HCEO WORKING PAPER SERIES

The Causal Impact of Removing Children from Abusive and Neglectful Homes

Anthony Bald
Eric Chyn
Justine S. Hastings
Margarita Machelett

Working Paper

2019-001

01/2019

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

THE CAUSAL IMPACT OF REMOVING CHILDREN FROM ABUSIVE AND NEGLECTFUL HOMES

Anthony Bald
Eric Chyn
Justine S. Hastings
Margarita Machelett

The opinions and conclusions expressed are solely those of the authors. We thank 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 Colleen Caron, Joseph Doyle, Sara Heller, Brian Jacob, Jonah Rockoff, Maya Rossin-Slater, Bruce Sacerdote, Emilia Simeonova, Sarah Turner, Christina Weiland, and seminar participants at the University of Michigan and the 2018 Annual Meeting of the Southern Economic Association. Finally, we are grateful for financial support from the Laura and John Arnold Foundation and the Smith Richardson Foundation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Causal Impact of Removing Children from Abusive and Neglectful Homes
Anthony Bald, Eric Chyn, Justine S. Hastings, and Margarita Machelett
January 2019
JEL No. H75,I21,I24,I28,I38,J12,J13,J24

ABSTRACT

This paper uses administrative data to measure causal impacts of removing children from families investigated for abuse or neglect. We use the removal tendency of quasi-experimentally assigned child protective service investigators as an instrument for whether authorities removed and placed children into foster care. Our main analysis estimates impacts on educational outcomes by gender and age at the time of an investigation. We find that removal significantly increases standardized test scores for young girls. There are no detectable impacts on the test scores of girls removed at older ages or boys of any age. For older children, we also find few significant impacts of removal on the likelihood of having a juvenile conviction, graduating from high school, enrolling in a postsecondary institution, or having a teenage birth. We investigate potential mechanisms driving heterogeneous impacts by gender and age. Our results do not appear to be driven by heterogeneous effects on foster care placement, school mobility and quality, or participation in special education programs. For girls, we find that removal significantly increases the likelihood of post-investigation criminal charges or incarceration for parents and caretakers who are the perpetrators of abuse or neglect.

Anthony Bald
Brown University
70 Waterman St.
Providence, RI 02912
anthony_bald@brown.edu

Eric Chyn
Department of Economics
University of Virginia
P.O. Box 400182
Charlottesville, VA 22904
ericchyn@virginia.edu

Justine S. Hastings
Brown University
Department of Economics
64 Waterman Street
Providence, RI 02912
and NBER
justine_hastings@brown.edu

Margarita Machelett
Department of Economics
Brown University
margarita_machelett@brown.edu

1 Introduction

Each year, child protective service agencies in the U.S. investigate more than 4 million allegations of abuse or neglect (U.S. HHS, 2016). As a result of these investigations, authorities annually remove nearly 200,000 children from their homes and place these children into foster care (U.S. HHS, 2016). The goal of removal is to protect children by reducing exposure to abuse and neglect. 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; Aizer and Doyle, 2018).[1]

There is relatively little evidence on the causal impact of child protective service removal on children. 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 instrumental variable (IV). He studied later-life outcomes of children subject to investigation between the ages of five and fifteen using data from Illinois, finding that removal increases delinquency and arrests and decreases labor market outcomes.

This paper 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. Our panel is well-suited to studying children removed at very young ages, allowing us to measure the causal impact of removal in early childhood (prior to the age of six). Nearly half of removed children are under the age of six (U.S. HHS, 2016), and the literature on child development suggests that early life events and interventions can have particularly strong influences on outcomes (Cunha et al., 2006; Cunha and Heckman, 2007; Almond and Currie, 2011a; Heckman and Mosso, 2014; Almond et al., 2017). Our results are the first to estimate causal impacts of removal for this important group of children.

We use the removal tendency of investigators of child abuse as an instrument for removal. We take two approaches. First, we use the standard leave-one-out measure of decision making of in-

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

vestigators. This type of measure has been used for judges and authorities in prior literature (Kling, 2006; Doyle, 2007, 2008; Aizer and Doyle, 2015; Bhuller et al., 2016; Sampat and Williams, 2017; Dobbie et al., 2018a,b; Bhuller et al., 2018). We calculate the removal rate for all other cases assigned to an investigator within a year using data from the Rhode Island Department of Children, Youth and Families (DCYF).[2] Second, we allow the removal rate to vary by child and case characteristics, and use this in an IV approach that relaxes the monotonicity assumption associated with the standard leave-one-out measure (Mueller-Smith, 2016). To do this parsimoniously and avoid overfitting, we use LASSO, a machine learning (ML) technique, to select the child and case characteristics that define subgroups used to calculate removal tendency (Belloni et al., 2014). We provide results from both the standard IV and ML-IV specifications by subgroups of children based on gender and age at the time of an investigation (before the age of six and six years old or later).[3] Our analysis of effects by gender is motivated by prior research, which shows that girls and boys may respond differently to family conditions and circumstances (Bertrand and Pan, 2013; Autor et al., 2016).

We find significant and positive effects of removal on standardized test scores (grades 3-8) for girls. The point estimate from the ML-IV approach indicates a 0.460 student-level standard deviation increase in average scores (math and reading) in the years after removal. The impacts are larger for young girls, although the standard errors also increase for this group (the subgroup estimates are statistically significant). For this subgroup, the point estimates indicate large impacts of 1 student-level standard deviation using either the IV or ML-IV specifications. These 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 test scores (Heckman et al., 2013). For older girls, we find non-significant impacts, and the point estimates suggest negative effects.

In contrast to the effects of removal for girls, we find no significant impacts on standardized test scores for the sample of all boys. The imprecise point estimates suggest negative impacts using either the IV or the ML-IV approach. There are also no statistically significant impacts for the

[2] In our sample, the leave-one-out removal rate is a statistically significant predictor of removal and is uncorrelated with child and case characteristics.

[3] Note that age six is the compulsory school starting age in Rhode Island during our sample period (Rhode Island General Law, 2016). Benson et al. (2018) provide evidence that reports of child maltreatment increase when children enroll in school. Their findings are consistent with the idea that educators are an important source for reporting abuse and neglect.

subgroups of younger or older boys. For young boys, the point estimates are near zero using either the standard IV or ML-IV specification. We can statistically reject the hypothesis that the effects are equal for young girls and young boys (p-value < 0.10).

We examine whether the test score 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 and age subgroups, we find that the estimates for young girls are significant at the 10 percent level using the FDR-adjusted q-values. We also study attrition and find no statistically significant impacts of removal on enrollment for young girls or young boys. The point estimates for these two groups using either the IV or ML-IV specification are not statistically different. This suggests that attrition due to changes in public school enrollment is unlikely to explain our main finding of gender-specific differences in test score impacts for young children.

Next, we conduct four investigations into potential mechanisms driving the impacts of removal on test scores by gender and age. First, we examine whether there are differences in case characteristics or foster care placement outcomes. Case characteristics are similar for girls and boys of the same age in our sample of investigations.[4] This does not align with the pattern of heterogeneous impacts of removal on test scores by gender. Similarly, we find no pattern of differences in foster care placement that corresponds to the test score results. Removal significantly increases the number of days spent in foster care for all subgroups, and we find statistically similar impacts for young girls and young boys.

Second, we consider the impact of removal on school stability and the types of schools that children attend. We find some evidence of significant and positive impacts on the number of schools that a child attends in the three years following removal, primarily among children investigated at ages six and older.[5] We cannot reject the hypothesis of equal impacts of removal for young girls and young boys. As measures of school characteristics, we use statewide enrollment and testing data to compute school-level value added, the fraction of students who receive free or subsidized

[4] We exclude sex abuse reports from all analyses because they make up only 5 percent of all investigations.

[5] For children who have not yet started school at the time of an investigation, we use the first three years of school observed.

lunch, and the fraction of enrolled students who are minorities.[6] We find no statistically significant impacts of removal on any measure of school type by gender and age subgroup.

Third, we examine child experiences within school by testing whether removal has impacts on participation in special education services. We measure participation based on whether the child has a written Individualized Education Program (IEP).[7,8] Our analysis shows no detectable impacts of removal on IEP participation during grades K-6 for any group of children. At the same time, the point estimates for girls and boys removed at young ages are negative and positive, respectively. We can reject the equality of the point estimates for young girls and young boys (p-value < 0.10), implying that removed girls and boys have different rates of IEP participation. To the extent that young removed boys have higher rates of IEP enrollment relative to girls, this could reflect heterogeneous impacts on underlying schooling ability (which is consistent with the pattern of test score impacts for young girls and young boys). Alternatively, the enrollment pattern could also be due to changes in the way school administrators, teachers, or parents differentially treat removed boys.

Fourth, we examine changes in household conditions and structure by studying criminal charges and incarceration for perpetrators of abuse and neglect. More than 95 percent of perpetrators in our sample are parents or caretakers such as relatives. For all girls, results from the IV specification show that removal significantly increases the likelihood of charges and incarceration for perpetrators in the four years following the conclusion of an investigation. There are no detectable impacts on perpetrators of abuse for male children, though we are unable to reject the hypothesis of equal impacts across subgroups. Impacts on perpetrators are not likely to mediate test score effects, unless they operate differentially through gender and the age of the child.

As a final analysis, we use additional administrative data to study later-life outcomes for children removed at ages six and older. This analysis focuses on older children because a child removed at a young age will not be old enough by the end of the period covered by the data. We estimate impacts on the likelihood of having a juvenile conviction, graduating from high school,

[6] We construct these measures after excluding children investigated by DCYF.

[7] 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.

[8] Note that having an IEP does not generally exempt a student from testing in Rhode Island. In academic year 2013, 89 percent of Rhode Island students with an IEP took standardized exams.

enrolling in a postsecondary institution, and having a teenage birth. For older children of either gender, we tend to find no statistically significant effects on any outcome. The only marginally significant effect that we find is a positive impact on teen fatherhood for older removed boys using the IV specification (p-value < 0.10). The point estimates also tend to suggest that removed boys have worse outcomes in terms of juvenile convictions and high school graduation. For older girls, the non-significant point estimates have signs that imply either positive (beneficial) or negative (detrimental) impacts. The estimates are sufficiently imprecise that we cannot reject that the effects of removal are equal for older girls and boys.

Overall, these findings contribute to a broad literature on the impact of interventions for children from disadvantaged backgrounds. In particular, the results add to a growing literature which shows that early-life interventions can have large, causal impacts on children's outcomes (Garces et al., 2002; Ludwig and Miller, 2007; Almond et al., 2010; Chetty et al., 2011; Bharadwaj et al., 2013; Heckman et al., 2013; Campbell et al., 2014; Aizer et al., 2016; Chetty et al., 2016; Hoynes et al., 2016; Isen et al., 2017; Chyn, 2018; Chyn et al., 2018; Currie et al., 2018).[9] Our results extend this literature by focusing on interventions for young children at risk of abuse and neglect. We do not find strong evidence of differential pathways in the foster care or school systems by age and gender, suggesting that impacts of removal found among young girls may be particular to age. This paper also complements a growing set of studies 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, 2006; Anderson, 2008; Angrist et al., 2009, 1996; Heckman et al., 2013; Deming et al., 2014; Hoynes et al., 2016).

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

An investigation into child abuse or neglect in Rhode Island begins when a suspected allegation is reported to the DCYF Child Protective Services (CPS) hotline.[10] Professional call staff at DCYF's

[9] See Almond and Currie (2011b) and Almond et al. (2017) for a review of the literature on the impact of interventions for children.

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

central office record details of the allegation, identify previous or pending investigations, and determine whether the report meets the criteria for an investigation. If it does not meet the criteria for an investigation, the report is kept in the system and expunged after three years.[11] If it does meet the criteria for an investigation, the call floor staff forward the case to the Investigative Unit where a supervisor assigns the case to an available Child Protective Investigator (CPI).

The Investigative Unit Supervisor assigns cases to CPIs randomly using an internal "rotation list." The rotation list is an ordered spreadsheet of CPIs maintained by the Investigative Unit supervisor. Each day, the supervisor assigns new cases as they arrive based on this list. CPIs usually receive one case per day. If all CPIs receive a case, the next day the rotation list remains the same. When there are sufficiently few calls, the supervisor places non-assigned CPIs as the first available for the next day's rotation. Thus, the rotation list order effectively assigns CPIs to cases randomly as they arrive. From our conversations with DCYF officials, there is generally no criteria for case assignments other than the rotation list. The only exception is in cases when a specific type of abuse is alleged and requires the attention of a particular CPI. For example, if there is a sex abuse allegation, the supervisor may attempt to assign a CPI of the same gender as the victim. Cases where a supervisor assigns a specific investigator based on case characteristics are flagged in the case management system (which allows us to remove these cases with non-random assignment from our estimation sample).

