sample that only includes the oldest (below age six) opposite sex siblings in the young children sample. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Figure 1: DCYF Process for Abuse and Neglect Allegations



Notes: This figure illustrates the process by which an allegation of abuse or neglect is processed by DCYF in Rhode Island. See Section 2 for further details.

Figure 2: CPI Removal Tendency



Notes: This figure reports the distribution of CPI removal tendency for the sample of young children investigated by DCYF. Section 4 describes how the measure is constructed. The total number of children is 13,674, and the number of unique CPIs is 102.

A Appendix Tables and Figures

Table A1: Exclusion Restriction Tests

	Panel A. Removed Young Girls (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Days in Any Foster care	Number of Placements	Placed with Relative (=1)	Police Notified (=1)
CPI removal tendency	-339.845	0.981	-0.192	0.071
	(266.559)	(0.989)	(0.262)	(0.110)
Mean of dependent variable	456.115	2.069	0.366	0.958
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
N	1,307	1,307	1,307	1,307
	Panel B. Removed Young Boys (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Days in Any Foster care	Number of Placements	Placed with Relative (=1)	Police Notified (=1)
CPI removal tendency	-81.710	1.051	-0.294	-0.057
	(220.454)	(1.052)	(0.255)	(0.091)
Mean of dependent variable	456.975	2.229	0.355	0.965
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
N	1,467	1,467	1,467	1,467

Notes: The table reports regression results testing whether placement and other investigation outcomes of removed children are correlated with CPI removal tendency for young girls (Panel A) and young boys (Panel B). The sample for this analysis is the set of removed children in the investigations sample described in Section 3.1. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A2: First-Stage Impact of CPI Removal Tendency, by Subgroup

	(1)	(2)
Dependent variable:	Removed (=1)	
White	0.512***	0.636***
	(0.121)	(0.092)
	[0.207]	[0.195]
	N=3,686	N=4,366
Black	0.408	0.278
	(0.276)	(0.278)
	[0.239]	[0.236]
	N=1,071	N=1,217
Hispanic	0.879***	0.469**
	(0.237)	(0.225)
	[0.168]	[0.159]
	N=993	N=1,220
Married couple	0.710***	0.744***
	(0.244)	(0.199)
	[0.116]	[0.110]
	N=713	N=952
Unmarried couple	0.621***	0.606***
	(0.193)	(0.155)
	[0.178]	[0.170]
	N=1,888	N=2,104
Single/other	0.587***	0.544***
	(0.115)	(0.115)
	[0.241]	[0.232]
	N=3,686	N=4,331
Neglect	0.537***	0.566***
	(0.118)	(0.087)
	[0.190]	[0.182]
	N=5,097	N=5,764
Physical abuse	0.979***	0.757***
	(0.309)	(0.226)
	[0.282]	[0.253]
	N=815	N=1,111
Professional reporter	0.610***	0.620***
	(0.107)	(0.074)
	[0.206]	[0.197]
	N=5,142	N=6,120
Family/friend reporter	0.713**	0.485**
	(0.276)	(0.245)
	[0.214]	[0.219]
	N=832	N=914
Immediate	0.858***	0.792***
	(0.128)	(0.098)
	[0.156]	[0.152]
	N=3,491	N=4,312
Routine	0.281*	0.343**
	(0.161)	(0.137)
	[0.177]	[0.159]
	N=2,154	N=2,298
Sample	Young Girls	Young Boys
Case controls	Yes	Yes
Investigation year FE	Yes	Yes

Notes: This table summarizes the first-stage relationship between removal and CPI removal tendency for subgroups. Subgroups are based on the characteristics listed in Table 1. The subgroups for physical neglect, other reporter, and emergency cases are not reported because these have relatively few observations. We also omit reporting results based on language since 97 percent of cases are English language. This analysis uses the investigations sample described in Section 3.1. Standard errors in parentheses are two-way clustered at the family and CPI level. Means for removal for each subgroup are reported in brackets. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A3: First-Stage Impact of CPI Removal Tendency, Reverse Sample Calculation for Subgroups

	(1)	(2)
Dependent variable:	Removed (=1)	
White	0.355***	0.407***
	(0.086)	(0.081)
	[0.207]	[0.196]
	N=3,678	N=4,347
Black	0.399	0.233
	(0.265)	(0.211)
	[0.239]	[0.236]
	N=1,071	N=1,217
Hispanic	0.862***	0.279
	(0.236)	(0.217)
	[0.168]	[0.158]
	N=993	N=1,219
Married couple	0.529***	0.636***
	(0.223)	(0.185)
	[0.116]	[0.109]
	N=713	N=950
Unmarried couple	0.582***	0.573***
	(0.173)	(0.133)
	[0.178]	[0.170]
	N=1,887	N=2,101
Single/other	0.614***	0.437***
	(0.142)	(0.130)
	[0.241]	[0.232]
	N=3,662	N=4,302
Neglect	0.347***	0.353***
	(0.093)	(0.088)
	[0.190]	[0.182]
	N=5,066	N=5,710
Physical abuse	0.913**	0.743***
	(0.287)	(0.221)
	[0.282]	[0.253]
	N=815	N=1,111
Professional reporter	0.326***	0.275***
	(0.092)	(0.078)
	[0.205]	[0.196]
	N=5,081	N=6,038
Family/friend reporter	0.632**	0.505*
	(0.277)	(0.245)
	[0.214]	[0.218]
	N=832	N=913
Immediate	0.545***	0.433***
	(0.135)	(0.110)
	[0.155]	[0.152]
	N=3,474	N=4,291
Routine	0.224	0.307**
	(0.127)	(0.113)
	[0.177]	[0.159]
	N=2,152	N=2,295
Sample	Young Girls	Young Boys
Case controls	Yes	Yes
Investigation year FE	Yes	Yes

Notes: This table summarizes the first-stage relationship between removal and CPI removal tendency for different subgroups. The instrument is recalculated for each subgroup with its complement ("reverse" sample definition). Subgroups are based on the characteristics listed in Table 1. The subgroups for physical neglect, other reporter, and emergency cases are not reported because these have relatively few observations. We also omit reporting results based on language since 97 percent of cases are English language. This analysis uses the investigations sample described in Section 3.1. Standard errors in parentheses are two-way clustered at the family and CPI level. Means for removal for each subgroup are reported in brackets. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A4: Characteristics of Compliers for Young Girls and Young Boys

		Young Girls (Age < 6)		Young Boys (Age < 6)	
		(1)	(2)	(3)	(4)
		$P(X = x)$	$P(X = x\|complier)$	$P(X = x)$	$P(X = x\|complier)$
Demographics	White	0.584	0.424	0.580	0.598
		(0.008)	(0.095)	(0.008)	(0.085)
	Black	0.171	0.183	0.166	0.109
		(0.006)	(0.080)	(0.005)	(0.067)
	Hispanic	0.177	0.307	0.189	0.197
		(0.006)	(0.075)	(0.006)	(0.066)
	Other race	0.067	0.099	0.065	0.101
		(0.004)	(0.048)	(0.004)	(0.047)
Family	Married couple	0.131	0.093	0.148	0.208
		(0.006)	(0.053)	(0.005)	(0.060)
	Unmarried couple	0.245	0.209	0.232	0.211
		(0.007)	(0.075)	(0.006)	(0.076)
	Single/other	0.624	0.709	0.620	0.591
		(0.008)	(0.084)	(0.007)	(0.078)
	English language	0.973	0.991	0.970	0.955
		(0.003)	(0.026)	(0.003)	(0.033)
	Other language	0.027	-0.001	0.030	0.043
		(0.003)	(0.027)	(0.002)	(0.029)
Allegation	Neglect	0.816	0.872	0.783	0.735
		(0.006)	(0.077)	(0.004)	(0.056)
	Physical neglect	0.070	0.059	0.077	0.054
		(0.004)	(0.048)	(0.004)	(0.056)
	Physical abuse	0.114	0.097	0.140	0.217
		(0.005)	(0.065)	(0.005)	(0.060)
Reporter	Professional	0.787	0.834	0.797	0.855
		(0.007)	(0.080)	(0.006)	(0.076)
	Family/friend	0.153	0.164	0.143	0.158
		(0.006)	(0.071)	(0.005)	(0.062)
	Other reporter	0.060	0.020	0.060	0.012
		(0.004)	(0.042)	(0.003)	(0.033)
Investigation	Emergency	0.093	0.126	0.101	0.080
		(0.004)	(0.057)	(0.004)	(0.069)
	Immediate	0.537	0.640	0.565	0.751
		(0.008)	(0.094)	(0.007)	(0.096)
	Routine	0.370	0.284	0.334	0.209
		(0.008)	(0.095)	(0.007)	(0.087)

Notes: This table reports the characteristics of compliers in the schooling outcomes sample described in Section 3.2. We define compliers as children whose removal decision would have been different had they been assigned the most strict versus the most lenient investigator. To identify compliers, we follow Abadie (2003), Dahl et al. (2014), and Dobbie, Goldin and Yang (2018). Let $\overline{z}$ denote the maximum value of the instrument (the most strict investigator) and $\underline{z}$ denote the minimum value of the instrument (the most lenient investigator). We can then express the share of compliers in our sample as: $p_c = Pr(Removed = 1|Z_i = \overline{z}) - Pr(Removed = 1|Z_i = \underline{z})$. In practice, we assign the top percentile of our instrument to $\overline{z}$ and the bottom percentile of our instrument to $\underline{z}$. As discussed in Dahl et al. (2014) and Dobbie, Goldin and Yang (2018), the share of compliers can be directly estimated as $p_c = \alpha$, where α is the coefficient on the instrument from the first stage regression (Equation 2). In this table, we report the average of a given characteristic (listed in each row) for compliers. See Appendix D for details. Standard errors in parentheses are obtained using 500 bootstrap replications.

Table A5: Impact of Removal on All Schooling Outcomes of Young Children, Pooled Sample Results

	All Young Children (Age < 6)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Average z-score	Math z-score	Reading z-score	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	0.739*	0.777**	0.701*	-0.226	-0.356*	-2.590	-0.577*
	(0.389)	(0.394)	(0.416)	(0.158)	(0.210)	(4.222)	(0.338)
Mean of dependent variable	-0.492	-0.493	-0.495	0.153	0.386	12.620	0.000
Complier mean if not removed	-1.281	-1.402	-1.163	0.397	0.775	12.029	0.616
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	27.704	26.977	28.038	26.390	24.917	24.917	24.917
N	22,324	22,391	22,420	5,727	5,756	5,756	5,756
Individuals	5,756	5,756	5,756	5,727	5,756	5,756	5,756

Notes: This table reports results for the impact of removal on schooling outcomes in a pooled sample (i.e., includes young girls and young boys). As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to the school test score and enrollment records. All results are from two-stage least squares models with a leave-out measure of CPI removal tendency as an instrument for removal. Columns 1-3 report impacts for the average of standardized math and reading scores, math, and reading scores, respectively. Columns 4-6 report impacts for whether the child was ever retained, ever participated in special education (i.e., had an IEP), and the average number of days absent during grades K-8, respectively. Column 7 reports results for an index that is constructed from standardized measures of retention, IEP, and absences. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A6: Impact of Removal on Test Scores of Young Children (Cohort Fixed Effects)

	Panel A. Young Girls (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score		Math z-score		Reading z-score	
Removed (= 1)	1.327**	1.356**	1.449***	1.459***	1.213*	1.213*
	(0.598)	(0.571)	(0.595)	(0.564)	(0.647)	(0.647)
Mean of dependent variable	-0.394	-0.394	-0.462	-0.462	-0.328	-0.328
Complier mean if not removed	-1.753	-1.753	-1.854	-1.854	-1.638	-1.638
Case controls	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	16.327	17.495	15.521	16.682	17.490	16.352
N	9,980	9,980	10,006	10,006	10,014	10,014
Individuals	2,614	2,614	2,614	2,614	2,614	2,614
	Panel B. Young Boys (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score		Math z-score		Reading z-score	
Removed (= 1)	0.131	0.037	0.089	-0.026	0.185	0.110
	(0.568)	(0.552)	(0.558)	(0.562)	(0.614)	(0.591)
Mean of dependent variable	-0.571	-0.571	-0.517	-0.517	-0.630	-0.630
Complier mean if not removed	-0.981	-0.981	-0.931	-0.931	-1.057	-1.057
Case controls	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	13.494	14.249	13.507	14.152	13.821	14.616
N	12,344	12,344	12,385	12,385	12,406	12,406
Individuals	3,142	3,142	3,142	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on schooling outcomes for young girls (Panel A) and young boys (Panel B). This robustness table includes cohort (year of birth) fixed effects in all specifications. As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to the school test score and enrollment records. We standardize scores at the grade-year level and construct a yearly panel of tests taken in grades 3-8 during school years 2005-2016. All results are from two-stage least squares models with a leave-out measure of CPI removal tendency as an instrument for removal. Columns 1-2 report impacts for the average of standardized math and reading scores. Columns 3-4 and 5-6 report results for reading scores and math scores, respectively. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A7: Impact of Removal on Additional Schooling Outcomes of Young Children (No Case Controls)

	Panel A. Young Girls (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	-0.403**	-0.496*	-3.493	-0.920**
	(0.165)	(0.299)	(5.503)	(0.439)
Mean of dependent variable	0.129	0.280	12.562	0.000
Complier mean if not removed	0.487	0.856	10.610	0.711
Case controls	No	No	No	No
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	13.539	12.839	12.839	12.839
N	2,604	2,614	2,614	2,614
Individuals	2,604	2,614	2,614	2,614
	Panel B. Young Boys (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	-0.005	-0.178	-1.149	-0.133
	(0.290)	(0.287)	(7.124)	(0.537)
Mean of dependent variable	0.173	0.473	12.667	0.000
Complier mean if not removed	0.357	0.946	14.055	0.495
Case controls	No	No	No	No
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	9.659	9.456	9.456	9.456
N	3,123	3,142	3,142	3,142
Individuals	3,123	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on schooling outcomes for young girls (Panel A) and young boys (Panel B). This robustness table omits controls for case characteristics in all specifications. As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to the school test score and enrollment records. Columns 1-3 report impacts on measures of whether an investigated child was ever retained, ever participated in special education (i.e., has an IEP), and the average number of days absent during grades 3-8. Column 4 reports results for an index that is constructed from standardized measures of the retention, IEP, and absence measures. All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$.

