

Temporary Stays and Persistent Gains: The Causal Effects of Foster Care*

Max Gross[†]

E. Jason Baron[‡]

November 29, 2020

Abstract

Six percent of children in the United States enter foster care by age 18. We estimate the effects of foster care on children's outcomes by exploiting the quasi-random assignment of child welfare investigators in Michigan. We find that foster care improved children's safety and educational outcomes. Gains emerged after children exited the foster system when most were reunified with their birth parents, suggesting that improvements made by their parents was an important mechanism. These results indicate that safely reducing the use of foster care, a goal of recent federal legislation, requires more effective in-home, prevention-focused efforts.

*We would like to thank Brian Jacob, Michael Mueller-Smith, Joseph Ryan, Charlie Brown, and Kevin Stange for their invaluable advice and guidance. We also benefited from feedback from Anna Aizer, Mark Courtney, Ashley Craig, Joseph Doyle, Susan Dynarski, Sara Heller, Sam Norris, Elizabeth Weigensberg, Fred Wulczyn, George Fenton, Matthew Gross, Parag Mahajan, Stephanie Owen, and Andrew Simon as well as feedback from seminar participants at Abt Associates, the Association for Education Finance and Policy, the Association of Public Policy and Management, Mathematica, and the University of Michigan. We appreciate the Child and Adolescent Data Lab for their generosity in sharing data, Andrew Moore and Daniel Hubbard for their help with record linkage, and Jasmina Camo-Biogradlija, Terri Gilbert, Andrea Plevck, and Nicole Wagner Lam for coordinating data access. We also thank the many child welfare employees across Michigan for help in understanding how the system works in practice and for bringing humanity to the data. We use data structured and maintained by the Michigan Consortium for Education Research (MCER). MCER data are modified for analysis purposes using rules governed by MCER and are not identical to those data collected and maintained by the Michigan Department of Education and the Center for Educational Performance and Information.

[†]Corresponding author. Mathematica. Email: mgross@mathematica-mpr.com.

[‡]Ford School of Public Policy, University of Michigan. E-mail: ejbaron@umich.edu.

“There are two powerful, emotional story lines in child welfare...There’s a strong pull for us to reject the disruption of families by governmental authorities. But children are sometimes harmed by their parents.”

— Dr. Matthew Stagner, Association for Public Policy Analysis & Management
Presidential Address, 2019

1 Introduction

About 250,000 children entered the foster system every year in the United States from 2000 to 2017 because they were abused or neglected at home (AECF, 2017; USDHHS, 2018a). By age 18, up to 6 percent of children—including up to 10 percent of Black children and 15 percent of Native American children—will have entered foster care at some point (Wildeman and Emanuel, 2014). Among historically vulnerable groups, foster children experience the worst life outcomes (Barrat and Berliner, 2013); however, there is little causal evidence on the impacts of foster care. Pathbreaking research in Doyle (2007, 2008) studied placements nearly two decades ago in Illinois and concluded that foster care was damaging for children. But the foster system in Illinois was not representative of other states at the time (USDHHS, 2003b) and nationwide child welfare policy and practice has since changed (ChildTrends, 2018). Especially given their increased use in response to the opioid epidemic (Talbot, 2017; Neilson, 2019), it is critical to understand the effectiveness of current foster care systems.

This paper makes several contributions. First, it provides new estimates of the causal effects of foster care on crucial indicators of child well-being: safety, education, and crime. Identifying causal impacts is challenging because foster children differ from their peers along a variety of dimensions. To overcome selection bias, we leverage exogenous variation in placement created by the quasi-random assignment of child welfare investigators who vary in their propensity to recommend foster care. Using administrative records from Michigan that link public school students to child welfare involvement and juvenile court filings, this study analyzes over 200,000 maltreatment investigations of school-age children between 2008 and 2016.

We find that foster care improved children’s outcomes. It reduced the likelihood that children were alleged as victims of abuse or neglect in the future by 13.2 percentage points, a 52 percent reduction relative to a baseline mean of 25.5 percent. In addition to improving child safety, placement had large, positive impacts on academic outcomes; it increased daily school attendance by 6.0 percent and standardized math test scores by 0.36 standard deviations. We also find a substantial but less precise reduction in juvenile delinquency.

Taken together, these estimates indicate that foster care had benefits in cases where investigators might disagree about placement, which is a critical population for child welfare policy (Berrick, 2018).

The results contrast with Doyle (2007, 2008), which used the same research design but found that foster care reduced earnings and increased crime for Illinois children investigated in the 1990s and early 2000s.¹ In fact, we can statistically reject that foster placement in Michigan during our sample period had the large negative impacts on children’s outcomes found in this earlier work. There are several possible explanations for this discrepancy. A likely reason is that children’s experiences while in the Illinois foster system were especially harmful. For example, foster children in Illinois remained in the system longer than in any other state at the time and changed foster homes at a higher rate than in all but two states (Figure 1). Therefore, placements in other states may have been less damaging than in Illinois, and perhaps beneficial. Importantly, evidence from our study is more likely to be representative, because the system in Michigan functions similarly to others across the country. Another explanation is that shifts in child welfare practice over time may have helped foster systems improve nationwide, such as increasing placements with relatives and decreasing length of stay in care (ChildTrends, 2018). We find less evidence for other potential reasons, such as differences between children at the margin of placement across settings.

A second contribution of this study is that it explores mechanisms by exploiting the fact that foster care is a temporary intervention. In our setting, children were in the foster system for 19 months, on average. During this initial period, there were no discernible differences in outcomes between children placed and not placed in foster care. Instead, the gains in safety and education emerged in the range of three to five years after placement, when most children were reunified with their birth parents.² One explanation for this surprising pattern is that birth parents, who worked closely with social workers following child removal, improved their parenting skills. Accordingly, we find that perpetrators of child maltreatment—almost always a parent—were less likely to abuse or neglect children even years later if their initial child victim entered foster care. We also rule out several alternative mechanisms that could,

¹They also differ from a sizable correlational literature that tends to find a negative association between foster placement and children’s outcomes (Pears and Fisher, 2005; Ryan and Testa, 2005; Pecora et al., 2006; Scherr, 2007; Trout et al., 2008; Wulczyn et al., 2009; Berzin, 2010; Zlotnick et al., 2012; Barrat and Berliner, 2013). Interestingly, however, they are consistent with recent evidence on parental incarceration in the U.S. from North Carolina (Billings, 2019) and Ohio (Norris et al., 2019), which is a somewhat analogous form of family separation.

²We refer to the adult/s with legal custody of the child before foster placement as the child’s birth parents throughout, even though in some cases the adult/s may not be their biological parent, for example, stepparents or grandparents.

in theory, drive impacts. For example, though by definition, children moved to new homes when they were removed, and prior work highlights the large impacts of geography on child outcomes (Chetty et al., 2016; Chyn, 2018), we find no evidence that placement caused lasting improvements to children’s neighborhoods or schools.

Third, this paper provides causal evidence on the impacts of child welfare interventions targeting adults, which are an understudied channel through which foster placement can impact children. Specifically, following child removal, the birth parents of foster children received community-based services, such as referrals to local drug rehabilitation groups or food pantries, as well as more intensive, targeted services like substance abuse treatment or parenting classes.³ A careful examination of mechanisms requires disentangling the role of these adult services from the dramatic changes that occurred in foster children’s own lives, yet doing so is challenging because they took place at the same time. To address this, we leverage the fact that quasi-randomly assigned investigators could also offer services to families whose children were not removed, and they differed in their propensity to do so. This allows us to identify the impact of child removal above and beyond the impact of family services alone. We find little evidence that family services without child removal impacted children’s outcomes. The impact of child removal above and beyond family services is much larger, particularly for educational outcomes. Since foster children’s gains appear only after they exited the foster system, when most were reunified with their birth parents, this analysis suggests that child removal enhanced the efficacy of child welfare interventions targeting adults.

The fourth contribution of this study is that it is the first to show that a common form of incomplete data coverage substantially biases estimates from the examiner assignment research design in practice. Specifically, this paper improves upon contemporaneous studies from Rhode Island (Bald et al., 2019) and South Carolina (Roberts, 2019) that offer quasi-experimental estimates of foster care, but do not follow children from the start of their child welfare investigation.⁴ The data in these studies contain only the subset of substantiated allegations—those in which investigators found a preponderance of evidence to support the maltreatment allegation—which represents just 40 percent of the caseload in Rhode Island and 25 percent in South Carolina (AECF, 2017). Since the same investigator who determines foster placement also makes subjective decisions around substantiation, the set of children in censored data

³In Section 5.4.1, we discuss the implications of these additional services for the exclusion restriction, and show that any bias introduced from these margins is likely to be small in our context.

⁴Bald et al. (2019) studied about 12,000 children 0–17 years old and found substantial gains for girls younger than 6 years old but imprecise null effects for other gender-age groups. Roberts (2019) examined about 17,000 children ages 2–17 and found positive impacts on on-time grade progression, yet noisy estimates on daily school attendance and test scores.

may not be balanced across investigators even if their cases were initially assigned at random. In Supplemental Appendix B.2, we replicate our primary analysis using only the sample of substantiated investigations and find estimates much smaller than the effects using the full data. As the examiner assignment design becomes increasingly common—and similar data restrictions appear in studies of crime and education—this exercise cautions against its application with incomplete data.⁵

This study is especially relevant given the dramatic changes to child welfare policy introduced in the Family First Prevention Services Act. The legislation, which took effect in 2019, makes reducing the use of foster care a federal priority by allowing states to redirect up to \$8 billion in federal funding from the foster system toward services aimed at preventing foster care entry (Wiltz, 2018). Our analysis finds that placement improved children’s outcomes, suggesting that current efforts to prevent child maltreatment in the home are falling short. To keep children safe at home without foster care, it is critical for states to identify and invest in more effective prevention services.

2 Overview of the Child Welfare System in Michigan

About one in five public school students in Michigan was the subject of a formal investigation of child abuse or neglect by 3rd grade (Ryan et al., 2018). One in 10 was the subject of more than one investigation and one in 60 experienced foster placement.⁶ In this section we review the maltreatment investigation process in Michigan and describe the state’s foster system.

2.1 Child Maltreatment Investigations

Figure 2 describes the maltreatment investigation process in Michigan, which is similar to most other states. It begins when someone calls an intake hotline to report child abuse (for example, bruises, burns, or sexual abuse) or neglect (for example, unmet medical needs, lack of supervision, or food deprivation).⁷ A hotline employee, who does not participate in

⁵Furthermore, we find that the method proposed in Arteaga (2019) to identify impacts when restricted to censored data does not resolve bias in the current context.

⁶These statistics reflect our calculations using the same sample as Ryan et al. (2018), which consists of over 700,000 third-grade students born between 2000 – 2006.

⁷The intake process is the same regardless of the reporter. Anyone can call the hotline to report suspected maltreatment, yet we do not observe the reporter in the administrative data. According to publicly available data, the most frequent reporters are people who are mandated by law to do so, such as education personnel (20.5%), legal and law enforcement personnel (18.7%), and social service workers (10.7%) (USDHHS, 2020). Reporters face little cost from calling the hotline since it is free and does not take much time. Yet they also face little cost from not reporting suspected maltreatment; mandatory reporters rarely face serious

the investigation process, transfers relevant reports to the child’s local child welfare office.⁸ The office assigns the report to a maltreatment investigator who has 24 hours to begin an investigation, 72 hours to establish face-to-face contact with the alleged child victim, and 30 days to complete the investigation.

Critical to our research design, maltreatment investigators are selected for cases according to a rotational assignment system rather than their particular skill set. Reports cycle through investigators based on who is next in the rotation.⁹ Since investigator assignment occurs within each local office—and within local geographic areas in some larger counties—all of the analyses include zip code by investigation year fixed effects, to compare children who could have been assigned the same investigator.¹⁰

Investigators make two crucial decisions that influence the intensity of child welfare’s involvement. First, they must decide whether there is enough evidence to substantiate the maltreatment allegation. Investigators interview the people involved, examine the home, and review any relevant police reports, medical records, or notes from prior maltreatment investigations. Seventy-five percent of reports in 2016 went unsubstantiated (USDHHS, 2018b, Tables 3-1 and 3-3), meaning child welfare offices did not follow up with the family further.

Second, investigators decide how much risk the child faces by continuing to live in the home. They complete a 22-question risk assessment to compute a risk score, which is used to determine whether foster placement is appropriate. Many of the items require simple yes or no answers, such as “primary caretaker able to put child’s needs ahead of own” and “primary caretaker views incident less seriously than the department.” Even with guidance on how to interpret these questions, some are inherently subjective. Moreover, Bosk (2015) offers detailed qualitative evidence that investigators often manipulate their responses to ensure risk scores that match their priors. Therefore, even with a standardized system in place,

consequences for not reporting (Hogelin, 2013). There are much larger costs to children and families from both under- and over-reporting. At worst, under-reporting can lead to child fatality. Yet over-reporting can place an unreasonable burden on families who must comply with child welfare.

⁸Reports are screened out if, for example, the perpetrator is younger than 18 years old or the victim is older than 18. We observe only screened-in reports, which does not affect the validity of the research design since investigator assignment occurs after this initial screening.

⁹Though investigators may vary slightly in completion time, even those who take somewhat longer are assigned new reports according to the rotation. In fact, despite a legal maximum caseload size of 12 instituted in 2013, two-thirds of investigators reported having a caseload of 13 or greater after 2014 (Ringle, 2018).

¹⁰There are two exceptions to the rotational assignment of investigators, which we exclude from the analysis. First, given their sensitivity, reports of sexual abuse tend to be assigned to more experienced investigators. Second, new reports involving a child for whom there was a recent prior report are usually assigned to the original investigator since he or she has familiarity with the family. Anecdotally, such reports tend to re-enter the rotation after a few months. We exclude from the analysis those within one year of a prior investigation to be conservative.

investigators yield immense discretion over foster placement.

Investigator judgment over both evidence and risk determines the outcome of the investigation. If the investigator substantiates the allegation and the risk level is low, the investigator must refer the family to community-based services like food pantries, support groups, or other local nonprofits. These cases require no further follow-up by child welfare. If the investigator substantiates the allegation and the risk level is high, the family also receives more intensive, targeted services, such as substance abuse treatment, parenting classes, or counseling. Local and state funding, and federal funding from Title IV-E, cover the costs of these targeted services. Lastly, substantiated allegations with especially high risk not only trigger targeted and community services but also require the investigator to file a court petition for child removal.¹¹ The main analysis in this study examines the combined effects of child removal and these adult interventions on children’s outcomes, yet additional analysis explores their individual contributions.

2.2 Foster Care System

Foster care is a family intervention: children are temporarily removed from their homes while their birth parents receive services to improve their parenting. Removal occurs quickly; just 10 days pass between the start of an investigation and the median placement. In Michigan and across the country, best practices recommend a strict ordering of placement settings: placement with relatives, with an unrelated family, and in group homes or institutions.¹² In many cases, though, children do not have suitable relatives available. In 2015, 41 percent of foster children in Michigan were living with an unrelated family, 35 percent lived with relatives, 9 percent lived in group homes or institutions, and 14 percent lived in other settings, such as pre-adoptive homes or supervised independent living.¹³ It is common to switch placement settings while in the foster system—60 percent of children in Michigan lived in more than one setting, and 17 percent lived in at least four. Michigan looks very similar to

¹¹Unlike investigators who no longer work with the family after completing the investigation, the same judge may interact with the family throughout the child’s stay in foster care. Since this repeated judge involvement violates the exclusion restriction, our research design leverages investigator discretion rather than judge discretion over foster placement.

¹²There is limited causal evidence on the effects of each placement type, and the instrumental variables design in this study cannot separately identify each effect. However, OLS analysis in Supplemental Appendix B.5 finds a larger positive association between kinship placement and children’s outcomes than other placement types.

¹³There are limited data available both nationwide and in Michigan on foster families (those who take in foster children). Estimates from the American Community Survey (ACS), which have known limitations, suggest that households with foster children tend to be larger and lower-income than other households with at least one member younger than 18 years old. Supplemental Appendix B.6 provides summary statistics and discusses the limitations of using ACS data to identify families with foster children.

the rest of the country along these statistics ([ChildTrends, 2017](#)).

After placement, child welfare caseworkers meet with birth parents to create a reunification plan stating the conditions under which the child can return home. Such plans might require the parent to secure housing, overcome drug addiction, or keep enough food in the home. Birth parents receive targeted services to address the challenges in their own lives, which can include substance abuse treatment, parenting classes, counseling, and job training.¹⁴ Caseworkers monitor their progress and make changes to the reunification plan as needed. Family reunification only occurs if a court decides that birth parents made sufficient changes for their child to be safe in the home.

Ultimately, children in Michigan, including those outside of the analysis sample, spent 17 months in the system on average, after which 47 percent were reunified with their birth parents, 34 percent were adopted or had legal guardianship transferred, and 9 percent exited the system as independent adults upon turning 18. The remaining 10 percent fell into less common exit categories, such as informal guardianship with relatives, incarceration, or transfer to another agency. Section 5.3.1 offers evidence on how the foster care experience of the overall population of foster children compares to children at the margin of placement.

2.3 Comparison to Other States

Child welfare systems in the U.S. vary across states; therefore, we compare the system in Michigan to other states in order to understand the generalizability of our findings. Overall, the experience of foster care in Michigan is similar to other states in terms of how long children remain in the system and the stability of their placements ([AECF, 2017](#)). Among all children in foster care in the U.S. in 2015, the average child spent 19 months in the system, ranging from just 12 months in Idaho to 35 months in Illinois. At 17 months, Michigan, along with eight other states, ranked 18th in this measure. Similarly, 35 percent of foster children in the U.S. lived in at least three different foster homes, ranging from 24 percent in Wyoming to 54 percent in Illinois. Michigan ranked 17th in this measure, at 31 percent. Furthermore, among all children in the US who exited the system in 2015, 51 percent reunified with their birth parents, compared to 47 percent of children in Michigan. Taken together, there is little reason to suspect that the findings from this study would not generalize to other child welfare systems across the country.

¹⁴Though limited local supply or high adult demand may constrain access to these services—for example, there may be a shortage of providers or long waitlists for care—caseworkers do their best to meet the needs of their families and sometimes have priority access.