The CPI who investigates the case decides whether there is sufficient evidence of child abuse or neglect to warrant out-of-home placement.[12] If there is sufficient evidence, the CPI petitions the Rhode Island Family Court (RIFC) for removal of the child and placement into DCYF custody. Conversations with DCYF staff suggest that RIFC almost always follows the recommendations made by investigators. 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 in a supervised environment such as a group home or shelter. After

[11] When deciding whether a report merits an investigation, call staff assess whether the report meets the following five criteria: 1) involves a victim of child abuse or neglect whose welfare is harmed or threatened with harm; 2) involves transfer of a child into the permanent care of an unrelated, unauthorized individual; 3) involves sexual abuse of a child by another child; 4) indicates a perpetrator of child abuse and/or neglect places other children at risk; 5) indicates a threat to the safety of an unborn child.

[12] In our analysis sample (described in Section 3.1) of investigations in Rhode Island, the average investigation lasts 22 days. In cases where the CPI recommends removal, the average time from the start of the investigation until the child is removed is 11 days.

a child is placed into DCYF custody, a CPI has no involvement in the case management associated with a removed child. When a child is in DCYF custody, parents may be permitted to visit, although the frequency of visitation varies depending on case-specific factors.

DCYF releases children from custody due to reunification with parents, adoption, or aging out of the child welfare system by reaching the legal adult age. Reunification with parents only occurs 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 case management (intake) workers monitor whether a parent complies with conditions for reunification.

3 Data

We use data from anonymized administrative records housed in a secure facility. 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 et al., 2018a).

3.1 DCYF Sample

There are 32,845 DCYF investigations that occurred between January 1, 2000 and December 31, 2015.[13] Using these data, we create a sample of investigated children (with substantiated reports of abuse or neglect) based on three main restrictions. First, we exclude sex abuse investigations and investigations where the Investigative Unit supervisor assigned the case without using the rotation list (N=7,536). Second, we drop investigations that occur after the first investigation associated with a child (age 0 to 18) (N=5,476).[14] Third, we exclude investigations assigned to CPIs who received less than 10 cases within a year (N=471). Our final sample consists of the remaining 19,362 investigations, which are associated with a sample of 26,977 children aged 0 to 18 at the

[13] See Appendix B for details on the process for data cleaning and constructing the analysis sample.

[14] Similarly, Doyle (2007; 2008) focuses on the first case manager associated with a child protective services investigation.

time of the investigation.

3.2 School-Age Academic Outcomes

We join the sample of children investigated by DCYF to records from the Rhode Island Department of Education (RIDE), which are available for the academic years 2003-2016. The RIDE data include records for school enrollment, identifiers for the school and grade enrolled in, and receipt of special education services as indicated by receipt of a written Individualized Education Program (IEP). Standardized test scores in reading and math are available in a subset of academic years (2005-2016).

Standardized test scores for exams taken during grades 3-8 are the main outcome that we study. We construct a panel at the academic year level, which contains 11,727 investigated children. Children who were born before 1991 or after 2008 will not have observations in the panel because they are either too old or young to enroll in the period (2005-2016) for the test score data. Likewise, children born in 2007 will have at most one year of test score data, and children born in 1992 will have at most one year of test score data. We focus on the average of the scores in math and reading (standardized by grade and academic year). We also examine impacts on reading and math scores separately.[15] Note that in a given academic year, there are no data for children who enrolled in a private school, although we do have test scores for children enrolled in charter schools.

We study additional school outcomes such as enrollment in a public school and the characteristics of the public schools that children attend. We use end-of-year enrollment files from RIDE to measure whether the child ever enrolls during grades 3-8 (this grade range corresponds to the period that we use to analyze test score outcomes). We also calculate the number of schools that a child attended in the first three academic years after the investigation. School mobility may be a mechanism by which removal affects children (Hanushek et al., 2004; Lyle, 2006).

For characteristics of schools attended, we measure standard school-level value-added for the average of standardized math and reading test scores, the fraction of students who receive a free or reduced price lunch, and the fraction of enrolled students who are minorities. Value-added for each

[15] As additional results, we examine the first test score observed after an investigation and the first three test scores observed after an investigation. These additional results are included in Appendix Table A.6.

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 on lagged test scores (including their square and cube), as well as indicators for a student's race, gender, special education status, English learner status, and free or reduced price lunch status; we use the mean residuals within a school as a single measure of value-added.[16,17] The fraction of minority and free/reduced lunch students at a school are calculated in each year. We join these school characteristics to a child-level panel to measure the impact of removal on the characteristics of the schools attended post-investigation.

Finally, the last schooling outcome that we study for children is a measure of participation in special education programs. Our measure is whether a child has a written IEP, which indicates that the child has at least one of the thirteen disability categories as defined by the Individuals with Disabilities Education Act (IDEA).[18] The determination of an IEP can start as early as pre-school, when the child is three to four years old. Over half of students with an IEP in Rhode Island are identified with special needs prior to entering first grade.[19] Since kindergarten (K) is mandatory, we measure IEP enrollment during grades K-6. In the RIDE enrollment data, we have a sample of 13,716 investigated children and measure the fraction of years for which a child has an IEP. Note that this measurement is conditional on enrollment in a Rhode Island public school. Not all children will have a complete set of years for which we can measure IEP enrollment. For example, if a child transfers (permanently) from public to private school in third grade, we would only observe IEP enrollment from kindergarten to second grade. We retain these children in our analysis and compute IEP participation for the grades available. Similarly, children born before 1997 or after 2011 can only be observed in a partial set of academic years due to the limited coverage of the RIDE data (from 2003-2016).

[16] Our approach is similar to prior studies such as Kane et al. (2008) and Chetty et al. (2014).

[17] See Appendix C for details on the estimation of school value-added and how we join this measure to the student-level data.

[18] The categories are: autism, deaf-blindness, deafness, developmental delay, emotional disturbance, hearing impairment, intellectual disability, learning disability, orthopedic impairment, speech or language impairment, traumatic brain injury, visual impairment including blindness, and other health impairment (Individuals with Disabilities Education Act, 2004).

[19] 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 development delays and disabilities, as mandated by IDEA.

3.3 Perpetrator Charges and Incarceration

For nearly all of the children in the sample of DCYF investigations, we observe the perpetrators associated with the allegations of abuse and neglect. We study impacts of removal on criminal charges and the likelihood of incarceration using records (1995-2017) from the Rhode Island Department of Corrections (RIDOC). The outcome in our analysis is whether a perpetrator is charged or incarcerated at any point in the four years following the conclusion of an investigation. For any investigation ending in 2013 or later, the four-year measure is partially censored because the data source ends in 2017.

3.4 Later-Life Outcomes

For children investigated at older ages (6 to 18), we estimate impacts on several later-life outcomes. We do not study effects for younger children since they will not be old enough by the final year covered in administrative data. We join the sample of older investigated children to juvenile conviction records from the Rhode Island Family Court (2000-2016), state birth records (2000-2016), and postsecondary institution enrollment files from the National Student Clearinghouse (2004-2016). The main outcomes are indicators for whether a child has any juvenile court convictions by age 18, graduates from high school by age 19 (based on RIDE records), is a teen parent by age 20, and enrolls in a two or four-year postsecondary institution by age 22. If a child does not reach the terminal age for each measure, they are not included in the analysis of later-life outcomes.[20]

3.5 Descriptive Statistics

Table 1 presents summary statistics for the DCYF sample of first investigations in Rhode Island.[21] Column 1 shows that the average age at the time of an investigation is 6.19, which is in line with national statistics for child abuse investigations (U.S. HHS, 2016). Children investigated are 62

[20]There are 7,997 children in the juvenile convictions sample, 5,267 children in the high school graduation sample, 4,753 children in the postsecondary enrollment sample, and 6,706 children in the teen birth sample.

[21]To parallel our main analysis, Appendix Table A.1 provides summary statistics for subgroups based on gender and child age at the time of an investigation (before age 6 and age 6 and older). Race and ethnicity are similar across each of these gender and age subgroups. The main distinction is that children investigated at young ages are less likely to live in households with a married couple and less likely to be associated with an investigation where physical abuse is one of the allegations. Finally, it is also worth noting that the removal rate is slightly higher for young children.

percent white and 17 percent Hispanic.[22] Race in the sample differs notably from Doyle (2007; 2008) which studied the impact of removal for a sample where 76 percent of investigated children were African-American. This contrast partly reflects differences in the demographics of Rhode Island, where 9 percent of children are African-American, and Illinois, where 15.8 percent of children are African-American (U.S. Census Bureau, 2018). In terms of family background, only a small fraction (21 percent) of children are from married households.

The DCYF data report all allegations associated with an investigation. An allegation of neglect occurs in 82 percent of investigations in the sample. Allegations of physical abuse or physical neglect (i.e., neglect that results in a physical injury) occur much less frequently in about 20 and 5 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. HHS, 2016).

We also observe the reporter associated with child investigations. For 82 percent of children, the reporter in the case was a professional such as a teacher, physician, social worker, or police officer.[23] The remaining fraction of reports are provided by family, friends, or other individuals such as neighbors or anonymous reports.

Column 1 shows that removal from home occurs in 16 percent of first investigations. This is less than the removal rate observed in Doyle (2007; 2008), which studied Illinois in the 1990s when the state's placement rate was one of the highest in the U.S. 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. Children who are removed are slightly younger than those who are not removed (4.92 years old versus 6.43). In addition to age, the statistics show that investigations that end in removal have significantly different child and case characteristics from investigations where removal does not occur. For example, children subject to removal live in households where marriage rates are lower by 11 percentage points (p-value < 0.01) and children are about 5 percentage points (p-

[22] Nationally, 45 percent of child abuse victims were white and 22 percent were Hispanic (U.S. HHS, 2016).

[23] As expected, there are differences in reporter type for children who are investigated at younger and older ages. For younger children, only 6 percent of reports are from school teachers, whereas the corresponding statistic for older children is 13 percent.

value < 0.01) more likely to be African-American in the investigations that result in removal. The final row of Table 1 shows that Rhode Island children who are removed spend roughly 440 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 model of outcomes for child i:

$$Y_{ijt} = \beta_0 + \beta_1 R_{ijt} + \beta_2 X_i + \varepsilon_{ijt} \tag{1}$$

where Y_{ijt} is the outcome for child i who is assigned to investigator j in year t, R_{ijt} is an indicator for removal, X_i is a vector of case characteristics, and ε_{ijt} is an error term. Standard OLS estimates of Equation (1) will be biased if home removal (R_{ijt}) is correlated with unobserved determinants of child outcomes (ε_{ijt}). Prior research suggests that unobserved family and home conditions affect both the likelihood of removal and child outcomes (Berger et al., 2009, 2014; Wildeman and Waldfogel, 2014).

To address the endogeneity concern in Equation (1), we rely on an instrumental variable (IV) strategy that is based on a measure of the removal tendency, denoted as Z_{ijt}, of the CPI assigned to child i. Our first-stage equation is:

$$R_{ijt} = \alpha_0 + \alpha_1 Z_{ijt} + \alpha_2 X_i + \nu_{ijt} \tag{2}$$

where Z_{ijt} is the simple leave-one-out removal tendency of investigators frequently calculated in the literature using judge decision tendencies as instruments for individual case decisions (Kling, 2006; Doyle, 2007, 2008; Aizer and Doyle, 2015; Bhuller et al., 2016; Mueller-Smith, 2016; Sampat and Williams, 2017; Dobbie et al., 2018a,b; Bhuller et al., 2018):

$$Z_{ijt} = \frac{\sum_{k \neq i} R_{kjt}}{N_{jt} - 1} \tag{3}$$

where Z_{ijt} is the removal tendency for investigator j assigned to child i in year t, R_{ijt} is removal status, and N_{jt} is the total number of children assigned to investigator j in year t.[24] Note that results report robust standard errors clustered at the investigator (CPI) level for non-panel outcomes and

[24] For our main analysis of test scores, we also provide results using the removal tendency for investigator j calculated over all of their cases for the entire sample period. These results using the alternative measure are similar to the findings in our main analysis. See Section 5 for further discussion.

two-way clustered at the child and CPI level for panel outcomes.

If there are heterogeneous impacts of removal, we must make two assumptions to interpret IV estimates of the parameter β_1 from Equation (1) as a local average treatment effect (LATE) of removal for marginal investigations (Angrist and Imbens, 1994). First, the measure of CPI removal tendency defined in Equation (3) must affect child outcomes only by changing the probability of removal. This assumption is plausible in our setting because supervisors assign CPIs to investigations based on the rotational assignment list described in Section 2. We test for random assignment of investigators, and show in Section 4.2 that the CPI tendency measure defined by Equation (3) does not cause baseline child and investigation characteristics.

Second, we must assume that there is a monotonic impact of CPI assignment on removal across children. A violation of this assumption may occur if CPI removal tendencies vary with observed case characteristics. For example, a given CPI may be relatively strict when it comes to removing African-American children, but lenient when it comes to removing all other children. If there is a non-monotone impact of removal tendency, the IV estimate would not identify a well-defined LATE.[25] As a test of the monotonicity assumption, Section 4.4 follows prior work and shows that the tendency measure defined in Equation (3) is positive in various subsamples (Bhuller et al., 2016; Dobbie et al., 2018b).