Table A8: Impact of Removal on Additional Schooling Outcomes of Young Children (Cohort Fixed Effects)

Panel A. Young Girls (Age < 6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Ever Retained (=1)		Ever IEP (=1)		Avg. Absences		School Index	
Removed (= 1)	-0.393**	-0.418**	-0.512*	-0.527*	-3.945	-4.700	-0.922**	-1.008**
	(0.165)	(0.171)	(0.301)	(0.303)	(5.349)	(5.318)	(0.435)	(0.453)
Mean of dependent variable	0.129	0.129	0.280	0.280	12.562	12.562	0.000	0.000
Complier mean if not removed	0.487	0.487	0.856	0.856	10.610	10.610	0.711	0.711
Case controls	No	Yes	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.528	15.314	14.816	14.771	14.816	14.771	14.816	14.771
N	2,604	2,604	2,614	2,614	2,614	2,614	2,614	2,614
Individuals	2,604	2,604	2,614	2,614	2,614	2,614	2,614	2,614

Panel B. Young Boys (Age < 6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Ever Retained (=1)		Ever IEP (=1)		Avg. Absences		School Index	
Removed (= 1)	-0.053	-0.075	-0.170	-0.185	-0.976	-0.425	-0.164	-0.174
	(0.259)	(0.259)	(0.285)	(0.285)	(6.742)	(6.443)	(0.502)	(0.492)
Mean of dependent variable	0.173	0.173	0.473	0.473	12.667	12.667	0.000	0.000
Complier mean if not removed	0.357	0.173	0.946	0.473	14.055	14.055	0.495	0.495
Case controls	No	Yes	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	13.250	13.250	11.928	12.904	11.928	12.904	11.928	12.904
N	3,123	3,123	3,142	3,142	3,142	3,142	3,142	3,142
Individuals	3,123	3,123	3,142	3,142	3,142	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on schooling outcomes for young girls (Panel A) and young boys (Panel B). This robustness table includes cohort (year of birth) fixed effects in all specifications. As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to the school test score and enrollment records. Columns 1-6 report impacts on measures of whether an investigated child was ever retained, ever participated in special education (i.e., has an IEP), and the average number of days absent during grades 3-8. Columns 7-8 report results for an index that is constructed from standardized measures of the retention, IEP, and absence measures. All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A9: Impact of Removal on Additional Schooling Outcomes of Young Children (Grades 3-8)

	Panel A. Young Girls (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	-0.222*	-0.644**	-3.669	-0.997**
	(0.116)	(0.309)	(6.004)	(0.484)
Mean of dependent variable	0.004	0.280	12.562	0.000
Complier mean if not removed	0.274	0.856	10.610	0.696
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.594	15.039	14.873	15.039
N	2,604	2,614	2,613	2,614
Individuals	2,604	2,614	2,613	2,614
	Panel B. Young Boys (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	-0.190	-0.316	-5.703	-0.666
	(0.147)	(0.308)	(7.148)	(0.492)
Mean of dependent variable	0.063	0.418	12.408	0.000
Complier mean if not removed	0.350	0.804	19.934	0.897
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	12.799	12.518	12.518	12.518
N	3,120	3,142	3,142	3,142
Individuals	3,120	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on schooling outcomes for young girls (Panel A) and young boys (Panel B). This robustness table defines additional schooling outcomes based on data from grades 3-8 (rather than grades K-8). As described in Section 3.2, the sample for this analysis is the set of investigated children who were matched to the school test score and enrollment records. Columns 1-3 report impacts on measures of whether an investigated child was ever retained, ever participated in special education (i.e., has an IEP), and the average number of days absent during grades 3-8, respectively. Column 4 reports results for an index that is constructed from standardized measures of the retention, IEP, and absence measures. All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A10: Adjusted p-values for Impact of Removal on Outcomes of Young Children

Panel A. Young Girls (Age < 6)			
	(1)	(2)	(3)
Dependent variable (below):	2SLS Estimate	p-value	FDR q-value
Average z-score	1.367**	0.016	0.064
	(0.567)		
Ever Retained (=1)	-0.426**	0.012	0.064
	(0.170)		
Ever IEP (=1)	-0.511*	0.083	0.223
	(0.295)		
Avg. Absences	-4.576	0.394	0.789
	(5.368)		
Panel B. Young Boys (Age < 6)			
	(1)	(2)	(3)
Dependent variable (below):	2SLS Estimate	p-value	FDR q-value
Average z-score	0.044	0.938	0.941
	(0.562)		
Ever Retained (=1)	-0.040	0.884	0.941
	(0.273)		
Ever IEP (=1)	-0.195	0.501	0.803
	(0.290)		
Avg. Absences	-0.496	0.940	0.941
	(6.610)		

Notes: This table reports adjusted p-values for the impact of removal on outcomes of young children. Column 1 of Panels A and B reproduce the results for young girls and young boys from Tables 4 and 5. Columns 2 and 3 report per-comparison (pairwise) and false discovery rate (FDR) adjusted p-values ("q-values"). The adjustment takes into account the fact that we tested the four listed outcomes for the gender subgroup. The FDR-adjusted p-values control for the number of false positives when multiple hypotheses are tested. These adjusted p-values are calculated using the two-step procedure in Benjamini et al. (2006).

Table A11: Test Score Results and Robustness to Changes in Sample Definition

	Panel A. Young Girls (Age < 6)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Average z-score						
Removed (= 1)	1.367**	1.356**	1.394**	0.981	1.232*	1.253**	1.004*
	(0.567)	(0.571)	(0.711)	(0.938)	(0.645)	(0.528)	(0.554)
Sample	Main sample	CPI > 100 cases	CPI > 200 cases	CPI > 300 cases	With sex cases	Ages 0-4	Ages 0-6
Mean of dependent variable	-0.394	-0.394	-0.402	-0.392	-0.384	-0.386	-0.388
Complier mean if not removed	-1.753	-1.603	-1.662	-1.741	-1.687	-1.746	-1.530
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	17.696	18.187	13.889	5.350	12.068	15.582	19.138
N	9,980	9,520	7,799	4,744	10,639	8,321	11,831
Individuals	2,614	2,496	2,060	1,388	2,770	2,196	3,089
	Panel B. Young Boys (Age < 6)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Average z-score						
Removed (= 1)	0.044	0.044	-0.482	-1.183	-0.148	-0.003	0.076
	(0.562)	(0.604)	(0.804)	(1.427)	(0.562)	(0.630)	(0.441)
Sample	Main sample	CPI > 100 cases	CPI > 200 cases	CPI > 300 cases	With sex cases	Ages 0-4	Ages 0-6
Mean of dependent variable	-0.571	-0.572	-0.569	-0.577	-0.567	-0.585	-0.572
Complier mean if not removed	-0.981	-1.070	-0.931	-1.187	-0.772	-0.925	-0.944
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	13.999	11.903	9.839	3.125	15.482	9.452	21.051
N	12,344	11,743	9,719	5,694	12,690	10,190	14,601
Individuals	3,142	2,987	2,469	1,588	3,217	2,619	3,718

Notes: This table reports results for the impact of removal on the average of standardized test scores for young girls (Panel A) and young boys (Panel B). For comparison, Column 1 reproduces estimates from our main sample and preferred specification (as reported in Table 3). Columns 2-7 report results using alternative samples. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$.

Table A12: School Index Results and Robustness to Changes in Sample Definition

	Panel A. Young Girls (Age < 6)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	School Index						
Removed (= 1)	-0.998**	-0.877**	-1.247**	-0.291	-1.180**	-0.831*	-0.787**
	(0.448)	(0.435)	(0.631)	(0.901)	(0.512)	(0.442)	(0.361)
Sample	Main sample	CPI > 100 cases	CPI > 200 cases	CPI > 300 cases	With sex cases	Ages 0-4	Ages 0-6
Mean of dependent variable	0.000	-0.003	0.002	-0.028	-0.007	-0.003	-0.005
Complier mean if not removed	0.711	0.558	0.897	-0.156	0.926	0.652	0.576
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.039	15.004	9.700	2.981	11.549	12.881	18.181
N	2,614	2,496	2,060	1,388	2,770	2,196	3,089
Individuals	2,614	2,496	2,060	1,388	2,770	2,196	3,089
	Panel B. Young Boys (Age < 6)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	School Index						
Removed (= 1)	-0.152	-0.031	0.528	-0.413	-0.162	-0.535	-0.112
	(0.513)	(0.511)	(0.547)	(0.730)	(0.507)	(0.600)	(0.395)
Sample	Main sample	CPI > 100 cases	CPI > 200 cases	CPI > 300 cases	With sex cases	Ages 0-4	Ages 0-6
Mean of dependent variable	0.000	-0.004	0.001	-0.011	-0.002	0.002	0.001
Complier mean if not removed	0.495	0.363	-0.032	0.715	0.507	0.900	0.344
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	12.518	13.126	14.551	5.355	13.192	7.651	20.987
N	3,142	2,987	2,469	1,633	3,217	2,619	3,718
Individuals	3,142	2,987	2,469	1,633	3,217	2,619	3,718

Notes: This table reports results for the impact of removal on schooling outcomes for young girls (Panel A) and young boys (Panel B). For comparison, Column 1 reproduces estimates from our main sample and preferred specification (as reported in Table 3). Columns 2-7 report results using alternative samples. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$.

Table A13: Test Score Results and Robustness Tests Using Flexible Measures of CPI Removal Tendency

	Panel A. Young Girls (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score					
Removed (= 1)	1.367**	1.032**	0.861	1.814	1.531**	1.219**
	(0.567)	(0.466)	(0.546)	(1.116)	(0.682)	(0.509)
IV version	Main	Varies by Gender	Varies by Minority	Varies by Marital Status	Varies by Allegation Type	Varies by Reporter
Mean of dependent variable	-0.394	-0.395	-0.394	-0.395	-0.393	-0.393
Complier mean if not removed	-1.753	-1.404	-0.799	-0.679	-1.542	-1.333
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	17.696	18.373	18.850	7.021	11.510	16.625
N	9,980	9,957	9,953	9,890	9,743	9,809
Individuals	2,614	2,610	2,606	2,594	2,539	2,572
	Panel B. Young Boys (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score					
Removed (= 1)	0.044	0.378	-0.523	-0.064	-0.377	-1.187
	(0.562)	(0.621)	(0.635)	(0.704)	(0.707)	(0.699)
IV version	Main	Varies by Gender	Varies by Minority	Varies by Marital Status	Varies by Allegation Type	Varies by Reporter
Mean of dependent variable	-0.571	-0.571	-0.572	-0.574	-0.572	-0.570
Complier mean if not removed	-0.981	-1.306	0.154	0.144	-0.710	-1.235
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	13.999	11.722	11.781	4.706	10.136	7.560
N	12,344	12,319	12,282	12,204	11,994	12,105
Individuals	3,142	3,134	3,125	3,104	3,046	3,076

Notes: This table reports results on test scores for young girls (Panel A) and young boys (Panel B) based on an IV approach where the CPI removal rate varies with case characteristics. Column 2 allows CPI removal rate to vary by gender. Column 3 allows CPI removal rate to vary by ethnicity/race (i.e., non-minority (white) and minority children). Column 4 allows CPI removal rate across three types of household marital status (married couples, unmarried couples, and single/other households). Column 5 allows CPI removal rate to vary by allegation types (neglect, physical neglect, and physical abuse). Column 6 allows CPI removal rate to vary by the type of reporter. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A14: School Index Results and Robustness Tests Using Flexible Measures of CPI Removal Tendency

	Panel A. Young Girls (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6
Dependent variable:	School Index					
Removed (= 1)	-0.998**	-0.747*	-0.858**	-1.452**	-1.437**	-1.114**
	(0.448)	(0.387)	(0.399)	(0.716)	(0.598)	(0.479)
IV version	Main	Varies by Gender	Varies by Minority	Varies by Marital Status	Varies by Allegation Type	Varies by Reporter
Mean of dependent variable	0.000	0.000	0.001	0.000	-0.001	-0.004
Complier mean if not removed	0.711	0.575	0.474	0.340	0.978	0.849
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.039	15.936	20.202	8.198	12.235	10.995
N	2,614	2,610	2,606	2,594	2,539	2,572
Individuals	2,614	2,610	2,606	2,594	2,539	2,572
	Panel B. Young Boys (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	School Index					
Removed (= 1)	-0.152	-0.317	-0.066	-0.313	-0.096	0.020
	(0.513)	(0.512)	(0.459)	(0.751)	(0.588)	(0.647)
IV version	Main	Varies by Gender	Varies by Minority	Varies by Marital Status	Varies by Allegation Type	Varies by Reporter
Mean of dependent variable	0.000	0.000	0.001	-0.001	0.000	0.000
Complier mean if not removed	0.495	0.634	0.251	-0.200	0.282	0.380
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	12.518	12.410	12.310	4.565	8.713	6.149
N	3,142	3,134	3,125	3,104	3,046	3,076
Individuals	3,142	3,134	3,125	3,104	3,046	3,076

Notes: This table reports results for impacts on the schooling index outcome for young girls (Panel A) and young boys (Panel B) based on an IV approach where the CPI removal rate varies with case characteristics. Column 2 allows CPI removal rate to vary by gender. Column 3 allows CPI removal rate to vary by ethnicity/race (i.e., non-minority (white) and minority children). Column 4 allows CPI removal rate across three types of household marital status (married couples, unmarried couples, and single/other households). Column 5 allows CPI removal rate to vary by allegation types (neglect, physical neglect, and physical abuse). Column 6 allows CPI removal rate to vary by the type of reporter. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A15: Test Scores Results and Robustness Tests Using Alternative Instruments

	Panel A. Young Girls (Age < 6)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Average z-score							
Removed (= 1)	1.367**	1.495**	1.053**	1.030**	1.386***	1.522***	1.126***	1.118***
	(0.567)	(0.605)	(0.427)	(0.404)	(0.524)	(0.556)	(0.413)	(0.394)
IV Version	All cases 8-year periods	First cases 8-year periods	All cases All (16) years	First cases All (16) years	All cases 8-year periods resid.	First cases 8-year periods resid.	All cases All (16) years resid.	First cases All (16) years resid.
Mean of dependent variable	-0.394	-0.394	-0.394	-0.394	-0.394	-0.394	-0.394	-0.394
Complier mean if not removed	-1.753	-1.783	-1.405	-1.369	-1.758	-1.802	-1.462	-1.429
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	17.696	18.216	16.981	22.170	19.402	20.118	18.790	25.077
N	9,980	9,963	9,980	9,980	9,980	9,963	9,980	9,980
Individuals	2,614	2,611	2,614	2,614	2,614	2,611	2,614	2,614
	Panel B. Young Boys (Age < 6)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Average z-score							
Removed (= 1)	0.044	-0.159	-0.416	-0.745	0.085	-0.163	-0.408	-0.784
	(0.562)	(0.505)	(0.625)	(0.610)	(0.617)	(0.537)	(0.661)	(0.652)
IV Version	All cases 8-year periods	First cases 8-year periods	All cases All (16) years	First cases All (16) years	All cases 8-year periods resid.	First cases 8-year periods resid.	All cases All (16) years resid.	First cases All (16) years resid.
Mean of dependent variable	-0.571	-0.571	-0.571	-0.571	-0.571	-0.571	-0.571	-0.571
Complier mean if not removed	-0.981	-0.847	-0.687	-0.407	-1.035	-0.882	-0.753	-0.444
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	13.999	15.737	12.537	14.662	10.983	12.872	9.801	11.743
N	12,344	12,329	12,344	12,344	12,344	12,329	12,344	12,344
Individuals	3,142	3,137	3,142	3,142	3,142	3,137	3,142	3,142

Notes: This table reports results for impacts on test scores based on an IV approach where the CPI removal rate is calculated using alternative definitions. Column 1 reproduces the estimates from our preferred measure, which calculates removal during an 8-year window using all (i.e., first and subsequent investigations for each child) cases. Column 2 reports estimates using a measure based on an 8-year window only using first cases. Column 3-4 report estimates using removal tendencies calculated during the entire sample period (2000-2015) using all cases, respectively. Columns 5-8 replicate the previous four columns using a version of each measure that is constructed from the predicted residuals from a regression of CPI removal on investigation year fixed effects. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A16: School Index Results and Robustness Tests Using Alternative Instruments

	Panel A. Young Girls (Age < 6)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	School Index							
Removed (= 1)	-0.998**	-1.019**	-0.688*	-0.656*	-0.893***	-0.907*	-0.638*	-0.625*
	(0.448)	(0.507)	(0.365)	(0.369)	(0.409)	(0.463)	(0.347)	(0.355)
IV Version	All cases 8-year periods	First cases 8-year periods	All cases All (16) years	First cases All (16) years	All cases 8-year periods resid.	First cases 8-year periods resid.	All cases All (16) years resid.	First cases All (16) years resid.
Complier mean if not removed	0.711	0.716	0.497	0.481	0.603	0.627	0.429	0.431
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	15.039	14.062	15.232	20.677	16.002	15.089	16.388	22.490
N	2,614	2,611	2,614	2,614	2,614	2,611	2,614	2,614
Individuals	2,614	2,611	2,614	2,614	2,614	2,611	2,614	2,614
	Panel B. Young Boys (Age < 6)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	School Index							
Removed (= 1)	-0.152	-0.241	-0.029	-0.289	-0.268	-0.346	-0.056	-0.334
	(0.513)	(0.463)	(0.532)	(0.517)	(0.546)	(0.491)	(0.549)	(0.533)
IV Version	All cases 8-year periods	First cases 8-year periods	All cases All (16) years	First cases All (16) years	All cases 8-year periods resid.	First cases 8-year periods resid.	All cases All (16) years resid.	First cases All (16) years resid.
Complier mean if not removed	0.495	0.566	0.426	0.701	0.574	0.865	0.447	0.746
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	12.518	12.968	13.266	12.227	10.761	11.296	11.039	10.341
N	3,142	3,137	3,142	3,142	3,142	3,137	3,142	3,142
Individuals	3,142	3,137	3,142	3,142	3,142	3,137	3,142	3,142