3 Data Sources and Sample Construction

3.1 Administrative Data Sources

This study uses administrative data from the Michigan Department of Health and Human Services (MDHHS), Michigan Department of Education (MDE), Center for Educational Performance and Information (CEPI), and Michigan State Court Administrative Office (SCAO) to test the effects of foster placement on children’s outcomes. Since there is no common identifier, we linked these files using a probabilistic matching algorithm based on first name, last name, date of birth, and gender. Overall, 84 percent of child welfare investigations of school-age children matched to a student enrolled in a Michigan public school in the year of their investigation. This match rate is quite high given that many investigated children should not have matched to an enrolled public school student (for example, private or homeschooled students, high school dropouts, and those who were not permanent Michigan residents). Specifically, we estimate that if there were a common identifier, just 87.1 percent of investigated children would have matched to a currently enrolled student.¹⁵ Supplemental Appendix B.4 describes the match process and match rate in greater detail.

Child welfare data from MDHHS consist of the universe of maltreatment investigations in Michigan between August 1996 and July 2017. They include details of each investigation, such as the allegation report date, allegation types as coded by the investigator, the child’s zip code, and substantiation. Importantly, the administrative data link investigations to placement records, which allows us to directly observe whether a child was removed following a specific investigation. We define treatment (foster placement) throughout the paper as removal due to a child welfare investigation.¹⁶ Conditional on placement, the data also contain limited information on placement settings and permanency outcome (reunified with birth parents, adopted, and so on). Critical to our analysis, the files also include unique investigator identifiers beginning in 2008. Crucially, unlike two recent studies that offer quasi-experimental estimates of foster placement, this dataset includes both substantiated and unsubstantiated cases (Roberts, 2019; Bald et al., 2019). Supplemental Appendix B.2 describes how incomplete data coverage can substantially bias estimates from the examiner

¹⁵We estimate that the remaining 12.9 percent of investigated children consist of private school students in Michigan (4.6 percent), non-Michigan residents (3.4 percent), homeschooled students in Michigan (2.6 percent), and students who dropped out of high school in Michigan (2.1 percent).

¹⁶Investigators are required by law to complete investigations within 30 days of a maltreatment report. Though it is possible for investigations to take slightly longer, the process moves much faster for cases that result in foster placement; the median amount of time between the start of an investigation and eventual placement is only 10 days.

assignment research design.

Education data from MDE and CEPI cover the universe of public school students in Michigan, including charter school students, between the 2002–2003 and 2016–2017 school years. These records include demographic information such as race/ethnicity, gender, and free or reduced-price lunch eligibility, as well as indicators of academic progress like daily attendance rate and standardized test scores. They also include the census block where a student lived during the school year, which we link to publicly available census block group characteristics from the U.S. Census Bureau.

Juvenile justice data from SCAO include all juvenile court petitions filed in almost every county in Michigan between 2008 and 2015. A court petition is an official document filed following juvenile arrest in cases where youth are not immediately diverted from the courts. Petitions can be dismissed by the court after filing and need not indicate that there was ever a formal court hearing. The SCAO data cover 75 of Michigan’s 83 counties, including Detroit and the metro-Detroit area but excluding the following five urban and three rural counties: Kent, Washtenaw, Ingham, Ottawa, Kalamazoo, Berrien, Delta and Keweenaw.¹⁷ We exclude the 19 percent of investigated children who lived in these eight counties from our analysis of juvenile delinquency; the conclusions on other outcomes are similar when these children are excluded.¹⁸

Using these administrative data sources, we construct an unbalanced panel at the investigation by school year level, and restructure non-educational outcomes to follow the school year calendar. For example, we define maltreatment reports and juvenile petitions occurring between September 2010 and August 2011 as the 2010-2011 school year. Children age out of the panel for certain outcomes; for example, the age at which young people are tried in the adult court system is 17 years old in Michigan, so 17-year-olds are ineligible for the juvenile delinquency outcome.

3.2 Child Safety, Academic, and Crime Outcomes

We assess the effects of foster care on child well-being across three dimensions: safety, schooling, and crime. Given that we study a variety of outcomes, multiple inference issues can be important. To address this, we construct a summary index of child well-being

¹⁷These counties include 3 of the state’s 10 most populated cities: Grand Rapids, Lansing, and Ann Arbor, and 3 more of the top 30: Kalamazoo, Wyoming, and Ypsilanti.

¹⁸Michigan’s juvenile arrest rate is quite representative of the average U.S. state. For instance, the U.S. Department of Justice’s Office of Justice Programs reports that—of the 48 contiguous states—in 2017 Michigan ranked 20th in the lowest number of aggravated assault arrests of persons under age 18 for every 100,000 persons aged 10-17. Illinois, the setting of [Doyle \(2007\)](#), ranked 18th. It is important to note, however, that our juvenile delinquency measure differs from [Doyle \(2007\)](#). We elaborate on this point in [Table 7](#).

so that the probability of a Type I error does not increase as additional outcomes are added. Furthermore, combining multiple outcomes into a single summary index reduces measurement error by averaging across outcomes (Deming, 2009). The index consists of six primary outcomes, described in detail below: two measures of child safety, three academic outcomes, and one indicator of juvenile delinquency.

We follow Kling et al. (2007) and Deming (2009) and normalize each of the outcomes to have a mean of zero and a standard deviation of one. We additionally reverse-code “bad” outcomes (juvenile delinquency and child safety indicators) so that positive values of the index represent “good” outcomes, and impute any outcomes with missing values as the average of the remaining non-missing standardized items in the index.¹⁹ Finally, we create a summary index variable that is the weighted average of all six outcomes—where the average is weighted by the inverse of the sample variance-covariance matrix to account for dependence across outcomes, as in O’Brien (1984).

To measure child safety, we create indicators for whether children were the alleged victim in a subsequent maltreatment investigation and whether they were a confirmed (substantiated) victim in a subsequent investigation. Second, we examine schooling by studying daily attendance rates and standardized math and reading test scores. Daily attendance rates are the fraction of days that a student showed up to school during the school year. Standardized test scores are normalized to have mean zero and standard deviation one within year-grade-subject cells across the full population of public school students.²⁰ Finally, we measure juvenile delinquency as the filing of a juvenile court petition.

3.3 Overview of Analysis Sample

The analysis sample consists of public school students who were the alleged victim in a maltreatment investigation between 2008 and 2016. We exclude cases where investigators were unlikely to have been quasi-randomly assigned—allegations of sexual abuse and those involving children from a recent prior report. We also restrict the sample to children enrolled in grades 1 through 11 in the school year of their investigation to observe baseline characteristics and at least one follow-up year.²¹ Appendix B.4 describes the sample restrictions

¹⁹The intuition for imputing missing values this way is that our best guess of the value of a missing outcome is the average of the remaining standardized outcomes in the index.

²⁰These educational outcomes are included in the analysis only if they occur after a child’s investigation. That is, we exclude scores from students investigated in the middle of the state testing cycle from the outcome analysis since the exact dates of test administration for a given school-grade-subject are not publicly available.

²¹The analysis sample excludes children who were too young to have entered school at the time of their investigation. Though these younger children appear in the child welfare data and, years later, may appear in public school records, we find that foster placement caused a large and statistically significant reduction in the likelihood that they ever enrolled in a Michigan public school. A likely explanation for this finding is that

in greater detail. Overall, we focus on 242,233 investigations of 186,250 students and follow students for at most nine years after their investigation.

Table 1 describes the sample. Column 1 consists of all public school students in Michigan during the 2016–2017 school year, while Column 2 consists of the investigations of children in the analysis sample. Black and low-income children were disproportionately involved in the child welfare system; 29 percent of investigations were of Black children and 83 percent were of low-income children, despite their making up just 21 percent and 49 percent of the population, respectively. Children with child welfare involvement had noticeably lower baseline daily attendance rates and scored about a quarter of a standard deviation worse on standardized math and reading tests. Column 3 describes children involved in the 2% of investigations that resulted in foster placement.²²

Relative to the overall sample in Column 2, foster children were also disproportionately Black and low-income, had much lower daily attendance rates, and scored about one-tenth of a standard deviation lower on math and reading tests. Overall, these descriptive statistics caution against a causal interpretation to mean comparisons between investigated children who were and were not removed.

4 Empirical Strategy

A naive analysis of foster care might regress children’s outcomes, such as daily school attendance rates or standardized test scores, on a binary treatment variable equal to one if the child’s investigation resulted in foster placement. Even with controls for a wide range of observable characteristics, estimates from such a regression would likely be biased because foster children differ along unobservable dimensions from those who were not removed. For example, they may have lived in more difficult home environments or been more severely maltreated. Such unobserved features would bias OLS estimates to understate the benefits of foster care and overstate the costs.

about one-third of foster children were adopted upon exiting the foster system and may have legally changed their last name prior to enrolling in school, meaning that the administrative child welfare and education records were unlikely to match. It is also possible, however, that young children differentially moved out of state or enrolled in private schools. Importantly, we find no evidence of differential attrition out of Michigan public schools for currently enrolled students (Table B1).

²²The 2 percent placement rate in the analysis sample is substantially lower than the one documented in Ryan et al. (2018) for two main reasons. First, our analysis sample consists of school-age children—those enrolled in at least 1st grade at the time of the investigation. Conditional on being the subject of an investigation, this older group is considerably less likely than younger children to be placed in foster care. Second, the analysis sample excludes sexual abuse cases, which are disproportionately likely to result in foster placement relative to other types of child abuse and neglect.

4.1 Research Design

In order to overcome omitted variable bias, we use the examiner assignment research design, which has been applied to other studies of foster care (Doyle, 2007, 2008) as well as research on incarceration (Kling, 2006; Aizer and Doyle, 2015; Mueller-Smith, 2015), disability insurance (Dahl et al., 2014), and evictions (Collinson and Reed, 2019; Humphries et al., 2019), among others. Specifically, we instrument for placement using the removal tendencies of quasi-randomly assigned investigators. Children assigned by chance to especially strict investigators—those with high propensities to remove—were more likely to enter foster care than they would have been if they happened to be assigned to a more lenient investigator.

In order to extract signal from noise in a measure of removal tendency, we restrict the analysis to children assigned to investigators who worked at least 50 cases, inclusive of quasi-randomly assigned cases outside of the analysis sample.²³ This restriction leaves 3,073 investigators assigned to 315 cases each, on average. Following the literature, we calculate the instrument as the fraction of all other investigations, both past and future, assigned to the same investigator that resulted in foster placement. Specifically, for investigation i assigned to investigator w :

$$Z_{iw}^R = \left(\frac{1}{n_w - 1}\right) \sum_{k \neq i}^{n_w - 1} (FC_{kw}) \quad (1)$$

where n_w equals the total number of cases assigned to investigator w and FC_{kw} is an indicator equal to one if investigation k resulted in foster care.²⁴ This instrument is equivalent to the investigator fixed effect from a leave-out regression where foster placement is the dependent variable.

The instrument has a mean of 0.030 and a standard deviation of 0.024, indicating considerable variation in investigator tendencies. Crucial to the research design, there is variation even among investigators who worked in the same local office. Figure 3 shows the distribution of the instrument net of child zip code by investigation year effects; an investigator at the 10th percentile removed at a rate 2.1 percentage points less than the average investigator in their local area whereas someone at the 90th percentile removed at a rate 2.4 percentage points greater. Relative to the average removal rate of 3 percent, this represents a 150 percent increase in the likelihood of foster placement.

We use the following instrumental variables specification to measure the causal effects of

²³Table A6 shows that the results are robust to larger thresholds.

²⁴There are other reasonable ways to measure removal stringency. For example, this approach does not allow for investigator tendencies to change over time. Section 5.4.2 describes several alternatives and shows that the results are robust across measures.

foster care:

$$FC_{iw} = \gamma_1 Z_{iw}^R + \gamma_2 X_{iw} + \Theta_r + \eta_{iw} \quad (2)$$

$$Y_{iw} = \beta_1 F\hat{C}_{iw} + \beta_2 X_{iw} + \theta_r + \epsilon_{iw} \quad (3)$$

where Y_{iw} is a child outcome, such as daily school attendance rate or score on a standardized math test, and X_{iw} is a vector of baseline covariates that includes a variety of socio-demographic and academic characteristics.²⁵ Θ_r and θ_r represent child zip code by investigation year fixed effects to control for the level of investigator rotational assignment, restricting the comparison to children who could have been assigned to the same investigator.²⁶ There are 7,534 unique rotation groups, consisting of 13 investigators on average. Finally, we cluster standard errors at the child level to account for the correlation in outcomes that arises mechanically by including the same child more than once in the panel.²⁷

β_1 is the local average treatment effect (LATE) of foster placement where compliers are children for whom investigators might disagree about removal. Given likely heterogeneous treatment effects, this study cannot speak to how foster care influences always takers—children so clearly in danger at home that all investigators would remove—and never takers—those so clearly safe that no investigators would remove them. As Dr. Jill Duerr Berrick points out, the debate around foster placement is “not in the cases that are black and white, but in the cases that occupy the center, gray area of child welfare” (Berrick, 2018). Thus, compliers

²⁵Specifically, it includes controls for socio-demographic features including gender, grade-level fixed effects, race/ethnicity, and free or reduced-price lunch receipt. It also controls for baseline academic characteristics measured in the year before the investigation including attendance rate and receipt of special education supports, as well as an indicator for ever expelled. It flexibly controls for a student’s most recent baseline standardized math and reading test scores by including linear, quadratic, and cubic terms, as well as the interaction of baseline math and reading performance. Furthermore, it contains information about the maltreatment report, such as whether the allegation was for physical abuse or neglect, the child’s relation to the perpetrator, and the number of investigations in which the child was previously the subject. It additionally controls for characteristics of the school that the child attended during the year prior to the investigation, including indicators for whether the child was enrolled in a charter or an urban school, the fraction of White, Black and Hispanic students, and the fraction who were eligible for free or reduced-price lunch. It also controls for characteristics of the child’s neighborhood in the year prior to the investigation, as defined by census block group, including median household income, employment rate, the fraction of adults with at least a bachelor’s degree, the fraction of White, Black, and Hispanic residents, an indicator for whether the child experienced homelessness, and the number of times the child moved neighborhoods. Lastly, it includes indicators for any missing values in the following covariates: female, free or reduced-price lunch receipt, and each of the prior schooling characteristics.

²⁶Child welfare staff from several local offices explained that some investigators only work in the northern part of the county, whereas others only work in the south, for example. However, such geographical boundaries are neither publicly available nor observed in administrative data. Importantly, Table A6 shows that the results are robust to instead defining rotational assignment at the child county by investigation year level.

²⁷There are other reasonable levels to cluster standard errors in our context, such as by investigator, by rotation group, or by child and rotation group. Section 5.4.5 details several alternative approaches and reports that the results are robust to the level of clustering.

represent a population that is especially relevant for child welfare policy.

4.2 Identifying Assumptions

Our estimates require several assumptions to be interpreted as local average treatment effects:

1. Relevance: $\gamma_1 \neq 0$. The instrument must predict foster placement. Figure 3 shows a histogram of the instrument, which varies in value from -0.04 to 0.06 after partialling out the child zip code by investigation year fixed effects. Superimposed over the histogram is the non-parametric regression of foster placement on investigator removal stringency. The relationship between the instrument and foster placement is highly positive. Table 2 shows the first-stage regression of foster placement on the removal stringency instrument. The correlation between the instrument and foster care is 0.48 (Column 1) and a one standard deviation (2.4 percentage points) increase in removal stringency increases the likelihood of placement by about one percentage point (Column 4). The F-statistic of 582 indicates that there is not a weak instruments problem.

2. Exogeneity: $\text{Cov}[Z^R, \epsilon] = 0$. The rotational assignment of child welfare investigators to cases suggests that unobserved determinants of children’s outcomes should be independent of investigator removal stringency. As is standard, we test an implication of exogeneity—that observable child and case characteristics are uncorrelated with the removal tendencies of the assigned investigator—by checking for balance along a range of observable child characteristics. Table 3 shows that a rich set of characteristics are not jointly predictive of the instrument despite being highly predictive of placement itself. As further evidence, the first stage F-statistic in Table 2 is stable with the inclusion of covariates.

3. Monotonicity: $\mathbb{E}[FC|X, \theta_r, Z^R = j] \geq \mathbb{E}[FC|X, \theta_r, Z^R = k]$ or $\mathbb{E}[FC|X, \theta_r, Z^R = j] \leq \mathbb{E}[FC|X, \theta_r, Z^R = k] \forall j, k$. Children who were removed by a particularly lenient investigator would also have been removed by a stricter one and vice versa. Recent advances note, however, that such pairwise monotonicity is neither realistic in most contexts nor necessary to estimate local average treatment effects (Norris, 2019; Frandsen et al., 2019). Instead, identifying the LATE requires the weaker assumption of average monotonicity, which states that for each child, the covariance between their investigator-specific removal treatment status and investigator stringency is weakly positive.²⁸

²⁸Specifically, while the pairwise monotonicity assumption ensures that the instrumental variables estimator aggregates treatment effects across complier groups using Imbens and Angrist (1994) weights, if the weaker assumption of average monotonicity holds then our estimates will still be a proper weighted average of treatment effects with the weights for each individual equal to the scaled covariance between foster care placement and investigator’s removal tendency (Norris et al., 2019; Frandsen et al., 2019).

It follows from average monotonicity that removal stringency and foster placement should be positively correlated for all child subgroups. There are two complementary ways to probe this implication. First, the first stage should be non-negative for all subgroups (Dobbie et al., 2018), which holds for gender, race/ethnicity, age, and prior child welfare involvement groups in our setting (Table B2, Panel A).²⁹ Second, investigators who were strict for certain groups should also have been strict for others (Bhuller et al., 2018). For example, amidst serious concerns of racism in maltreatment investigations (Clifford and Silver-Greenberg, 2017), monotonicity asserts that investigators who were particularly likely to remove children of color should also have been weakly stricter than their colleagues in their investigations of White children. In support of the assumption, the first stage remains positive and statistically significant when we re-calculate the instrument as a leave-subgroup-out measure (Table B2, Panel B).

Our analysis also requires an exclusion restriction in order for the estimates to be interpreted as local average treatment effects. We discuss exclusion in detail in Section 5.4.1.

5 Causal Effects of Foster Care on Children’s Outcomes

Table 4 shows the OLS (Panel A) and 2SLS (Panel B) effects of foster care on several critical indicators of child well-being covering the areas of safety, education, and crime. Standard errors clustered at the child level are shown in parentheses. Control means for OLS and control complier means for 2SLS are reported in curly brackets.

The OLS results suggest that removal had a precise but near-zero impact on the index of child well-being. This may be surprising, particularly in light of the sizable correlational literature that tends to find a negative association between foster placement and children’s outcomes. However, the OLS results shown in Table 4 include a rich set of lagged covariates and fixed effects, and recent work by Berger et al. (2014) shows that controlling for lagged outcomes substantially reduces the negative association. We replicate the negative relationship using simple bivariate regressions in Table B3.