To further address concern about monotonicity and rigorously identify instruments that maximize the accuracy of predicted removal, we allow CPI tendency and its impact on removal to vary with baseline case characteristics. We do this by creating a set of potential instruments based on leave-one-out measures of removal tendency calculated for different categorizations of child and case characteristics (sex, age, minority, allegation type, and investigation level). This creates a large number of potential instruments. Following Belloni et al. (2012, 2014), we use the machine-learning (ML) algorithm, Least Absolute Shrinkage and Selection Operator (LASSO), to select the instruments with greatest predictive power for removal in the first stage equation.[26] We refer to IV estimates using these instruments as ML-IV.

[25] 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).

[26] The use of LASSO for regularization is necessary since there are a large number of potential case characteristics by which CPI tendencies may vary. An unrestricted model would likely result in too many instruments and potentially weak instruments, creating challenges for causal inference (Bound et al., 1995).

4.1 Variation in Child Protective Investigator (CPI) Removal Tendency

Figure 1 plots the distribution of the leave-one-out CPI removal tendency based on Equation (3) for our sample of first investigations. We observe 43 CPIs in an average year, and there are 103 unique CPIs across the entire sample period. The average number of cases seen in a year by CPIs in our sample is 39. The mean of the removal tendency is 0.16, while the 25th and 75th percentiles of the distribution are 0.10 and 0.20, respectively. The standard deviation is 0.09, which is generally similar to the sample of child abuse investigations studied in Doyle (2007).[27]

4.2 Instrument Validity (Testing Random Assignment)

According to the assignment process described in Section 2, investigations in our sample should be quasi-randomly assigned to CPIs. We test this implication by regressing the removal tendency measure on baseline child and investigation characteristics. The upper section of Table 2 reports results from an F-test for joint significance of baseline characteristics in determining investigator removal tendency. We report this for all investigations and subgroups based on gender and age of children at the time of an investigation. Baseline characteristics include the child demographics and case characteristics listed in Table 1. We consistently fail to reject the null hypothesis that the coefficients for investigation characteristics are jointly zero. For example, Column 1 shows that the F-statistic is 1.06 with a p-value of 0.40 in the sample of all investigations. Appendix Table A.2 shows the full regression results corresponding to each F-test statistic presented.

4.3 First-Stage Impact

The lower section of Table 2 reports results from Equation 2, measuring the impact of our instruments on whether or not an investigation resulted in removal of the child from the home. Column 1 shows that the leave-one-out measure of mean CPI removal tendency is highly predictive of removal. 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 5.33 percentage

[27] In Doyle (2007), the standard deviation is 9 percent in the delinquency sample, 10 percent in the teen motherhood sample, and 7 percent in the labor market outcomes sample.

points ($= 0.533 * 0.10$).[28] Columns 2 through 7 show that the effects by subgroups for gender and the age of children are somewhat similar in magnitude. The point estimates suggest that removal tendency has a larger impact for male children, though we cannot reject the hypothesis of equal first-stage impacts between girls and boys.

4.4 Monotonicity

To interpret IV estimates from Equations 1 and 2 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. (2016) and Dobbie et al. (2018b), one testable implication of monotonicity is that the first-stage estimates should be non-negative for any subgroup of the investigations sample. Columns 2 through 7 of the bottom of 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 and age at the time of investigation. Appendix Table A.3 expands on these results by providing additional results for narrower subgroups defined based on gender, age at the time of investigation, and various case characteristics. These results show that the first-stage impacts of removal tendency are consistently positive.[29]

4.5 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. To conduct this analysis, we follow the approach

[28] 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.

[29] The magnitudes of the first-stage estimates for subgroups defined by each case characteristic (shown in the rows of Appendix Table A.3) are generally similar to the impact in the sample of all investigations. The exception is when we consider the results for the subgroup of emergency cases. This is not entirely surprising given that removal recommendations may not vary substantially across investigators in these severe cases of abuse and neglect. Our results are robust to excluding emergency cases from our analysis.

based on Abadie (2003), Dahl et al. (2014), and Dobbie et al. (2018b).[30]

Appendix Table A.4 provides information on complier characteristics for the entire sample of first investigations and subgroups defined by gender and age at the time of an investigation. Specifically, we report the relative likelihood of being a complier based on child and case characteristics. Across the samples that we consider, compliers are consistently less likely to be associated with an investigation where there is an allegation of physical neglect or physical abuse. We also find that compliers are generally more likely to be Hispanic or associated with a case where the reporter was a family member or friend.

5 Main Results

5.1 Standardized Test Scores

Table 3 presents estimates of the impact of removal on standardized test scores by gender and age subgroups. Panel A provides estimates for the average of reading and math scores, while Panels B and C provide separate estimates for reading and math, respectively. For each outcome, we report the mean scores of non-removed children next to the IV and ML-IV estimates that we obtain from the second stage model (Equation 1). Robust standard errors that are two-way clustered at the individual and investigator level are reported throughout.

The ML-IV results in Panel A show that the marginal removal has a significant and positive impact on the average standardized test scores for all girls. The point estimate for removal is 0.460. The IV estimate is nearly identical although not statistically significant. We cannot reject the hypothesis that the IV and ML-IV point estimates are equal, which is expected given the large standard errors for each estimate.

The remaining rows in Table 3 show that the pooled results for girls are driven by significant and large improvements in test score performance for girls removed before the age of six. For this subgroup, removal increases average test scores by just slightly more than 1 student-level

[30] Similar to Dahl et al. (2014) and Dobbie et al. (2018b), 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 most strict 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 notes to Appendix Table A.4 for further details on our calculation of complier characteristics.

standard deviation using either the IV or ML-IV specification. For older girls, the point estimates are not significant and the sign is consistently negative. Appendix Table A.5 shows that we obtain similar estimated impacts of removal when we use an alternative definition for our instruments that combines all removal decisions observed for an investigator during our sample period (2000-2015).

The point estimates that we find are large in magnitude, although it is worth noting that the sample means show that children subject to DCYF investigations have low standardized test scores. The average non-removed child in our sample scores nearly half a standard deviation below the Rhode Island statewide testing average (within grade and year). Relative to prior studies, one point of comparison for these effects are evaluations of high-quality early education programs targeting disadvantaged children. Heckman et al. (2013) found that the Perry Preschool program increased female standardized test scores by 0.806 standard deviations.

In contrast to the results for girls, Panel A shows that there are no detectable impacts on post-investigation test scores of boys. The IV and ML-IV estimates are consistently negative for this group of children. There are also no detectable effects by age subgroups for boys, and we can reject the hypothesis of equal impacts of removal between young girls and young boys (p-value < 0.10).

Panels B and C of Table 3 show that the estimated impact of removal is similar when we look at reading and math scores separately. The IV and ML-IV estimates are consistently significant and positive for girls removed prior to the age of six. For boys and older children, we find no detectable impact of removal and the sign of the point estimates remains consistent across the IV and ML-IV estimates.

In Appendix Table A.6, we estimate impacts on average standardized test scores by years since removal. We find that impacts are similar during the first year and first three years after removal: there are significant and large positive impacts among young girls, and negative impacts for boys. These results suggest that impacts are persistent over time and may be caused by permanent changes in child ability prior to the first test.

Given that the analysis tests for impacts in a large number of subgroups, one may be concerned that our finding of significant effects in age and gender subgroups is 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 A.7 shows that the IV and ML-IV estimated impact of removal on test scores for young girls is still significant at the 10 percent level after adjusting for multiple comparisons across gender and age subgroups.

5.2 Attrition due to Changes in Enrollment to Public Schools

A concern for interpreting the test score results is that removal may affect whether a child attends public school in Rhode Island. This would generate selection into the panel of observed standardized test scores that we use for our analysis. Table 4 presents estimates of the impact of removal on a measure of child enrollment in public school during grades 3-8. These are the grades studied in the test score sample.

For girls removed before age six, we do not find significant impacts of removal on enrollment, and the point estimate changes sign depending on the specification. For young boys, the IV and ML-IV estimates are not significant and consistently negative. We cannot reject the hypothesis of equal impacts of removal on enrollment between young girls and young boys. Table 4 also shows that there are no detectable or consistently-signed impacts for girls removed at older ages. In contrast to these results, the point estimates for older boys are positive and precisely estimated (p-value < 0.05).

These findings suggest that attrition from public school is unlikely to explain the pattern of heterogeneous impacts by gender and age. Enrollment effects for girls do not consistently vary by the age at the time of an investigation. This does not align with the estimated impacts on test scores where we find positive impacts for young removed girls that are statistically different from the estimates we obtain for their older counterparts. For young children, we do not find evidence that changes in enrollment affect the interpretation of test score results that are heterogeneous by gender.

6 Exploring Mechanisms

6.1 Case Characteristics and Foster Care Placement Outcomes

Differences in case characteristics or foster care placements by age and gender are a potential explanation for heterogeneous impacts of removal on educational outcomes. Appendix Table A.1 reports descriptive statistics by subgroups based on gender and age. The demographic, family, allegation, reporter, and investigation statistics are consistently similar for girls and boys of the same age group. The primary differences in average case characteristics occurs between investigated children of different ages. For example, approximately 14 percent of younger children are associated with cases of physical abuse. The corresponding statistic is approximately 27 percent for older children. This does not provide evidence that heterogeneity in test score impacts is due to differences in the types of cases across gender and age subgroups.

Table 5 reports the impact of removal on the number of days spent in each type of foster care and the likelihood of adoption.[31] We find that removal has significant and large positive impacts on the number of days spent in foster care for all subgroups. Table 5 also shows no detectable impacts of removal on the likelihood of adoption in most subgroups. For older boys, the ML-IV estimate suggests removal increases the likelihood of adoption by 2.6 percentage points. These results for adoption reflect the fact that this is a relatively rare outcome in our sample. Only 9 percent of children in our DCYF sample of first investigations are later adopted.

Across gender and age subgroups, the main distinction is the effect of removal on non-relative foster family placement. For example, the IV point estimates suggest that removal for older girls and older boys increases time spent with foster non-relatives by 227 and 120 days, respectively. We can reject that these point estimates are equal (p-value < 0.05). For the remaining types of placement outcomes, we find little evidence of detectable differences in the IV estimated impact of removal when making comparisons within gender by age or within age by gender. The ML-IV estimates are consistently similar to the IV estimates in terms of all placement outcomes.

These findings provide little evidence that heterogeneous treatment within the foster care system

[31] Note that the mean number of days spent in each type of foster care is zero by definition for children who are not removed from home.

explains the pattern of test score impacts from Section 5, which showed that removal has significant and positive impacts for young girls. While the IV estimates suggest that younger girls spend less time in non-relative placements, the magnitude of this difference is less than three months in this type of care. Further, when we compare across young girls and boys, we cannot reject the null hypothesis that removal has equal impacts on any type of placement outcome.

6.2 School Mobility and Characteristics of Schools Attended

We next measure if removal has impacts on the number or types of schools that children attend. Findings from several strands of research motivate these analyses. Hanushek et al. (2004) and Lyle (2006) provide evidence that increased school mobility has negative impacts on academic performance. Prior research has also linked a number of school characteristics to student-level outcomes. Deming et al. (2014) and Chetty et al. (2014) show that teacher value-added has causal impacts on a number of child outcomes. Several studies show that school segregation negatively impacts academic test scores and other educational outcomes (Guryan, 2004; Lutz, 2011; Billings et al., 2014). We use statewide enrollment and standardized testing data to construct three school characteristics: school value-added, free or reduced price lunch share, and minority share.[32] We exclude children in the DCYF sample of investigations from these calculations.

Table 6 provides estimates for impacts on the number of schools attended within the first three enrollment years after an investigation. These results provide no strong evidence that would suggest the differential impacts by gender and age at removal are driven by differences in school mobility effects. For all groups of children, we find positive point estimates for the impact of removal on school mobility. For young children, we cannot reject the hypothesis that positive estimated effects of removal are equal for young girls and boys.

Next, Table 7 provides impacts of removal on the three measures of school characteristics. We find no detectable impacts of removal on any measure. In some cases, the signs of estimates align with the test score results; for example, the point estimates for school value-added are consistently positive for younger girls and negative for older girls. Overall, however, these results do not

[32]See Appendix C for details on the construction of these measures.

provide strong evidence of heterogeneous impacts on school characteristics that vary by age or gender. This suggests that changes in school type are not likely to mediate impacts on test scores.

6.3 Participation in Special Education Programs

Another possibility is that changes in a child's experience within school mediate impacts on test scores. One measure of school experience is whether a child participates in special education programs (*e.g.*, receives an Individualized Education Program, or IEP). Prior research suggests that enrollment in special education programs can improve test score outcomes of children (Hanushek et al., 2002). In addition to providing specialized instruction, enrollment in specialized education could impact test scores because assisted students may receive accommodations like extra time ("time and a half") to finish exams, providing extra support for children with specialized needs.[33]

The DCYF system could directly impact child participation in special education because children who enter foster care may receive an educational surrogate to make decision on the child's behalf (RI Department of Children, Youth, and Families, 2018). The appointment of an educational surrogate is required if the child is suspected of having a disability or learning difficulties. These policies could affect special education participation even for young children given that enrollment in special education can begin as early as pre-school.