Notes: This table reports results for impacts on the school index outcome based on an IV approach where the CPI removal rate is calculated using alternative definitions. Column 1 reproduces the estimates from our preferred measure, which calculates removal during an 8-year window using all (i.e., first and subsequent investigations for each child) cases. Column 2 reports estimates using a measure based on an 8-year window only using first cases. Column 3-4 report estimates using removal tendencies calculated during the entire sample period (2000-2015) using all cases, respectively. Columns 5-8 replicate the previous four columns using a version of each measure that is constructed from the predicted residuals from a regression of CPI removal on investigation year fixed effects. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A17: Anderson-Rubin Confidence Interval Results

Panel A. Young Girls (Age < 6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Average *z*-score	Math *z*-score	Reading *z*-score	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	1.367**	1.471***	1.271**	-0.426**	-0.511*	-4.576	-0.998**
	(0.567)	(0.561)	(0.615)	(0.170)	(0.295)	(5.368)	(0.448)
Standard Wald CIs	(0.255, 2.479)	(0.371, 2.570)	(0.066, 2.477)	(-0.759, -0.092)	(-1.089, 0.067)	(-15.096, 5.944)	(-1.876, -0.119)
Anderson-Rubin CIs	(0.457, 3.560)	(0.571, 3.636)	(0.285, 3.646)	(-0.982, -0.112)	(-1.614, -0.038)	(-18.922, 7.219)	(-2.569, -0.226)
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	17.696	16.901	17.697	15.594	15.039	15.039	15.039
N	9,980	10,006	10,014	2,604	2,614	2,614	2,614
Individuals	2,614	2,614	2,614	2,604	2,614	2,614	2,614

Panel B. Young Boys (Age < 6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Average *z*-score	Math *z*-score	Reading *z*-score	Ever Retained (=1)	Ever IEP (=1)	Avg. Absences	School Index
Removed (= 1)	0.044	-0.003	0.102	-0.040	-0.195	-0.496	-0.152
	(0.562)	(0.574)	(0.601)	(0.273)	(0.290)	(6.610)	(0.513)
Standard Wald CIs	(-1.057, 1.146)	(-1.129, 1.122)	(-1.075, 1.280)	(-0.573, 0.496)	(-0.762, 0.373)	(-13.452, 12.460)	(-1.157, 0.854)
Anderson-Rubin CIs	(-1.572, 1.280)	(-1.607, 1.327)	(-1.575, 1.423)	(-0.673, 0.788)	(-0.900, 0.614)	(-18.164, 16.386)	(-1.400, 1.341)
Case controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	13.999	13.896	14.397	12.760	12.518	12.518	12.518
N	12,344	12,385	12,406	3,123	3,142	3,142	3,142
Individuals	3,142	3,142	3,142	3,123	3,142	3,142	3,142

Notes: This table reports Anderson-Rubin 95-percent confidence interval results for the test score and additional schooling outcomes. The standard 95-percent confidence intervals and main estimates from Tables 4 and 5 are reproduced for comparison. Andrews et al. (2019) recommend reporting Anderson-Rubin (AR) confidence intervals. The AR confidence intervals are robust to weak identification and are efficient in the just-identified case. These intervals are calculated by inverting the weak-instrument test of Anderson and Rubin (1949). Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A18: Bootstrap First-stage F-Statistic Results

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	Investigations Sample		Average z-score Sample		School Index Sample	
F-statistic (instrument)	37.706	69.163	17.696	13.999	15.039	12.518
Bootstrap results						
Mean	28.661	50.994	14.599	11.669	12.748	10.369
Median	28.243	49.558	14.216	11.291	12.710	10.096
Age/gender group	Young Girls	Young Boys	Young Girls	Young Boys	Young Girls	Young Boys
N	6,287	7,387	9,980	12,344	2,614	3,142
Individuals	6,287	7,387	2,614	3,142	2,614	3,142

Notes: Our main analysis is based on estimating two-stage least square models with a leave-out measure of CPI removal tendency as an instrument for removal. This table reports the F-statistics associated with the excluded instrument for the investigation sample examined in Table 3, the average z-score outcome examined in Table 4, and the school index outcome examined in Table 5. In addition, the table reports the mean and median from a bootstrap procedure. For each case worker, the bootstrap procedure samples (with replacement) their investigations and calculates leave-out removal rates within the sampled data. We create 250 bootstrap samples which we use to estimate first stage models.

Table A19: Impact of Removal on Test Scores for Young Children, By Subgroup

	Panel A. Young Girls (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
	White	Minority	Single Parent	Not-single Parent	Neglect Allegation	Professional Reporter
Dependent variable:	Average z-score					
Removed (= 1)	1.697	1.048**	1.237**	1.687	1.352**	1.707**
	(1.203)	(0.424)	(0.549)	(1.600)	(0.632)	(0.669)
Mean of dependent variable	-0.261	-0.560	-0.454	-0.292	-0.406	-0.396
Complier mean if not removed	-1.907	-1.432	-1.401	-2.675	-1.437	-1.950
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	6.611	13.802	17.575	4.927	15.881	13.130
N	5,538	4,442	6,270	3,710	8,077	7,602
Individuals	1,414	1,200	1,604	1,010	2,137	2,026
	Panel B. Young Boys (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
	White	Minority	Single Parent	Not-single Parent	Neglect Allegation	Professional Reporter
Dependent variable:	Average z-score					
Removed (= 1)	-0.692	0.478	-0.131	0.845	-0.824	-0.103
	(1.301)	(0.597)	(0.552)	(3.223)	(1.042)	(0.506)
Mean of dependent variable	-0.444	-0.730	-0.614	-0.493	-0.584	-0.578
Complier mean if not removed	-0.362	-1.286	-0.921	-1.450	-0.253	-0.682
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	4.294	8.935	12.104	0.826	4.884	17.199
N	6,888	5,456	7,908	4,436	9,666	9,554
Individuals	1,719	1,423	1,951	1,191	2,485	2,026

Notes: This table reports results for the impact of removal on test score outcomes for young girls (Panel A) and young boys (Panel B) by subgroups. All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A20: Impact of Removal on the School Index for Young Children, By Subgroup

	Panel A. Young Girls (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
	White	Minority	Single Parent	Not-single Parent	Neglect Allegation	Professional Reporter
Dependent variable:	School Index					
Removed (= 1)	-1.266	-0.788*	-0.986**	-0.824	-0.996*	-1.032*
	(0.818)	(0.427)	(0.489)	(0.935)	(0.433)	(0.529)
Mean of dependent variable	-0.009	0.009	0.029	-0.048	0.001	-0.015
Complier mean if not removed	1.108	0.398	0.395	1.689	0.575	0.780
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	5.142	10.611	13.584	4.553	14.014	13.519
N	1,414	1,200	1,604	1,010	2,137	2,026
Individuals	1,414	1,200	1,604	1,010	2,137	2,026
	Panel B. Young Boys (Age < 6)					
	(1)	(2)	(3)	(4)	(5)	(6)
	White	Minority	Single Parent	Not-single Parent	Neglect Allegation	Professional Reporter
Dependent variable:	School Index					
Removed (= 1)	-0.698	0.304	-0.405	0.610	-0.455	0.018
	(0.808)	(0.581)	(0.524)	(1.474)	(0.860)	(0.491)
Mean of dependent variable	-0.025	0.030	0.025	-0.042	0.003	-0.011
Complier mean if not removed	0.998	0.105	0.722	0.155	0.965	0.323
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	4.863	6.015	10.060	1.869	4.067	13.790
N	1,719	1,423	1,951	1,191	2,485	2,026
Individuals	1,719	1,423	1,951	1,191	2,485	2,026

Notes: This table reports results for the impact of removal on the school index outcomes for young girls (Panel A) and young boys (Panel B) by subgroups. All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A21: Impact of Removal on Criminal Justice Outcomes for Parent Perpetrators

Panel A. Parent Perpetrators of Young Girls (Age < 6)

	(1)	(2)	(3)	(4)	(5)
	Criminal Justice Outcomes: Charged/Incarcerated, 4-year Post Invs.				
Dependent variable:	Any (=1)	Property (=1)	Drug (=1)	Public Offense (=1)	Sex Offense (=1)
Removed (= 1)	0.300	-0.107	0.231	0.365	-0.022
	(0.278)	(0.168)	(0.198)	(0.264)	(0.043)
Mean of dependent variable	0.282	0.085	0.086	0.201	0.006
Complier mean if not removed	0.156	0.238	0.068	0.033	0.038
Case controls	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	9.756	9.756	9.756	9.756	9.756
N	2,333	2,333	2,333	2,333	2,333
Individuals	2,333	2,333	2,333	2,333	2,333

Panel B. Parent Perpetrators of Young Boys (Age < 6)

	(1)	(2)	(3)	(4)	(5)
	Criminal Justice Outcomes: Charged/Incarcerated, 4-year Post Invs.				
Dependent variable:	Any (=1)	Property (=1)	Drug (=1)	Public Offense (=1)	Sex Offense (=1)
Removed (= 1)	-0.034	0.051	-0.108	-0.270	0.006
	(0.319)	(0.188)	(0.198)	(0.321)	(0.062)
Mean of dependent variable	0.268	0.076	0.078	0.198	0.011
Complier mean if not removed	0.286	0.037	0.077	0.333	0.062
Case controls	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	8.034	8.034	8.034	8.034	8.034
N	2,777	2,777	2,777	2,777	2,777
Individuals	2,777	2,777	2,777	2,777	2,777

Notes: This table reports results for the impact of removal on criminal justice outcomes for the parents of young girls (Panel A) and young boys (Panel B). Information on parent perpetrators comes from DCYF records. In the sample of young investigated children, 95 percent of children have at least one perpetrator who is a parent. As described in Section 3, we construct samples of parent perpetrators of young girls and young boys and measure whether parents are charged or incarcerated within 4-year windows after the conclusion of an investigation and by type of offense. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A22: Impact of Removal on Subsequent CPS Contact

	Panel A. Young Girls (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Subseq. Investigation (=1)		Subseq. Removal (=1)	
Removed (= 1)	-0.356	-0.350	-0.066	-0.047
	(0.291)	(0.285)	(0.195)	(0.188)
Mean of dependent variable	0.279	0.279	0.085	0.085
Complier mean if not removed	0.584	0.584	0.185	0.185
Case controls	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	12.839	15.039	12.839	15.039
N	2,614	2,614	2,614	2,614
Individuals	2,614	2,614	2,614	2,614
	Panel B. Young Boys (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Subseq. Investigation (=1)		Subseq. Removal (=1)	
Removed (= 1)	-0.215	-0.238	0.153	0.140
	(0.332)	(0.332)	(0.209)	(0.206)
Mean of dependent variable	0.267	0.267	0.075	0.075
Complier mean if not removed	0.690	0.690	0.169	0.169
Case controls	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes
F-statistic (instrument)	9.456	12.518	9.456	12.518
N	3,142	3,142	3,142	3,142
Individuals	3,142	3,142	3,142	3,142

Notes: This table reports results for the impact of removal on subsequent investigation and removal outcomes for young girls (Panel A) and young boys (Panel B). All measures are based on DCYF records for investigations that occur after the first investigation. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table A23: Tests of Random Case Assignment (Full Regression Results), Older Children Sample

	(1)	(2)	(3)
Dependent variable:	CPI removal tendency		
Female	-0.001		
	(0.001)		
Black	-0.001	0.000	0.001
	(0.002)	(0.002)	(0.003)
Hispanic	-0.000	-0.003	0.003
	(0.002)	(0.002)	(0.004)
Other race	0.001	0.001	-0.000
	(0.003)	(0.003)	(0.004)
Age	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Married couple	0.000	0.001	-0.001
	(0.002)	(0.003)	(0.002)
Unmarried couple	-0.000	0.002	-0.002
	(0.002)	(0.002)	(0.002)
English language	-0.003	-0.005*	-0.001
	(0.003)	(0.003)	(0.004)
Neglect	-0.001	-0.004	0.001
	(0.002)	(0.002)	(0.001)
Physical neglect	-0.007*	0.015***	0.001
	(0.004)	(0.003)	(0.005)
Professional reporter	-0.003	-0.005*	-0.001
	(0.003)	(0.004)	(0.004)
Family/friend reporter	-0.002	-0.004	0.000
	(0.003)	(0.003)	(0.004)
Emergency investigation	0.014**	0.016**	0.013
	(0.006)	(0.007)	(0.008)
Immediate investigation	0.002	0.000	0.004
	(0.001)	(0.002)	(0.001)
Chi-squared statistic	23.640	31.590	12.740
p-value of joint significance	0.051	0.003	0.469
Sample	Older Children	Older Girls	Older Boys
Mean of CPI removal tendency	0.177	0.177	0.177
Investigation year FE	Yes	Yes	Yes
N	13,120	6,643	6,477

Notes: This table reports regression results testing the random assignment of cases to CPIs. Results are from a regression of CPI removal tendency on the case characteristics listed and investigation year fixed effects. Column 1 reports estimates for all older children (investigated at ages 6-18). Columns 2-3 report estimates for older girls and older boys, respectively. The chi-square statistic and p-value reported are from a test of joint significance of all variables except investigation year fixed effects (FE). Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

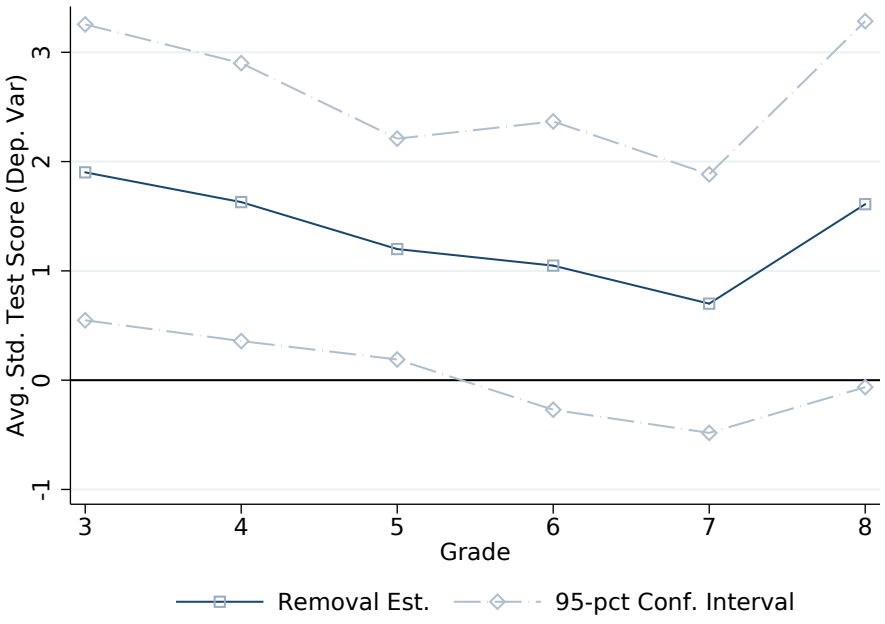
Table A24: Impact of Removal on Outcomes of Older Children