The 2SLS estimate reveals that removal improved the index of child well-being by 39.2 percent of a standard deviation, an effect statistically significant at the 5 percent level. Two expected findings stand out from comparing the OLS and 2SLS results on the index of child well-being. First, the OLS estimate is substantially smaller than the 2SLS estimate, suggesting that unobserved features, like the severity of maltreatment, for example, lead

²⁹We do not create groups based on the type of maltreatment such as abuse or neglect because investigators code this information after they begin their investigation. However, to the extent that different types of maltreatment are related to observable child subgroups, the exercise offers an indirect test for non-monotonic tendencies based on these features.

OLS to understate the benefits of removal. Second, the control mean, the mean outcome among all investigated children who were not removed, is larger than the control complier mean, the estimated outcome for compliers who were not removed—indicating that children at-risk of placement were worse off by remaining in the home than the average investigated child. The index provides a useful summary, but in order to understand what drives the improvement, as well as to more easily interpret the magnitudes, we turn next to the effects on each of the six components.

5.1 Effects on Child Safety, Academics, and Crime

Table 4 shows that foster children were safer than they would have been had they remained at home, indicating that the foster system achieved its primary objective. The 2SLS estimates show that removal reduced the likelihood of being an alleged victim of maltreatment in a subsequent investigation by 13.2 percentage points, a 52 percent reduction relative to a complier mean of 25.5 percent. Similarly, it reduced the likelihood of being a confirmed victim of maltreatment by 5.3 percentage points, a 56 percent reduction.

Although in theory these effects may represent a reduction in reporting behavior without a change in underlying safety, the data do not support this interpretation. For example, suppose that teachers were less likely to report minor bruises to child welfare if they knew that the bruised student was, or had been, in foster care. We would still expect them to report especially severe abuse against foster children, since teachers and other mandated reporters are required by law to report suspected maltreatment. Therefore, if placement only reduced reporting, then the reported abuse against foster children should be more serious than the reported incidents against children who were not removed. However, we find no clear evidence that foster placement influenced the likelihood of substantiation among children with a subsequent investigation.³⁰ Moreover, caseworkers, who are also mandatory reporters, visited foster children regularly, both during their time in the system and after they exited, suggesting that actual maltreatment against foster children would have been reported (USDHHS, 2016c).

Consistent with an improvement in child safety, we find large gains in academic outcomes. Removal increased daily school attendance rates by 5.5 percentage points; for the 180-day school year, this is equivalent to showing up for 10 additional days of school. Furthermore, removal had a very large positive effect on standardized math test scores, equal to 36 percent

³⁰Specifically, we estimate Equations 2 and 3 using an indicator for substantiation as an outcome and limit the analysis to students with a subsequent investigation. We obtain a point estimate on foster placement of -0.14 and a standard error of 0.12 .

of a standard deviation.³¹ This estimate is statistically significant at the 10 percent level, yet we can rule out decreases greater than 4 percent of a standard deviation. Although the point estimate on standardized reading test scores is positive and substantively large, about half the size of the effect on math, it is not statistically significant. This is not particularly surprising, because reading skills are considered less malleable than math at older ages.³² Lastly, we examine the effect of removal on juvenile delinquency, defined by the filing of a juvenile court petition. The point estimate suggests a large decrease in juvenile crime—a 55 percent drop relative to a control complier mean of 5.1 percent—but the estimate is imprecise.

Given that prior work in Doyle (2007, 2008), as well as decades of correlational studies, find that placement harmed children’s outcomes, we also examine whether placement worsened children’s outcomes in our setting using a one-sided hypothesis test where the null hypothesis is that foster care worsened outcomes. We can statistically reject that placement reduced the index of child well-being at the 1 percent level. We also find that foster care did not increase children’s likelihood of being confirmed as victims of maltreatment and can rule out that placement reduced student attendance and math test scores. Overall, the results across dimensions of safety, academics, and criminality consistently suggest that foster care improved children’s outcomes.

5.2 Mechanisms

We examine the channels through which placement improved child outcomes through two complementary exercises. First, we explore the impacts of placement over time, which help understand the mechanisms at work because foster care is a temporary intervention. We focus on four key dynamics: whom children lived with, where they lived, where they went to school, and how their outcomes evolved. Second, we evaluate interventions targeted at birth parents as a potential channel through which placement influences children.

5.2.1 Evidence from the Timing of Impacts

Forty percent of children who were removed had exited the foster system after one year and nearly all had exited after two years (Figure 4).³³ We create an index of neighborhood

³¹As a benchmark, Goodman (2014) estimates that each additional student absence reduces math achievement by 0.05 standard deviations, suggesting that the estimated math score effect is roughly in line with the increase in daily school attendance.

³²Removal did not influence the likelihood of taking standardized tests (Table B4). We also find imprecise impacts on high school graduation and college enrollment; because the sample of students old enough to be eligible for these outcomes is small, we cannot rule out large positive or negative effects (Table B5).

³³They spent 19 months in foster care, on average (Table B6).

and school characteristics according to [Kling et al. \(2007\)](#) in order to explore the effects of placement on childhood environment. The index consists of three neighborhood components: median household income, the fraction of adults with a bachelor’s degree, and employment rate. It also includes two school components: average math and reading test scores and the share of students eligible for free or reduced-price lunch. There was a large and statistically significant increase in the index during the first year after placement (Table 5, Panel A).³⁴ Given that moving to areas with lower-poverty levels can improve child well-being ([Chetty et al., 2016](#); [Kawano et al., 2017](#); [Chyn, 2018](#)), such exposure might lead to contemporaneous gains in children’s outcomes. However, there were no discernible differences in year one outcomes between children placed and not placed in foster care (Table 5, Panel A).³⁵ That foster children were no more or less likely to be abused or neglected in the first year may be especially surprising since maltreatment in foster homes is extremely rare.³⁶ It is possible, however, that the threat of child removal reduced the maltreatment of children who were not removed in the short run.

Nearly all (85 percent) marginal foster children had exited the system after two years and reunified with their birth parents.³⁷ Upon exiting, foster children returned to neighborhoods and schools similar to those of untreated compliers; we do not detect differences in the characteristics of their neighborhoods or schools after the first year (Table 5, Panel B). Despite this, gains in safety and academic outcomes emerged several years after removal. Specifically, the index of child well-being increased by 45 percent of a standard deviation across all years after the first, driven by gains in safety, daily school attendance rates, and standardized math test scores (Table 5, Panel B). Figure 5 shows the effects separately by year, revealing steady improvements in most outcomes that persist for several years. For example, the likelihood of being the victim of maltreatment only began to decrease after four years and continued to decrease every year for three more.³⁸

A likely explanation for this surprising pattern is that children returned to safer and more nurturing homes after exiting the system. Given that most children were reunified with their birth parents, this can largely be interpreted as parental improvement. There are

³⁴This was driven by exposure to more highly educated neighborhoods and higher-income classmates (Table B7).

³⁵For ease of exposition, Table 5 and most further analyses report only a set of the outcomes that were statistically significant at the 10 percent level from Table 4. Additional results are available upon request.

³⁶Though there continue to be far too many tragic stories of abuse and neglect in foster care—and any amount of maltreatment within the system is too much—according to the most recent national data, 0.71% of foster children in Michigan were maltreated while in the foster system ([USDHHS, 2018c](#)).

³⁷Table A1 shows that of the remaining 15 percent who exited, 8 percent were adopted, 5 percent had guardianship transferred, and 2 percent turned eighteen years old and legally exited foster care as adults.

³⁸These estimates represent time-since-treatment effects rather than age-of-treatment effects because all specifications include fixed effects for student grade level at the time of the investigation.

several institutional features that support this channel. For example, after their children were removed, birth parents worked closely with social workers to address challenges in their own lives, such as confronting drug addiction, finding stable employment, securing housing, or strengthening parenting skills. Second, birth parents received fully funded services to help with these challenges, such as substance abuse treatment, parenting classes, or counseling. Moreover, a judge needs to approve that it is safe for children to return home before they can be reunified with their birth parents. Although we do not observe the types of services that birth parents receive in our context, [Grimon \(2020\)](#) finds that child removal led to the increased use of mental health and substance abuse services among birth mothers in Allegheny County, Pennsylvania, in the first follow-up year. [Grimon \(2020\)](#) also finds that the impact on substance abuse services persists even three to six years later, which is consistent with the pattern of parental improvement in our setting. Furthermore, in our context, we find statistical evidence of birth parent improvement. Perpetrators of child maltreatment, almost always a birth parent, were less likely to abuse or neglect children even years later if their initial child victim entered foster care (Figure 5e).

Though we can not definitively rule them out, we find little evidence for two alternative explanations of the pattern of impacts. First, it is possible that moving to areas with lower-poverty levels during placement improved child outcomes. However, credibly identified studies of mobility find that such effects increase with duration ([Chetty et al., 2016](#); [Chyn, 2018](#)), whereas exposure in our context was only temporary. Furthermore, [Sanbonmatsu et al. \(2006\)](#) and [Jacob \(2004\)](#) find that the long-run benefits of moving do not run through schooling channels, yet foster care had large impacts on educational outcomes.

Second, it could be that children’s experiences while in foster care benefited them only years later; that is, foster care could trigger additional supports whose benefits take time to manifest. In particular, if the costs related to family separation are high in the short run yet fade over time, even the benefits from channels that have more immediate impacts may appear only years later. However, we find no evidence that foster care increased supports in school either during placement or after exiting, as proxied by receipt of special education services (Table 5, Column 6). Foster children were also no more or less likely to be retained in grade (Table 5, Column 7). Moreover, although children may have benefited from placement in other ways, perhaps through access to better counseling, new role models, or more nutritious meals, credible estimates of these channels for school-age children consistently find effects on standardized test scores of less than one-tenth of a standard deviation, much smaller than the 0.36 standard deviation increase in math test scores found in this study.³⁹

³⁹See, for example, [Carrell and Hoekstra \(2014\)](#) and [Mulhern \(2019\)](#) for the effects of school counselors, [Dee \(2004\)](#) for the effects of teacher role models, [Heller \(2014\)](#) for the effects of summer jobs and mentors,

Therefore, evidence from the timing of impacts suggests that positive changes made by birth parents were a key channel through which foster placement improved children’s safety and schooling. This finding begs the question of how child removal influenced birth parents, which we describe in detail in the next section.

5.2.2 Evidence from Adult Interventions

Following child removal, birth parents received two broad types of services: light-touch, community-based services, such as referrals to food pantries and local drug rehabilitation groups, and intensive, targeted services, like funded substance abuse treatment, parenting classes, and employment programs. What were the roles of these services in explaining the large, positive effects of foster care on children? It is challenging to disentangle this channel from the dramatic changes occurring in children’s own lives because both happen at the same time. However, a useful comparison group exists because quasi-randomly assigned investigators could offer these services to parents even if their children were not removed.

As shown in Figure 2, investigators placed families on one of four tracks based on the strength of evidence that maltreatment occurred and the child’s risk of future harm: (1) no services, (2) community-based services, (3) community-based and targeted services, and (4) child removal plus community-based and targeted services.⁴⁰ As such, we create two new instruments according to Equation 1: investigator propensity to recommend community-based services alone (Z^C) and investigator propensity to recommend both community-based and targeted services without child removal (Z^{TC}). Together with the main removal stringency measure (denoted here by Z^{RTC}), we use these new measures to simultaneously instrument for tracks two, three, and four according to the following three first-stage and one second-stage equations:

$$RTC_{iw} = \gamma_1 Z_{iw}^{RTC} + \gamma_2 Z_{iw}^{TC} + \gamma_3 Z_{iw}^C + \gamma_4 X_{iw} + \kappa_r + \mu_{iw} \quad (4)$$

$$TC_{iw} = \alpha_1 Z_{iw}^{RTC} + \alpha_2 Z_{iw}^{TC} + \alpha_3 Z_{iw}^C + \alpha_4 X_{iw} + \chi_r + \nu_{iw} \quad (5)$$

$$C_{iw} = \delta_1 Z_{iw}^{RTC} + \delta_2 Z_{iw}^{TC} + \delta_3 Z_{iw}^C + \delta_4 X_{iw} + \pi_r + \zeta_{iw} \quad (6)$$

$$Y_{iw} = \beta_1 \hat{RTC}_{iw} + \beta_2 \hat{TC}_{iw} + \beta_3 \hat{C}_{iw} + \beta_4 X_{iw} + \Pi_r + \xi_{iw} \quad (7)$$

Anderson et al. (2018) for the effects of healthier meals, and Figlio and Winicki (2005), Leos-Urbel et al. (2013), Imberman and Kugler (2014), Frisvold (2015), and Schwartz and Rothbart (2017) for the effects of greater access to food.

⁴⁰Although we observe track assignment in the child welfare records, we do not observe the specific types of services received—for example, substance abuse treatment or parenting classes.

where RTC_{iw} is a binary variable equal to one if the child was removed.⁴¹ Similarly, TC_{iw} is a binary indicator equal to one if the family was referred to both targeted and community-based services and C_{iw} equals one if the family was only referred to community-based services. Since the families of children who were removed also received services, by construction, RTC_{iw} can only equal one whenever TC_{iw} and C_{iw} equal one. Therefore, β_1 in Equation 7 represents the additional impact of child removal relative to both targeted and community-based services without removal.⁴²

The first-stage relationships in Equations 4 through 6 are strong, with F-statistics above 200 (Table B8), and balance tests indicate that each of the three instruments is unrelated to a rich set of baseline child characteristics (Table B9). Table 6 shows estimates of β_1 , β_2 , and β_3 . We report impacts separately for the first follow-up year and for later years. Consistent with the analysis in Section 5.2.1, we find no additional impact of child removal on children’s outcomes relative to adult interventions without placement in the year following the investigation. In later years, however, the additional impact of child removal on the index of child well-being is nearly three times larger than the effect of receiving targeted and community services alone—which is mostly small and statistically insignificant. We also find that removal has large positive effects on daily school attendance rates and math test scores above and beyond the impacts of adult interventions alone. Though limited somewhat by statistical power in this more demanding specification, these results offer suggestive evidence that child removal was a crucial component of the foster care intervention.⁴³

There are at least two explanations as to why adult interventions may have been less effective while children remained in the home. First, they might have lacked sufficient intensity. Unlike services offered to the birth parents of foster children, which lasted up to 24 months and were monitored by both the child welfare office and the courts, services lasted at most 12 months and participation was tracked only by the local child welfare office when the child was not removed. A second potential explanation is that child removal was critical to ensure adult compliance. For example, temporary relief from parenting may have provided birth parents with the time and space needed to overcome challenges in their own lives. Child removal might also have increased adult incentives to engage with these

⁴¹ RTC_{iw} is equivalent to FC_{iw} from Equation 2 and Z_{iw}^{RTC} is equivalent to Z_{iw}^R . We use different notation here to facilitate the interpretation of this alternative specification.

⁴²The three instruments are positively, but not perfectly, correlated with each other, indicating that there is independent identifying variation from each. Conditional on zip code by year fixed effects, $\text{Corr}(Z^{RTC}, Z^C)=0.14$, $\text{Corr}(Z^{RTC}, Z^{TC})=0.24$, and $\text{Corr}(Z^{TC}, Z^C)=0.60$.

⁴³Of note, however, the three estimates are not necessarily comparable because they are for different complier groups. For example, the LATE for community-based services identifies effects for families at the margin of receiving any services, which is a much lower-risk group than families at the margin of child removal. However, to the extent that these services have similar impacts for struggling families, the results indicate that child removal was necessary for adult interventions to be effective.

programs.⁴⁴

Overall, this section highlights the importance of framing foster placement as a family intervention. Evidence from the timing of impacts as well as from an analysis of adult services suggests that improvements made by birth parents were an important mechanism to explain gains in children’s outcomes.

5.3 Compliers Analysis and Subgroup Effects

5.3.1 Contextualizing Children at the Margin of Foster Placement

The estimates in this study represent effects for children at the margin of placement, those cases in which investigators might disagree over whether foster care is appropriate. In order to better understand how these children compare to the overall population of foster children, we report complier characteristics in Table A2, estimated according to the methodology in Dahl et al. (2014). We find that 5 percent of investigated children in the sample were compliers. Compliers were younger than the average foster child—61 percent were 10 years old or younger at the start of their investigation, relative to just 51 percent of foster children overall—but otherwise looked similar in terms of demographic and baseline academic characteristics.

To contextualize the positive effects of placement, it is also useful to explore how marginal children experienced the foster system. Table A3 compares the experiences of these children, calculated using the instrumental variables design, to the overall population of foster children, defined as the mean among all foster children in the sample. Although these groups were initially placed in similar types of homes—for example, 57.2 percent of children at the margin were initially placed with relatives compared to 58.2 percent of all placements—their experiences varied in terms of placement stability, length of time in foster care, and permanency outcomes. Children at the margin had more stable placements: 51 percent experienced just one or two different placements, compared to 44 percent of all foster children in the sample. They also spent about 38 fewer days in the system and were more likely to be reunified with their birth parents. Therefore, marginal children had more stable placements, quicker exits, and higher reunification rates than the overall population of foster children, all of which are important objectives for foster care systems.

⁴⁴Prior work highlights that services like drug rehabilitation or job training programs often have high failure rates overall (SAMHSA, 2009; Barnow and Smith, 2015), but birth parents of foster children may have put in more effort than the average participant.

5.3.2 Heterogeneity by Child Age and Gender

Previous work highlights disparities in how children respond to environmental changes by age, finding that young children benefit from moving to lower-poverty areas more than older youth (Chetty et al., 2016; Chyn, 2018).⁴⁵ We find similar effects for foster placement. Table A4 shows that foster care improved the index of child well-being for young children—those ages 10 and younger at the beginning of the investigation—by 66.6 percent of a standard deviation, an effect statistically significant at the 1 percent level. This was substantially larger than the point estimate for older youth, which was not statistically significant. These estimates are statistically different from each other at the 1 percent level.

In addition, previous work shows that males are often more vulnerable than females to childhood disadvantage or disruption (Kling et al., 2005; Bertrand and Pan, 2013; Autor et al., 2019). However, Table A4 shows that the impacts of placement were similarly positive for both groups. Further exploring this heterogeneity using age by gender groups reveals that the effects were qualitatively much larger for young male children than for young female children, but qualitatively larger for older females than for older males. Taken together, this exercise indicates that the intersection between age and gender may play an important role in determining the impacts of placement.⁴⁶

5.4 Threats to the Research Design and Robustness Checks

5.4.1 Multi-dimensionality of treatment

The exclusion restriction requires that the removal stringency instrument can only influence outcomes through foster placement. Though the exclusion restriction is inherently not testable, both institutional and empirical evidence lend credence to this assumption in a number of ways. First, a potential concern is that investigators might have influenced children’s experiences in the foster system, conditional on placement. However, investigators did not work with children after the investigation; cases that required follow-up were transferred to other child welfare caseworkers. Investigators were not involved in determining foster placement type, that is, whether children were placed with relatives, an unrelated family, or in a group home. Accordingly, the instrument does not predict the initial placement setting

⁴⁵Previous studies of foster care using a similar research design also find heterogeneous impacts by age and gender. For instance, while Warburton et al. (2014) find that foster placement harms the outcomes of 16- to 18-year-old males, Bald et al. (2019) find that placement significantly improves outcomes for young girls, but has no impacts on young boys.