Table 8 provides estimates of the impact of removal on special education participation for children in our sample. As mentioned in Section 3.2, the measure that we focus on is whether a child is recorded as having a written IEP in grades K-6. Although we find no statistically significant impacts, the pattern of point estimates suggests there are heterogeneous effects by gender. The IV and ML-IV point estimates suggest removal has a negative impact on the likelihood that a young girl has an IEP. These results contrast with the positive point estimates observed for young boys. For the IV estimates, we can reject the hypothesis that effects for young girls and young boys are equal (p-value < 0.10).

These results do not provide support for the idea that receipt of specialized education services and testing accommodations drive the impact on test scores that we observe for young girls. Rather,

[33] An IEP does not automatically imply that a student is exempt from testing. In academic year 2013, 89% of students with an IEP took a standardized exam.

the pattern of the point estimates suggests removed young girls have lower rates of IEP enrollment relative to removed young boys. One interpretation is that this pattern of effects could reflect differential changes in schooling ability, which correspond to the effects that we detect for test scores.

6.4 Perpetrator Criminal Charges and Incarceration

Our last examination of mechanisms looks at household structure and conditions by studying post-investigation outcomes for adult household members. Specifically, we study perpetrators of child abuse or neglect and the impact of removal on criminal charges and incarceration within the four years after an investigation concludes.[34,35] In our sample, approximately 90 percent of perpetrators are parents and 60 percent of perpetrators are women.

Panel A of Table 9 shows that the IV-estimated impact of removal is significant and positive in the sample of investigations for girls. The point estimate shows that removal increases the likelihood of charges or incarceration by 15.4 percentage points. In the subgroups of younger and older girls, the IV point estimates are sufficiently similar in magnitude that we cannot reject the hypothesis of equal impacts. The ML-IV estimates are not statistically significant and consistently smaller than the IV estimates.

In contrast to the results for girls, we find no detectable impacts on perpetrators of abuse and neglect against boys. The point estimates for subgroups of boys are also consistently smaller than the respective estimates for girls. Given the large standard errors for the estimates, we cannot reject the hypothesis that impacts are equal for young girls and young boys.

We expand on these results by estimating effects separately for male perpetrators associated with the abused or neglected child. For all girls, Panel B of Table 9 shows that removal has a significant and positive effect on charges and incarceration. These results provide suggestive evidence that removal may have stronger effects on family structure when men within a household

[34] Note that all perpetrators in the sample are associated with an investigation where DCYF has substantiated the report of abuse or neglect.

[35] Appendix Table A.8 shows results where we define the measure of charges and incarceration using one and two year windows for the post-investigation period. While these results are less precise than those using the four year window, the pattern of estimates is similar. We cannot reject the hypothesis that the estimated impacts for females using the one or two year measures are equal to the estimate for females using the four year measure.

perpetrate abuse or neglect.[36] In the remaining subgroups based on gender and age, the only statistically significant result occurs for older girls where the IV estimate is a 40 percentage point effect. This estimate is larger than the positive point estimate for younger removed girls. Results for boys are consistently smaller in magnitude than the estimates for girls.[37]

Overall, this analysis of perpetrators provides evidence that removal has impacts on parents and household structure. There are at least three reasons why charges and incarceration of 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 child protective service investigators) as part of a child reunification plan. Third, removal may adversely affect the mental health of perpetrators resulting in changes in criminal behavior.

7 Results for Later-Life Outcomes

Due to data constraints, we can only study effects of removal on later-life outcomes of children removed from home after the age of six.[38] As discussed in Section 3.4, we study the following outcomes: having any juvenile court convictions by age 18, graduation from high school by age 19, teen births, and enrollment in any postsecondary institution by age 22. We continue to conduct all of our analysis separately for girls and boys.

Table 10 reports the effects of removal for all later-life outcomes. In general, we find no statistically significant impact of removal using either the IV or ML-IV specifications. The exception in our results is the IV-estimated impact of removal for teen fatherhood, which is marginally significant (p-value < 0.10). The point estimates for the remaining outcomes for older males tend to suggest negative impacts of removal (*e.g.*, increases in the likelihood of having a juvenile convic-

[36] Prior research suggests that changes in household structure are significant predictors of investigations for abuse and neglect. Hastings et al. (2018b) show that the entry of a new man within a household in a baseline year predicts increases in the probability of a substantiated DCYF investigation and finding of abuse (or neglect) in the following year.

[37] We do not find meaningful differences between male and female perpetrators in the type of abuse charge. (Recall that we exclude the small fraction of sex abuse investigations from the analysis sample.) Male and female perpetrators are similarly likely to have a physical abuse charge (21 versus 18 percent) and similarly likely to have an an immediate type of investigation (61 versus 62 percent).

[38] As mentioned in Section 3, we focus on older children since a child removed before age six will 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.

tion and lower probability of high school graduation).

For older females, the non-significant point estimates are opposite signed. The results suggest reductions in juvenile convictions and increased likelihood of attending a postsecondary institution. This contrasts with remaining point estimates that suggest reductions in the likelihood of high school graduation and increases in the probability of having a child before age 20.

Note that we do not find strong evidence that attrition due to moving from the state of Rhode Island affects the interpretation of these results. Appendix Table A.9 reports results for the impact of removal on two proxies of attrition based on receipt of social assistance (Medicaid, TANF or SNAP) provided by the Rhode Island Department of Human Services. We focus on the share of years within a given time period (5 or 10 years post-investigation) that a child is *not* confirmed to be living in Rhode Island because they do not receive some form of social assistance. Using the 5-year measure of attrition, the point estimate from the ML-IV specification is marginally statistically significant for older girls (p-value < 0.10). We also find a marginally significant IV estimate for older boys using the 10-year measure (p-value < 0.10). All other estimates for older children are not statistically significant.

One benchmark for this analysis are the results from Doyle (2007), which analyzed the impact of foster care for children ages 5 to 15 in Cook County, Illinois. Using a similar IV approach based on child protective investigator tendencies, he reports significant and large positive impacts on teenage pregnancy (29 percentage points) and juvenile delinquency (47 percentage points) for older girls. In our sample, the positive point estimate for teenage pregnancy for girls is much smaller in magnitude, but the standard error is sufficiently large that we cannot rule out the effect size from Doyle (2007). The IV and ML-IV point estimates and standard errors for juvenile convictions allow us to rule out effect sizes larger than 10 percentage points for older girls in Rhode Island.

8 Conclusion

Despite the fact that child protection authorities remove more than 200,000 children from their homes each year, there is relatively little research on the causal impacts of this policy (U.S. HHS, 2016). We use administrative data to estimate effects of home removal by gender and child age at the time of an investigation. Following prior studies, we use the removal tendency of quasi-experimentally assigned child protective investigators as an instrument.

We find that removal causes significant improvements in performance on standardized exams for girls. This results is driven by large impacts on test scores for young girls. The estimates show similar impacts on test scores starting from the first testing year after removal and onward, which suggests a permanent change in ability prior to when a young girl begins taking exams. While the academic performance of girls investigated before age six appears to improve due to removal, there is some suggestive evidence of negative impacts on schooling outcomes for boys. In our analysis of potential mechanisms, we find that case characteristics across gender and age subgroups are similar, and we do not find any evidence of differential impacts of removal on the path that a child experiences in the foster care or public school systems. We examine household structure by studying impacts of removal on charges and incarceration of parents and caretakers who are the perpetrators of abuse and neglect. For girls, we find some evidence of significant and positive impacts of removal on perpetrators, but the estimates do not differ by investigation age.

Our findings are consistent with the hypothesis that both gender and age may be strong mediators for the impact of removal in cases of abuse and neglect on childhood development. The fact that girls experience stronger positive impacts of removal on academic achievement compared to boys is a pattern that echoes prior studies of schooling and social program interventions (Hastings et al., 2006; Kling et al., 2007; Anderson, 2008; Angrist et al., 2009; 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 only to girls removed before age six complements the recent literature on the importance of early-life conditions (Cunha and Heckman, 2007; Almond and Currie, 2011b; Almond et al., 2017).

References

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

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

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

Aizer, A., Shari, E., Ferrie, J., and Lleras-Muney, A. (2016). The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review*, 106(4):935–971.

Almond, D. and Currie, J. (2011a). Human Capital Development Before Age Five. In *Handbook of Labor Economics*, volume 4b, pages 1316–1476. Elsevier.

Almond, D. and Currie, J. (2011b). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, 25(3):153–172.

Almond, D., Currie, J., and Duque, V. (2017). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature (forthcoming)*.

Almond, D., Doyle, J. J., Kowalski, A., and Williams, H. (2010). Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns. *Quarterly Journal of Economics*, 125(2):591–634.

Anderson, M. (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.

Angrist, J. and Imbens, G. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Society*, 91(434):444–455.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 1(1):136–163.

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

Autor, D., Figlio, D., Karbownik, K., Roth, J., and Wasserman, M. (2016). Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. Working Paper 22267, National Bureau of Economic Research.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2):29–50.

Benjamini, Y., Krieger, A., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Benson, C., Fitzpatrick, M., and Bondurant, S. (2018). Beyond Reading, Writing, and Arithmetic: The Role of Teachers and Schools in Reporting Child Maltreatment.

Berger, L., Bruch, S., Johnson, E., James, S., and Rubin, D. (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, L., Cancian, M., Han, E., Noyes, J., and Rios-Salas, V. (2014). Children's Academic Achievement and Foster Care. *Pediatrics*, 135(1):110–116.

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

Bharadwaj, P., Loken, K., and Neilson, C. (2013). Early Life Health Interventions and Academic Achievement. *American Economic Review*, 103(5):1862–1891.

Bhuller, M., Dahl, G., Loken, K., and Mogstad, M. (2016). Incarceration, Recidivism, and Employment. Working Paper 22648, National Bureau of Economic Research.

Bhuller, M., Dahl, G., Loken, K., and Mogstad, M. (2018). Intergenerational Effects of Incarceration. *AEA Papers and Proceedings*, 108:234–240.

Billings, S., Deming, D., and Rockoff, J. (2014). School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg. *Quarterly Journal of Economics*, 129(1):435–476.

Bound, J., Jaeger, D., and Baker, R. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association*, 90(430):443–450.

Campbell, F., Conti, G., Heckman, J., Moon, S. H., Pinto, R., and Pungello, E. (2014). Early Childhood Investments Substantially Boost Adult Health. *Science*, 343(6178):1478–1485.

Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D., and Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, 126(4):1593–1660.

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

Chetty, R., Hendren, N., and Katz, L. (2016). The Effects of Exposure to Better Neighborhoods on Children: Evidence from the Moving to Opportunity Project. *American Economic Review*, 106(4):855–902.

Chyn, E. (2018). Moved to Opportunity: The Long-Run Effect of Public Housing Demolition on Labor Market Outcomes of Children. *American Economic Review*, 108(10): 3028-3056.

Chyn, E., Gold, S., and Hastings, J. (2018). Short- and Long-run Impacts of Health Interventions for Very Low Birth Weight Children.

Cunha, F. and Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2):31–47.

Cunha, F., Heckman, J., Lochner, L., and Masterov, D. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education*, volume 1, pages 697–812. North Holland.

Currie, J., Mueller-Smith, M., and Rossin-Slater, M. (2018). Violence while in Utero: The Impact of Assaults During Pregnancy on Birth Outcomes. Working Paper 24802, National Bureau of Economic Research.

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

Dahl, G., Kostol, A., and Mogstad, M. (2014). Family Welfare Cultures. *Quarterly Journal of Economics*, 129(4):1711–1752.

Deming, D., Hastings, J., Kane, T., and Staiger, D. (2014). School Choice, School Quality, and Postsecondary Attainment. *American Economic Review*, 104(3):1–25.

Dobbie, W., Goldin, J., and Yang, C. (2018a). The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. *American Economic Review*, 108(2):201–240.

Dobbie, W., Gronqvist, H., Niknami, S., Palme, M., and Priks, M. (2018b). The Intergenerational Effects of Parental Incarceration. Working Paper 24186, National Bureau of Economic Research.

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

Doyle, J. 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.

Fantuzzo, J. and Mohr, W. (1999). Prevalance and Effects of Child Exposure to Domestic Violence. *The Future of Children*, 9(3):21–32.

Garces, E., Thomas, D., and Currie, J. (2002). Longer-Term Effects of Head Start. *American Economic Review*, 92(4):999–1012.

Guryan, J. (2004). Desegregation and Black Dropout Rates. *American Economic Review*, 94(4):919–943.

Hanushek, E., Kain, J., and Rivkin, S. (2002). Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities? *Review of Economics and Statistics*, 84(4):584–599.