	Panel A. Older Girls (Age ≥ 6)					
	School-age outcomes		Later-life outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score	School Index	Delinquent (=1)	HS Grad. (=1)	Teen Birth (=1)	College (=1)
Removed (= 1)	-0.230	-0.373	-0.030	-0.010	0.089	0.133
	(0.582)	(0.326)	(0.261)	(0.187)	(0.162)	(0.222)
Mean of dependent variable	0.068	-0.005	0.055	0.351	0.194	0.303
Complier mean if not removed	-0.337	0.138	0.101	0.263	0.210	0.032
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	22.129	35.092	9.296	49.213	33.333	38.718
N	7,517	3,029	1,829	4,136	2,956	3,326
Individuals	2,581	3,029	1,829	4,136	2,956	3,326
	Panel B. Older Boys (Age ≥ 6)					
	School-age outcomes		Later-life outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score	School Index	Delinquent (=1)	HS Grad. (=1)	Teen Birth (=1)	College (=1)
Removed (= 1)	-0.237	0.323	-0.016	-0.144	0.119	-0.127
	(0.429)	(0.216)	(0.156)	(0.157)	(0.115)	(0.187)
Mean of dependent variable	0.053	-0.003	0.147	0.319	0.059	0.239
Complier mean if not removed	-0.414	-0.297	0.096	0.385	0.000	0.367
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	30.911	34.273	24.610	44.810	26.860	41.145
N	8,838	3,440	2,185	3,770	3,025	2,953
Individuals	2,965	3,440	2,185	3,770	3,025	2,953

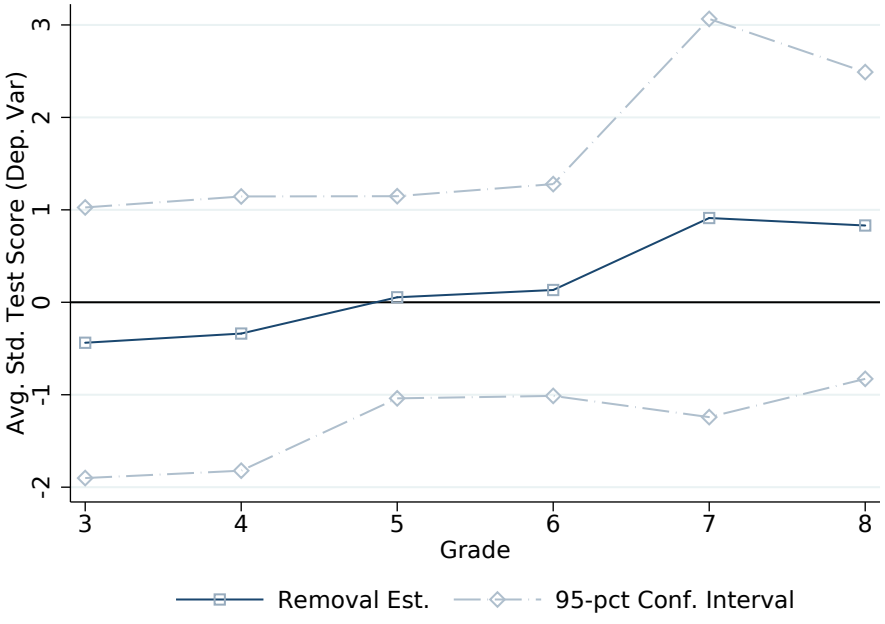
Notes: This table reports results for the impact of removal on outcomes for older girls (Panel A) and older boys (Panel B). Older is defined as being investigated at ages six or later (up to age 18). All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. The school index is constructed based on standardized measures of whether an investigated child was ever retained, ever participated in special education (i.e., has an IEP), and the average number of days absent during grades 3-8. All outcomes are measured after the first investigation. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Figure A1: Impact of Removal on Test Scores of Young Children, by Grade

(a) Young Girls



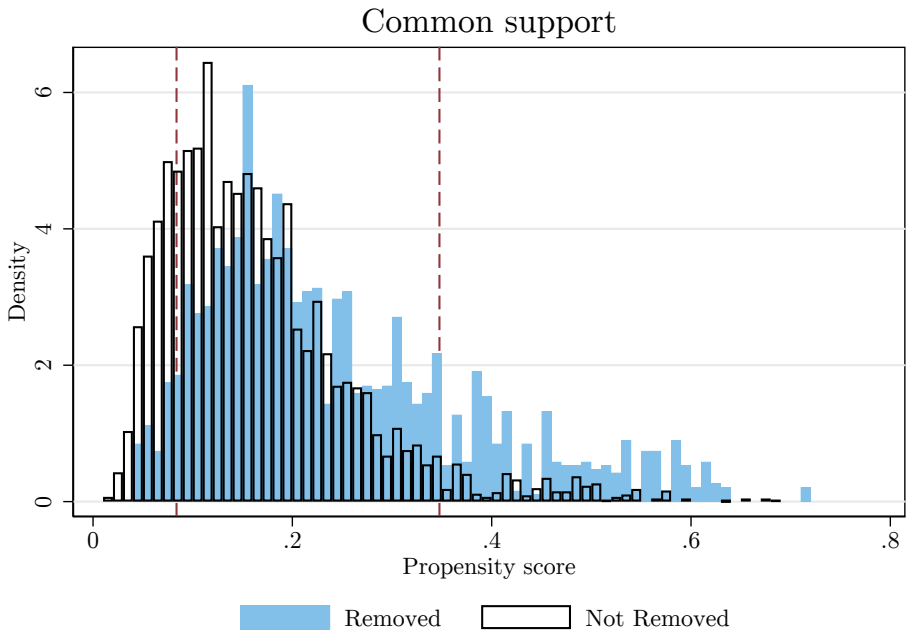
(b) Young Boys



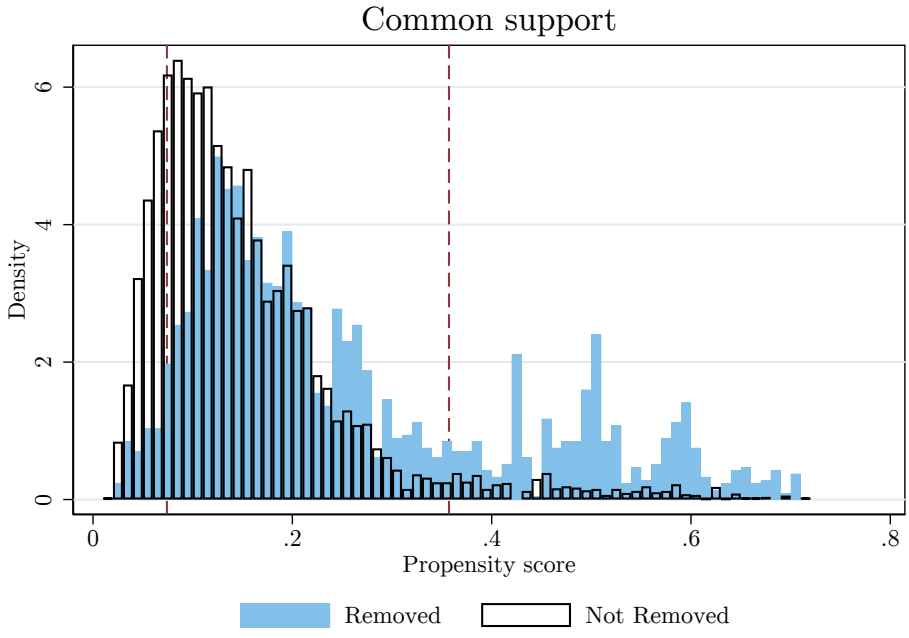
Notes: These figures show results for the impact of removal on test scores estimated in separate regressions for grades 3-8 for young girls (Panel A) and young boys (Panel B). All results are from two-stage least squares models with the leave-out measure of CPI removal tendency as an instrument for removal. All models include controls for the case characteristics in Table 1 and investigation year fixed effects. Confidence intervals are based on standard errors that are two-way clustered at the family and CPI levels.

Figure A2: Common Support of CPI Removal Tendency

(a) Young Girls



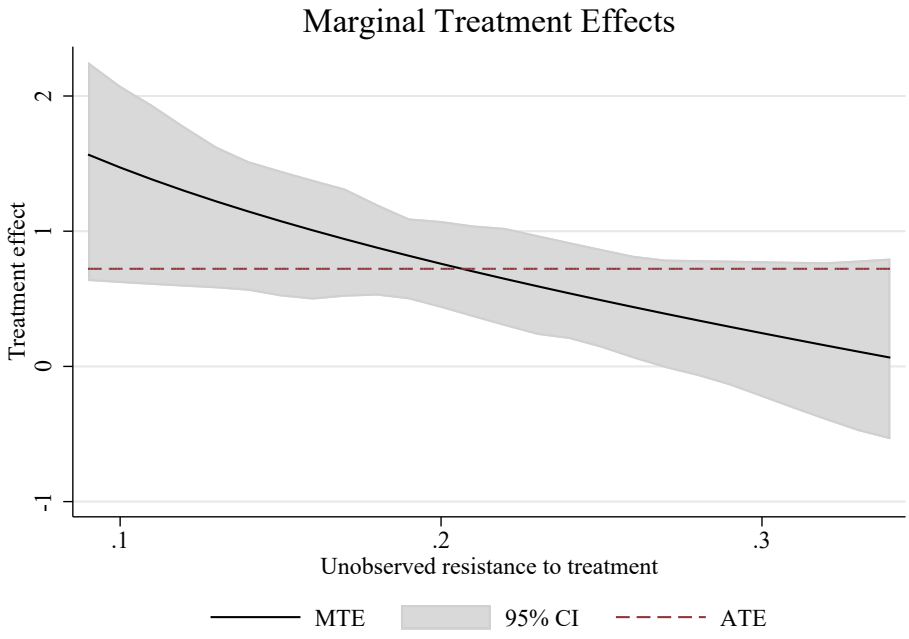
(b) Young Boys



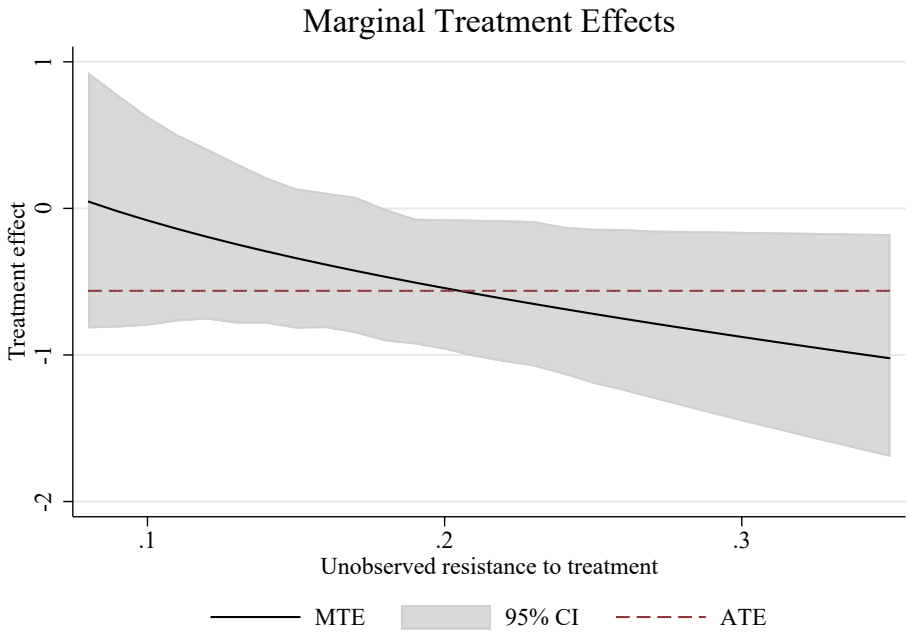
Notes: These figures show the distribution of the propensity score for treated (removed) and non-treated (non-removed) children. The dashed red lines in each figure indicate the upper and the lower points of the propensity score with common support (based on five percent trimming).

Figure A3: MTE for Test Scores of Young Children

(a) Young Girls



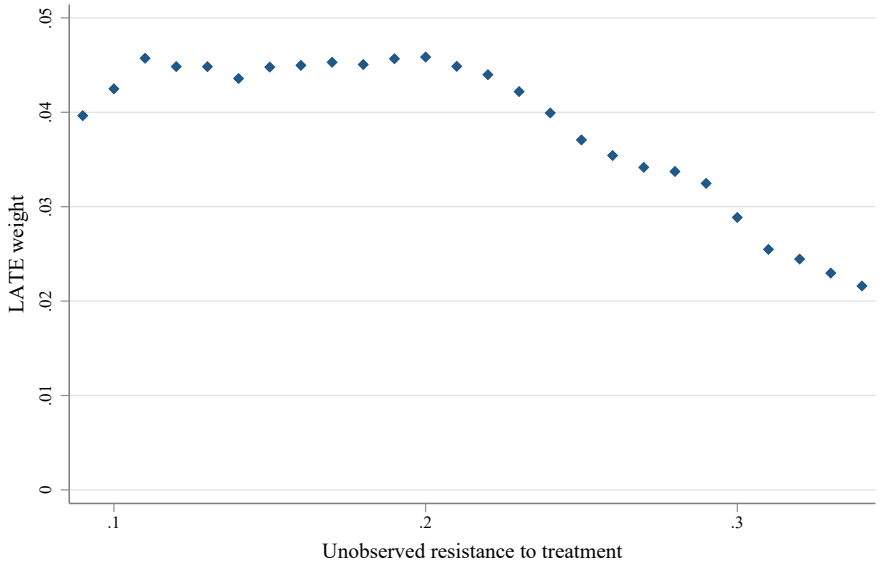
(b) Young Boys



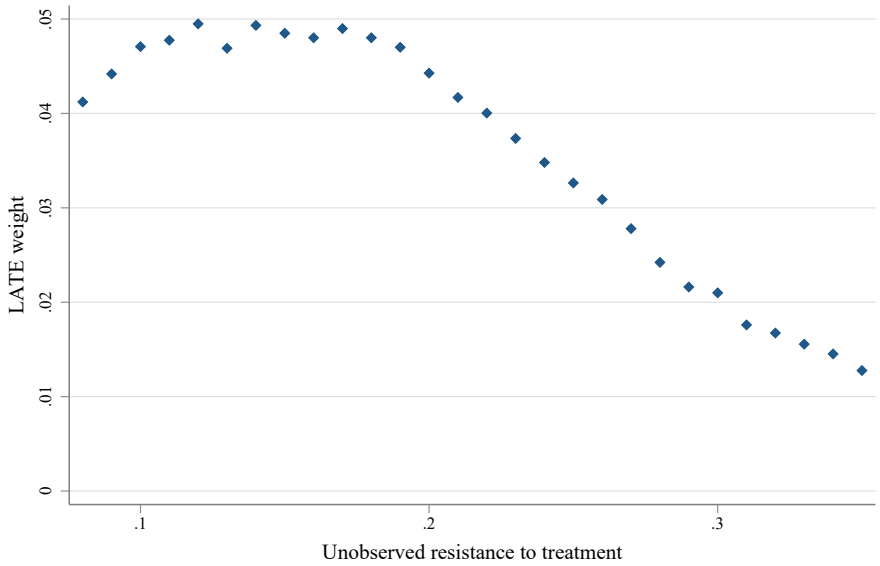
Notes: These figures plot MTEs for the impact of removal on young children based on a local instrumental variables (IV) approach using a global quadratic polynomial specification for the trimmed sample with common support. Standard errors are constructed based on 100 bootstrap replications.