⁴⁶Though LGBTQ youth are over-represented in foster care (HRC, 2015) and have especially traumatic experiences in the system (Sullivan et al., 2001), this study is unable to examine differences along this margin because the administrative data sources do not include information on sexual orientation or gender identity.

(Table A5, Columns 1–3).

Second, one might worry that investigators have discretion over the length of foster placement. However, unlike the criminal justice context in which judges make sentencing decisions at the extensive (sentence or not) and intensive (length of sentencing) margins, child welfare investigators play no role in decision making at the intensive margin.⁴⁷ Consistent with this claim, we find no evidence that the instrument is related to the number of days that children spend in the foster system or the number of foster homes in which they are placed (Table A5, Columns 4–5).⁴⁸

Though investigators do not influence children’s experiences in foster care, they may affect children and families during the investigation in ways that could potentially impact outcomes. For example, investigators could vary in their sensitivity to a family’s schedule or in how they conduct themselves during the investigation process, which could alter outcomes. Even with detailed survey data on family experiences, we would not be able to address all of the potential channels through which investigators could impact children’s outcomes.

We can empirically account for the provision of family services, however, which is perhaps the most important way in which investigators could influence children’s outcomes other than removal. As described in Section 5.2.2, in addition to placement, investigators also refer families to community and targeted services. Thus, the exclusion restriction could be violated if investigators who were more likely to remove children were also more likely to recommend targeted services, and tendencies over targeted services are not included in the estimation.⁴⁹

Table 6 reports the results of a specification that explicitly accounts for these additional margins, as described in Section 5.2.2. Two main findings emerge from this table. First, there is little evidence that family services (either community-based services alone or community-based and targeted services) without child removal impacted children’s outcomes. Estimates of the

⁴⁷The length of placement depends on variety of factors, none of which involve the initial investigator. First and foremost, it depends on the progress that birth parents make on their reunification plan, which details the steps they must take to regain child custody. This progress is monitored by both a child welfare case worker, who works in a different department than the initial investigator, and a judge. If parental rights are terminated either by the birth parents or the judge then the length of placement depends on the supply of adoptive homes and, after an adoptive family is identified, the speed of the adoption process.

⁴⁸We also find no evidence that the removal stringency instrument is jointly related to children’s experiences in the foster care system like placement setting or length; the p-value from an F-test for joint significance is 0.52.

⁴⁹It is worth noting, however, that there is little evidence that these types of services have meaningful impacts on children’s outcomes. For example, light-touch community services include simple referrals to community resources with no follow-up, such as giving a pamphlet about local resources. It is very unlikely that these “information-only” treatments have large effects on children. Although investigators refer more targeted services in higher-risk cases like substance abuse treatment, parenting classes, or job training, these programs tend to have small, if any, impacts on the outcomes of adult participants, let alone inter-generational effects on children (Barnow and Smith, 2015; SAMHSA, 2009; Wood et al., 2014).

impacts of these additional margins are mostly small and statistically insignificant. This evidence indicates that investigator stringency over family services was largely unrelated to children’s outcomes.

Second, we find that the impacts of child removal above and beyond family services are nearly identical in magnitude to the estimates in Table 5. Although the main estimate on the index of child well-being is less precise than in the main specification, we can reject that placement hurt children’s outcomes at the 10 percent level using a one-sided hypothesis test. We also find positive impacts on educational outcomes that are statistically significant at the 10 percent level. Thus, this exercise suggests that the removal stringency instrument operates through foster placement.⁵⁰

Finally, one may also be concerned that—prior to the placement decision—investigators make a decision on whether or not to substantiate the investigation. Therefore, the exclusion restriction could be violated if the decision of whether or not to substantiate influences children’s outcomes regardless of the decision to place children in foster care. To address this, we include an analysis that explicitly accounts for investigator discretion over substantiation. Specifically, we instrument for the decision to substantiate using an investigator’s leave-one-out propensity to substantiate. Together with the main removal stringency measure, we use this new measure to simultaneously instrument for substantiation and foster care placement. These estimates are shown in Table B15 and yield two important findings. First, the additional impact of foster placement relative to substantiation without removal is nearly identical to the main results shown in Table 4. Second, the impact of substantiation alone is small and statistically insignificant. Thus, even though investigators can in theory influence children’s outcomes through the decision to substantiate, any bias from this “treatment” is likely small in our context.

5.4.2 Robustness to Alternative Samples

Table A6 shows that the main results are robust in both sign and magnitude to a variety of design decisions. We conduct the analysis using alternative samples (Panel A). First, we limit the sample to only the first investigation of each child. Next, we test sensitivity to

⁵⁰A remaining concern is whether investigator placement tendencies are not only related to the likelihood of family services but also to the match quality of families and services. However, it is unlikely that this threatens identification in our context. First, when only community services are offered, investigators choose from a standard set of options (e.g., if the family needs help with bills, investigators refer to a community organization that deals with these issues; if the family needs food, they refer to a food pantry, and so on). Therefore, it is unlikely that investigator tendencies are correlated with match quality when community services are offered alone. Second, when targeted services are offered in addition to community services, investigators do not make specific referrals. Once the investigation is completed, the case is transferred to a caseworker who then connects the family with the targeted services. Thus, it is the caseworker and not the investigator who makes the recommendation for the specific types of targeted services.

the number of cases an investigator must have been assigned to be included in the sample. The main analysis excludes children assigned to investigators who worked fewer than 50 cases, so we strengthen this threshold to 75. The results are similar to those in the main analysis across these three alternative samples. Moreover, because we observe outcomes for a different number of years for each child, one may worry that the time pattern of the effects reflects sample-composition changes rather than the true dynamics of the treatment effects. Thus, we restrict the analysis to a balanced panel consisting of the first five follow-up years for students who could be observed in the public school system for five years after their investigation based on their grade level and year of investigation—those investigated in 7th grade or below in 2012 or earlier. The impacts of foster placement are substantially larger in the balanced panel than in the main analysis.⁵¹ Figure A2 shows the time patterns of the effects for this balanced panel. The results are nearly identical to those shown in Figure 5.

5.4.3 Robustness to Alternative Instruments

We also check for robustness using other reasonable ways to measure investigator removal tendencies (Panel B). First, we randomly split the sample in half and define the instrument as the investigator’s removal rate from the other half of the sample. Second, we allow tendencies to vary over time by creating a leave-out-other-years measure. Third, we address concerns that removal decisions occurring around the same time may be correlated by constructing a leave-out-same-year measure.

Fourth—as in Mueller-Smith (2015) and Bald et al. (2019)—we allow the removal rate to vary by child and case characteristics, and implement an IV approach that relaxes the monotonicity assumption inherent in specifications that use standard leave-out measures. We use Least Absolute Shrinkage and Selection Operator (LASSO) regressions to select the instruments with the greatest predictive power of foster care placement in the first-stage equation and estimate the second-stage specification with the chosen instruments.⁵² Lastly,

⁵¹This is consistent with the previously discussed results. Specifically, the effects in the main analysis could not have been driven by the placement of older children or those investigated later in the sample period because Section 5.2.1 shows that impacts appear only several years after removal. Similarly, the subgroup analysis in Section 5.3.2 shows considerably larger effects for younger children.

⁵²Specifically, we created five potential instruments based on leave-out measures of investigator removal tendency calculated for five types of child and case characteristics: (1) gender (female or male); (2) minority status (white or not); (3) allegation type (physical abuse or not); (4) perpetrator type (parent or not); (5) age at investigation (younger than 10 years old or not). For each characteristic, we defined mutually exclusive groups of children and calculated the leave-one-out measure of removal tendency based on the investigator tendency for the group. For instance—for gender—we calculate for each investigator the tendency measure for male and female cases. A male child assigned to investigator w will get investigator w ’s removal rate from all other cases involving male children. Similarly, a female child assigned to investigator w will get investigator w ’s removal rate from all other cases involving female children. We also create tendency measures using the other four binary covariates. This yields a total of five potential instruments. We then use LASSO

since including many controls (e.g., over 7,500 rotation fixed effects) can induce bias in leave-out instrument approaches, we implement the unbiased jackknife instrumental variables (UJIVE) approach of [Kolesar \(2013\)](#) which is robust to this issue.⁵³ Though they vary in precision, we find large, positive effects of foster care across all of the alternative instruments.

We also test sensitivity to the definition of rotational assignment (Panel C). The main analysis includes zip code by investigation year fixed effects because some of the local offices in Michigan divide investigators into teams based on small regions. A tiny fraction of zip codes in Michigan cross county lines, however, which could create measurement error in the main analysis. Importantly, the results are very similar when we instead include county by investigation year fixed effects.

5.4.4 Robustness to Control Selection

In addition, we test the sensitivity of our main results to control selection. Given that the 2SLS regressions in Table 4 include a variety of additional controls beyond rotation fixed effects, one may worry that the main results of the paper are unique to a particular specification. Table B10 shows the robustness of the main estimates to alternative sets of control variables. As the table indicates, 2SLS estimates of the impact of foster placement on the index of child well-being consistently show large and positive effects (ranging from 30 to 40 percent of a standard deviation), regardless of the control variables we include.

5.4.5 Robustness to Alternative Clustering Levels

Finally, we test the sensitivity of the results presented in Table 4 to alternative clustering levels. As discussed above, we cluster standard errors at the child level in our main results in order to account for the correlation in outcomes that arises mechanically by including the same child more than once in the panel. Table B11 shows that our main results are robust to clustering at the investigator level to account for potential correlations in the error term within investigator but across investigations and time (since the same investigator is assigned to multiple investigations throughout the panel). The results are also robust to clustering at the zip-year level to account for potential neighborhood- or child welfare office-level shocks. Lastly, the results are robust to two-way clustering by child and investigator—as well as two-way clustering by child and zip-year level—to account for all of the potential correlations in outcomes described above.

regressions to select the instruments with the greatest predictive power of foster care placement in the first-stage equation and estimate the second-stage specification using the selected instruments.

⁵³The UJIVE approach of [Kolesar \(2013\)](#) uses a leave-out approach to estimate investigator removal tendency conditional on the control variables included in the first and second stages in Equations 2 and 3.

6 Comparison to Doyle (2007, 2008)

The analysis in this study contrasts with the findings in Doyle (2007, 2008), which conclude that foster care placement had large, negative impacts on children’s long-term outcomes. These two studies also apply the examiner assignment design but use administrative data on children investigated between 1990 and 2003 in Illinois. Table 7 compares our estimates to these earlier results. Column 1 of Panel A reports our main estimate on standardized math test scores from Table 4; placement increased math scores by 0.36 standard deviations, an effect that is statistically significant at the 10 percent level. Since Doyle (2007, 2008) do not examine test scores, but do explore earnings, we use estimates from Deming et al. (2016) that link test scores and earnings as a benchmark. Specifically, Doyle (2007) finds that foster care reduced adult quarterly earnings by about \$1,300, which, as shown in Column 2, translates to a decrease in math scores of about 0.83 standard deviations. We can statistically reject that placement in our setting led to this large reduction in math achievement at the 1 percent level (Column 3). Columns 4 and 5 also report that we can rule out the large decline in math scores using estimates from Table 6, which explicitly account for investigator propensity to refer families to services. Panel B shows that we can also statistically rule out the 300 percent increase in juvenile delinquency found in Doyle (2007) at the 1 percent level using our main specification and at the 10 percent level using the more demanding specification from Table 6. Therefore, the evidence strongly suggests that foster care in Michigan during our sample period did not have the same large and lasting negative effects as it did for foster children in Illinois in earlier years.

There are two likely reasons as to why our findings differ from these earlier results, which we summarize next and offer a more detailed explanation in Supplemental Appendix B.3. First, children’s experiences in foster care were tremendously different across study settings. Median placement length in Illinois during the earlier period was over two years longer than in Michigan more recently (Wulczyn et al., 2000; USDHHS, 2003a, 2017b).⁵⁴ These long placements were also less stable: 45 percent of foster children in Illinois lived in three or more different homes in 1998 compared to an average of 31 percent across our 10-year panel in Michigan (USDHHS, 2003a; AECF, 2017). It is perhaps unsurprising that a setting with considerably shorter and more stable placements would have less harmful impacts. A second potential reason for the difference in findings is that national changes to child welfare over time likely improved foster care across the country. For example, the Adoption and Safe Families Act of 1997—in the middle of Doyle (2007, 2008)’s sample period—sought to

⁵⁴We are grateful to Mark Courtney and Fred Wulczyn for pointing us to a Chapin Hall report with statistics on foster care placement in the early 1990s.

reduce the length of foster placements. There has also been a considerable push to increase placements with relatives in the last two decades ([ChildTrends, 2018](#)). Both of these national trends could have contributed to improvements in foster care over time in both Michigan and Illinois.

We find less evidence for other plausible explanations for the contrast in findings. As discussed in much greater detail in [Appendix B.3](#), we find mixed evidence that observable characteristics of compliers like age, gender, or race/ethnicity drive the differences. We also find little evidence that compliers differed along unobserved dimensions such as risk in the home. Lastly, it is unlikely that the counterfactual for children placed (e.g., the quality of available alternatives) was considerably worse in our setting than in earlier work; in fact, family prevention services have grown and improved substantially over time ([USDHHS, 2016a](#)). Overall, the most likely explanation seems to be the large institutional differences in placement length and stability between foster care in Illinois during the earlier period and Michigan more recently.

7 Conclusion

This paper offers some of the only causal estimates of foster care on crucial indicators of child well-being: safety, education, and crime. To do so, we leverage the quasi-random assignment of child welfare investigators who vary in their propensity to recommend placement. Using detailed administrative data from Michigan to study over 200,000 child welfare investigations between 2008 and 2016, we find that placement improved a variety of children’s outcomes. Foster children were 50 percent less likely to be abused or neglected in the future. Placement also increased daily school attendance by 6 percent and improved standardized math test scores by one-third of a standard deviation, the equivalent of moving from the 33rd to the 46th percentile in the state. We also estimate a substantively large, yet imprecise reduction in juvenile delinquency. Even so, we can statistically reject that foster placement had the large negative impacts on children’s outcomes found in [Doyle \(2007, 2008\)](#).

The new research findings from this paper have important implications for public policy, especially in light of the Family First Prevention Services Act that took effect in October 2019. The legislation introduced massive changes to the child welfare system. Most relevant for this study, it made reducing the use of foster care a federal priority by allocating up to \$8 billion of federal Title IV-E funds for states to spend on alternatives to placement. Previously reserved for foster care and adoption budgets, except for waivers permitted in special cases, states can now use this funding stream on services to prevent foster care entry among children at risk of placement.

The results from this study, which indicate that foster care placement improved children's outcomes, suggest that current efforts to prevent child maltreatment in the home are falling short. To keep vulnerable children safe at home without foster care, state and local governments must focus on implementing more effective prevention programs. Identifying what works to ensure the safety and well-being of abused and neglected children who remain in their homes—and learning how to scale these interventions—is a crucial frontier for future research.

References

- AECF (2017). Kids count data center. Technical report, The Annie E. Casey Foundation.
<https://datacenter.kidcount.org>.
- Aizer, A. and J. J. Doyle (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics* 130(2), 759–803.
- Anderson, M. L., J. Gallagher, and E. R. Ritchie (2018). School meal quality and academic performance. *Journal of Public Economics* 168, 81–93.
- Arteaga, C. (2019). The cost of bad parents: Evidence from the effects of parental incarceration on children’s education. Working paper.
- Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economics Journal: Applied Economics* 11(3), 338–381.
- Bald, A., E. Chyn, J. S. Hastings, and M. Machelett (2019). The causal impact of removing children from abusive and neglectful homes. National Bureau of Economic Research Working Paper 25419.
- Barnow, B. S. and J. Smith (2015). Employment and training programs. In *Economics of Means-Tested Transfer Programs in the United States, Volume 2*.
- Barrat, V. X. and B. Berliner (2013). The invisible achievement gap, part 1: Education outcomes of students in foster care in california’s public schools, part one. Technical report, WestEd.
- Berger, L. M., M. Cancian, E. Han, J. Noyes, and V. Rios-Salas (2014). Children’s academic achievement and foster care. *Pediatrics* 135(1), 109–116.
- Berrick, J. D. (2018). *The Impossible Imperative: Navigating the competing principles of child protection*. Oxford University Press.
- Bertrand, M. and J. Pan (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economics Journal: Applied Economics* 5(1), 32–64.
- Berzin, S. C. (2010). Understanding foster youth outcomes: Is propensity scoring better than traditional methods? *Research on Social Work Practice* 20(1), 100–111.
- Bhuller, M., G. B. Dahl, K. V. Loken, and M. Mogstad (2018). Intergenerational effects of incarceration. In *AEA Papers and Proceedings*, Volume 108, pp. 234–40.
- Billings, S. B. (2019). Parental arrest and incarceration: How does it impact the children? Working paper.
- Bosk, E. A. (2015). *All Unhappy Families: Standardization and Child Welfare Decision-Making*. Ph. D. thesis, University of Michigan.
- Carrell, S. E. and M. Hoekstra (2014). Are school counselors an effective education input? *Economics Letters* 125(1), 66–69.
- Chetty, R., N. Hendren, and L. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review* 106(4), 855–902.
- ChildTrends (2017). Michigan foster care, federal fiscal year 2015. Technical report, Child Trends.
https://www.childtrends.org/wp-content/uploads/2017/01/Michigan-Foster-Care-Factsheet_2015.pdf.