Hanushek, E., Kain, J., and Rivkin, S. (2004). Disruption versus Tiebout improvement: the costs and benefits of switching schools. *Journal of Public Economics*, 88:1721–1746.

Hastings, J., Howison, M., Lawless, T., Ucles, J., and White, P. (2018a). Integrating Administrative Data for Policy Insights.

Hastings, J., Howison, M., and Ucles, J. (2018b). Predicting Child Maltreatment.

Hastings, J., Kane, T., and Staiger, D. (2006). Gender and Performance: Evidence from School Assignment by Random Lottery. *American Economic Review*, 96(2):232–236.

Heckman, J. and Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics*, 6:689–733.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, 103(6):2052–2086.

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

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

Hoynes, H., Schanzenbach, D., and Almond, D. (2016). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review*, 106(4):903–934.

Individuals with Disabilities Education Act (2004). 20 U.S. Code 1400. Accessed 8 August 2018.

Isen, A., Rossin-Slater, M., and Walker, W. R. (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, T., Rockoff, J., and Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6):615–631.

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

Kling, J. R., Liebman, J., and Katz, L. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

Ludwig, J. and Miller, D. (2007). Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *Quarterly Journal of Economics*, 122:159–208.

Lutz, B. (2011). The End of Court-Ordered Desegregation. *American Economic Journal: Economic Policy*, 3(2):130–168.

Lyle, D. (2006). Using Military Deployments and Job Assignments to Estimate the Effect of Parental Absences and Household Relocations on Children's Academic Achievement. *Journal of Labor Economics*, 24(2):319–350.

Mueller-Smith, M. (2016). The Criminal and Labor Market Impacts of Incarceration.

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

RI Department of Children, Youth, and Families (2018). Policy Manual.

Sampat, B. and Williams, H. (2017). How do Patents Affect Follow-on Innovation? Evidence from the Human Genome. Working Paper 21666, National Bureau of Economic Research.

U.S. Census Bureau (2018). S0901: Children Characteristics, 2012-16 American Community Survey 5-year Estimates. Accessed 26 July 2018.

U.S. HHS (2016). Child Maltreatment 2016.

Wildeman, C. and Waldfogel, J. (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, D., Crooks, C., Lee, V., McIntyre-Smith, A., and Jaffe, P. (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.

Table 1: Descriptive Statistics

		All	Non-removed	Removed	*p*-value	N
Variable		(1)	(2)	(3)	(4)	(5)
Demographics	Female	0.49	0.48	0.50	0.02	26,977
		(0.50)	(0.50)	(0.50)		
	White	0.62	0.62	0.60	0.01	26,288
		(0.49)	(0.48)	(0.49)		
	African American	0.17	0.16	0.21	0.00	26,288
		(0.37)	(0.36)	(0.41)		
	Hispanic	0.17	0.18	0.15	0.00	26,288
		(0.38)	(0.38)	(0.36)		
	Age at investigation	6.19	6.43	4.92	0.00	26,977
		(5.22)	(5.12)	(5.53)		
Family	Married couple	0.21	0.23	0.12	0.00	22,594
		(0.41)	(0.42)	(0.32)		
	Unmarried couple	0.28	0.29	0.26	0.00	22,594
		(0.45)	(0.45)	(0.44)		
	Single parent/other	0.50	0.48	0.62	0.00	22,594
		(0.50)	(0.50)	(0.48)		
	English language	0.97	0.97	0.97	0.91	26,771
		(0.17)	(0.17)	(0.17)		
	Spanish language	0.03	0.03	0.03	0.91	26,771
		(0.17)	(0.17)	(0.17)		
Allegation	Neglect	0.82	0.82	0.80	0.05	26,977
		(0.39)	(0.39)	(0.40)		
	Physical neglect	0.05	0.04	0.06	0.00	26,977
		(0.21)	(0.20)	(0.25)		
	Physical abuse	0.20	0.20	0.20	0.93	26,977
		(0.40)	(0.40)	(0.40)		
Reporter	Professional	0.82	0.82	0.80	0.00	26,620
		(0.38)	(0.38)	(0.40)		
	Family/friend	0.14	0.14	0.16	0.00	26,620
		(0.35)	(0.34)	(0.36)		
	Other	0.04	0.04	0.04	0.58	26,620
		(0.19)	(0.19)	(0.19)		
Investigation	Emergency	0.06	0.03	0.20	0.00	26,977
		(0.24)	(0.18)	(0.40)		
	Immediate	0.63	0.65	0.52	0.00	26,977
		(0.48)	(0.48)	(0.50)		
	Routine	0.31	0.31	0.28	0.00	26,977
		(0.46)	(0.46)	(0.45)		
	Removed	0.16	0.00	1.00		26,977
		(0.37)	(0.00)	(0.00)		
	Days in foster care	72.45	0.00	442.29		26,977
		(226.12)	(0.00)	(385.47)		
Observations		26,977	22,558	4,419		

Notes: This table reports descriptive statistics for children in the DCYF sample described in Section 3.1. Column 1 reports the mean and standard deviation (in parentheses) for all children. Columns 2 and 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 and 3. Column 5 reports total non-missing observations for each variable.

Table 2: Tests of Random Case Assignment and First-Stage Results

Tests of Randomization

Dep. var.:	All	Female	Male	Young Female	Young Male	Old Female	Old Male
CPI removal tendency	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F-statistic	1.058	0.982	1.250	0.862	1.143	1.287	1.525
p-value	0.405	0.484	0.241	0.619	0.326	0.216	0.101
Observations	26,977	13,100	13,877	6,450	7,400	6,650	6,477

First-Stage Impact of CPI Removal Tendency

	All	Female	Male	Young Female	Young Male	Old Female	Old Male
Dep. var.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Removed (= 1)	0.533***	0.496***	0.568***	0.431***	0.511***	0.556***	0.636***
	(0.030)	(0.041)	(0.044)	(0.062)	(0.063)	(0.047)	(0.057)
Observations	26,977	13,100	13,877	6,450	7,400	6,650	6,477

Notes: This table summarizes tests of random case assignment and the first-stage impact of CPI removal tendency. Column 1 reports results for the DCYF sample described in Section 3.1. Columns 2 and 3 report results for female and male children, respectively. Columns 4 and 5 report results for young (aged < 6) females and young males. Columns 6 and 7 report results for old (aged 6+) females and old males. *Tests of Randomization*: Joint test statistics are from an OLS regression of CPI removal tendency on the set of case characteristics listed in Table 1. All models include indicators for missing information and investigation year fixed effects. The F-statistic and p-value reported are from an F-test for joint significance of all variables except investigation year fixed effects. Standard errors are clustered at the CPI level. *First-Stage Impact*: First-stage results are from an OLS regression of removal on CPI removal tendency and the set of case characteristics listed in Table 1. Removed is an indicator for home removal at the child's first investigation. All models include indicators for missing information and investigation year fixed effects. Standard errors clustered at the CPI level in parentheses. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 3: Impact of Removal on Test Scores

Dep. var.:		Panel A: Average Test Score			Panel B: Reading Score			Panel C: Math Score		
		Mean of non-removed	IV	ML-IV	Mean of non-removed	IV	ML-IV	Mean of non-removed	IV	ML-IV
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	All	-0.390	0.464	0.460*	-0.325	0.382	0.370	-0.458	0.553*	0.556*
			(0.286)	(0.272)		(0.299)	(0.282)		(0.299)	(0.284)
			[N=5468]	[N=5468]		[N=5477]	[N=5477]		[N=5472]	[N=5472]
	Young (Aged < 6)	-0.385	1.068**	1.088***	-0.321	0.889*	0.917**	-0.451	1.229***	1.249***
			(0.431)	(0.418)		(0.457)	(0.436)		(0.449)	(0.435)
			[N=2699]	[N=2699]		[N=2703]	[N=2703]		[N=2700]	[N=2700]
	Old (Aged 6+)	-0.396	-0.250	-0.134	-0.330	-0.198	-0.101	-0.465	-0.266	-0.143
			(0.312)	(0.305)		(0.336)	(0.334)		(0.318)	(0.303)
			[N=2769]	[N=2769]		[N=2774]	[N=2774]		[N=2772]	[N=2772]
Male	All	-0.557	-0.176	-0.129	-0.625	-0.166	-0.114	-0.496	-0.159	-0.121
			(0.205)	(0.194)		(0.211)	(0.200)		(0.226)	(0.212)
			[N=6259]	[N=6259]		[N=6267]	[N=6267]		[N=6265]	[N=6265]
	Young (Aged < 6)	-0.557	0.003	0.002	-0.620	0.064	0.082	-0.501	-0.034	-0.049
			(0.334)	(0.306)		(0.344)	(0.314)		(0.356)	(0.331)
			[N=3135]	[N=3135]		[N=3136]	[N=3136]		[N=3136]	[N=3136]
	Old (Aged 6+)	-0.556	-0.383	-0.173	-0.632	-0.433	-0.259	-0.489	-0.302	-0.084
			(0.291)	(0.230)		(0.299)	(0.231)		(0.314)	(0.256)
			[N=3124]	[N=3124]		[N=3131]	[N=3131]		[N=3129]	[N=3129]

Notes: This table reports results for the impact of removal on test scores. We standardize scores at the grade-year level and construct a yearly panel of tests taken in grades 3-8 and in school years 2005-2016. We report results by gender and age at investigation. Column 1 reports the mean of average test score for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. Columns 4-6 and 7-9 report results for reading scores and math scores, respectively. All models include the case characteristics listed in Table 1 as controls, as well as grade and school year fixed effects. Standard errors two-way clustered at the child-CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 4: Impact of Removal on School Enrollment

Dep. var.:		Mean of non-removed	IV	ML-IV
Ever observed in grades 3-8 (= 1)		(1)	(2)	(3)
Female	Young (Aged < 6)	0.700	0.061	-0.171
			(0.179)	(0.180)
			[N=4096]	[N=4096]
	Old (Aged 6+)	0.777	0.105	-0.002
			(0.110)	(0.103)
			[N=3967]	[N=3967]
Male	Young (Aged < 6)	0.703	-0.169	-0.217
			(0.136)	(0.142)
			[N=4755]	[N=4755]
	Old (Aged 6+)	0.762	0.283**	0.296**
			(0.131)	(0.129)
			[N=4575]	[N=4575]

Notes: This table reports results for the impact of removal on school enrollment. School enrollment is measured as an indicator for whether the child was ever enrolled in grades 3-8. We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. All models include the case characteristics listed in Table 1 as controls. Standard errors clustered at the CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 5: Impact of Removal on Foster Care Placement Outcomes

	Young (Aged < 6)		Old (Aged 6+)	
	IV	ML-IV	IV	ML-IV
Dep. var.	(1)	(2)	(3)	(4)
Female				
Days in foster care (all)	394.963***	484.834***	520.208***	464.558***
	(74.000)	(77.012)	(59.705)	(56.624)
Foster family (relatives)	214.087***	229.750***	164.105***	133.207***
	(44.334)	(55.678)	(27.487)	(24.877)
Foster family (non-relatives)	144.733**	242.308***	227.836***	213.645***
	(69.222)	(68.010)	(41.380)	(36.545)
Group home	8.458	7.799*	62.288***	49.978***
	(5.957)	(4.488)	(15.830)	(14.630)
Other	27.685***	4.976	65.979***	67.728***
	(9.107)	(9.045)	(20.540)	(20.082)
Adoption (= 1)	0.000	0.019	0.037	0.032
	(0.050)	(0.057)	(0.024)	(0.020)
N (Obs.)	6,450	6,450	6,650	6,650
Male				
Days in foster care (all)	456.099***	469.425***	374.115***	368.005***
	(87.502)	(101.645)	(63.709)	(58.471)
Foster family (relatives)	207.753***	208.999***	144.392***	131.953***
	(42.886)	(50.501)	(28.557)	(25.884)
Foster family (non-relatives)	190.752***	218.564***	120.555***	129.769***
	(57.265)	(67.160)	(35.066)	(26.703)
Group home	25.406	19.356*	38.175***	31.279***
	(16.915)	(11.298)	(14.470)	(11.196)
Other	32.187	22.506	70.992*	75.004*
	(21.025)	(14.141)	(36.917)	(38.724)
Adoption (= 1)	0.017	0.022	0.015	0.026**
	(0.042)	(0.064)	(0.014)	(0.012)
N (Obs.)	7,400	7,400	6,477	6,477

Notes: This table reports results for the impact of removal on foster care placement outcomes associated with the child's first investigation. We report results by gender and age at investigation. Days in foster care is a measure of total time spent in foster care as a result of the child's first investigation. We 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. Adoption is an indicator for whether the child is adopted upon discharge from foster care. Note that the means of placement outcomes are zero for non-removed children. Columns 1 and 3 report two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Columns 2 and 4 report two-stage least squares estimates and add LASSO selection of instruments to the models in Columns 1 and 3, respectively. All models include the case characteristics listed in Table 1 as controls. Standard errors clustered at the CPI level in parentheses. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 6: Impact of Removal on School Mobility