Figure A4: Instrumental Variable Weights

(a) Young Girls



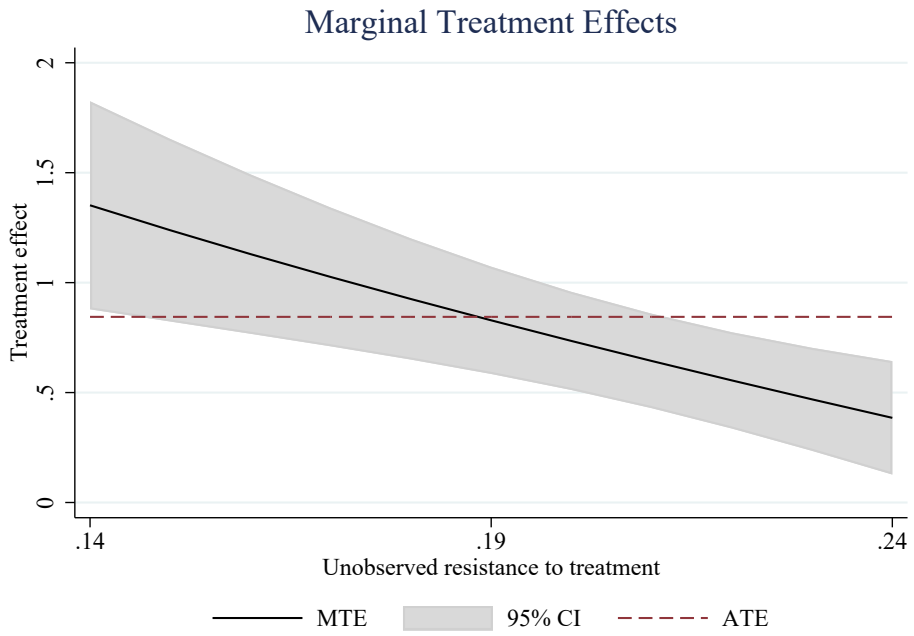
(b) Young Boys



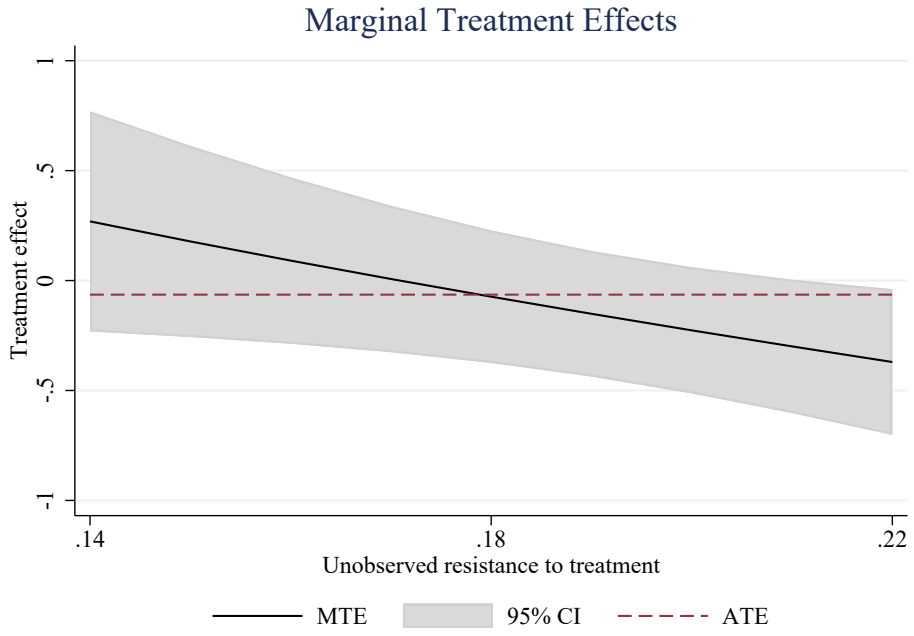
Notes: This figure plots instrumental variable (IV) weights for the samples of young girls (Panel A) and young boys (Panel B) that are included in the MTE analysis of the effects of removal on average test scores. To calculate these weights, we use the estimation approach from Andresen (2018).

Figure A5: MTE for Test Scores of Young Children, Robustness (Results without Covariates)

(a) Young Girls



(b) Young Boys



Notes: These figures plot MTEs for the impact of removal on young children based on a local instrumental variables (IV) approach using a global quadratic polynomial specification for the trimmed sample with common support. The unobserved resistance to treatment range differs from the MTE results in Appendix Figure A3 where the region of common support is based on a first stage model that includes covariates. Standard errors are constructed based on 100 bootstrap replications.

B Analysis of Unfounded and Founded Allegations

As noted in Section 2, the assigned CPI also makes decisions about whether an allegation of abuse or neglect is founded or unfounded (see Figure 1). When DCYF dismisses unfounded allegations, these cases are closed, and there is no active choice on whether to remove the child or not. The reports associated with unfounded cases are kept in the DCYF system and removed after a specified period. Thus, we do not have data on unfounded cases for the full period covered by the cases in our main investigation sample.

One concern for our analysis is that the assigned CPI's removal rate may be correlated with the rate of determining whether an allegation is unfounded. This could introduce sample selection bias. For example, one possibility is that a high removal rate investigator has a low rate of determining that allegations are unfounded. This could imply that their cases are less serious on average.

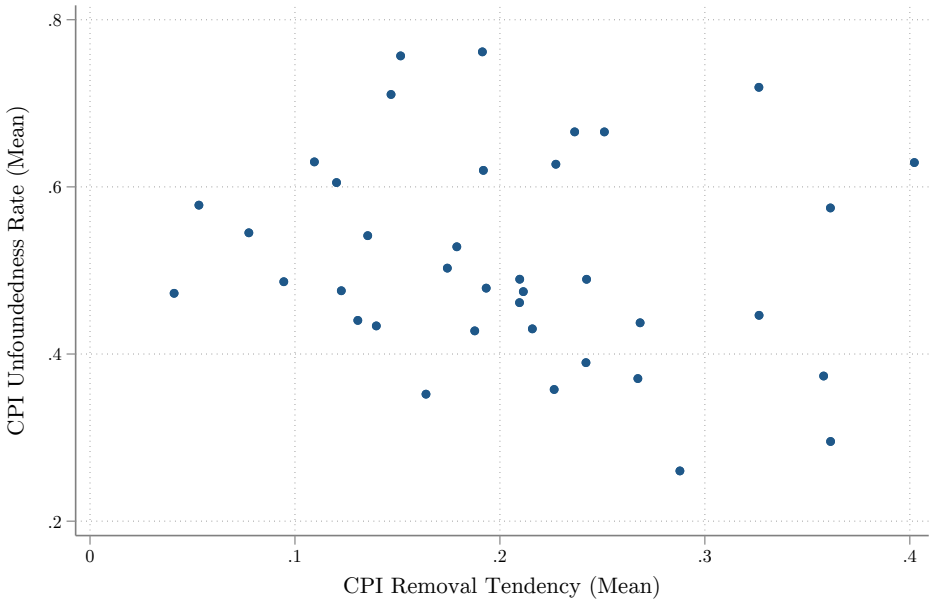
To investigate this concern, we obtained a limited sample of unfounded records from DCYF for the 2015-2017 period. We combined these records with founded cases from the same period. The resulting sample of founded and unfounded investigations contains records for $4,821$ children who were less than age six at the time of their investigation. Note that we focus on children less than age six to parallel our main analysis. For each investigated child, we calculate a leave-out measure of removal for their assigned CPI. Similarly, we calculate a leave-out measure of whether the assigned CPI determined whether an investigation is unfounded. The mean of the unfounded tendency measure is approximately 0.50, and the standard deviation of this measure is 0.12.

There are two main findings from analyzing this sample of founded and unfounded records. First, we find no statistically significant correlation between the assigned CPI's removal tendency and their rate of determining whether allegations were unfounded. The correlation between the CPI removal and unfoundedness rates is -0.17 and is not statistically significant. The p-value on the correlation coefficient is 0.28. To further explore the measure of unfounded tendency, we take each assigned CPI and calculate the mean unfounded and removal rates in this limited sample. Appendix Figure B1 plots the mean unfoundedness and removal rates. This figure is in line with the insignificant correlation and does not suggest a non-linear pattern between each CPI's unfoundedness and removal rates.

Second, we find that the child and case characteristics in this sample of founded and unfounded investigations are not significantly correlated with either the CPI removal tendency for founded investigations or the CPI rate of determining that allegations were unfounded. These results are demonstrated in Appendix Table B1. Note that this analysis is at the child level, which differs from Appendix Figure B1 where the unit of the analysis is at the CPI

level. The table reports point estimates from a regression where the dependent variable is a leave-out measure for the CPI's unfoundedness rate or the CPI's removal tendency. The independent variables in the regressions include characteristics for the investigated child and their case. We estimate these models separately for young girls (Columns 1 and 3) and young boys (Columns 2 and 4). The point estimates are generally small in magnitude and not statistically significant. In each of the four models that we estimate, we consistently fail to reject the null hypothesis that the coefficients for independent variables are jointly zero. For example, the final row in Column 1 reports that the p-value on the joint test is equal to 0.968.

Figure B1: Unfoundedness and Removal Rates at the CPI Level (2015-2017 DCYF Records)



Notes: This figure is a plot of each CPI's unfoundedness and removal rates. Statistics are computed using a sample of founded and unfounded records from DCYF for the 2015-2017 period.

Table B1: Randomization Tests Using Combined Founded and Unfounded Investigations Sample

	(1)	(2)	(3)	(4)
Dependent variable:	CPI Unfoundedness Rate		CPI Removal Tendency	
Black	0.007	0.006	-0.006	0.005
	(0.007)	(0.006)	(0.004)	(0.004)
Hispanic	-0.004	0.005	0.005	-0.001
	(0.008)	(0.006)	(0.005)	(0.003)
Other race	0.002	0.004	-0.001	-0.002
	(0.003)	(0.007)	(0.002)	(0.004)
Age	-0.001	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)
Married couple	0.006	0.004	-0.004	0.002
	(0.004)	(0.006)	(0.004)	(0.003)
Unmarried couple	0.006	-0.000	-0.003	0.001
	(0.004)	(0.003)	(0.003)	(0.002)
English language	-0.006	-0.000	0.004	-0.004
	(0.004)	(0.005)	(0.005)	(0.006)
Neglect	-0.001	-0.010	-0.003	0.002
	(0.007)	(0.007)	(0.004)	(0.004)
Physical neglect	0.002	-0.010	-0.000	0.003
	(0.007)	(0.007)	(0.004)	(0.004)
Professional reporter	0.007	0.016**	0.000	-0.007
	(0.006)	(0.008)	(0.004)	(0.004)
Family/friend reporter	0.009	0.013	0.001	-0.005
	(0.006)	(0.008)	(0.005)	(0.004)
Emergency investigation	0.005	-0.004	0.005	0.000
	(0.009)	(0.008)	(0.005)	(0.004)
Immediate investigation	-0.001	0.005	0.003	-0.004
	(0.005)	(0.005)	(0.003)	(0.003)
Age/gender group	Young Girls	Young Boys	Young Girls	Young Boys
Chi-squared statistic	4.670	13.860	15.900	12.970
p-value of joint significance	0.968	0.384	0.254	0.450
Mean of dependent variable	0.501	0.502	0.202	0.199
N (Individuals)	2,207	2,614	2,207	2,614

Notes: This table reports regression results where the dependent variable is a measure of the rate at which a CPI determines that an allegation is unfounded (Columns 1-2) or the rate at which a CPI recommends removal in an investigation (Columns 3-4). The independent variables in the regression are the characteristics of the child associated with the case. The sample includes children who were investigated during the 2015-2017 period. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

C Data Sources, File Descriptions, and Description of Sample Construction

This section describes the data sources, data files, and samples that we use for the analysis of the main text.

C.1 Data Sources

Our analysis relies on data from several administrative sources. Table C1 lists each administrative source, files provided, and the time period covered by the associated files.

C.2 Description of Files

C.2.1 Child Protective Services and foster care placement files

Child Protective Services (CPS) files (2000-2017) identify victims and perpetrators of child abuse or neglect. These data contain the CPS reports created when a suspected abuse or neglect allegation is reported via the Rhode Island (RI) Department of Children, Youth and Families (DCYF) hotline. Note that CPS functions as the investigative arm of DCYF. The CPS files a report with family structure, primary language, reporter type, allegation type for each victim-perpetrator combination, and designated investigation level. The investigation and placement files include all substantiated investigations resulting from CPS reports, and the assignment history of investigations to field Child Protection Investigators (CPIs). The Adoption and Foster Care Analysis and Reporting System (AFCARS) data file contains information on all children in foster care in RI.

C.2.2 Juvenile delinquency records

The DCYF houses the Division of Juvenile Corrections, which oversees youth located at the Rhode Island Training School (RITS) or sentenced to probation by the RI Family Court. The RI Family Court handles wayward or delinquent offenses for youth ages 10-17, while youth can remain at RITS through age 18. Records of juvenile delinquency (2000-2016) contain the dates of sentencing for each person.

C.2.3 Criminal justice records

The RI Department of Corrections (DOC) records contain the population of charged and incarcerated individuals in Rhode Island (1995-2017). The dates of each unique charge or sentence are observed, as well as the type of charge (e.g., assault, property crime) and the total sentence length.

Table C1: List of Data Sources

Source	Data	Time Period
RI Dept. of Children, Youth, and Families	Child Protective Services (CPS) files – CPS report (allegations) – substantiated investigations – case assignments (field CPIs)	2000-2017
	Adoption and Foster Care Analysis and Reporting System (AFCARS) – foster care placements	2000-2017
	Juvenile delinquency records – sentences to the Rhode Island Training School (RITS) – placement on probation	2000-2016
RI Dept. of Corrections	Criminal justice records – criminal charges – incarceration history	1995-2017
RI Dept. of Education	End-of-Year enrollment records – school, enrollment dates, grade – Individualized Education Program (IEP), free/reduced price lunch status, grade retention, absences – high school graduation	2003-2016
	Standardized testing records – testing school and year – NECAP reading and math test scores (grades 3-8, school years 2005-2013) – PARCC reading and math test scores (grades 3-8, school years 2014-2016)	2005-2016
National Student Clearinghouse	Postsecondary enrollment records – college-going	2004-2015
RI Dept. of Health	Vital records – teen births	2000-2016
RI 360 Database	Demographics – birth date, gender, race	1997-2016

Notes: This table lists data sources, files, and the time period covered by the associated files.

C.2.4 End-of-year enrollment records

The RI Department of Education (RIDE) maintains records of all students enrolled in RI public and charter schools; we have access to data from school years 2003-04 through 2016-17. These data include enrollment dates, grade and school attended, Individualized Education Program status (which identifies participation in special education services), free and reduced-price lunch status, yearly absences, and high school graduation status.

C.2.5 Standardized testing records

RIDE reports standardized mathematics and reading test score results for enrolled students in grades 3-8. Rhode Island administered the New England Common Assessment Program (NECAP) test from school years 2005-06 to 2013-14 and the Partnership for Assessment of Readiness for College and Careers (PARCC) test from 2014-15 to 2016-17. Participation rates for standardized exams in RI have historically been high (more 95 percent of students take exams). In 2014, participation rates fell to roughly 90 percent, but rose to previous levels by 2016.

C.2.6 Post-secondary enrollment records

The National Student Clearinghouse (NSC) reports post-secondary enrollment dates for RI high school students (2004-2015), regardless of high school completion.

C.2.7 Vital (birth) records

The RI Department of Health (DOH) vital (birth) records contain all Rhode Island births (2000-2016) and include identifiers for the mother and father, as well as mother's date of birth.

C.2.8 Demographics

The RI 360 Database joins records associated with an individual across a range of social programs and government services (see Hastings et al. (2019)). The database provides demographic information (birth date, gender, and race) for all children in the DCYF sample born between 1982 and 2015 and appearing in administrative records between 1997 and 2016.

C.3 Samples and Key Outcomes

C.3.1 Sample of DCYF Investigated Children

We use CPS records to construct a sample with children involved in abuse or neglect investigations. As an initial step, we link alleged abuse or neglect investigation records

to a file containing assignment records. This allows us to determine the Child Protective Investigator (CPI) assigned to each investigation, and whether the assignment was via the rotation list (see Section 2). We also link investigations to the AFCARS foster care placement history file to determine whether DCYF placed investigated children into foster care due to an investigation.

Using the assembled CPS investigation records, we impose the following restrictions to create the sample of DCYF investigated children.

1. Restrictions related to data cleaning:

   (a) Restrict to children ages 0-18 with known demographics. We join children in CPS case files to the RI 360 database to obtain a global identifier and verifiable demographic information (see Hastings et al. (2019)). To be included, children must have an observed birth date and gender.

   (b) Restrict to allegations reported via the DCYF hotline. Allegations are primarily reported via the hotline.

   (c) Restrict to allegations in which the alleged perpetrator is a family member. In the full CPS case files, 93 percent of neglect or abuse reports are alleged to have been perpetrated by a member of the child's family. The remaining seven percent involve DCYF providers of care or institutional abuse allegations, but these investigations follow a different set of procedures.[66]

   (d) Drop allegations reported after the initial DCYF hotline call.