- ChildTrends (2018). Foster care. Technical report, Child Trends Databank.
- 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.
- Clifford, S. and J. Silver-Greenberg (2017). Foster care as punishment: The new reality of ‘jane crow’. *The New York Times*.
<https://www.nytimes.com/2017/07/21/nyregion/foster-care-nyc-jane-crow.html>.
- Collinson, R. and D. Reed (2019). The effects of evictions on low-income households. Working paper.
- CRHE (2017). Homeschooling by the numbers. Technical report, Coalition for Responsible Home Education.
<https://www.responsiblehomeschooling.org/homeschooling-101/homeschooling-numbers/>.
- Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014). Family welfare cultures. *The Quarterly Journal of Economics* 129(4), 1711–1752.
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *The Review of Economics and Statistics* 86(1), 195–210.
- Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics* 1(3), 111–34.
- Deming, D., S. Cohodes, J. Jennings, and C. Jencks (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics* 98(5), 848–862.
- Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review* 108(2), 201–40.
- Doyle, 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.
- Eren, O. and N. Mocan (2017). Juvenile punishment, high school graduation and adult crime: Evidence from idiosyncratic judge harshness. National Bureau of Economic Research Working Paper 23573.
- Figlio, D. and J. Winicki (2005). Food for thought: The effects of school accountability plans on school nutrition. *Journal of Public Economics* 89, 381–394.
- Frandsen, B. R., L. J. Lefgren, and E. C. Leslie (2019). Judging judge fixed effects. National Bureau of Economic Research Working Paper 25528.
- Frisvold, D. E. (2015). Nutrition and cognitive achievement: An evaluation of the school breakfast program. *Journal of Public Economics* 124, 91–104.
- Goodman, J. (2014). Flaking out: Student absences and snow days as disruptions of instructional time. National Bureau of Economic Research Working Paper 20221.
- Grimon, M.-P. (2020). Effects of the child protection system on parents. Working paper.
- Hayduk, I. (2017). The effect of kinship placement laws on foster children’s well-being. *The B.E. Journal of Economic Analysis & Policy* 17(1), 1–23.
- Heller, S. (2014). Summer jobs reduce violence among disadvantaged youth. *Science* 346.

- Herbst, D. (2018). Liquidity and insurance in student loan contracts: Estimating the effects of income-driven repayment on default and consumption. Working paper.
- Hogelin, J. M. (2013). To prevent and to protect: The reporting of child abuse by educators. *Brigham Young University Education and Law Journal*, 225.
- HRC (2015). Lgbtq youth in the foster care system. Technical report, Human Rights Campaign.
- Humphries, J. E., N. S. Mader, D. I. Tannenbaum, and W. L. van Dijk (2019). Does eviction cause poverty? quasi-experimental evidence from cook county, il. National Bureau of Economic Research Working Paper 26139.
- Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Imberman, S. A. and A. D. Kugler (2014). The effect of providing breakfast on student performance. *Journal of Policy Analysis and Management* 33(669-699).
- Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *American Economic Review* 94(1), 233–258.
- Kawano, L., B. Sacerdote, W. Skimmyhorn, and M. Stevens (2017). On the determinants of young adult outcomes: An examination of random shocks to children in military families. Working paper.
- Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review* 96(3), 863–876.
- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica* 75(1), 83–119.
- Kling, J. R., J. Ludwig, and L. F. Katz (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Quarterly Journal of Economics* 120(1), 87–130.
- Kolesar, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. *Unpublished Working Paper*.
- Leos-Urbel, J., A. E. Schwartz, M. Weinstein, and S. Corcoran (2013). Not just for poor kids: The impact of universal free school breakfast on meal participation and student outcomes. *Economics of Education Review* 36, 88–107.
- Lovett, N. and Y. Xue (2018). Family first or the kindness of strangers? foster care placements and adult outcomes. Working paper.
- Mack, J. (2017). Where michigan children attended school in 2016-2017 – public and private. Technical report, MLive. https://www.mlive.com/news/2017/09/where_michigan_children_attend.html.
- Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. Working paper.
- Mulhern, C. (2019). Beyond teachers: Estimating individual guidance counselors’ effects on educational attainment. Working paper.
- Neilson, S. (2019). More kids are getting placed in foster care because of parents’ drug use. *NPR*. <https://www.npr.org/sections/health-shots/2019/07/15/741790195/more-kids-are-getting-placed-in-foster-care-because-of-parents-drug-use>.
- Norris, S. (2019). Examiner inconsistency: Evidence from refugee appeals. Working paper.

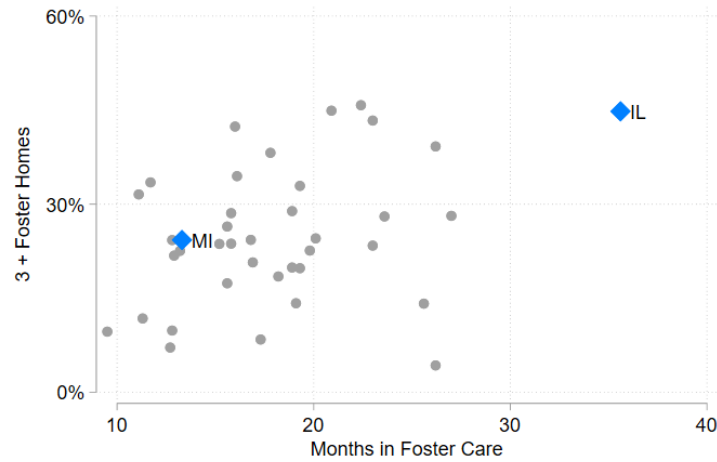
- Norris, S., M. Pecenco, and J. Weaver (2019). The effects of parental and sibling incarceration: Evidence from ohio. Working paper.
- O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, 1079–1087.
- OCA (2018). Examining connecticut's safety net for children withdrawn from school for the purpose of homeschooling. Technical report, Office of the Child Advocate, State of Connecticut.
- O'Hare, W. P. (2007, January). *Census Bureau Plans to Eliminate 'Foster Child' Category*. Population Reference Bureau. <https://www.prb.org/censusbureau/fosterchildcategory/>.
- Pears, K. and P. A. Fisher (2005). Developmental, cognitive, and neuropsychological functioning in preschool-aged foster children: Associations with prior maltreatment and placement history. *Journal of Developmental & Behavioral Pediatrics* 26(2), 112–122.
- Pecora, P. J., R. C. Kessler, K. O'Brien, C. R. White, J. Williams, E. Hiripi, D. English, J. White, and M. A. Herrick (2006). Educational and employment outcomes of adults formerly placed in foster care: Results from the northwest foster care alumni study. *Children and youth services review* 28(12), 1459–1481.
- Ringler, D. A. (2018). Office of the auditor general, performance audit report, children's protective services investigations. Technical report, Michigan Department of Health and Human Services.
- Roberts, K. V. (2019). Foster care and child welfare. Working paper.
- Ryan, J. P., B. A. Jacob, M. Gross, B. E. Perron, A. Moore, and S. Ferguson (2018). Early exposure to child maltreatment and academic outcomes. *Child maltreatment* 23(4), 365–375.
- Ryan, J. P. and M. F. Testa (2005). Child maltreatment and juvenile delinquency: Investigating the role of placement and placement instability. *Children and youth services review* 27(3), 227–249.
- SAMHSA (2009). Treatment episode dataset (teds) highlights 2007, national admissions to substance abuse treatment services. Technical report, Substance Abuse and Mental Health Services Administration, Office of Applied Studies.
- Sanbonmatsu, L., J. R. Kling, G. J. Duncan, and J. Brooks-Gunn (2006). Neighborhoods and academic achievement: Results from the moving to opportunity experiment. *Journal of Human Resources* 41(4), 649–691.
- Scherr, T. G. (2007). Educational experiences of children in foster care: Meta-analyses of special education, retention and discipline rates. *School Psychology International* 28(4), 419–436.
- Schwartz, A. E. and M. W. Rothbart (2017). Let them eat lunch: The impact of universal free meals on student performance. Working paper.
- Stagner, M. (2019, November). Getting closer: Embracing the emotional aspects of our craft to help policy research matter more. <https://www.mathematica.org/commentary/getting-closer-embracing-the-emotional-aspects-of-our-craft-to-help-policy-research-matter-more>.
- Sullivan, C., S. Sommer, and J. Moff (2001). Youth in the margins: A report on the unmet needs of lesbian, gay, bisexual, and transgender adolescents in foster care. Technical report, Lambda Legal Defense and Education Fund.
- Talbot, M. (2017). The addicts next door. *The New Yorker*. <https://www.newyorker.com/magazine/2017/06/05/the-addicts-next-door>.

- Trout, A. L., J. Hagaman, K. Casey, R. Reid, and M. H. Epstein (2008). The academic status of children and youth in out-of-home care: A review of the literature. *Children and Youth Services Review* 30(9), 979–994.
- USDHHS (2003a). Child welfare outcomes. Technical report, Administration for Children and Families, Administration on Children, Youth and Families, Children’s Bureau.
- USDHHS (2003b). Child welfare outcomes 2001: Annual report. Technical report, United States Department of Health and Human Services.
- USDHHS (2008). Child welfare outcomes. Technical report, Administration for Children and Families, Administration on Children, Youth and Families, Children’s Bureau.
- USDHHS (2016a). Child maltreatment prevention: Past, present, and future. Technical report, Children’s Bureau, Department of Health and Human Services.
- USDHHS (2016b). Child welfare outcomes. Technical report, Administration for Children and Families, Administration on Children, Youth and Families, Children’s Bureau.
- USDHHS (2016c). *Reunification: Bringing Your Children Home From Foster Care*. Children’s Bureau. <https://www.childwelfare.gov/pubPDFs/reunification.pdf#page=9&view=What%20can%20I%20expect%20after%20my%20children%20come%20home?>
- USDHHS (2017a). Child maltreatment 2015. Technical report, United States Department of Health and Human Services.
- USDHHS (2017b). Child welfare outcomes report data. Technical report, Children’s Bureau, Administration for Children and Families, U.S. Department of Health and Human Services. <https://cwoutcomes.acf.hhs.gov/cwdatasite/byState>.
- USDHHS (2018a). The afcars report: Preliminary fy 2017 estimates. Technical report, United States Department of Health and Human Services, Administration for Children and Families.
- USDHHS (2018b). Child maltreatment 2016. Technical report, Administration for Children and Families, Administration on Children, Youth and Families, Children’s Bureau.
- USDHHS (2018c). Child welfare outcomes report data. Technical report, Children’s Bureau, Administration for Children and Families, U.S. Department of Health and Human Services. <https://cwoutcomes.acf.hhs.gov/cwdatasite/recurrence/index>.
- USDHHS (2020). Child maltreatment 2020. Technical report, United States Department of Health and Human Services.
- Warburton, W. P., R. N. Warburton, A. Sweetman, and C. Hertzman (2014). The impact of placing adolescent males into foster care on education, income assistance, and convictions. *Canadian Journal of Economics* 47(1), 35–69.
- White, R. and B. DeGrow (2016). A survey of michigan’s private education sector. Technical report, Mackinac Center for Public Policy.
- Wildeman, C. and N. Emanuel (2014). Cumulative risks of foster care placement by age 18 for us children, 2000–2011. *PloS one* 9(3), e92785.
- Wiltz, T. (2018). This new federal law will change foster care as we know it. *PEW Charitable Trusts*. <https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2018/05/02/this-new-federal-law-will-change-foster-care-as-we-know-it>.

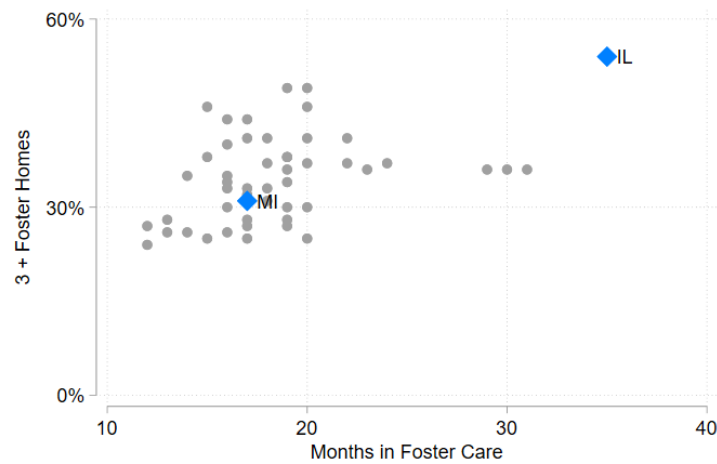
- Wood, R. G., Q. Moore, A. Clarkwest, and A. Killewald (2014). The long-term effects of building strong families: A program for unmarried parents. *Journal of Marriage and Family* 76(2), 446–463.
- Wulczyn, F., C. Smithgall, and L. Chen (2009). Child well-being: The intersection of schools and child welfare. *Review of research in education* 33(1), 35–62.
- Wulczyn, F. H., K. B. Hislop, and R. M. Goerge (2000). Foster care dynamics 1983-1998. Technical report, Chapin Hall Center for Children at the University of Chicago.
- Zlotnick, C., T. W. Tam, and L. A. Soman (2012). Life course outcomes on mental and physical health: the impact of foster care on adulthood. *American Journal of Public Health* 102(3), 534–540.

Figure 1: Comparison of State Foster Care Systems

(a) 1998 Statistics

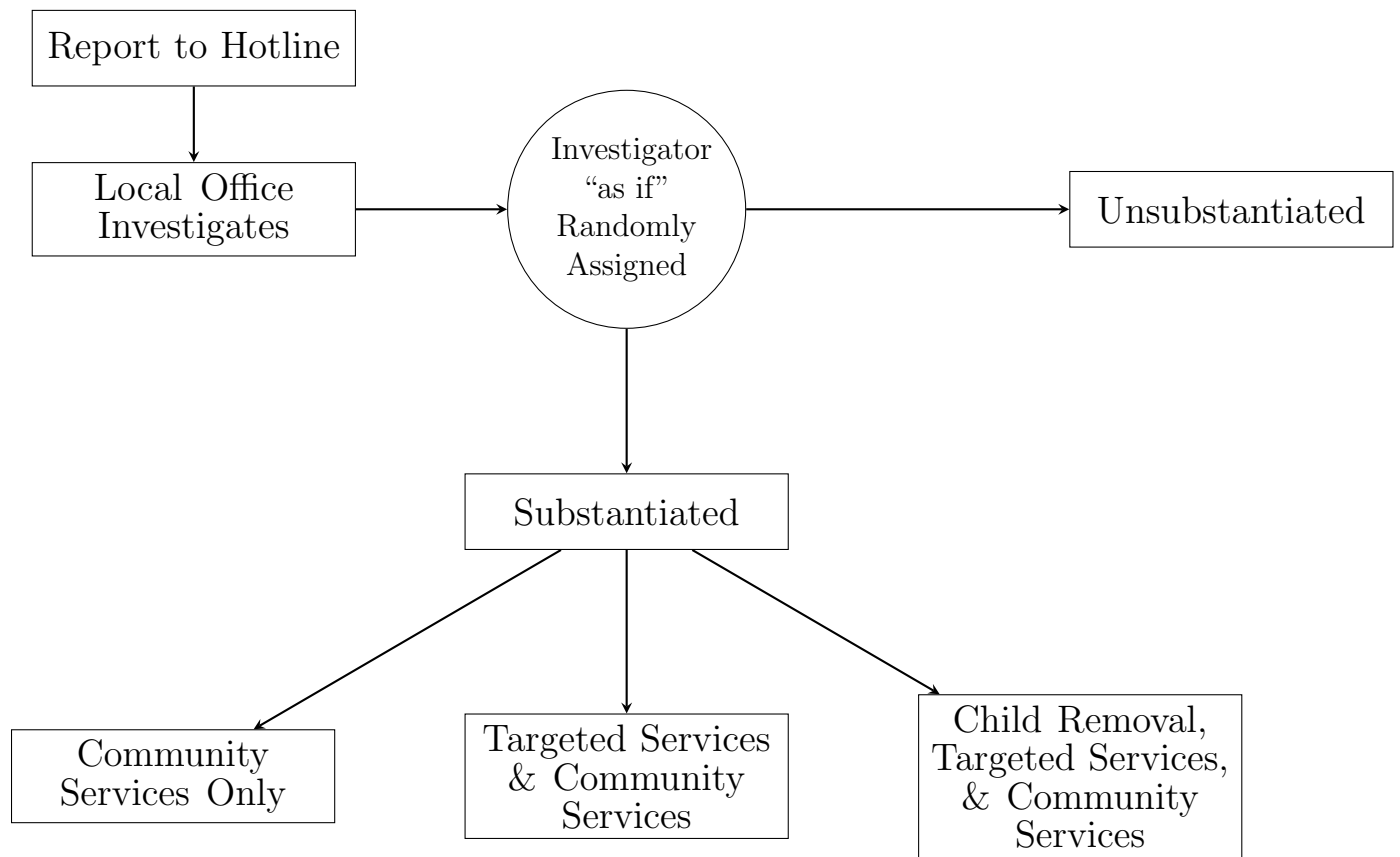


(b) 2015 Statistics



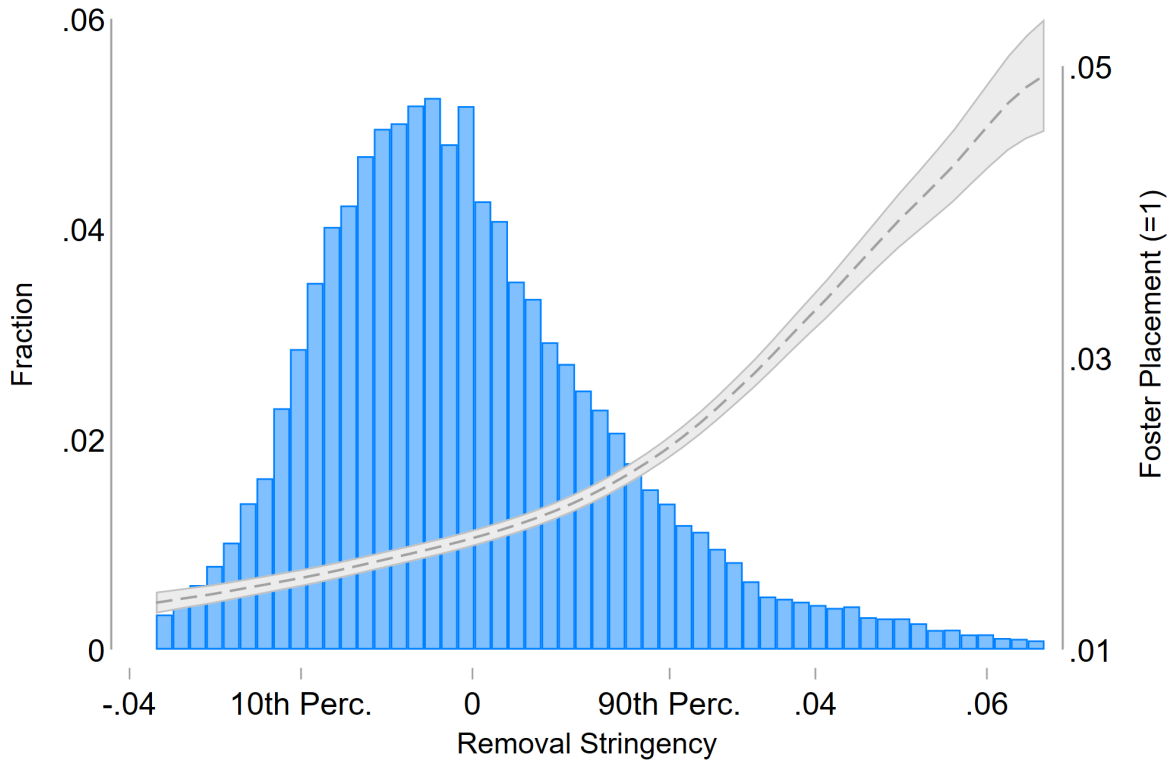
Notes. These figures show statistics about state foster care systems from 1998, the first year of publicly available data, reported in [USDHHS \(2003b\)](#), and from 2015, reported in [USDHHS \(2017a\)](#). Due to a change in reporting, the horizontal axis shows the median number of months spent in foster care for each state in 1998 and the average number of months in 2015. The vertical axis shows the share of foster children who lived in at least three different foster homes in both periods. In 1998, 10 states did not report either of these statistics.