Dep. var.:		Mean of non-removed	IV	ML-IV
School moves in first 3 yrs.		(1)	(2)	(3)
Female	Young (Aged < 6)	1.012	0.953*	0.418
			(0.491)	(0.406)
			[N=3624]	[N=3624]
	Old (Aged 6+)	1.220	1.477***	1.284***
			(0.364)	(0.345)
			[N=4653]	[N=4653]
Male	Young (Aged < 6)	1.113	0.609	0.638
			(0.443)	(0.400)
			[N=4264]	[N=4264]
	Old (Aged 6+)	1.235	1.076***	1.075***
			(0.322)	(0.290)
			[N=4633]	[N=4633]

Notes: This table reports results for the impact of removal on school mobility. We measure mobility as the total number of school changes in the first three years of school post-investigation. We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. All models include the case characteristics listed in Table 1 as controls. Standard errors clustered at the CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 7: Impact of Removal on School Characteristics

			Mean of non-removed	IV	ML-IV	N (Obs.)
Dep. var. (Avg. over K-6)			(1)	(2)	(3)	(4)
Female	Young (Aged < 6)	School value-added	-0.051	0.027	0.028	3,857
				(0.025)	(0.030)	
		Free lunch share	0.599	-0.098	-0.014	3,947
				(0.090)	(0.116)	
		Minority share	0.492	-0.146	0.076	3,947
				(0.096)	(0.142)	
	Old (Aged 6+)	School value-added	-0.047	-0.028	-0.030	2,351
				(0.031)	(0.027)	
		Free lunch share	0.557	0.060	0.046	2,363
				(0.109)	(0.098)	
		Minority share	0.453	0.047	0.052	2,363
				(0.120)	(0.106)	
Male	Young (Aged < 6)	School value-added	-0.053	0.004	0.003	4,487
				(0.025)	(0.024)	
		Free lunch share	0.587	0.035	0.096	4,581
				(0.076)	(0.070)	
		Minority share	0.481	0.014	0.083	4,581
				(0.084)	(0.079)	
	Old (Aged 6+)	School value-added	-0.048	0.012	0.011	2,805
				(0.020)	(0.020)	
		Free lunch share	0.548	-0.110	-0.050	2,825
				(0.069)	(0.064)	
		Minority share	0.444	-0.058	-0.002	2,825
				(0.078)	(0.071)	

Notes: This table reports results for the impact of removal on school characteristics. School characteristics are measured as a mean over grades K-6. We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. All models include the case characteristics listed in Table 1 as controls. Column 4 reports total observations. Standard errors clustered at the CPI level in parentheses. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 8: Impact of Removal on Special Education Status (IEP)

Dep. var.:		Mean of non-removed	IV	ML-IV
Pct. of years on IEP (K-6)		(1)	(2)	(3)
Female	Young (Aged < 6)	0.175	-0.200	-0.116
			(0.133)	(0.120)
			[N=3947]	[N=3947]
	Old (Aged 6+)	0.211	-0.147	-0.184
			(0.132)	(0.128)
			[N=2363]	[N=2363]
Male	Young (Aged < 6)	0.319	0.157	0.133
			(0.164)	(0.149)
			[N=4581]	[N=4581]
	Old (Aged 6+)	0.366	0.183	0.119
			(0.162)	(0.134)
			[N=2825]	[N=2825]

Notes: This table reports results for the impact of removal on special education status (IEP). We measure special education status in grades K-6 as the percent of years with an IEP. We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. All models include the case characteristics listed in Table 1 as controls. Standard errors clustered at the CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 9: Impact of Removal on Perpetrators

		Panel A: All perpetrators			Panel B: Male perpetrators		
Dep. var.:		Mean of non-removed	IV	ML-IV	Mean of non-removed	IV	ML-IV
Charged within 4 yrs. of investigation		(1)	(2)	(3)	(4)	(5)	(6)
Female	All	0.223	0.154*	0.071	0.319	0.375**	0.099
			(0.081)	(0.077)		(0.154)	(0.166)
			[N=12982]	[N=12982]		[N=6101]	[N=6101]
	Young (Aged < 6)	0.265	0.182	-0.058	0.398	0.307	0.255
			(0.140)	(0.124)		(0.298)	(0.252)
			[N=6403]	[N=6403]		[N=2913]	[N=2913]
	Old (Aged 6+)	0.185	0.144*	0.090	0.249	0.407***	0.153
			(0.088)	(0.084)		(0.156)	(0.145)
			[N=6579]	[N=6579]		[N=3188]	[N=3188]
Male	All	0.213	0.038	-0.050	0.295	0.133	-0.038
			(0.085)	(0.074)		(0.144)	(0.131)
			[N=13760]	[N=13760]		[N=6804]	[N=6804]
	Young (Aged < 6)	0.247	0.102	0.014	0.365	0.216	-0.182
			(0.106)	(0.104)		(0.189)	(0.273)
			[N=7338]	[N=7338]		[N=3402]	[N=3402]
	Old (Aged 6+)	0.176	-0.026	-0.111	0.229	0.037	-0.140
			(0.097)	(0.081)		(0.163)	(0.145)
			[N=6422]	[N=6422]		[N=3402]	[N=3402]

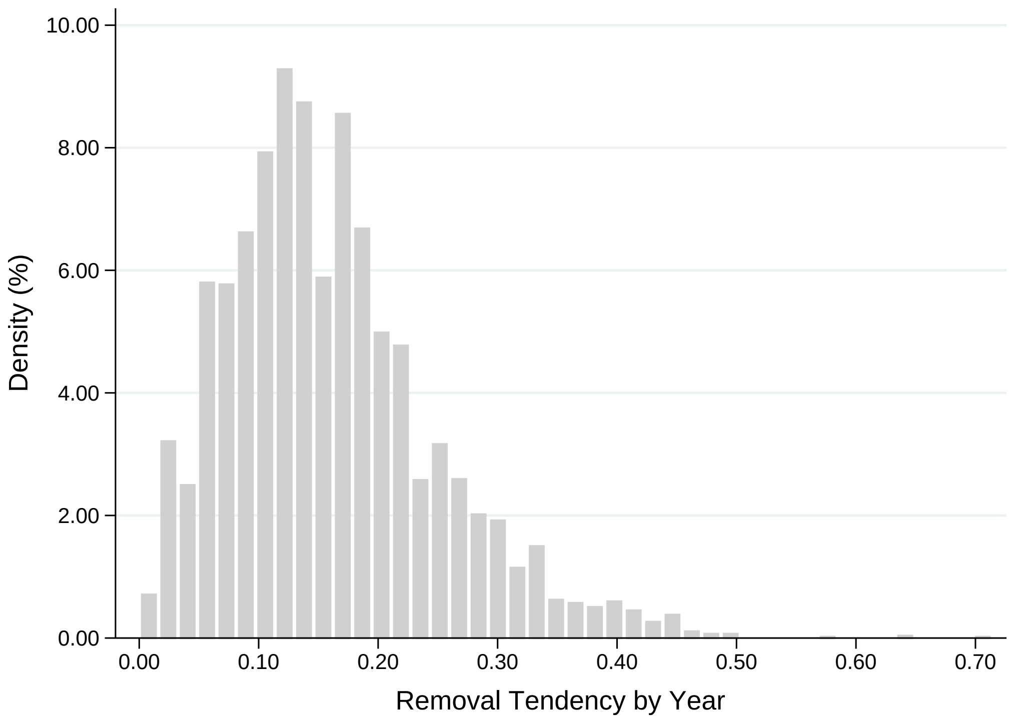
Notes: This table reports results for the impact of removal on perpetrators. We measure whether perpetrators were ever charged or sentenced with a crime in the four years after an investigation. We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. Columns 4-6 report results using the sample of male perpetrators only. All models include the case characteristics listed in Table 1 as controls. Standard errors clustered at the CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table 10: Impact of Removal on Later-Life Outcomes for Older Children

	Mean of non-removed	IV	ML-IV	N (Obs.)
Dep. var.	(1)	(2)	(3)	(4)
Female (Aged 6+)				
Juvenile conviction (= 1)	0.062	-0.059	-0.065	4,231
		(0.066)	(0.051)	
High school grad (= 1)	0.505	-0.131	-0.084	2,770
		(0.169)	(0.161)	
Any postsecondary (= 1)	0.305	0.177	0.135	2,564
		(0.173)	(0.118)	
Teen mother (= 1)	0.220	0.040	0.063	3,563
		(0.165)	(0.124)	
Male (Aged 6+)				
Juvenile conviction (= 1)	0.169	0.057	-0.032	3,766
		(0.118)	(0.102)	
High school grad (= 1)	0.451	-0.084	-0.020	2,497
		(0.194)	(0.176)	
Any postsecondary (= 1)	0.248	0.014	-0.027	2,189
		(0.163)	(0.158)	
Teen father (= 1)	0.065	0.124*	0.074	3,143
		(0.073)	(0.071)	

Notes: This table reports results for the impact of removal on later-life outcomes. We report results for older children (aged 6+), split by gender. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. All models include the case characteristics listed in Table 1 as controls. Column 4 reports total observations. Standard errors clustered at the CPI level in parentheses. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Figure 1: CPI Removal Tendency



Notes: This figure reports the distribution of CPI removal tendency for the DCYF sample described in Section 3.1. We estimate a leave-one-out removal tendency using data from other children assigned to a CPI in the same year, as described in Section 4. We restrict to CPIs who handle at least 10 cases per year. The total number of observations is 26,977 and the number of unique CPIs is 103.

Appendix A: Supplementary Tables

Table A.1: Descriptive Statistics, by Subgroup

		All	Female	Male	Young Female	Young Male	Old Female	Old Male
Variable		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Demographics	Female	0.49	1.00	0.00	1.00	0.00	1.00	0.00
		(0.50)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	White	0.62	0.62	0.62	0.61	0.62	0.62	0.63
		(0.49)	(0.49)	(0.48)	(0.49)	(0.49)	(0.49)	(0.48)
	African American	0.17	0.17	0.16	0.18	0.17	0.16	0.15
		(0.37)	(0.37)	(0.37)	(0.38)	(0.38)	(0.37)	(0.36)
	Hispanic	0.17	0.17	0.17	0.17	0.17	0.18	0.17
		(0.38)	(0.38)	(0.38)	(0.37)	(0.38)	(0.38)	(0.38)
	Age at investigation	6.19	6.55	5.84	1.78	1.83	11.18	10.42
		(5.22)	(5.43)	(4.99)	(1.76)	(1.76)	(3.42)	(3.24)
Family	Married couple	0.21	0.21	0.21	0.14	0.16	0.28	0.27
		(0.41)	(0.41)	(0.41)	(0.34)	(0.36)	(0.45)	(0.44)
	Unmarried couple	0.28	0.28	0.29	0.36	0.34	0.21	0.22
		(0.45)	(0.45)	(0.45)	(0.48)	(0.48)	(0.40)	(0.42)
	Single parent/other	0.50	0.51	0.50	0.50	0.50	0.51	0.51
		(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
	English language	0.97	0.97	0.97	0.98	0.98	0.96	0.97
		(0.17)	(0.18)	(0.16)	(0.15)	(0.15)	(0.20)	(0.17)
	Spanish language	0.03	0.03	0.03	0.02	0.02	0.04	0.03
		(0.17)	(0.18)	(0.16)	(0.15)	(0.15)	(0.20)	(0.17)
Allegation	Neglect	0.82	0.82	0.81	0.86	0.83	0.78	0.79
		(0.39)	(0.39)	(0.39)	(0.35)	(0.37)	(0.41)	(0.41)
	Physical neglect	0.05	0.04	0.05	0.06	0.07	0.02	0.02
		(0.21)	(0.20)	(0.22)	(0.24)	(0.26)	(0.14)	(0.15)
	Physical abuse	0.20	0.20	0.20	0.13	0.15	0.27	0.26
		(0.40)	(0.40)	(0.40)	(0.34)	(0.36)	(0.44)	(0.44)
Reporter	Professional	0.82	0.82	0.83	0.83	0.84	0.80	0.81
		(0.38)	(0.39)	(0.38)	(0.38)	(0.37)	(0.40)	(0.39)
	Family/friend	0.14	0.15	0.14	0.13	0.12	0.16	0.15
		(0.35)	(0.35)	(0.34)	(0.34)	(0.33)	(0.36)	(0.35)
	Other	0.04	0.04	0.04	0.04	0.04	0.04	0.04
		(0.19)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)	(0.20)
Investigation	Emergency	0.06	0.06	0.06	0.10	0.10	0.02	0.02
		(0.24)	(0.24)	(0.25)	(0.30)	(0.30)	(0.13)	(0.14)
	Immediate	0.63	0.62	0.63	0.56	0.59	0.69	0.69
		(0.48)	(0.48)	(0.48)	(0.50)	(0.49)	(0.46)	(0.46)
	Routine	0.31	0.32	0.30	0.34	0.31	0.29	0.29
		(0.46)	(0.46)	(0.46)	(0.47)	(0.46)	(0.46)	(0.46)
	Removed	0.16	0.17	0.16	0.21	0.20	0.13	0.12
		(0.37)	(0.37)	(0.37)	(0.41)	(0.40)	(0.34)	(0.32)
Observations		26,977	13,100	13,877	6,450	7,400	6,650	6,477

Notes: This table reports descriptive statistics for children in the DCYF sample described in Section 3.1. Column 1 reports the mean and standard deviation (in parentheses) for all children. Columns 2 and 3 report statistics for female and male children, respectively. Columns 4 and 5 report statistics for young (aged < 6) females and young males. Columns 6 and 7 report statistics for old (aged 6+) females and old males.