   (e) Drop allegations that do not meet the investigation criteria (internally designated as "info/referral" reports). These reports would not be forwarded to the Investigative Unit.

   (f) Drop investigations that are unfounded (i.e., there was no preponderance of evidence that child abuse or neglect occurred). These records are only available for the period 2015-2017. Unfounded records from prior years are *not* available because DCYF removes older records from their database periodically.

   (g) Restrict to investigations from 2000 to 2015. We remove investigations that began after 2015 to avoid censored foster care placement outcomes.

   (h) Drop children involved in at most one investigation per day. CPS may receive more than one report of abuse or neglect on the same day for the same child; in such instances, the child could be affiliated with more than one CPI. We exclude these cases.

[66] Following DCYF Operating Procedure 500.0035.

(i) Restrict to investigations matched to a CPI assignment. The link between investigations and case assignment history is imperfect, and we are sometimes unable to identify the CPI assigned to the investigation following the initial hotline call. We ignore these unmatched observations.

2. Restrictions related to the research design:

(j) Restrict to investigations assigned via the rotation list. We do not consider investigations that the CPI supervisor assigns to CPIs "off-rotation." For example, CPIs can volunteer to take an investigation. To identify full-time CPIs who received their daily case assignment via the rotation list, we impose additional restrictions and do not consider investigations where CPIs were working primarily as hotline workers or investigations where CPIs had already received their daily assignment via the rotation list.

(k) Drop investigations based on alleged sex abuse. From conversations with DCYF, we understand that the Investigative Unit supervisor attempts to assign sex abuse cases to CPIs of the same gender as the child. This violates random assignment, and, therefore, we do not consider these investigations.

(l) Restrict to the first investigation observed for each child. We do not consider later investigations where the child reappears in the DCYF caseload.

(m) Drop if the associated CPI's removal tendency (see definition in Section 4) is calculated using less than 10 cases. We impose this restriction to avoid concerns regarding small cell sizes.

(n) Drop outliers based on the top or bottom one percent of CPI removal tendency.

The items listed (a)–(n) in Table C2 provide the number of distinct allegations, investigations, and children present in CPS case files after imposing the above restrictions. The first row shows that initial CPS records contain $187,023$ allegations of abuse or neglect associated with $54,119$ investigations and $63,351$ children (more than one child can be part of the same investigation). The subsequent rows report the remaining number of observations after imposing data restrictions. For example, the row labeled (a) shows there are $176,034$ allegations of abuse or neglect associated with $51,864$ investigations and $58,429$ children. The rows under the header for the DCYF investigation sample report the final statistics for the number of young (investigated before age six) and old (investigated at age six or after) children. The last two rows report the statistics for the main analysis sample by gender. This corresponds to the schooling outcomes sample by gender. This main analysis sample includes the investigated children matched to the school test score and enrollment records.

Table C2: Summary and Statistics for Data Restrictions

	(1) Allegations	(2) Investigations	(3) Children
Full DCYF data	187,023	54,119	63,351
1. Restrictions related to data cleaning			
a. Restrict to children ages 0-18 with known demographics	176,034	51,864	58,429
b. Restrict to the first allegations reported via the DCYF hotline	154,809	51,585	56,508
c. Restrict to allegations involving a family	146,372	49,103	54,427
d. Drop additional info. allegations	134,684	48,943	54,079
e. Drop allegations not investigated	102,005	48,026	46,036
f. Drop unfounded investigations	81,134	38,120	38,730
g. Restrict to investigations from 2000 to 2015	71,451	33,492	34,364
h. Drop if child in multiple investigations on the same date	71,278	33,418	34,348
i. Restrict to investigations matched to a CPI assignment	70,039	32,845	33,971
2. Restrictions related to the research design			
j. Restrict to investigations assigned via the rotation list	57,986	27,050	29,286
k. Drop investigations involving sex abuse	54,697	25,312	27,798
l. Restrict to the first investigation for each child	39,813	19,838	27,606
m. Drop if the CPI removal tendency is calculated with ≤ 10 obs.	39,636	19,758	27,484
n. Drop outliers in CPI removal tendency	38,631	19,270	26,794
DCYF investigation sample			
Young children (age < 6)	19,001	11,411	13,674
Older children (age ≥ 6)	19,630	9,853	13,120
Sample for main analysis of young children			
Young girls	3,690	2,407	2,614
Young boys	4,421	2,886	3,142

Notes: This table summarizes the data restrictions and the resulting number of allegations, investigations and children present in the CPS case files after imposing the associated restriction. The unfounded cases removed in Part (f) of Step 1 are from the period 2015-2017. Older records on unfounded cases are not available from DCYF.

C.3.2 Schooling Outcomes Sample and Key Outcomes

As discussed in Section 3, we create a schooling outcomes sample by joining the DCYF sample of investigated young children to records from RIDE. The sample is defined as the set of all investigated young children who are observed in both the test score and enrollment records after an investigation occurred. This definition ensures that we have a consistent sample for whom we observe testing performance and non-testing outcomes (e.g., whether the child participated in special education). Note that investigated children who were born before 1995 or after 2008 are *not* in the sample because they are either too old or young to be enrolled in the testing grades (3-8) during the period 2005-2016. In addition, investigated children who move from Rhode Island or enrolled in a private school are also not included. There are $2,614$ young girls and $3,142$ young boys in our schooling outcomes sample. The main outcomes for our analysis are the following test score measures:

- Reading z-score: Reading test score, standardized with mean equal to zero and standard deviation equal to one at the grade and year level among the full population of tested students in Rhode Island.
- Math z-score: Math test score, standardized with mean equal to zero and standard deviation equal to one at the grade and year level among the full population of tested students in Rhode Island.
- Average z-score: The mean of a child's reading and math z-scores.

Note that we standardized these scores by grade and academic year to maintain comparability across testing years.

We also study additional post-investigation non-test score outcomes for this sample. Specifically, we measure the additional schooling outcomes listed below:

- Ever Retained (=1): Indicator for ever repeating a grade over two consecutive years in grades K-8. (This is missing for students not observed in two consecutive years.)
- Ever IEP (=1): Indicator for enrollment in special education services, identified by having an Individualized Education Program (IEP) in grades K-8.
- Avg. Absences: Average yearly absences (excused and unexcused) in grades K-8. We set the top percentile in school absences to missing as these students were likely not enrolled.

- School Index: The weighted sum of standardized retention, IEP, and absences outcomes, where we standardize each outcome by gender and age group (e.g., those younger than six at the time of an investigation).

Note that there are 10 young girls and 19 young boys for whom we cannot measure grade repetition because they are only enrolled in one academic year during our sample period. For these children, we compute the School Index measure using only the IEP and average attendance outcomes.

Finally, we study several factors that may mediate the impact of removal from home on child outcomes. We study two types of mediating factors for our schooling outcomes sample using data from DCYF and RIDE. First, we study the foster care outcomes. The key variables for foster care outcomes are:

- Total days in foster care: Days spent in foster care due to the child's first investigation, from removal date to discharge date (also applies to days spent with relatives, with foster families, in group homes, and in other care).
- Adopted (=1): Indicator for child adopted upon discharge from foster care.
- Number of placements: Number of foster care placements resulting from the child's first investigation.
- Placed with relative (=1): Indicator for any placement with a relative due to the child's first investigation.
- Police notified (=1): Indicator for whether police were notified during the investigation.
- Subseq. Investigation (=1): Indicator for any future child abuse or neglect investigation within the 4-year post-investigation period after the conclusion of the first investigation.
- Subseq. Removal (=1): Indicator for any future home removal due to a future child abuse or neglect investigation within the 4-year post-investigation period after the conclusion of the first investigation.

We focus on these outcomes measured for investigations from 2000-2015 to ensure an uncensored foster care placement measure. For children still in care as of January 1, 2018, foster care variables (e.g., total days in care, days spent with relatives) are measured as of January 1, 2018.

Second, we also study school mobility (i.e., school change) and the characteristics of schools attended during grades 3-8 for children in our schooling outcomes analysis sample.

We do this using a panel at the child-academic-year level. The key variables for children in the mobility and school characteristics analysis are:

- Moved Schools (=1): Indicator for changing schools.
- School Value-Added: We construct a school-level value-added measure that considers tests taken by RI students in grades 4-8. We restrict to students not in the DCYF sample. We exclude test scores for students who repeat grades or are missing any of the baseline controls used in the value-added estimation. We estimate a school's value-added measure (μ) from the following student-level regression:

  $$A_{ijt} = X_{ijt}\beta + \nu_{ijt}$$

  where

  $$\nu_{ijt} = \mu_j + \epsilon_{ijt}.$$

  For each child i in school j in year t, we observe the dependent variable A_{ijt} as the child's test score (standardized by grade and year). We include a vector of control variables X_{ijt} that includes race, gender, special education status, English learner status, free or reduced-price lunch (FRPL) status, and a cubic in lagged test scores. The residual v_{ijt} is composed of the school j's value-added measure (μ) and an error term. To match the students in the DCYF sample to school value-added measures, we assign the value-added measure to the first school attended in each of grades 3-8. The final outcome is the mean of the school value-added measure for schools that a child attends in grades 3-8.
- School avg. test scores: The raw average standardized test score for each school, used in the calculation of the value-added measure described above. We restrict to students not in the DCYF sample.
- School % Black: Fraction of Black students in the child's school, measured at the school-year level. We restrict to students not in the DCYF sample.
- School % IEP: Fraction of students who participate in special education (i.e., have an IEP) subsidies at the child's school, measured at the school-year level. We restrict to students not in the DCYF sample.

Note that we consider the characteristics only of the first school attended in cases where a child attends multiple schools in an academic year.

C.3.3 Enrollment Outcome Sample and Key Variables

For children in the DCYF sample, we create a panel from 2005 to 2016 that is balanced at the child and school-year level. This panel includes all investigated children who are expected to be enrolled in grades 3-8 based on their age. Using this definition, there are 3,971 girls and 4,470 boys who were investigated before age six. We create a yearly panel for this sample and join these data to RIDE public school records to determine enrollment in public school during grades 3-8. As in the schooling outcomes sample, children who were born before 1994 or after 2008 will not have observations because they are too old or too young. We consider only post-investigation years and create an indicator for whether a child was enrolled in that year. We also use the RIDE records to create an indicator for whether a child took a standardized test during grades 3-8. The key variables for children in the enrollment and test-taking sample are:

- Enrolled (=1): Indicator for enrollment in RI public school, defined as a panel outcome for children who are ages 8-13 or 9-14 (depending on date of birth) in a given school year.
- Tested (=1): Indicator for having taken a standardized test, defined as a panel outcome for children who are ages 8-13 or 9-14 (depending on date of birth) in a given school year.

C.3.4 Samples for Older Investigated Children

We also create several samples to analyze the outcomes of older investigated children (ages 6-18 at the time of an investigation). Each sample differs because some cohorts of older children may not be sufficiently old to be observed in several of the administrative data records that we use to measure outcomes. We provide full details on each of the distinct samples below.

For short-run outcomes, we examine test score and non-test score school outcomes for older children. The matched DCYF-RIDE sample for test scores contains 2,581 older girls and 2,965 older boys who have both standardized math and reading testing records in at least one year. The matched DCYF-RIDE sample for non-test-scores (grade retention, participation in special education, and absences) contains 3,029 older girls and 3,440 older boys. Note that this sample comprises only older children who have records in years *after* their first investigation. For example, we *do not* study a child's third grade standardized test score if the child was enrolled and took an exam in third grade at the time of the DCYF investigation. Instead, we focus on their post-investigation exams in grades 4-8.

For older children investigated at age 6-18, we also create samples to study the following later-life outcomes: delinquency, high school graduation, teen births, and college enrollment. We construct a different sample for each outcome, based on the time period available for each outcome and the investigated child's expected age. The restrictions ensure that outcomes are uncensored and that children are observable in the post-investigation period. (See the list below for further details on restrictions for each outcome.) The variables used in the analysis of outcomes for older investigated children are:

- Average z-score: The mean of a child's reading and math z-scores. All measures are based on scores observed after the year of the investigation.
- School index: The mean of the retention, IEP, and absences outcomes, where each outcome has been standardized by gender and age group (e.g., less than age six at the time of an investigation). All of the components of the index are based on outcomes observed after the year of the investigation.
- Delinquent (=1): Indicator for RITS enrollment or probation for wayward or delinquent offenses at ages 12-18. Eligible children are those investigated prior to the age of 12 and are born between 1988 and 1998 so that they are observable at ages 12-18 in juvenile delinquency records.
- HS Grad. (=1): Indicator for graduation from a RI high school at ages 18-19. Eligible children are investigated prior to the age of 18 and are born between 1985 and 1997 so that they are observable at ages 18-19 in RIDE public school records.
- Teen Birth (=1): Indicator for presence in the DOH vital records as a teen parent at ages 15-19. Eligible children are those investigated prior to the age of 15 and those born between 1985 and 1997 (ensuring that they are observable at ages 15-19 in vital records). Note that there are important limitations to how we measure teen parenthood for boys. Father information is often missing in birth records. Overall, 82 percent of all birth records from 2000-2016 have personally identifiable information for a father.
- College (=1): Indicator for any post-secondary educational institution enrollment at ages 18-20. Eligible children are investigated prior to the age of 18 and are born between 1986 and 1995 so that they are observable at ages 18-20 in NSC records.

C.3.5 Sample of Parent Perpetrators

For nearly all children in the DCYF sample (99 percent), we observe the set of perpetrators associated with allegations of abuse or neglect. We focus on parent perpetrators, which make

up 95 percent of the perpetrators for children in the schooling outcomes sample.[67] We join the DCYF parent perpetrators associated with victims in the schooling outcomes sample to criminal justice records (1995-2017). The outcomes that we consider are whether the perpetrators were ever charged with a crime or incarcerated in the post-investigation years. We construct a 4-year post-investigation measure that is partially censored for perpetrators investigated in 2014. The variables used in the analysis of perpetrators are described below:

- Charged/incar., 4-year post: Indicator for whether the parent perpetrator of abuse or neglect was charged with any crime or incarcerated in the 4-year post-investigation period.
- Charged/incar. for property crime, 4-year post: Indicator for whether the parent perpetrator of abuse or neglect was charged or incarcerated with a property-related crime in the 4-year post-investigation period.
- Charged/incar. for drug crime, 4-year post: Indicator for whether the parent perpetrator of abuse or neglect was charged or incarcerated with a drug-related crime in the 4-year post-investigation period.
- Charged/incar. for public offense crime, 4-year post: Indicator for whether the parent perpetrator of abuse or neglect was charged or incarcerated with a public order offense (e.g., disorderly conduct) in the 4-year post-investigation period.
- Charged/incar. for sex offense crimes, 4-year post: Indicator for whether the parent perpetrator of abuse or neglect was charged or incarcerated with sex-related crime in the 4-year post-investigation period.

We define charge and incarceration categories following guidelines from the Bureau of Justice Statistics (Durose et al., 2014).

C.4 Description and Statistics for Child Protection Investigators (CPI)

As detailed in Section 3 and Appendix C.3.1, we create a sample of 13,674 young children subject to a substantiated (founded) DCYF investigation. There are 102 Child Protective Investigators (CPIs) associated with these investigations. Table C3 reports statistics for the first and repeat investigations assigned to the 102 CPIs. First refers to whether the investigation is the initial case that we see for the associated child. We provide statistics for first and subsequent investigations because we use both in our preferred definition for the

[67] Note that restricting the sample to children with parent perpetrators does not imply perpetrators live in the same location (or home) as the child.

Table C3: CPI Summary Statistics

	(1) Mean	(2) p10	(3) p50	(4) p90	(5) N
All Years					
# Years	8.07	2.00	8.00	14.00	102
# Children	386.99	55.00	304.50	796.00	102
# Removed Children	69.61	10.00	60.00	142.00	102
Year Start	2003	2000	2000	2009	102
Year End	2011	2004	2013	2015	102
Period 2000-2007					
# Children	188.61	0.00	182.50	384.00	102
# Removed Children	35.87	0.00	32.50	73.00	102
Period 2008-2015					
# Children	198.38	0.00	66.00	575.00	102
# Removed Children	33.73	0.00	13.50	100.00	102

Notes: This table presents summary statistics for the sample of 102 CPIs that are associated with the children in the DCYF investigations sample.

instrumental variable. By using first and subsequent investigations, we have more information to use to infer removal tendencies.