Figure 2: Overview of Child Maltreatment Investigations in Michigan



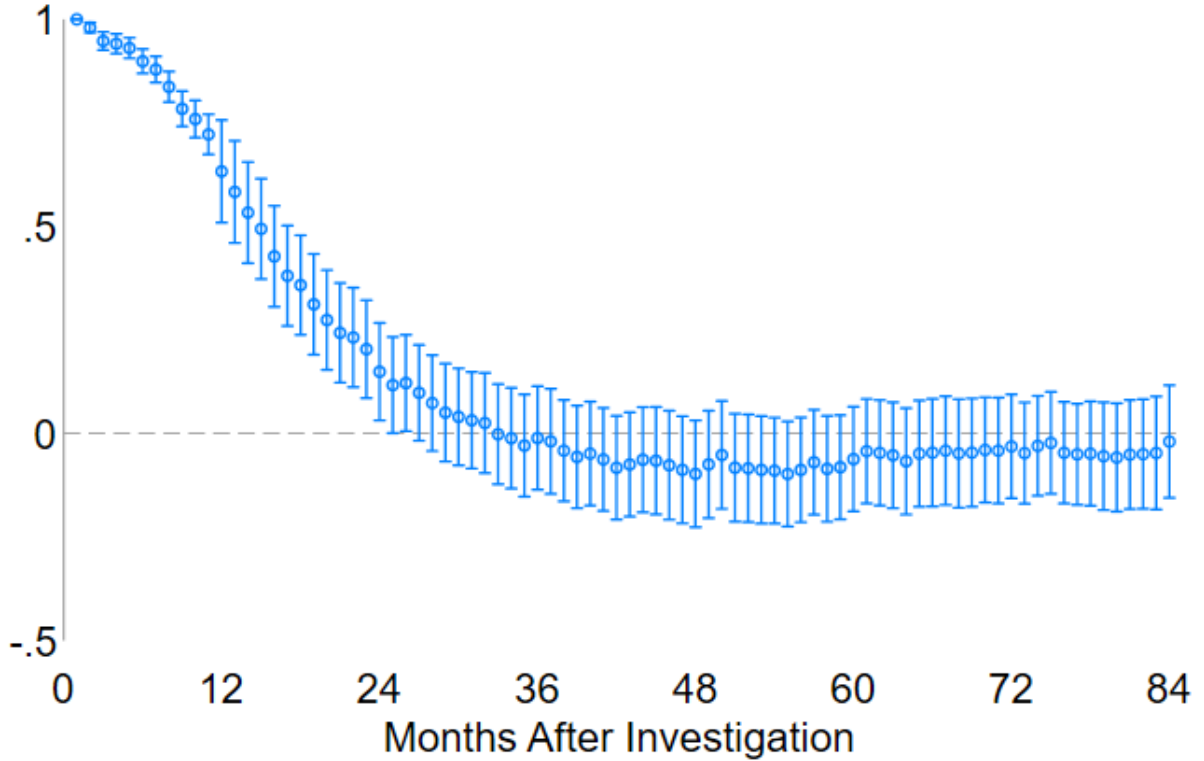
Notes. This figure describes the child maltreatment investigation process in Michigan. “Substantiated” means that investigators found enough evidence to support the abuse or neglect allegation. Conditional on substantiation, low-risk families receive a referral to community-based services like a local food pantry or drug rehabilitation group; high-risk families additionally receive targeted services like substance abuse treatment or parenting classes. In cases with the most intensive risk, the child is also removed from the home and placed in foster care.

Figure 3: Distribution of Investigator Removal Stringency Instrument



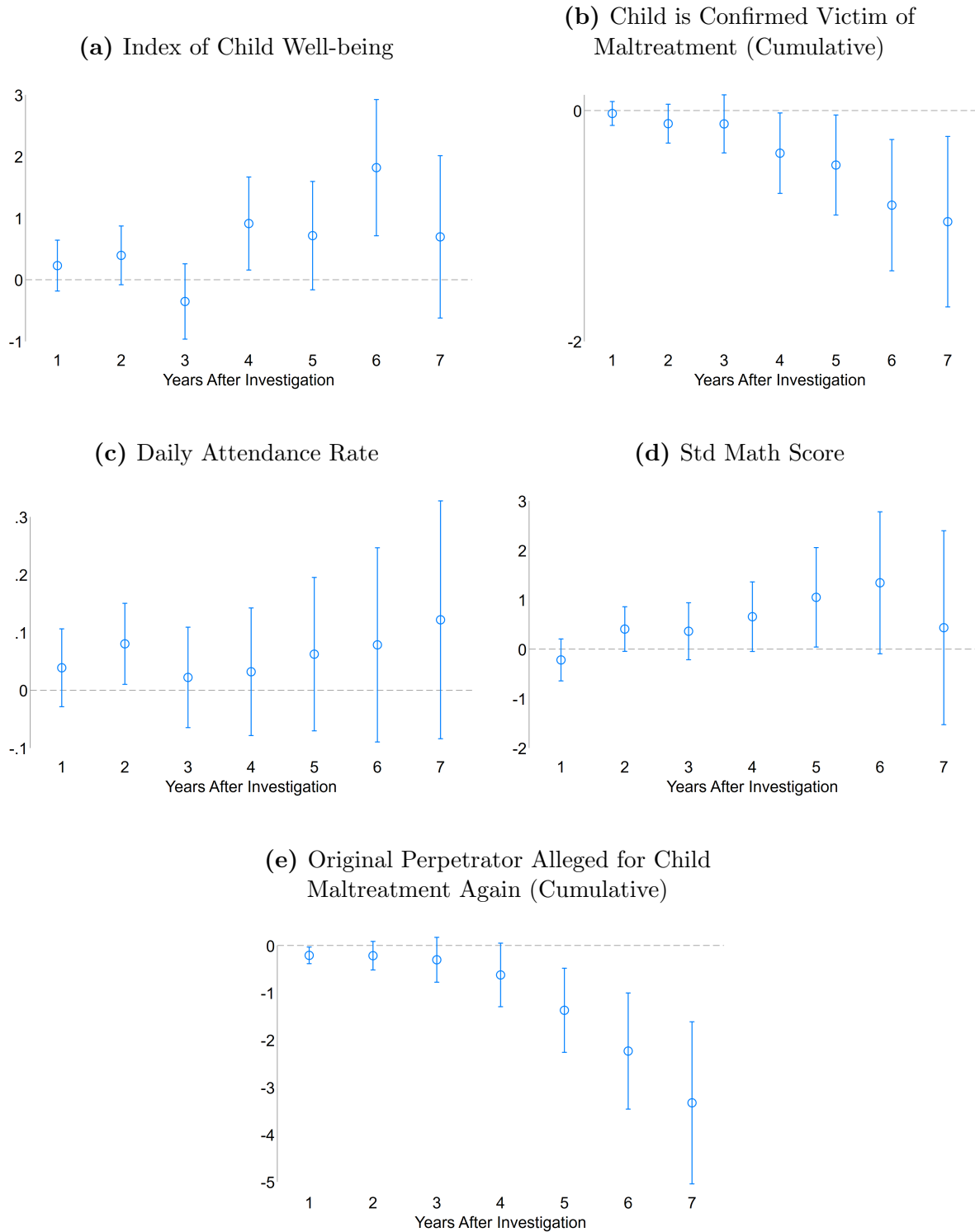
Notes. This figure shows the distribution of the removal stringency instrument residualized by the level of rotational assignment. That is, the instrument is shown net of child zip code by investigation year fixed effects in order to show that there is variation in propensity to remove within local offices. The instrument is calculated as the fraction of all other investigations—both past and future—assigned to the same investigator that resulted in foster placement. Superimposed over the histogram is the non-parametric regression of foster placement on investigator tendencies, residualizing out child zip code by investigation year fixed effects. The shaded area represents the 95 percent confidence interval.

Figure 4: Effects of Foster Care on Likelihood of Being in Foster System Over Time



Notes. This figure reports the results from 2SLS regressions of the likelihood of being in the foster system on an indicator for foster placement using removal stringency to instrument for placement. It plots both the point estimates and their 95 percent confidence intervals. All specifications include the covariates as listed in the text, as well as zip code by investigation year fixed effects. Standard errors are clustered by child. Children are defined as being in the foster system during a given month if they were ever in foster care during that month. The figure shows the results from an unbalanced panel where children who turn 18 years old exit from the analysis. The point estimate can be negative in the rare case that control compliers eventually enter foster care.

Figure 5: Effects of Foster Care Over Time



Notes. These figures report the results from 2SLS regressions of the outcome variable on foster care using removal stringency to instrument for foster care. They plot both the point estimates and their 95 percent confidence intervals. All specifications include the covariates as listed in the text, as well as zip code by investigation year fixed effects. Standard errors are clustered by child. Follow-up years are defined by school years even for non-schooling outcomes. Figure 5e represents the effect of child removal on the cumulative number of future allegations of child maltreatment against the original perpetrator. Since multiple perpetrators can be involved in the original case, this represents the mean effect across all perpetrators. For reference, 56 percent of investigations involved a single perpetrator, 97 percent involved one or two, and 99.4 percent involved three or fewer.

Table 1: Summary Statistics

		Analysis Sample	
	(1) All Michigan Students	(2) All	(3) Foster Care
<i>Child Socio-Demographics</i>			
Female	0.49	0.49	0.47
White	0.67	0.62	0.52
Black	0.21	0.29	0.39
Hispanic	0.08	0.07	0.07
Other Race	0.05	0.03	0.02
Age	11.70	10.34	10.59
Grade in School	6.15	4.76	4.93
Low Income	0.49	0.83	0.87
<i>Prior Schooling Characteristics</i>			
Attendance Rate	0.95	0.81	0.74
Special Education	0.14	0.22	0.23
Ever Retained in Grade	0.20	0.36	0.39
Std Math Score	0.00	-0.27	-0.36
Std Reading Score	0.00	-0.25	-0.34
<i>Investigation Characteristics</i>			
Had Prior Investigation	0.23	0.59	0.68
Abuse		0.32	0.26
Neglect		0.68	0.74
Substantiated		0.20	1.00
Foster Care		0.02	1.00
Observations	1,262,665	242,233	4,809

Notes. This table reports summary statistics for three groups of students. Column 1 consists of the cross-section of Michigan public school students during the 2016–2017 academic year enrolled in grades 1 through 11. All variables listed in Column 1 are measured during the 2016–2017 school year, and age is defined as of September 1, 2016. Column 2 contains all investigations in the analysis sample and Column 3 contains the subset of investigations that resulted in foster placement. The socio-demographic variables in Columns 2 and 3 are measured in the school year of the investigation. Low income is measured by free or reduced-price lunch eligibility. The prior schooling characteristics are measured in the school year prior to the investigation. Math and reading test scores are normalized for the entire state to have mean zero and standard deviation of one within every subject by grade by year cell. The abuse and neglect categories are coded to be mutually exclusive indicators such that abuse is equal to one for any investigation that involved physical abuse and neglect is equal to one for all investigations that did not involve physical abuse.

Table 2: First Stage Effect of Removal Stringency on Foster Placement

	(1) Foster Care	(2) Foster Care	(3) Foster Care	(4) Foster Care
Removal Stringency	0.480*** (0.019)	0.451*** (0.021)	0.450*** (0.021)	0.449*** (0.021)
Observations	242,233	242,233	242,233	242,233
F-Statistic	898.405	586.569	584.34	582.28
Zipcode by Year FE		✓	✓	✓
Socio-Demographic Controls			✓	✓
Academic Controls				✓

Notes. This table reports the results from regressions of foster placement on the leave-out measure of removal stringency. Each column includes a different set of covariates. Socio-demographic controls include gender, race/ethnicity, indicators for grade in school, an indicator for whether the child was the subject of a prior investigation, and the number of prior investigations. Academic controls include an indicator for free or reduced-price lunch eligibility, an indicator for receipt of special education services, an indicator for ever expelled, daily attendance rate—measured in the school year prior to the investigation—as well as the most recent pre-investigation score from standardized math and reading test scores. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Balance Tests for the Conditional Random Assignment of Investigators

	Full Sample		4th Grade and Above	
	(1) Foster Care	(2) Investigator Removal Stringency	(3) Foster Care	(4) Investigator Removal Stringency
Dependent Variable:				
F-Statistic from Joint Test	24.421	1.092	14.434	1.030
P-Value from Joint Test	0.000	0.341	0.000	0.421
Observations	242,233	242,233	144,032	144,032

Notes. This table reports the results from regressions of the dependent variable (either foster care placement or investigator removal stringency) on a variety of socio-demographic and academic covariates as described in the main text, as well as zip code by investigation year fixed effects. Columns 1 and 2 include the full sample of investigations and exclude standardized test scores in the vector of covariates. As students in Michigan begin taking statewide standardized tests in grade 3, Columns 3 and 4 report results for students enrolled in at least grade 4 during the maltreatment investigation and include standardized test scores. Standard errors are clustered by child.

Table 4: Effects of Foster Care on Child Outcomes

	(1) Index of Child Well-being	(2) Alleged Victim of Maltreatment	(3) Confirmed Victim of Maltreatment	(4) Daily Attendance Rate	(5) Std Math Score	(6) Std Reading Score	(7) Juvenile Delinquency
<i>Panel A: OLS</i>							
Foster Care	0.026** (0.011) {0.002}	-0.032*** (0.004) {0.177}	-0.007*** (0.002) {0.046}	0.011*** (0.002) {0.912}	0.057*** (0.013) {-0.501}	0.065*** (0.014) {-0.479}	0.041*** (0.004) {0.025}
<i>Panel B: 2SLS</i>							
Foster Care	0.392** (0.164) {-0.123}	-0.132** (0.058) {0.255}	-0.053* (0.028) {0.094}	0.055** (0.026) {0.893}	0.356* (0.203) {-0.429}	0.175 (0.219) {-0.234}	-0.028 (0.040) {0.051}
One-Sided P-Value	0.008	0.011	0.029	0.019	0.040	0.212	0.241
Observations	242,233	242,233	242,233	224,925	177,118	177,084	134,076

Notes. Panel A reports the results from OLS regressions of the outcome variable on foster care while Panel B reports the results from 2SLS regressions using removal stringency to instrument for foster care. Standard errors clustered by child are shown in parentheses. The curly brackets below the standard error represent the control mean in Panel A and the control complier mean in Panel B. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. The education and crime outcomes do not include all of the observations in the sample. Specifically, some grade-level and attendance records are missing and students may not have taken a standardized math or reading test if they were too young or old to be in grades 3–8, were absent from school on a test day, or were exempt. Furthermore, juvenile delinquency data are missing for eight counties, available only through 2015, and relevant only for children younger than Michigan’s age of majority of 16. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Effects of Foster Care Over Time

	(1) Index of Neighborhood & School Characteristics	(2) Index of Child Well-being	(3) Confirmed Victim of Maltreatment	(4) Daily Attendance Rate	(5) Std Math Score	(6) Received Special Education Services	(7) Retained in Grade
<i>Panel A: One Year After Investigation</i>							
Foster Care	0.257** (0.100) {-0.147}	0.231 (0.211) {0.028}	-0.024 (0.053) {0.068}	0.040 (0.034) {0.912}	-0.207 (0.217) {0.062}	-0.013 (0.063) {0.099}	-0.035 (0.049) {0.065}
<i>Panel B: Two+ Years After Investigation</i>							
Foster Care	0.066 (0.125) {-0.011}	0.446** (0.196) {-0.159 }	-0.064** (0.032) {0.102}	0.060* (0.031) {0.885}	0.579** (0.239) {-0.624}	0.014 (0.105) {0.035}	-0.008 (0.036) {0.062}
Observations	242,233	242,233	242,233	224,925	177,118	242,233	242,204

Notes. This table reports the results from 2SLS regressions of the outcome variable on foster care using removal stringency to instrument for foster care. Panel A reports results for outcomes measured during the first school year after the investigation and Panel B reports results across all school years after the first. Standard errors are in parentheses and clustered by child. The curly brackets below the standard error represent the control complier mean. The index of neighborhood and school characteristics is made up of neighborhood median income, educational attainment, and employment rate, as well as school average test scores and income level. The effects on each component of the index of neighborhood and school characteristics is shown in Table B7. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Effects of Adult Interventions on Child Outcomes

	(1) Index of Child Well-being	(2) Alleged Victim of Maltreatment	(3) Confirmed Victim of Maltreatment	(4) Daily Attendance Rate	(5) Std Math Score	(6) Std Reading Score	(7) Juvenile Delinquency
<i>Panel A: One Year After Investigation</i>							
Child Removal, Targeted Services, and Community Services	0.033 (0.294)	0.031 (0.128)	-0.038 (0.074)	0.015 (0.046)	-0.337 (0.296)	-0.288 (0.318)	-0.005 (0.068)
Targeted Services and Community Services	0.075 (0.114)	-0.108** (0.049)	-0.005 (0.028)	-0.000 (0.014)	0.157 (0.105)	0.040 (0.112)	-0.007 (0.027)
Community Services	0.009 (0.073)	0.046 (0.032)	0.011 (0.018)	0.009 (0.009)	-0.110 (0.069)	0.050 (0.075)	0.008 (0.017)
<i>Panel B: Two+ Years After Investigation</i>							
Child Removal, Targeted Services, and Community Services	0.350 (0.260)	-0.069 (0.087)	-0.037 (0.043)	0.075* (0.040)	0.564* (0.313)	0.188 (0.334)	-0.013 (0.064)
Targeted Services and Community Services	0.128 (0.093)	-0.045 (0.032)	-0.024 (0.016)	0.003 (0.013)	0.081 (0.113)	0.031 (0.119)	-0.006 (0.024)
Community Services	-0.090 (0.060)	0.018 (0.020)	0.012 (0.010)	-0.010 (0.009)	-0.086 (0.074)	-0.005 (0.079)	-0.006 (0.015)
One-Sided P-Value	0.089	0.215	0.192	0.031	0.036	0.286	0.421