Table A.2: Tests of Random Case Assignment (Full Regression Results)

Dep. var.:	All	Female	Male	Young Female	Young Male	Old Female	Old Male
CPI removal tendency	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.001						
	(0.001)						
White	-0.003	-0.002	-0.003	0.003	-0.001	-0.009*	-0.006
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)
African American	0.000	-0.001	0.001	-0.002	0.002	-0.002	0.000
	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)	(0.006)	(0.005)
Hispanic	-0.001	-0.003	0.001	0.003	0.002	-0.013**	-0.002
	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)
Age at investigation	-0.000*	-0.000*	-0.000	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Married couple	-0.001	-0.001	-0.001	-0.008	-0.002	0.003	-0.001
	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
Unmarried couple	-0.002	0.000	-0.004	-0.003	-0.006*	0.005	-0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Single parent/other	-0.002	-0.001	-0.003	-0.002	-0.003	-0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
English language	0.003	-0.004	0.011	-0.001	0.008	-0.003	0.014
	(0.007)	(0.008)	(0.009)	(0.014)	(0.014)	(0.010)	(0.011)
Spanish language	0.007	0.004	0.012	0.004	-0.005	0.006	0.026*
	(0.009)	(0.010)	(0.011)	(0.015)	(0.015)	(0.012)	(0.014)
Allegation - neglect	0.003	0.002	0.004	0.006	0.004	-0.001	0.004
	(0.003)	(0.004)	(0.004)	(0.006)	(0.005)	(0.006)	(0.005)
Allegation - phys. neglect	0.003	0.003	0.003	0.010*	0.004	-0.010	0.002
	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)	(0.007)	(0.009)
Allegation - phys. abuse	0.002	0.000	0.004	0.001	0.004	0.000	0.004
	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)	(0.006)	(0.004)
Reporter - professional	-0.004	-0.003	-0.004	0.002	0.004	-0.006	-0.012
	(0.006)	(0.006)	(0.007)	(0.012)	(0.009)	(0.007)	(0.009)
Reporter - family/friend	0.001	0.000	0.002	0.004	0.012	-0.003	-0.007
	(0.006)	(0.008)	(0.007)	(0.013)	(0.009)	(0.009)	(0.010)
Reporter - other	-0.001	0.003	-0.005	0.007	0.005	-0.000	-0.013
	(0.006)	(0.008)	(0.007)	(0.014)	(0.010)	(0.009)	(0.010)
Emergency investigation	0.003	0.002	0.004	0.001	0.002	0.004	0.014
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)	(0.010)
Immediate investigation	0.002	0.003	0.001	0.002	0.001	0.004	0.001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)
F-statistic	1.058	0.982	1.250	0.862	1.143	1.287	1.525
p-value	0.405	0.484	0.241	0.619	0.326	0.216	0.101
Observations	26,977	13,100	13,877	6,450	7,400	6,650	6,477

Notes: This table reports regression results testing the random assignment of cases to CPIs. Results are from an OLS regression of CPI removal tendency on the case characteristics listed, indicators for missing information, and investigation year fixed effects. Column 1 reports estimates for the DCYF sample. Columns 2 and 3 report estimates for female and male children, respectively. Columns 4 and 5 report estimates for young (aged < 6) females and young males. Columns 6 and 7 report estimates for old (aged 6+) females and old males. The F-statistic and p-value reported are from an F-test for joint significance of all variables except investigation year fixed effects. Standard errors clustered at the CPI level in parentheses. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table A.3: First-Stage Impact of CPI Removal Tendency, by Subgroup

	All	Female	Male	Young Female	Young Male	Old Female	Old Male
Dep. var.: Removed (= 1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Subgroup:							
All	0.533***	0.496***	0.568***	0.431***	0.511***	0.556***	0.636***
	(0.030)	(0.041)	(0.044)	(0.062)	(0.063)	(0.047)	(0.057)
	[26,977]	[13,100]	[13,877]	[6,450]	[7,400]	[6,650]	[6,477]
White	0.487***	0.435***	0.541***	0.383***	0.528***	0.470***	0.546***
	(0.044)	(0.050)	(0.061)	(0.071)	(0.083)	(0.065)	(0.077)
	[16,304]	[7,855]	[8,449]	[3,769]	[4,374]	[4,086]	[4,075]
African American	0.548***	0.611***	0.482***	0.495***	0.368***	0.751***	0.691***
	(0.091)	(0.138)	(0.107)	(0.184)	(0.135)	(0.163)	(0.161)
	[4,364]	[2,153]	[2,211]	[1,108]	[1,228]	[1,045]	[983]
Hispanic	0.665***	0.634***	0.678***	0.426***	0.603***	0.826***	0.821***
	(0.083)	(0.109)	(0.106)	(0.121)	(0.125)	(0.169)	(0.141)
	[4,538]	[2,212]	[2,326]	[1,029]	[1,216]	[1,183]	[1,110]
Married couple	0.443***	0.444***	0.453***	0.437**	0.391***	0.428***	0.492***
	(0.075)	(0.093)	(0.095)	(0.174)	(0.133)	(0.090)	(0.114)
	[4,751]	[2,332]	[2,419]	[741]	[955]	[1,591]	[1,464]
Unmarried couple	0.593***	0.553***	0.631***	0.532***	0.588***	0.551***	0.695***
	(0.069)	(0.101)	(0.092)	(0.117)	(0.120)	(0.154)	(0.142)
	[6,437]	[3,112]	[3,325]	[1,951]	[2,101]	[1,161]	[1,224]
Single parent/other	0.548***	0.518***	0.573***	0.493***	0.490***	0.570***	0.669***
	(0.056)	(0.062)	(0.073)	(0.096)	(0.104)	(0.084)	(0.088)
	[11,406]	[5,557]	[5,849]	[2,681]	[3,035]	[2,876]	[2,814]
Allegation - neglect	0.573***	0.539***	0.605***	0.462***	0.536***	0.628***	0.683***
	(0.036)	(0.048)	(0.050)	(0.068)	(0.069)	(0.059)	(0.070)
	[21,989]	[10,726]	[11,263]	[5,528]	[6,176]	[5,198]	[5,087]
Allegation - phys. neglect	0.309*	0.296	0.281	0.303	0.316	0.384	0.339
	(0.160)	(0.197)	(0.197)	(0.265)	(0.238)	(0.265)	(0.216)
	[1,233]	[551]	[682]	[411]	[529]	[140]	[153]
Allegation - phys. abuse	0.408***	0.317***	0.484***	0.304*	0.399***	0.333***	0.541***
	(0.072)	(0.090)	(0.087)	(0.175)	(0.142)	(0.098)	(0.125)
	[5,446]	[2,639]	[2,807]	[833]	[1,109]	[1,806]	[1,698]
Emergency investigation	0.125	0.138	0.070	0.136	0.003	0.607	0.268
	(0.199)	(0.247)	(0.227)	(0.242)	(0.241)	(0.562)	(0.372)
	[1,669]	[775]	[894]	[659]	[762]	[116]	[132]
Immediate investigation	0.591***	0.562***	0.620***	0.551***	0.623***	0.555***	0.606***
	(0.037)	(0.052)	(0.051)	(0.081)	(0.079)	(0.060)	(0.069)
	[16,970]	[8,187]	[8,783]	[3,609]	[4,334]	[4,578]	[4,449]
Routine investigation	0.496***	0.395***	0.594***	0.309***	0.494***	0.513***	0.703***
	(0.054)	(0.078)	(0.067)	(0.098)	(0.090)	(0.106)	(0.105)
	[8,338]	[4,138]	[4,200]	[2,182]	[2,304]	[1,956]	[1,896]

Notes: This table summarizes the first-stage relationship between removal and CPI removal tendency for different subgroups. Removed is an indicator for home removal at the child's first investigation. Results are from an OLS regression of removal on CPI removal tendency and the set of case characteristics listed in Table 1. All models include indicators for missing information and investigation year fixed effects. Column 1 reports estimates for the DCYF sample. Columns 2 and 3 report estimates for female and male children, respectively. Columns 4 and 5 report estimates for young (aged < 6) females and young males. Columns 6 and 7 report estimates for old (aged 6+) females and old males. Standard errors clustered at the CPI level in parentheses. Sample size in brackets. Significance reported as *** $p<0.01$; ** $p<0.05$; * $p<0.10$.

Table A.4: Characteristics of Compliers, by Subgroup

	All	Female	Male	Young Female	Young Male	Old Female	Old Male
Subgroup:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.931						
Male	1.066						
Young (Aged < 6)	0.889	0.869	0.897				
Old (Aged 6+)	1.135	1.143	1.133				
White	0.914	0.879	0.953	0.887	1.032	0.846	0.849
African American	1.045	1.214	0.895	1.069	0.720	1.403	1.140
Hispanic	1.264	1.301	1.190	1.005	1.184	1.486	1.254
Married couple	0.848	0.913	0.813	0.999	0.768	0.800	0.809
Unmarried couple	1.097	1.099	1.096	1.216	1.141	0.970	1.066
Single parent/other	1.043	1.058	1.006	1.158	0.962	1.025	1.049
English language	0.997	0.996	0.997	1.018	0.998	0.975	0.990
Allegation - neglect	1.076	1.088	1.064	1.069	1.052	1.128	1.061
Allegation - phys. neglect	0.615	0.632	0.521	0.778	0.642	0.766	0.584
Allegation - phys. abuse	0.766	0.640	0.849	0.669	0.771	0.599	0.856
Reporter - professional	0.954	0.944	0.962	0.857	0.968	1.000	0.959
Reporter - family/friend	1.234	1.265	1.172	1.551	1.089	1.141	1.165

Notes: This table reports characteristics of compliers in the DCYF sample. 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 use our instrument Z_i, calculated as the leave-one-out removal tendency of the investigator assigned to each child. Following Dobbie et al. (2018), let $\bar{z}$ 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: $\pi_c = Pr(Removed = 1|Z_i = \bar{z}) - Pr(Removed = 1|Z_i = \underline{z})$. In practice, we assign the top percentile of our instrument to $\bar{z}$ and the bottom percentile of our instrument to $\underline{z}$. As discussed in Dahl et al. (2014) and Dobbie et al. (2018), the share of compliers can be directly estimated as $\pi_c = \hat{\alpha}_1(\bar{z} - \underline{z})$, where $\hat{\alpha}_1$ is the coefficient on the instrument from the first stage regression (Equation 2). In this table, we estimate the share of compliers in each subgroup (row) and report the ratio of this share to the overall complier share in each column. For example, the value reported for the Female subgroup is the ratio of the complier share among females to the complier share in the full DCYF sample.