To summarize, the average CPI handles investigations in eight of the years covered by the DCYF records (2000-2015). The average CPI makes decisions for 387 children and removed 70 children over the entire period that we observe them. The average CPI is first observed (in the administrative records) in 2003, and the median CPI is first observed in 2000. The average CPI is last observed in 2011, and the median CPI is last observed in 2013.

Note that we calculate the main instrument separately for the 2000-2007 and 2008-2015 periods. Table C3 also provides the average CPI statistics in each of these eight-year periods. (When a CPI is not observed in one of the two periods, we include a zero in computing these summary statistics. There are 31 CPIs who only appear in the 2000-2007 period. There are 12 CPIs who only appear in the 2008-2015 period.) In each period, the average CPI handles nearly 200 cases and removed around 35 children.

C.5 Sibling Statistics

As detailed in Section 4 of the main text and our discussion above, the sample created from the DCYF investigations records contains 13,674 total children investigated before age

six. These children are associated with 9,675 cases. In 6,760 of these cases (70 percent), there is only a single young child. The remaining 2,915 cases contain siblings. At the case level, the average number of young children is 1.41.

D Complier Calculations

This section provides details on how we estimate the characteristics and outcomes of compliers in our sample.

D.1 Estimating Complier Characteristics

In the child protective service context, we define compliers as children whose removal decision would have been different if they had been assigned to the most lenient (i.e., less likely to recommend a removal from home) instead of the strictest investigator (CPI). We follow the approaches developed by Abadie (2003), Dahl et al. (2014), and Dobbie, Goldin and Yang (2018) to characterize compliers in the sample of investigated children.

Let $\overline{z}$ denote the maximum value of the instrument (the most strict investigator) and $\underline{z}$ denote the minimum value of the instrument (the most lenient investigator). By the monotonicity and independence assumptions, we define the share of compliers as:

$$p_c = Pr(R_i = 1|Z_i = \overline{z}) - Pr(R_i = 1|Z_i = \underline{z}) = Pr(R_i(\overline{z}) > R_i(\underline{z})), \tag{D1}$$

where R_i is an indicator for removal. In practice, we assign the top percentile of our instrument to $\overline{z}$ and the bottom percentile of our instrument to $\underline{z}$. As discussed in Dahl et al. (2014) and Dobbie, Goldin and Yang (2018), the share of compliers can be directly estimated as $p_c = \alpha$, where α is the coefficient on the instrument from the first stage regression (Equation 2).

This is useful for studying the characteristics of compliers. For binary characteristic x_i, we know that:

$$\begin{aligned}\frac{Pr(x_i = 1|R_i(\overline{z}) > R_i(\underline{z}))}{Pr(x_i = 1)} &= \frac{Pr(R_i(\overline{z}) > R_i(\underline{z})|x_i = 1)}{Pr(R_i(\overline{z}) > R_i(\underline{z}))} \\ &= \frac{\mathbb{E}(R_i|Z_i = \overline{z}, x_i = 1) - \mathbb{E}(R_i|Z_i = \underline{z}, x_i = 1)}{\mathbb{E}(R_i|Z_i = \overline{z}) - \mathbb{E}(R_i|Z_i = \underline{z})}\end{aligned} \tag{D2}$$

This expression shows that the relative characteristics of compliers can be recovered by computing a ratio where the numerator is obtained by estimating the first stage coefficient for the subgroup $x_i = 1$ and constructing $\alpha_x(\overline{z} - \underline{z})$. The denominator is constructed similarly using the entire sample to estimate a first stage equation. (In Appendix Table A4, we multiply this ratio by $Pr(x_i = 1)$ to compute the average of a given characteristic for compliers.)

D.2 Estimating Complier Outcomes When Not-Removed

Our IV estimates are the causal impact of removal for compliers (i.e., the children whose removal decision would have been different if they had been assigned the most lenient instead

of the strictest investigator). In other words, the estimates tell us about the impact of removal for a child on the marginal case. To better understand this impact, it is helpful to have a benchmark comparison by estimating complier outcomes when removal *does not* occur. To answer this question, we need to estimate the untreated potential outcome (denoted Y_{i0}) for compliers:

$$\mathbb{E}(Y_{0i}|R_i(\overline{z}) > R_i(\underline{z})) \tag{D3}$$

As discussed in Dahl et al. (2014), this can be obtained by examining children who are assigned to lenient and strict investigators: For non-removed children (i.e., those with $R_i = 0$) assigned to $\underline{z}$, we know:

$$\begin{aligned}\mathbb{E}(Y_i|R_i = 0, Z_i = \underline{z}) = &\left(\frac{p_c}{p_c + p_n}\right)\mathbb{E}(Y_{0i}|R_i(\overline{z}) > R_i(\underline{z})) \\ &+ \left(\frac{p_n}{p_c + p_n}\right)\mathbb{E}(Y_{0i}|R_i(\overline{z}) = R_i(\underline{z}) = 0)\end{aligned} \tag{D4}$$

where Y_i is the observed outcome, p_c is the share of compliers, and p_n is the share of never-takers (i.e., children who would never be removed by the most or least strict investigator).

The outcomes for never-takers can be inferred from the outcomes of the non-removed children who are assigned the strictest investigator:

$$\mathbb{E}(Y_{0i}|R_i(\overline{z}) = R_i(\underline{z}) = 0) = \mathbb{E}(Y_i|R_i = 0, Z_i = \overline{z}) \tag{D5}$$

Equation D5 allows us to disentangle the mixture from Equation D4. Specifically, we can re-write Equation D4 as:

$$\begin{aligned}\mathbb{E}(Y_{0i}|R(\overline{z}) > R(\underline{z})) = &\left(\frac{p_c + p_n}{p_c}\right)\mathbb{E}(Y_i|R_i = 0, Z_i = \underline{z}) \\ &- \left(\frac{p_n}{p_c}\right)\mathbb{E}(Y_i|R_i = 0, Z_i = \overline{z})\end{aligned} \tag{D6}$$

To evaluate this expression, we estimate the share of always-takers, never-takers and compliers in the sample.[68] With these quantities, we solve Equation D6 by estimating a linear model for Y_i and z_i in the subsample of non-removed children (i.e., $R_i = 0$). In this specification, we control for investigation year fixed effects.

[68] Recall that $p_a = Pr(R_i = 1|Z_i = \underline{z})$ and $p_n = Pr(R_i = 0|Z_i = \overline{z})$.

E Machine Learning IV Approach and Results

We conduct additional robustness tests using a machine learning (ML) approach to re-estimate the impact of removal on child outcomes. The approach follows Mueller-Smith (2015), which studies the impact of adult incarceration by creating multiple instruments and uses an ML approach to choose instruments with the highest predictive power. In our case, we create a range of CPI removal tendency measures that vary with different case characteristics. We use LASSO regressions to select from the set of potential instruments and use the chosen instruments in our two-stage least squares models. The following sections describe the implementation of this approach and reports these results.

E.1 Constructing Flexible Leave-out Measures for Machine Learning

As a first step, we build CPI removal tendency measures that vary with case characteristics. We focus on the following five characteristics:

1. gender;
2. minority (ethnic/race) status (non-minority and minority, respectively);
3. marital status;
4. reporter type;
5. allegation type.

We define mutually exclusive groups for each of the following case characteristics: the gender of the child, whether the child belongs to a minority group (as measured by being black or Hispanic), the parent's marital status, the type of reporter making the allegation, and the type of allegation. Then, we re-calculate the instrument for each CPI and case characteristic. For example, each CPI will have a leave-out removal tendency calculated separately for minority (non-white) and non-minority (white) children. We do this for five characteristics and create five versions of leave-out removal tendency measures. (We do not consider any interactions between case characteristics.)

To parallel our main measure of CPI removal tendency, we create the case characteristic-specific instruments over two eight-year periods (2000-2007 and 2008-2015). By calculating the measures using an eight-year period, we address concerns that a CPI may see relatively few children with a given case characteristic in a shorter period (e.g., one year). Table E1 provides statistics for CPIs on the number of investigated children by types of case characteristics. For example, the first rows show that the average CPI investigates about 387 children over

the entire 2000-2015 period. In addition, the average CPI investigated 150 and 237 and minority (non-white) and non-minority (white) children, respectively. In our implementation, we address concerns over small cell-sizes by defining a given CPI tendency measure to be missing when there are fewer than 10 children available to construct the leave-out measure. For example, if a CPI investigates only nine other children whose cases involved a physical neglect allegation over the relevant period, then we define the instrument that varies at the allegation type level to be missing for this CPI and child.

E.2 Machine Learning Implementation Details

We test the robustness of our main results using an alternative machine-learning (ML) approach that allows the instrument to vary with case characteristics. We consider five types of case characteristics: (1) gender, (2) minority (ethnic/race) status (non-minority and minority, respectively), (3) marital status, (4) reporter type, and (5) allegation type. We have five potential instruments to use in our first-stage removal equation. Following Mueller-Smith (2015), we use LASSO to select the instruments with the greatest predictive power (Belloni et al., 2014). For each outcome, note that we estimate separate LASSO regressions of removal on the five removal tendencies by case characteristic measures to select instruments for the first stage. Each of these regressions always includes investigation year fixed effects and case characteristic controls (as these controls are included in our two-stage least squares specifications). Table E2 reports which instruments are selected by LASSO for each outcome for young girls and young boys. For young girls, our approach selects multiple instruments (i.e., the ones based on gender, minority status, and reporter status) for the analysis of test scores.[69] For young girls and their additional schooling outcomes (i.e., the school index, retention and IEP outcomes), the approach only selects the instrument that varies based on minority status. For young boys, the test score outcomes (average, math, and reading) as well as the school index, retention and IEP outcomes only use the instrument that varies by minority status.

Appendix Table E3 presents the ML IV estimates for impacts of removal on test scores, grade retention, special education (IEP), and the school index measure. For young girls, the ML IV estimates are similar to the main results that we report in Tables 4 and 5. That is, Panel A shows that we consistently find significant and positive impacts of removal on test scores for young girls. Similarly, the ML IV estimates indicate that removal significantly reduces the likelihood of ever being retained and the likelihood of ever participating in IEP for young girls. In Panel B, the ML IV estimates for boys are never statistically significant.

[69]Note that in this case, where multiple instruments are used, the 2SLS estimates can retain a causal interpretation as a positively weighted average of LATEs (Imbens and Angrist, 1994; Mogstad et al., 2019).

Relative to our main results for test scores, the ML IV results for boys differ from Table 4 in that they are relatively large negative point estimates.

Table E1: CPI Summary Statistics by Case Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	2000-2015 (# CPIs=102)				2000-2007 (# CPIs=90)				2008-2015 (# CPIs=71)			
	Mean	p10	p50	p90	Mean	p10	p50	p90	Mean	p10	p50	p90
All children	387.0	55.0	304.5	796.0	213.8	47.0	201.0	390.5	285.0	12.0	213.0	638.0
Girls	193.6	27.0	149.5	405.0	108.2	24.0	103.0	196.5	140.9	5.0	100.0	309.0
Boys	193.5	28.0	152.5	395.0	105.6	21.0	101.5	199.5	144.1	6.0	109.0	328.0
Non-Minority	236.7	34.0	196.0	488.0	137.8	28.0	134.5	258.5	165.3	8.0	119.5	351.0
Minority	150.3	18.0	108.0	352.0	75.9	16.5	68.5	139.0	119.7	2.0	83.0	275.0
Married couple	63.5	8.0	56.0	129.0	40.5	6.0	38.0	75.0	39.9	2.0	26.5	94.0
Unmarried couple	89.9	8.0	55.0	221.0	35.4	4.5	30.5	69.5	83.2	3.0	56.5	193.0
Single/Other	233.5	40.0	202.0	468.0	137.8	33.0	131.5	262.0	160.9	5.0	121.0	349.0
Neglect	288.5	38.0	208.5	624.0	150.0	33.5	138.5	267.5	224.4	10.0	152.0	505.0
Physical neglect	14.2	1.0	12.0	31.0	9.3	0.5	9.0	18.0	8.6	1.0	7.0	20.0
Physical abuse	69.8	11.0	64.0	141.0	44.3	9.5	43.5	85.5	44.1	1.0	39.0	96.0
Professional reporter	309.5	37.0	227.0	629.0	164.1	32.5	157.0	310.5	236.5	12.0	165.0	517.0
Other reporter	20.6	4.0	19.0	43.0	14.1	1.0	11.0	27.0	11.6	0.0	8.0	27.0
Family/friend reporter	57.0	9.0	49.0	115.0	35.5	7.0	35.0	66.0	36.0	0.0	28.0	88.0
Routine	111.4	14.0	81.0	235.0	60.7	11.5	55.0	119.0	83.01	4.0	57.0	174.0
Immediate	237.9	33.0	186.5	534.0	128.3	26.5	123.0	237.5	179.0	8.0	134.0	393.0
Emergency	37.8	7.0	32.5	70.0	24.7	6.0	23.5	46.0	23.0	1.0	20.0	50.0

Notes: This table presents summary statistics for the sample of 102 CPIs that are associated with the children in the DCYF investigations sample. The rows provide summary statistics based on case characteristics. For example, the second row provides summary statistics for the number of girls involved in a CPI's cases during different time periods. Column 1 shows that the average CPI had 193 girls in their cases during 2000-2015.

Table E2: Instrument(s) Selected by LASSO for ML Approach

	Young Girls (Age < 6)		Young Boys (Age < 6)	
	(1)	(2)	(3)	(4)
Instruments by case characteristics:	Test Score Outcomes	Schooling Outcomes	Test Score Outcomes	Schooling Outcomes
Gender	Yes			
Minority	Yes	Yes	Yes	Yes
Marital status				
Reporter type	Yes			
Allegation type	Yes			
Investigation level			Yes	

Notes: This table reports the versions of the instruments selected by LASSO in a regression of removal on five instruments, where each instrument varies based on the case characteristic listed. Columns 1-2 report the selected instruments (denoted by "Yes") for the test score and schooling outcomes of young girls. Columns 3-4 report the selected instruments for the test score and schooling outcomes of young boys. The LASSO regressions always specify investigation year fixed effects and case characteristic controls as variables selected.

Table E3: Robustness to Estimating Impacts using ML-IV Approach

	Panel A. Young Girls (Age < 6)					
	Test Score Outcomes			Additional Schooling Outcomes		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score	Math z-score	Reading z-score	School Index	Ever Retained	Ever IEP
Removed (= 1)	1.177**	1.283**	1.074*	-0.930*	-0.232**	-0.503**
	(0.469)	(0.470)	(0.499)	(0.390)	(0.108)	(0.244)
Mean of dependent variable	-0.392	-0.460	-0.327	-0.001	0.045	0.248
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	7.058	6.873	7.053	25.869	25.869	25.869
N	9,577	9,598	9,610	2,485	2,485	2,485
Individuals	2,496	2,496	2,496	2,485	2,485	2,485
	Panel B. Young Boys (Age < 6)					
	Test Score Outcomes			Additional Schooling Outcomes		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score	Math z-score	Reading z-score	School Index	Ever Retained	Ever IEP
Removed (= 1)	-0.308	-0.498	-0.098	-0.399	-0.146	-0.111
	(0.567)	(0.606)	(0.566)	(0.410)	(0.123)	(0.295)
Mean of dependent variable	-0.571	-0.519	-0.630	0.003	0.064	0.418
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	6.488	6.401	6.655	14.809	14.809	14.809
N	12,204	12,245	12,266	3,076	3,076	3,076
Individuals	3,098	3,098	3,098	3,076	3,076	3,076

Notes: This table reports results for young girls (Panel A) and young boys (Panel B) based on an IV approach where the CPI removal rates vary with case characteristics. See Appendix E for details on the IV calculations. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

F Additional Discussion of Impacts for Older Children

As discussed in Section 7, we hope to estimate the causal impact of home removal for older children investigated at ages 6-18. To assess the validity of our IV approach, we examined the relationship between CPI removal tendency and case characteristics for older children. The randomization test results in Appendix Table A23 show that, while we do not reject our null hypothesis of no joint significance of case characteristics in the sample of older boys, we reject the null hypothesis in the sample of older girls. Examining the regression results for older girls in Column 2 shows that there are four case characteristics (out of fourteen) that have coefficients that are significant at the 10-percent level or lower. The largest statistically significant coefficient is equal to roughly one quarter of a standard deviation of CPI removal tendency.