Notes. This table reports estimates of β_1 , β_2 , and β_3 from Equation 7. One-sided p-values are for estimates of β_1 after the first year following the investigation. Standard errors are in parentheses and clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Comparison to [Doyle \(2007\)](#)

(1) Main Estimate Gross and Baron (2020)	(2) Comparable Estimate Doyle (2007)	(3) P-Value (1) = (2)	(4) Table 6, Panel B Gross and Baron (2020)	(5) P-Value (5) = (2)
<i>Panel A: Standardized Math Test Scores</i>				
0.356* (0.203)	-0.83	0.000	0.564* (0.313)	0.000
<i>Panel B: Juvenile Delinquency</i>				
-0.028 (0.040)	0.10	0.002	-0.013 (0.064)	0.077

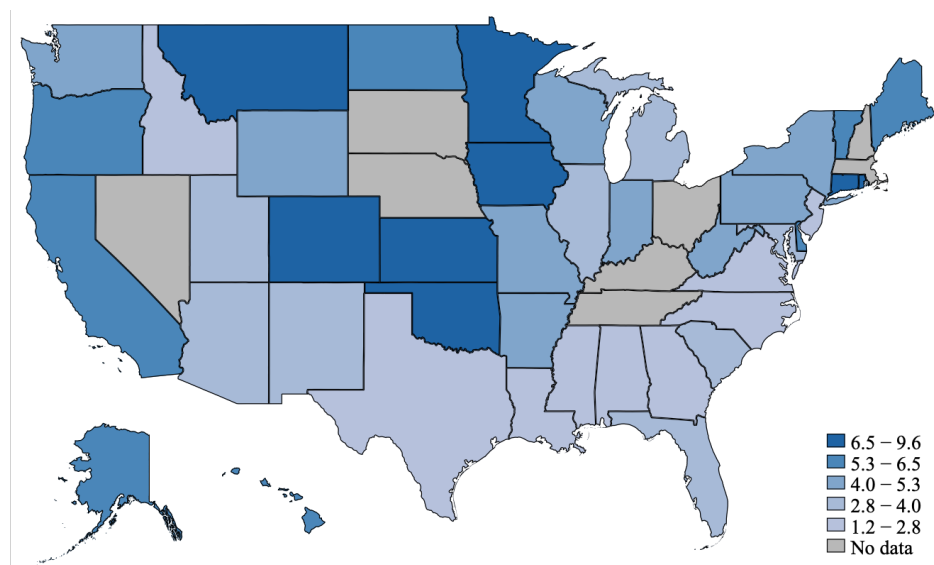
Notes. This table compares the main results of this paper to those in [Doyle \(2007\)](#)’s seminal study. The first column of the table presents the main 2SLS estimates (from Table 4) of foster care placement on standardized math test scores (Panel A) and juvenile delinquency (Panel B). Column 2 presents what we call “comparable” estimates from [Doyle \(2007\)](#). We calculate these two estimates as follows: [Doyle \(2007\)](#) finds that removal reduced annual earnings by \$1,300 for 18- to 28-year-olds. We rely on estimates from [Deming et al. \(2016, Table 2\)](#) to link test scores to future earnings. [Deming et al. \(2016\)](#) show that a school accountability program increased 10th grade math scores for students who had failed their 8th grade exam by 0.19 standard deviations and increased earnings at age 25 by \$298. We use this subgroup of students to mirror the low average baseline performance of children with a report of abuse or neglect. Based on these estimates, [Doyle \(2007\)](#)’s estimate on earnings would imply a decline in test scores of roughly 83 percent of a standard deviation. In terms of juvenile delinquency, [Doyle \(2007\)](#) finds that removal increased juvenile delinquency by about 300 percent relative to the sample mean. It is important to note that we use a slightly different measure of juvenile delinquency than [Doyle \(2007\)](#). Specifically, our outcome measures the filing of a juvenile court petition, which occurs following arrest so long as youth are not immediately diverted from the courts. In contrast, [Doyle \(2007\)](#) examines appearance before a juvenile court, which the study notes, “usually entails three juvenile arrests or an arrest for a serious charge.” Since the outcome in [Doyle \(2007\)](#) indicates greater involvement with the juvenile justice system than the filing of a court petition, our analysis likely underestimates a potential reduction in juvenile court appearances in this comparison. Nevertheless, given that average juvenile delinquency in our sample is roughly 2.5 percent, [Doyle \(2007\)](#)’s estimates would imply an increase of roughly 10 percentage points, or 0.10, in the juvenile delinquency rate. Column 3 tests whether our main estimates are equal to [Doyle \(2007\)](#)’s comparable estimates. Column 4 presents the estimates of foster placement on math test scores and juvenile delinquency shown in Table 6, which controls directly for targeted and community services. Finally, Column 5 tests whether the estimates in Table 6 are equal to [Doyle \(2007\)](#)’s comparable estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Print Appendix

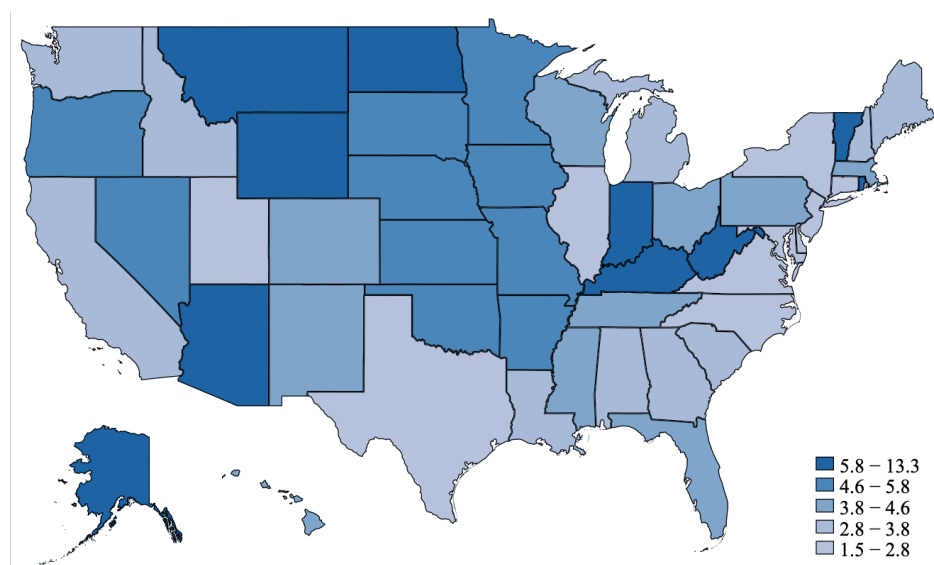
A.1 Supplemental Print Figures and Tables

Figure A1: Foster Care Entry Per 1000 Children

(a) 1998 Statistics

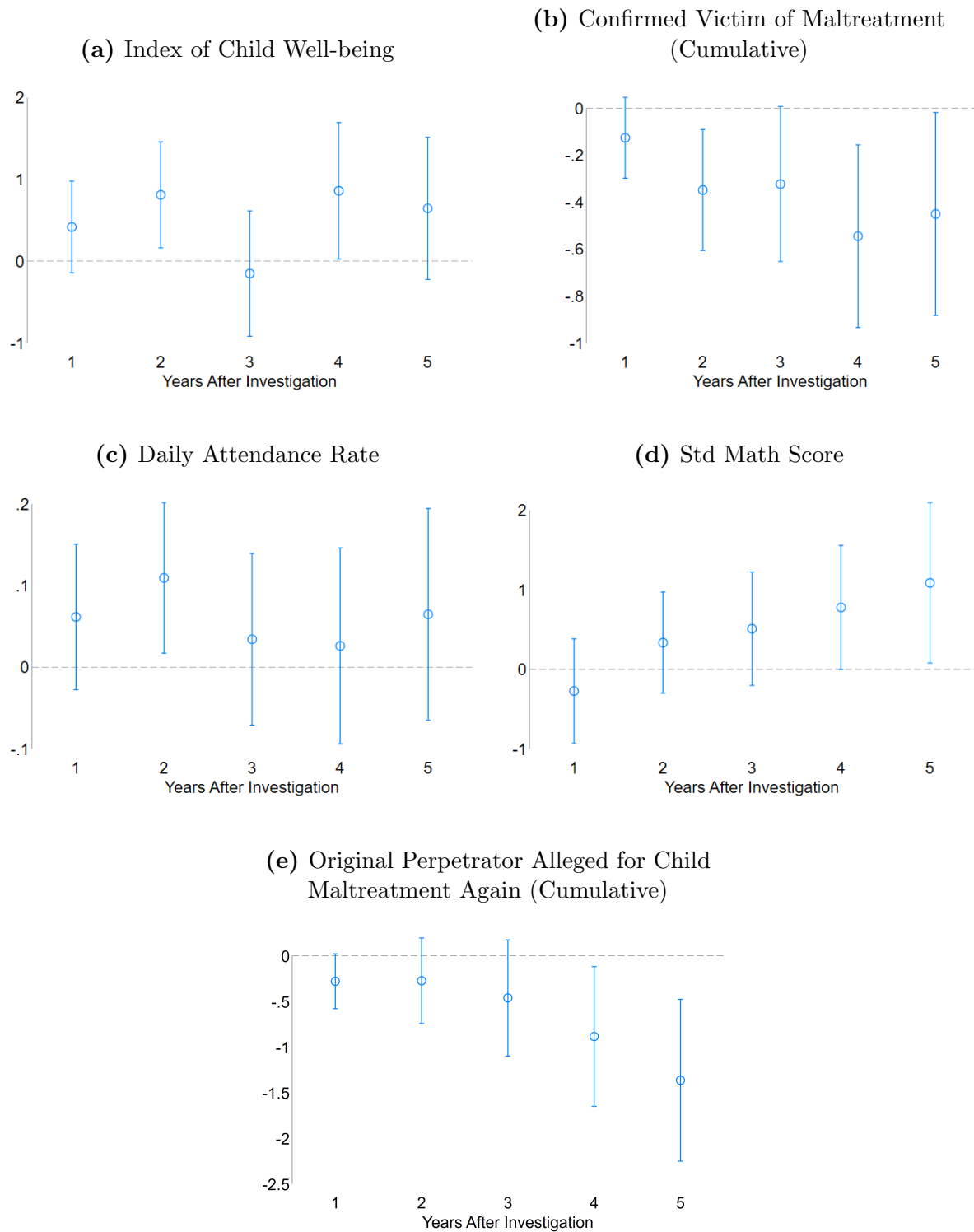


(b) 2017 Statistics



Notes. These figures show the rate of foster care entry per 1000 children for each state in the first and last year of available data as of March 2019. There are five different shades of blue representing the quantile of the foster care entry rate for each state, with darker shading indicating higher rates of entry. Eight states do not report the number of children who entered foster care in 1998 and are shaded in gray in Figure A1a. The 1998 information comes from [USDHHS \(2003b\)](#) and the 2017 information comes from [USDHHS \(2017b\)](#).

Figure A2: Effects of Foster Care on Child Outcomes Over Time for Balanced Sample



Notes. These figures report the results from 2SLS regressions of the outcome variable on foster care using removal stringency to instrument for foster care. They plot both the point estimates and their 95 percent confidence intervals. The sample is restricted to children in grades 1 through 7 with investigations between 2008 to 2012 so that we observe five follow-up years for all children. All specifications include the covariates as listed in the text, as well as zip code by investigation year fixed effects. Standard errors are clustered by child. Follow-up years are defined by school years even for non-schooling outcomes.

Table A1: Effects of Foster Care on Permanency Placements

	(1) Reunified	(2) Adopted	(3) Guardianship	(4) Emancipated	(5) Still in FC in Sep 2017
Foster Care	0.703*** (0.020)	0.064*** (0.011)	0.040*** (0.009)	0.017*** (0.006)	0.176*** (0.016)
% Conditional on Exiting Observations	85.3% 242,233	7.8% 242,233	4.9% 242,233	2.1% 242,233	2.1% 242,233

Notes. This table reports the results from 2SLS regressions of the outcome variable on foster care, using removal stringency to instrument for foster care. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. Each permanency outcome is mutually exclusive. Some students were still in the foster system at the end of the sample period in September 2017; these students are coded as such for their permanency outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Characteristics of Compliers at the Margin of Foster Placement

	(1) All	(2) Foster Care	(3) Compliers
Female	0.49	0.47	0.52
White	0.62	0.52	0.52
Student of Color	0.38	0.48	0.47
10 Years Old & Younger	0.55	0.51	0.61
11 Years Old & Older	0.45	0.49	0.39
Urban/Suburban County	0.64	0.63	0.63
Rural County	0.36	0.37	0.37
Low Income	0.83	0.87	0.89
Ever Retained in Grade	0.36	0.39	0.38
Above-Median Math Score	0.50	0.41	0.39
Above-Median Reading Score	0.50	0.42	0.38
Share of Sample	1.00	0.02	0.05

Notes. We follow [Dahl, Kostøl, and Mogstad \(2014\)](#) to calculate the share and characteristics of compliers. Specifically, we compute the share of compliers as the difference in the first-stage effect between children assigned to investigators with removal stringency at the 99th and the 1st percentiles. Then, we calculate the characteristics of compliers as the fraction of compliers across each characteristic subgroup. Above-median math and reading scores are indicators for scoring higher than the median child in the sample on baseline standardized math and reading tests.

Table A3: Effects of Foster Care on Children's Experience in Foster System

	(1) All	(2) Marginal Placements
<i>Initial Placement</i>		
With Relatives	0.582	0.572
With Unrelated Family	0.320	0.344
In Group Home	0.098	0.085
<i>Placement Stability</i>		
Number of Different Placements	3.121	3.085
One or Two Different Placements	0.441	0.512
Three or More Different Placements	0.559	0.488
Days in Foster System	619	581
<i>Permanency Outcomes</i>		
Reunified	0.666	0.703
Adopted	0.076	0.064
Guardianship	0.048	0.040
Emancipated	0.021	0.017
Still in Foster Care in Sep 2017	0.188	0.176
Observations	242,233	242,233

Notes. This table compares the experiences of the average foster placement and the marginal foster placement while in the foster system. Column 1 reports the mean outcome among all foster placements while Column 2 reports the results from 2SLS regressions of the outcome variable on foster care, using removal stringency to instrument for foster care. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. For initial placement details, group homes include institutions. Some students were still in the foster system at the end of the sample period in September 2017; these students are coded as such for their permanency outcome.

Table A4: Effects of Foster Care on Index of Child Well-being, by Age and Gender

	(1) Young	(2) Old	(3) Male	(4) Female	(5) Young Male	(6) Young Female	(7) Old Male	(8) Old Female
Foster Care	0.666*** (0.174)	-0.203 (0.285)	0.405* (0.222)	0.384** (0.194)	0.657*** (0.243)	0.284 (0.211)	0.043 (0.466)	0.562 (0.407)
P-value		0.003		0.935		0.168		0.331
Observations	133,476	108,757	123,715	118,518	70,438	63,038	53,277	55,480

Notes. This table reports the results from 2SLS regressions of the index of child well-being on foster care for a variety of subgroups, using removal stringency to instrument for foster care. The young subgroup includes children ages 10 and younger at the start of the child welfare investigation while the old subgroup includes children ages 11 and older. The p-value reports whether the subgroup estimates are statistically different from each other. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Testable Implications of the Exclusion of Removal Stringency Instrument

	(1) First Placed with Relatives	(2) First Placed with Unrelated Family	(3) First Placed in Group Home	(4) Days in Foster Care	(5) # Foster Homes
Removal Stringency	0.166 (0.465)	-0.021 (0.438)	-0.145 (0.277)	25.797 (607.685)	0.241 (3.283)
Joint P-Value	0.519				
Observations	4,809	4,809	4,809	4,809	4,809

Notes. This table reports the results from a regression of the dependent variable on the removal stringency instrument. The dependent variable in Columns 1 through 5 is conditional on foster placement. Standard errors are clustered by child. We find no evidence that the removal stringency instrument is jointly predictive of children's experiences in foster care using the outcomes in Columns 1 through 5. The p-value from an F-test for joint significance is 0.519.