Table A.5: Robustness: Impact of Removal Using Alternative Measure of Removal Tendency

Dep. var.:		Mean of non-removed	IV	ML-IV
Average test score		(1)	(2)	(3)
Female	All	-0.390	0.563***	0.376**
			(0.160)	(0.153)
			[N=5468]	[N=5468]
	Young (Aged < 6)	-0.385	1.121***	1.147***
			(0.236)	(0.283)
			[N=2699]	[N=2699]
	Old (Aged 6+)	-0.396	-0.271	0.044
			(0.241)	(0.199)
			[N=2769]	[N=2769]
Male	All	-0.557	-0.226	-0.004
			(0.149)	(0.141)
			[N=6259]	[N=6259]
	Young (Aged < 6)	-0.557	-0.662**	-0.135
			(0.270)	(0.243)
			[N=3135]	[N=3135]
	Old (Aged 6+)	-0.556	0.117	0.007
			(0.174)	(0.155)
			[N=3124]	[N=3124]

Notes: This table reports robustness results for the impact of removal on test scores. CPI removal tendency is calculated here using all years in which the CPI is observed. We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 3. All models include the set of case characteristics listed in Table 1 as controls, as well as grade and school year fixed effects. Standard errors two-way clustered at the child-CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table A.6: Impact of Removal on First Observed Test Scores

Dep. var.:		Panel A: First test score			Panel B: First three test scores		
		Mean of non-removed	IV	ML-IV	Mean of non-removed	IV	ML-IV
		(1)	(2)	(3)	(4)	(5)	(6)
Female	All	-0.398	0.518*	0.537**	-0.393	0.504*	0.457*
			(0.272)	(0.248)		(0.267)	(0.256)
			[N=5468]	[N=5468]		[N=5468]	[N=5468]
	Young (Aged < 6)	-0.395	1.479***	1.367***	-0.393	1.296***	1.224***
			(0.544)	(0.526)		(0.468)	(0.427)
			[N=2699]	[N=2699]		[N=2699]	[N=2699]
	Old (Aged 6+)	-0.401	-0.083	0.040	-0.392	-0.152	-0.075
			(0.337)	(0.316)		(0.317)	(0.315)
			[N=2769]	[N=2769]		[N=2769]	[N=2769]
Male	All	-0.544	-0.316	-0.291	-0.551	-0.319	-0.265
			(0.201)	(0.199)		(0.203)	(0.198)
			[N=6259]	[N=6259]		[N=6259]	[N=6259]
	Young (Aged < 6)	-0.537	-0.293	-0.340	-0.547	-0.295	-0.309
			(0.316)	(0.304)		(0.334)	(0.327)
			[N=3135]	[N=3135]		[N=3135]	[N=3135]
	Old (Aged 6+)	-0.550	-0.308	-0.194	-0.555	-0.308	-0.155
			(0.316)	(0.272)		(0.278)	(0.241)
			[N=3124]	[N=3124]		[N=3124]	[N=3124]

Notes: This table reports results for the impact of removal on first observed test scores. Test scores are the average of a child's reading and math scores. We standardize scores at the grade-year level and construct a yearly panel of tests taken in grades 3-8 and in school years 2005-2016. We report results by gender and age at investigation. Column 1 reports the mean of the first observed test score for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. Columns 4-6 report results using the first three observed test scores. All models include the case characteristics listed in Table 1 as controls, as well as grade and school year fixed effects. Standard errors two-way clustered at the child-CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table A.7: Adjusted p-values for Impact of Removal on Test Scores

(a) Average Test Score (IV Estimates)

	Estimate	Std. Error	p-value	FDR q-value
Subgroup:	(1)	(2)	(3)	(4)
Young Female	1.068	0.431	0.015	0.060
Old Female	-0.250	0.312	0.425	0.567
Young Male	0.003	0.334	0.993	0.994
Old Male	-0.383	0.291	0.191	0.383

(b) Average Test Score (ML-IV Estimates)

	Estimate	Std. Error	p-value	FDR q-value
Subgroup:	(1)	(2)	(3)	(4)
Young Female	1.088	0.418	0.011	0.043
Old Female	-0.134	0.305	0.661	0.881
Young Male	0.002	0.306	0.996	0.996
Old Male	-0.173	0.230	0.455	0.881

Notes: This table reports adjusted p-values for the impact of removal on test scores (see Table 3). Subtable (a) reports results for IV estimates while Subtable (b) reports results for ML-IV estimates. Columns 1 and 2 report estimates and standard errors. Columns 3 and 4 report per-comparison (pairwise) and false discovery rate (FDR) adjusted p-values for the four subgroups considered in the analysis of children: young (aged < 6) females, old (aged 6+) females, young males, and old males. 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 A.8: Impact of Removal on Perpetrators: Additional Measures

(a) Within 1 Year of Investigation

		Panel A: All perpetrators			Panel B: Male perpetrators		
Dep. var.: Charged within 1 year of investigation		Mean of non-removed	IV	ML-IV	Mean of non-removed	IV	ML-IV
		(1)	(2)	(3)	(4)	(5)	(6)
Female	All	0.125	0.108	0.063	0.207	0.246	0.242
			(0.077)	(0.066)		(0.163)	(0.155)
			[N=12982]	[N=12982]		[N=6101]	[N=6101]
	Young (Aged < 6)	0.153	0.101	0.026	0.265	0.285	0.479*
			(0.127)	(0.106)		(0.315)	(0.265)
			[N=6403]	[N=6403]		[N=2913]	[N=2913]
	Old (Aged 6+)	0.101	0.115	0.070	0.156	0.215	0.164
			(0.072)	(0.065)		(0.147)	(0.121)
			[N=6579]	[N=6579]		[N=3188]	[N=3188]
Male	All	0.119	-0.016	-0.070	0.186	-0.017	-0.115
			(0.069)	(0.062)		(0.138)	(0.125)
			[N=13760]	[N=13760]		[N=6804]	[N=6804]
	Young (Aged < 6)	0.141	-0.016	-0.051	0.237	-0.097	-0.172
			(0.093)	(0.087)		(0.200)	(0.210)
			[N=7338]	[N=7338]		[N=3402]	[N=3402]
	Old (Aged 6+)	0.095	-0.015	-0.094	0.137	0.037	-0.123
			(0.074)	(0.069)		(0.143)	(0.132)
			[N=6422]	[N=6422]		[N=3402]	[N=3402]

(b) Within 2 Years of Investigation

		Panel A: All perpetrators			Panel B: Male perpetrators		
Dep. var.: Charged within 2 yrs. of investigation		Mean of non-removed	IV	ML-IV	Mean of non-removed	IV	ML-IV
		(1)	(2)	(3)	(4)	(5)	(6)
Female	All	0.175	0.107	0.040	0.268	0.227	0.088
			(0.076)	(0.069)		(0.167)	(0.158)
			[N=12982]	[N=12982]		[N=6101]	[N=6101]
	Young (Aged < 6)	0.211	0.118	-0.006	0.343	0.255	0.317
			(0.129)	(0.110)		(0.333)	(0.268)
			[N=6403]	[N=6403]		[N=2913]	[N=2913]
	Old (Aged 6+)	0.142	0.104	0.038	0.202	0.207	0.063
			(0.082)	(0.078)		(0.152)	(0.131)
			[N=6579]	[N=6579]		[N=3188]	[N=3188]
Male	All	0.167	0.012	-0.041	0.247	0.056	-0.105
			(0.078)	(0.068)		(0.152)	(0.137)
			[N=13760]	[N=13760]		[N=6804]	[N=6804]
	Young (Aged < 6)	0.195	0.067	0.030	0.310	0.110	-0.292
			(0.103)	(0.096)		(0.212)	(0.245)
			[N=7338]	[N=7338]		[N=3402]	[N=3402]
	Old (Aged 6+)	0.137	-0.039	-0.115	0.188	-0.012	-0.180
			(0.091)	(0.080)		(0.159)	(0.149)
			[N=6422]	[N=6422]		[N=3402]	[N=3402]

Notes: These tables report additional results for the impact of removal on perpetrators. We measure whether perpetrators were ever charged or sentenced with a crime in (a) the year or (b) the two years after an investigation. We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. Columns 4-6 report results using the sample of male perpetrators only. All models include the case characteristics listed in Table 1 as controls. Standard errors clustered at the CPI level in parentheses. Sample size in brackets. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Table A.9: Impact of Removal on Attrition from Rhode Island

			Mean of non-removed	IV	ML-IV	N (Obs.)
Dep. var.: Not observed (Pct. of years)			(1)	(2)	(3)	(4)
Female	Young (Aged < 6)	5 yrs. after	0.235	-0.132	-0.013	4,596
				(0.139)	(0.147)	
		10 yrs. after	0.307	-0.187	0.089	2,383
				(0.172)	(0.180)	
	Old (Aged 6+)	5 yrs. after	0.335	-0.129	-0.176*	5,209
				(0.105)	(0.103)	
		10 yrs. after	0.380	-0.143	-0.140	3,412
				(0.153)	(0.128)	
Male	Young (Aged < 6)	5 yrs. after	0.236	0.147	0.008	5,351
				(0.135)	(0.167)	
		10 yrs. after	0.309	0.029	0.010	2,891
				(0.127)	(0.128)	
	Old (Aged 6+)	5 yrs. after	0.336	-0.159	-0.129	5,009
				(0.129)	(0.110)	
		10 yrs. after	0.416	-0.222*	-0.176	3,199
				(0.123)	(0.109)	

Notes: This table reports results for the impact of removal on attrition from Rhode Island. We measure attrition as the share of years post-investigation (in a balanced panel) in which a child is not observed in social assistance records (Medicaid, SNAP, or TANF). We report results by gender and age at investigation. Column 1 reports the mean of the dependent variable for non-removed children. Column 2 reports two-stage least squares estimates with the standard leave-one-out measure of CPI removal tendency as an instrument for removal. Column 3 reports two-stage least squares estimates and adds LASSO selection of instruments to the model in Column 2. All models include the case characteristics listed in Table 1 as controls. Column 4 reports total observations. Standard errors clustered at the CPI level in parentheses. Significance reported as *** p<0.01; ** p<0.05; * p<0.10.

Appendix B: Data Construction

We create our analysis sample of children subject to investigations based on all data on DCYF allegations that occurred between 2000 and 2015. The creation of the sample proceeds in two main steps, which are outlined in Appendix Table B.1 (below):

- First, we clean the initial data on 187,023 allegations. This process involves (a) dropping all allegations where a child has incomplete information (*e.g.*, missing an identifier or demographic information), (b) dropping allegations not reported via the DCYF hotline (*e.g.*, reported by fax, by mail, or during an investigation), (c) dropping allegations that do not involve a family (*e.g.*, allegations of abuse or neglect for children in DCYF care), (d) dropping non-standard allegations (*e.g.*, "additional info." allegations that are reported after the initial hotline call), (e) dropping allegations that do not meet the criteria for an investigation, (f) dropping unfounded investigations (*e.g.*, instances in which a CPI found no evidence of child abuse or neglect), (g) dropping investigations that occur before 2000 (because the data is incomplete) or after 2015 (because we cannot observe any outcome after this point), (h) dropping children who are part of more than one investigation on the same day, and (i) dropping any investigation that we cannot match to a CPI in the DCYF case assignments file.
- Second, we create our main analysis sample after imposing a series of restrictions to the cleaned DCYF data on 30,279 investigations for 31,883 children. This process involves (a) dropping any "augmented" case assignments (*e.g.*, instances where the CPI either volunteered for the case or was mandated to take the case), (b) dropping investigations in which the CPI primarily worked as a hotline worker in that month, (c) dropping investigations that occur after an initial investigation for the associated CPI (since these may not be assigned via the rotation list), (d) dropping sex abuse cases (not always randomly assigned), (e) dropping investigations that are not the first investigation for the associated child, and (f) dropping investigations where the CPI has fewer than 10 cases in the calculation of CPI removal tendency.[39]

[39] We generate several measures of removal tendency that vary by child and case characteristics. These additional instruments are undefined if a CPI saw no other cases within a given cell, and we drop investigations with any missing instruments.

Table B.1: Data Construction

		Allegations	Investigations	Children
Full DCYF data		187,023	54,119	63,351
1. Data cleaning				
a.	Drop allegations with incomplete/invalid child info.	176,034	51,864	58,429
b.	Drop allegations not reported via the DCYF hotline	154,809	51,585	56,508
c.	Drop allegations that do not involve a family	146,372	49,103	54,427
d.	Drop non-standard 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.	Drop investigations before 2000 or after 2015	71,451	33,492	34,364
h.	Drop if child is in multiple investigations on the same day	71,278	33,418	34,348
i.	Drop investigations not matched to a CPI assignment	70,039	32,845	33,971
2. Restrictions for DCYF sample				
a.	Drop augmented case assignments	64,848	30,279	31,883
b.	Drop if CPI is flagged as a hotline worker	64,728	30,226	31,848
c.	Drop non-first investigations at the CPI level	57,978	27,046	29,276
d.	Drop sex abuse investigations	54,691	25,309	27,788
e.	Drop non-first investigations at the child level	39,799	19,833	27,596
f.	Drop if CPI has fewer than 10 cases within the year	38,886	19,362	26,977

Appendix C: School Characteristics

We use enrollment records from the Rhode Island Department of Education (RIDE) to measure the following characteristics of Rhode Island schools:

School Value-Added

We construct one school value-added measure for each school that considers all tests taken by children in grades 4-8. Children in Rhode Island took the New England Common Assessment Program (NECAP) test in school years 2005-2013. The state switched to the Partnership for Assessment of Readiness for College and Careers (PARCC) test in 2014, and so we also have PARCC test scores for school years 2014-2016. We restrict to students not in our DCYF sample discussed in Section 3.1. We also drop test scores for students repeating grades and students who are missing any of the baseline controls we use in our estimation. We estimate school value-added (μ) from the following student-level regression:

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

where

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

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 status, and lagged test scores (their square and cube). The residual ν_{ijt} is composed of school value-added (μ) and an error term. To match students in our DCYF sample to measures of school value-added, we identify school years after the child's first investigation and assign the value-added measure to the first school attended in a given year.

Free Lunch and Minority Share

We construct additional characteristics for the schools attended by children in our DCYF sample. Free lunch share is the fraction of students with free or reduced price lunch subsidies measured at the school-year level. Minority share is the fraction of non-white students measured at the school-year level. We restrict to students not in our DCYF sample discussed in Section 3.1 and we drop students with missing demographics. To match students in our DCYF sample to these characteristics, we identify school years after the child's first investigation and assign measures for the first school attended in a given year.