To help assess whether this imbalance threatens the validity of IV estimates for older children, we conduct two tests. First, we examine estimates of the impact of removal with and without controls for case characteristics. Altonji et al. (2005) suggest that assessing whether point estimates are sensitive to the inclusion of controls provides information on the extent of selection bias. In Table F1, we restrict our analysis to older (ages 6-18) investigated children and present IV estimates for test scores, the school index (which is based on retention, special education participation and absences) and college attendance with and without case characteristic controls. For older girls, there is no strong pattern of coefficient sensitivity for these outcomes. For example, the point estimates for the school index are -0.341 and -0.373, respectively. The point estimates for older boys also display no strong pattern of sensitivity, which is expected given that the results in Appendix Table A23 provide no evidence of a relationship between case characteristics and CPI removal tendency for older boys.

In our second test, we assess the validity of our IV approach by examining test scores in the periods before an investigation begins for older children.[70] Due to the random assignment of cases, we expect that there should be no statistically significant relationship between removal (and our instrument) and the "pre-treatment" test score outcomes. To conduct this test, we construct a panel of test scores for older investigated children that includes observations from school years that *precede* the year of the first investigation. For most older removed children, we observe two test scores that precede the year of the first investigation. For the purpose of comparison, we also include observations in the panel for the year of the investigation and three school years that follow.

Using the panel of test scores for older children, we estimate separate IV models where the dependent variable is the average of standardized test scores in a given school year.

[70]Note that we cannot conduct this type of analysis for young children because their first investigation occurs before they enter testing grades.

We estimate six models starting with observations that are two years before the year of a DCYF investigation and ending with the school year that is three years after a DCYF investigation. Figure F1 displays the point estimates and confidence intervals associated with these estimates. The x-axis displays the year relative to the investigation. For example, the left-most point estimate for older girls shows that there is an insignificant 0.19 standard deviation impact of removal on test scores that occur two years prior to the investigation. Across the school years that we examine, there are no statistically significant impacts of removal (including in the years that follow an investigation). The results for test scores that occur two years and one year before an investigation provide no strong evidence that CPI removal tendency is correlated with child characteristics, although the standard errors associated with our estimates are large and the confidence intervals span from -1 to 1 standard deviation.

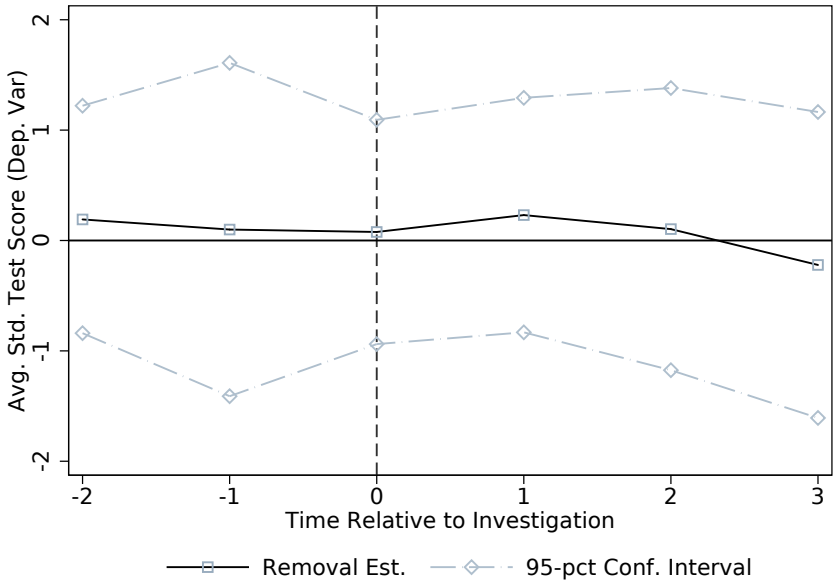
Table F1: Impact of Removal on Outcomes of Older Children, Sensitivity Test

	School-age outcomes		Later-life outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Average z-score		School Index		College (=1)	
Panel A. Older Girls (Age $\geq$ 6)						
Removed (= 1)	-0.109	-0.230	-0.341	-0.373	0.155	0.133
	(0.625)	(0.582)	(0.347)	(0.326)	(0.228)	(0.222)
Mean of dependent variable	0.068	0.068	-0.005	-0.005	0.303	0.303
Complier mean if not removed	-0.337	-0.337	0.138	0.138	0.302	0.032
Case controls	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	24.050	22.129	37.376	35.092	36.367	38.718
N	7,517	7,517	3,029	3,029	3,326	3,326
Individuals	2,581	2,581	3,029	3,029	3,326	3,326
Panel B. Older Boys (Age $\geq$ 6)						
Dependent variable:	Average z-score		School Index		College (=1)	
Removed (= 1)	-0.250	-0.237	0.352	0.323	-0.147	-0.127
	(0.458)	(0.429)	(0.219)	(0.216)	(0.194)	(0.187)
Mean of dependent variable	0.053	0.053	-0.003	-0.003	0.239	0.239
Complier mean if not removed	-0.414	-0.414	-0.297	-0.297	0.367	0.367
Case controls	No	Yes	No	Yes	No	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic (instrument)	27.910	30.911	30.435	34.273	37.069	41.145
N	8,838	8,838	3,440	3,440	2,953	2,953
Individuals	2,965	2,965	3,440	3,440	2,953	2,953

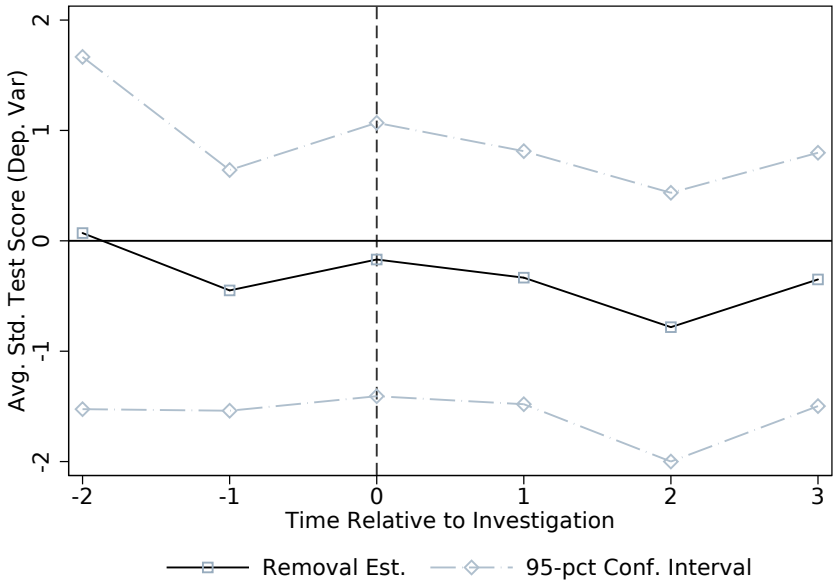
Notes: This table reports results for the impact of removal on outcomes for older girls (Panel A) and older boys (Panel B). Older is defined as being investigated at ages six or later (up to age 18). All results are two-stage least squares models with the standard leave-out measure of CPI removal tendency as an instrument for removal. The school index is constructed based on standardized measures of whether an investigated child was ever retained, ever participated in special education (i.e., has an IEP), and the average number of days absent during grades 3-8. Standard errors are two-way clustered at the family and CPI level in parentheses. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Figure F1: Impact of Removal on Test Scores of Older Children, by Time Relative to Investigation

(a) Older Girls



(b) Older Boys



Notes: These figures show results for the impact of removal on test scores estimated in separate regressions by time relative to the year of investigation for older girls (Panel A) and older boys (Panel B). All results are two-stage least squares models with the standard leave-out measure of CPI removal tendency as an instrument for removal. All models include controls for the case characteristics in Table 1 and investigation year fixed effects. Confidence intervals are based on standard errors that are two-way clustered at the family and CPI levels.

G Additional Tables for Enrollment and Schooling Outcomes Samples

Table G1: First-Stage Results (Additional Results for Young Children)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	Investigations Sample		Enrollment Outcomes Sample		Schooling Outcomes Sample	
Dependent variable:			Removed (=1)			
CPI removal tendency	0.594***	0.582***	0.629***	0.500***	0.649***	0.403***
	(0.096)	(0.069)	(0.127)	(0.092)	(0.166)	(0.113)
Age/gender group	Young Girls	Young Boys	Young Girls	Young Boys	Young Girls	Young Boys
Mean of dependent variable	0.208	0.199	0.201	0.195	0.181	0.174
Case controls	Yes	Yes	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
N (Individuals)	6,287	7,387	3,971	4,770	2,614	3,142

Notes: This table summarizes the first-stage impact of CPI removal tendency. Columns 1-2 report results for all young children (investigated before age six) included in the investigations sample described in Section 3.1. Columns 3-4 report results for the children in the enrollment outcome sample. These are the cohorts of children who were age-eligible to attend grades 3-8 during the period in which we observe test scores (i.e., the academic years 2005-2016). Columns 5-6 report results for the investigated children who matched to the school test score and enrollment records. The first-stage results are from a regression of removal on CPI removal tendency, controls for case characteristics, and investigation year fixed effects (FE). Removed is an indicator for home removal at the child's first investigation. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table G2: Tests of Randomization (Additional Results for Young Children)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	Investigations Sample		Enrollment Outcome Sample		Schooling Outcomes Sample	
Dependent variable:	CPI Removal Tendency					
Black	-0.001	0.000	-0.002	0.002	-0.004	-0.000
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)
Hispanic	0.003	0.001	0.002	0.001	0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.006)
Other race	0.001	0.004	-0.001	0.004	-0.003	-0.001
	(0.003)	(0.003)	(0.005)	(0.004)	(0.005)	(0.006)
Age	-0.000	0.000	-0.000	0.001	0.000	0.000
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Married couple	-0.002	0.002	-0.002	0.000	-0.005	-0.003
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)
Unmarried couple	-0.001	-0.001	0.000	0.001	-0.002	0.000
	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
English language	-0.002	0.003	-0.001	0.002	0.000	0.001
	(0.004)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)
Neglect	0.001	-0.002	0.002	-0.002	-0.001	-0.004
	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)
Physical neglect	0.002	0.005*	0.000	0.003	0.000	0.005
	(0.004)	(0.003)	(0.005)	(0.004)	(0.006)	(0.005)
Professional reporter	-0.004	-0.004	0.004	-0.002	0.002	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)
Family/friend reporter	-0.006	-0.002	-0.002	0.001	-0.004	-0.002
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)
Emergency investigation	-0.001	0.001	0.000	0.001	0.001	-0.002
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)
Immediate investigation	0.001	0.002	0.004**	0.003	0.005*	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Age/gender group	Young Girls	Young Boys	Young Girls	Young Boys	Young Girls	Young Boys
Chi-squared statistic	10.249	17.679	14.228	14.931	18.696	16.974
p-value of joint significance	0.673	0.170	0.358	0.312	0.133	0.201
Mean of CPI removal tendency	0.176	0.180	0.178	0.184	0.178	0.183
Investigation year FE	Yes	Yes	Yes	Yes	Yes	Yes
N (Individuals)	6,287	7,387	3,971	4,770	2,614	3,142

Notes: This table summarizes tests of random case assignment. Columns 1-2 report results for the young children (investigated before age six) included in the investigations sample described in Section 3.1. Columns 3-4 report results for the children in the enrollment outcome sample. These are the cohorts of children who were age-eligible to attend grades 3-8 during the period in which we observe test scores (i.e., the academic years 2005-2016). Columns 5-6 report for the investigated children who matched to the school test score and enrollment records. The test statistics are from a regression of CPI removal tendency on the set of case characteristics and investigation year fixed effects. The chi-square test-statistic and p-value reported are from a test for joint significance of all variables except investigation year fixed effects. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table G3: Exclusion Restriction Tests (Schooling Outcomes Sample)

	Panel A. Removed Young Girls (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Days in any foster care	Number of placements	Placed with relative (=1)	Police Notified (=1)
CPI removal tendency	-253.238	0.985	-0.289	-0.095
	(363.579)	(1.512)	(0.364)	(0.132)
Mean of dependent variable	412.541	2.068	0.425	0.949
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
N	473	473	473	473
	Panel B. Removed Young Boys (Age < 6)			
	(1)	(2)	(3)	(4)
Dependent variable:	Days in any foster care	Number of placements	Placed with relative (=1)	Police Notified (=1)
CPI removal tendency	-130.882	0.051	-0.101	-0.146
	(375.356)	(1.547)	(0.344)	(0.170)
Mean of dependent variable	457.461	2.233	0.353	0.966
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes
N	546	546	546	546

Notes: The sample for this analysis is the set of removed children. This analysis is based on the schooling outcomes sample described in Section 3.1. The table reports regression results testing whether placement and other investigation outcomes of removed children are correlated with CPI removal tendency. Standard errors in parentheses are two-way clustered at the family and CPI level. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table G4: First-Stage Impact of CPI Removal Tendency, by Subgroup (Schooling Outcomes Sample)

	(1)	(2)	(3)	(4)
IV Version:	Subgroup-specific Measure		Subgroup-specific Measure, Reverse Sample Calculation	
Dependent Variable:	Removed (=1)		Removed (=1)	
White	0.483**	0.296**	0.294**	0.171
	(0.210)	(0.133)	(0.148)	(0.108)
	[0.170]	[0.163]	[0.170]	[0.163]
	N=1,414	N=1,719	N=1,408	N=1,707
Non-white	0.856***	0.503**	0.662***	0.400**
	(0.258)	(0.202)	(0.169)	(0.165)
	[0.193]	[0.187]	[0.193]	[0.187]
	N=1,200	N=1,423	N=1,194	N=1,411
Single/other parent	0.755***	0.483***	0.690***	0.438***
	(0.202)	(0.150)	(0.231)	(0.151)
	[0.204]	[0.198]	[0.204]	[0.198]
	N=1,604	N=1,951	N=1,592	N=1,930
Non-single/other parent	0.481**	0.241	0.279	0.134
	(0.221)	(0.173)	(0.247)	(0.188)
	[0.144]	[0.134]	[0.143]	[0.134]
	N=1,010	N=1,191	N=1,001	N=1,178
Neglect	0.730***	0.268**	0.606***	0.212**
	(0.193)	(0.131)	(0.142)	(0.107)
	[0.170]	[0.163]	[0.170]	[0.163]
	N=2,137	N=2,485	N=2,114	N=2,446
Professional reporter	0.668***	0.480***	0.391***	0.285***
	(0.179)	(0.128)	(0.112)	(0.092)
	[0.177]	[0.171]	[0.175]	[0.170]
	N=2,026	N=2,476	N=1,991	N=2,427
Sample:	Young Girls	Young Boys	Young Girls	Young Boys
Case controls	Yes	Yes	Yes	Yes
Investigation year FE	Yes	Yes	Yes	Yes

Notes: This table summarizes the first-stage relationship between removal and CPI removal tendency for subgroups. Subgroups are based on the case characteristics listed in Table 1. The subgroups for physical abuse, physical neglect, non-professional reports, and emergency cases are not reported because these have relatively few observations. We also omit reporting results based on language since 97 percent of cases are English language. This analysis uses the schooling outcomes sample described in Section 3.1. Standard errors in parentheses are two-way clustered at the family and CPI level. Means for removal for each subgroup are reported in brackets. Significance reported as $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.