Table A6: Robustness Checks

	Index of Child Well-being
<i>Panel A: Alternative Samples</i>	
Child's First Investigation (N=180,859)	0.339* (0.184)
Investigator Assigned ≥ 75 Investigations (N=232,818)	0.318* (0.169)
Balanced Panel (N=96,156)	0.520** (0.227)
<i>Panel B: Alternative Removal Stringency Instruments</i>	
Split Sample (N=242,233)	0.391** (0.188)
Leave-out Other Years (N=242,233)	0.228** (0.097)
Leave-out Same Year (N=242,233)	0.672* (0.353)
LASSO (N=242,233)	0.348*** (0.122)
UJIVE (N=242,233)	0.476*** (0.162)
<i>Panel C: Alternative Level of Rotational Assignment</i>	
County by Year (N=242,233)	0.562*** (0.171)

Notes. Panel A reports the results from 2SLS regressions using alternative sample definitions, Panel B uses alternative measures of removal stringency to instrument for foster care, and Panel C reports the results using the main stringency instrument but replaces zip code by investigation year fixed effects with county by investigation year fixed effects. All regressions include the covariates as listed in the text and, except for Panel C, zip code by investigation year fixed effects. Standard errors are clustered by child. In Panel A, the balanced panel sample is restricted to the first five follow-up years for children investigated in 7th grade or below in 2012 or earlier. In Panel B, the split sample measure is the removal rate of the assigned investigator from a random half of the sample. The leave-out other years measure is the leave-out removal rate of the assigned investigator from other children who had investigations in the same calendar year. The leave-out same year measure is the leave-out removal rate of the assigned investigator from other children who had investigations in different calendar years. For the LASSO approach—of the five potential instruments described in the main text—the algorithm selected instruments that vary based on minority status, allegation type, and perpetrator type. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Online Appendix

B.1 Supplemental Online Tables

Table B1: Effects of Foster Care on Michigan Public School Enrollment

	(1) Ever Enrolled After	(2) Enrolled One Year After	(3) Enrolled Two Years After	(4) Enrolled Three Years After	(5) Enrolled Four Years After	(6) Enrolled Five Years After
<i>Panel A: Children 6 Years Old and Younger During Investigation</i>						
Foster Care	-0.191*** (0.057)					
Observations	236,925					
<i>Panel B: Analysis Sample, Enrolled in Grades 1 to 11 During Investigation</i>						
Foster Care	-0.033 (0.035)	-0.017 (0.042)	0.002 (0.061)	-0.123 (0.082)	0.004 (0.102)	0.042 (0.121)
Observations	248,730	248,730	212,718	168,711	133,268	99,014

Notes. This table reports the results from 2SLS regressions of the outcome variable on foster care using removal stringency to instrument for foster care. Panel A consists of children ages 6 years old and younger at the time of their investigation and Panel B consists of children in the analysis sample—those enrolled in public school in grades 1 through 11 during the investigation. Only children eligible for school enrollment in a given year are included in the analysis. For example, a 3-year-old who was investigated in 2016 is not included in Panel A because the child was not eligible to enroll in a public school by 2017, the last year of available education data. Similarly, students in 11th grade during the investigation are not included in the analysis of enrollment three years later in Panel B. This explains why the sample size decreases with every follow-up year in Panel B. All regressions include zip code by investigation year fixed effects, Panel A also includes non-academic socio-demographic covariates, and Panel B further includes the full set of covariates as listed in the text. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Testable Implications of Monotonicity of the Removal Stringency Instrument

	(1) Female	(2) Male	(3) White	(4) Student of Color	(5) Age ≤ 10	(6) Age > 10	(7) Had Prior Inv	(8) No Prior Inv
<i>Panel A: Main Leave-One-Out Instrument</i>								
Removal Stringency	0.481*** (0.027)	0.422*** (0.026)	0.399*** (0.022)	0.515*** (0.032)	0.481*** (0.024)	0.411*** (0.029)	0.544*** (0.027)	0.323*** (0.026)
<i>Panel B: Leave-Subgroup-Out Instrument</i>								
Removal Stringency	0.365*** (0.023)	0.305*** (0.021)	0.161*** (0.013)	0.226*** (0.024)	0.195*** (0.017)	0.318*** (0.026)	0.269*** (0.022)	0.160*** (0.019)
Observations	118,436	123,715	149,527	92,706	133,476	108,757	142,034	100,199

Notes. Panel A reports the first-stage effect of removal stringency on foster placement separately by student subgroup. Panel B reports the first-stage effect using the leave-subgroup-out instrument. The leave-subgroup-out instrument is the fraction of an investigator's cases other than those in the same subgroup that resulted in foster placement. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Relationship Between Foster Care and Child Outcomes

	(1) Index of Child Well-being	(2) Alleged Victim of Maltreatment	(3) Confirmed Victim of Maltreatment	(4) Daily Attendance Rate	(5) Std Math Score	(6) Std Reading Score	(7) Juvenile Delinquency
Foster Care	-0.088*** (0.012)	-0.014*** (0.004)	-0.002 (0.002)	-0.003 (0.002)	-0.040*** (0.015)	-0.025 (0.016)	0.047*** (0.004)
Observations	242,233	242,233	242,233	224,925	177,118	177,084	134,076

Notes. The table reports the results of bivariate OLS regressions of the outcome variable on foster care placement. Standard errors are shown in parentheses and clustered by child. The education and crime outcomes do not include all of the observations in the sample. Specifically, some grade level and attendance records are missing and students may not have taken a standardized math or reading test if they were too young or old to be in grades 3–8, were absent from school on a test day, or were exempt. Furthermore, juvenile delinquency data are missing for eight counties, available only through 2015, and relevant only for children younger than Michigan’s age of majority of 16. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Effects of Foster Care on Taking Standardized Tests

	(1)	(2)
	Took Std Math Test	Took Std Reading Test
<i>Panel A: OLS</i>		
Foster Care	0.007* (0.004)	0.008* (0.004)
<i>Panel B: 2SLS</i>		
Foster Care	0.023 (0.063)	-0.025 (0.064)
Observations	189,084	189,084

Notes. Panel A reports the results from OLS regressions of the outcome variable on foster care while Panel B reports the results from 2SLS regressions using removal stringency to instrument for foster care. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. Students may not take standardized tests if they are absent from school during the testing dates or took an alternative state assessment for students who require special accommodations. Children who were too young or too old to have been in grades 3–8 after their investigation are also excluded from this analysis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Effects of Foster Care on High School Graduation and College Enrollment

	(1) Graduated High School	(2) Ever Enrolled in College	(3) Ever Enrolled in a Two-Year College	(4) Ever Enrolled in a Four-Year College
<i>Panel A: OLS</i>				
Foster Care	-0.024* (0.014)	0.001 (0.017)	-0.009 (0.015)	0.012 (0.013)
<i>Panel B: 2SLS</i>				
Foster Care	0.106 (0.296)	0.177 (0.392)	-0.016 (0.365)	0.024 (0.292)
Observations	60,776	36,661	36,661	36,661

Notes. Panel A reports the results from OLS regressions of the outcome variable on foster care while Panel B reports the results from 2SLS regressions using removal stringency to instrument for foster care. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. Only students expected to be in 12th grade by 2017 based on an on-time grade progression from the school year of their investigation are included in the analysis of high school graduation. The analysis of college enrollment is similarly restricted to students expected to be in 12th grade by 2016. Some colleges are missing information on their type, so the two and four-year college enrollment estimates need not add up to the overall college enrollment estimate. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Effects of Foster Care on Type of Foster Placement

	(1) Days in Foster Care	(2) Days in Kinship Care	(3) Days with Unrelated Family	(4) Days in Group Home
Foster Care	581*** (40)	345*** (24)	185*** (22)	50*** (16)
Observations	242,233	242,233	242,233	242,233

Notes. This table reports the results from 2SLS regressions of the outcome variable on foster care, using removal stringency to instrument for foster care. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B7: Effects of Foster Care on Neighborhood and School Environment Over Time

		Neighborhood			School	
	(1) Index of Neighborhood & School Characteristics	(2) Median Income (\$100,000)	(3) BA Degree or Higher	(4) Employment Rate	(5) Test Scores	(6) Low Income
<i>Panel A: One Year After Investigation</i>						
Foster Care	0.257** (0.100) {-0.147}	0.071* (0.037) {0.406}	0.084*** (0.026) {0.121}	0.021 (0.022) {0.848}	-0.003 (0.082) {-0.119}	-0.100** (0.039) {0.649}
<i>Panel B: Two+ Years After Investigation</i>						
Foster Care	0.066 (0.125) {-0.011}	0.055 (0.048) {0.411}	0.034 (0.033) {0.157}	-0.011 (0.026) {0.875}	0.086 (0.102) {-0.239}	-0.021 (0.049) {0.538}
Observations	242,233	209,446	209,446	209,446	217,956	241,267

Notes. This table reports the results from 2SLS regressions of the outcome variable on foster care using removal stringency to instrument for foster care. Panel A reports results for outcomes measured during the first school year after the investigation and Panel B reports results across all school years after the first. Standard errors are in parentheses and clustered by child. The curly brackets below the standard errors represent the control complier mean. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Neighborhoods are defined by census block groups. A child's school in each follow-up year is defined as the school where they spent the most time during the school year and their neighborhood is defined as where they lived while enrolled in that school. School test scores represent the average of standardized math and reading scores and low income represents the fraction of students in the school who qualify for free or reduced-price lunch. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B8: First-Stage Effects of Investigator Tendencies Over Removal and Family Services

Dependent Variable:	(1) Child Removal, Targeted Services, and Community Services	(2) Targeted Services and Community Services	(3) Community Services
Tendency Over Child Removal	0.365*** (0.023)	-0.297*** (0.042)	-0.462*** (0.054)
Tendency Over Targeted and Community Services	0.052*** (0.011)	0.790*** (0.023)	0.229*** (0.031)
Tendency Over Community Services	-0.001 (0.007)	0.032** (0.014)	0.666*** (0.020)
Observations	242,233	242,233	242,233
F-Statistic	208.810	1100.450	1293.020
Zip code by Year FE	✓	✓	✓
Socio-Demographic Controls	✓	✓	✓
Academic Controls	✓	✓	✓

Notes. This table reports the results from regressions of each of the three dependent variables (child removal plus targeted and community services, targeted and community services, and community services) on three instruments: investigator propensity to remove, investigator propensity to recommend both community-based and targeted services but without child removal, and investigator propensity to recommend community services alone. Socio-demographic controls include gender, race/ethnicity, indicators for grade in school, an indicator for whether the child was the subject of a prior investigation, and the number of prior investigations. Academic controls include an indicator for free or reduced-price lunch eligibility, an indicator for receipt of special education services, an indicator for ever expelled, and daily attendance rate—measured in the school year prior to the investigation—as well as the most recent pre-investigation score from standardized math and reading test scores. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B9: Balance Tests of the Conditional Random Assignment of Investigators

Dependent Variable:	(1) Child Removal, Targeted Services and Community Services	(2) Targeted Services and Community Services	(3) Community Services	(4) Tendency Over Child Removal	(5) Tendency Over Targeted and Community Services	(6) Tendency Over Community Services
<i>Panel A: Full Sample</i>						
F-Stat from Joint Test	21.517	79.401	10.489	1.123	1.381	1.199
P-Value from Joint Test	0.000	0.000	0.000	0.296	0.083	0.212
Observations	242,233	242,233	242,233	242,233	242,233	242,233
<i>Panel B: 4th Grade and Above</i>						
F-Stat from Joint Test	14.434	52.683	6.630	1.030	1.210	1.289
P-Value from Joint Test	0.000	0.000	0.000	0.421	0.205	0.140
Observations	144,032	144,032	144,032	144,032	144,032	144,032

Notes. This table reports the results from regressions of the dependent variable on a variety of socio-demographic and academic covariates as described in the main text, as well as zip code by investigation year fixed effects. Panel A includes the full sample of investigations and exclude standardized test scores in the vector of covariates. As students in Michigan begin taking standardized tests in grade 3, Panel B reports the results for students enrolled in at least grade 4 during the maltreatment investigation and includes standardized test scores. Standard errors are clustered by child.

Table B10: Robustness of the Main Results to Control Selection

	(1) Index of Child Well-being	(2) Index of Child Well-being	(3) Index of Child Well-being	(4) Index of Child Well-being	(5) Index of Child Well-being
Foster Care	0.290* (0.169)	0.308* (0.167)	0.360** (0.164)	0.388** (0.164)	0.392** (0.164)
<i>Baseline Controls</i>					
Grade 2	0.044*** (0.005)	0.039*** (0.005)	0.023*** (0.005)	0.021*** (0.005)	0.022*** (0.005)
Grade 3	0.033*** (0.005)	0.032*** (0.005)	0.013*** (0.005)	0.011** (0.005)	0.015*** (0.005)
Grade 4	-0.060*** (0.010)	-0.042*** (0.010)	-0.015 (0.009)	-0.013 (0.009)	-0.007 (0.009)
Grade 5	-0.098*** (0.010)	-0.076*** (0.010)	-0.049*** (0.010)	-0.048*** (0.010)	-0.039*** (0.010)
Grade 6	-0.130*** (0.010)	-0.107*** (0.010)	-0.081*** (0.010)	-0.080*** (0.010)	-0.070*** (0.010)
Grade 7	-0.177*** (0.011)	-0.152*** (0.011)	-0.123*** (0.011)	-0.126*** (0.011)	-0.112*** (0.011)
Grade 8	-0.198*** (0.012)	-0.174*** (0.012)	-0.142*** (0.011)	-0.146*** (0.012)	-0.129*** (0.012)
Grade 9	-0.184*** (0.012)	-0.159*** (0.012)	-0.114*** (0.012)	-0.119*** (0.012)	-0.100*** (0.012)
Grade 10	-0.127*** (0.014)	-0.105*** (0.014)	-0.069*** (0.014)	-0.080*** (0.014)	-0.058*** (0.014)
Grade 11	-0.048*** (0.018)	-0.025 (0.018)	0.004 (0.018)	-0.007 (0.018)	0.015 (0.018)
Std Math Score	0.133*** (0.003)	0.125*** (0.003)	0.127*** (0.004)	0.126*** (0.004)	0.124*** (0.004)
Std Reading Score	0.086*** (0.003)	0.076*** (0.003)	0.070*** (0.004)	0.068*** (0.004)	0.067*** (0.004)
Female	0.032*** (0.003)	0.034*** (0.003)	0.022*** (0.003)	0.021*** (0.003)	0.021*** (0.003)
White	-0.016 (0.177)	0.007 (3.606)	0.001 (0.027)	0.005 (0.151)	-0.019 (0.101)
Black	-0.099 (0.178)	-0.093 (3.605)	-0.080*** (0.028)	-0.049 (0.151)	-0.070 (0.101)
<i>Investigation Controls</i>					
# Prior Investigations		-0.069*** (0.001)	-0.060*** (0.001)	-0.059*** (0.001)	-0.056*** (0.001)
Allegation was for Physical Abuse		-0.004 (0.003)	-0.022*** (0.003)	-0.024*** (0.003)	-0.025*** (0.003)
Perpetrator was a Parent		-0.021*** (0.007)	-0.023*** (0.007)	-0.024*** (0.007)	-0.025*** (0.007)
<i>Prior Academic Characteristics</i>					
Attendance Rate			1.197*** (0.023)	1.160*** (0.023)	1.140*** (0.023)
Special Education			-0.086*** (0.004)	-0.088*** (0.004)	-0.087*** (0.004)
Ever Expelled			-0.207*** (0.063)	-0.203*** (0.062)	-0.200*** (0.063)
Free or Reduced Price Lunch			-0.132*** (0.004)	-0.119*** (0.004)	-0.107*** (0.004)

Std Math Score X Std Reading Score	0.007 (0.004)	0.007 (0.004)	0.007 (0.004)		
Std Math Score Squared	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)		
Std Reading Score Squared	-0.004** (0.002)	-0.004** (0.002)	-0.004** (0.002)		
Std Math Score Cubed	-0.007*** (0.001)	-0.007*** (0.001)	-0.007*** (0.001)		
Std Reading Score Cubed	-0.003*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)		
<i>School Controls</i>					
Urban		-0.028*** (0.005)	-0.027*** (0.005)		
Charter		0.098*** (0.006)	0.098*** (0.006)		
% White		-0.010 (0.020)	0.006 (0.021)		
% Black		-0.064*** (0.019)	-0.057*** (0.020)		
%Free or Reduced Price Lunch		-0.131*** (0.012)	-0.089*** (0.012)		
<i>Neighborhood Controls</i>					
# Neighborhoods Lived in Before Investigation			-0.009*** (0.001)		
Household Median Income			0.001*** (0.000)		
Employment Rate			0.016 (0.019)		
% Bachelor's Degree or Higher			0.094*** (0.018)		
% White			-0.003 (0.029)		
% Black			0.005 (0.030)		
Homeless in SY Before Investigation			-7.625*** (0.738)		
Observations	242,233	242,233	242,233	242,233	242,233
Rotation Group FE	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓
Investigation Controls		✓	✓	✓	✓
Academic Controls		✓	✓	✓	✓
School Controls				✓	✓
Neighborhood Controls					✓

Notes. The table shows the robustness of the 2SLS results shown in Table 4 to alternative selections of control variables. Column 1 includes only baseline controls including gender, race/ethnicity, grade-level fixed effects, and controls for a student's most recent baseline standardized math and reading test scores. Column 2 adds investigation controls including whether the allegation was for physical abuse or neglect, the child's relation to the perpetrator, and the number of prior investigations that the child was previously the subject of. Column 3 includes academic controls measured in the year before the investigation. Column 4 includes the characteristics of the school that the child attended during the investigation. Finally, Column 5 includes characteristics of the child's neighborhood. All columns include indicators for any missing covariates. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B11: Robustness of the Main Results to Alternative Clustering Decisions

	(1) Index of Child Well-being	(2) Alleged Victim of Maltreatment	(3) Confirmed Victim of Maltreatment	(4) Daily Attendance Rate	(5) Std Math Score	(6) Std Reading Score	(7) Juvenile Delinquency
<i>Panel A: Baseline (by child)</i>							
Foster Care	0.392** (0.164)	-0.132** (0.058)	-0.053* (0.028)	0.055** (0.026)	0.356* (0.203)	0.175 (0.219)	-0.028 (0.040)
<i>Panel B: By Investigator</i>							
Foster Care	0.392** (0.173)	-0.132** (0.066)	-0.053* (0.031)	0.055** (0.027)	0.356* (0.204)	0.175 (0.217)	-0.028 (0.038)
<i>Panel C: By Rotation</i>							
Foster Care	0.392** (0.183)	-0.132* (0.069)	-0.053 (0.033)	0.055* (0.030)	0.356* (0.210)	0.175 (0.229)	-0.028 (0.043)
<i>Panel D: By Child and Investigator</i>							
Foster Care	0.392** (0.173)	-0.132** (0.066)	-0.053* (0.031)	0.055** (0.027)	0.356* (0.204)	0.175 (0.218)	-0.028 (0.038)
<i>Panel E: By Child and Rotation</i>							
Foster Care	0.392** (0.184)	-0.132* (0.070)	-0.053 (0.033)	0.055* (0.030)	0.356* (0.211)	0.175 (0.230)	-0.028 (0.043)
Observations	242,233	242,233	242,233	224,925	177,118	177,084	134,076

Notes. The table shows the robustness of 2SLS results shown in Table 4 to alternative clustering levels. Panel A shows the baseline results in which standard errors are clustered by child. Panels B and C show standard errors clustered at the investigator and rotation levels, respectively. Finally, Panels D and E show standard errors two-way clustered by child and investigator and by child and rotations, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.2 Censored Data and the Examiner Assignment Research Design

The examiner assignment research design used in this study has been widely applied recently as increased access to large administrative datasets allows researchers to exploit discretionary decision-making. It has been used to study a variety of interventions other than foster care, such as juvenile incarceration (Aizer and Doyle, 2015; Eren and Mocan, 2017), adult incarceration (Kling, 2006; Mueller-Smith, 2015), disability insurance (Dahl et al., 2014), student loan repayment (Herbst, 2018), and evictions (Collinson and Reed, 2019; Humphries et al., 2019), among others. In many of these settings, treatment assignment is a two-step selection process in which individuals are assigned to treatment only after crossing an initial decision threshold. For example, in the context of foster care, children can only be removed if their maltreatment allegation is first substantiated.⁵⁵ Similarly, in the criminal justice setting, defendants can only be incarcerated conditional on being convicted. Whether due to restrictions from data partners or privacy considerations, some studies apply this design using partially censored data that contain only individuals who cross the initial decision threshold—for example, only substantiated investigations or only convicted defendants. Such restrictions appear in two recent studies of foster care (Bald et al., 2019; Roberts, 2019) as well as in other contexts (Kling, 2006; Eren and Mocan, 2017; Herbst, 2018), and may introduce bias.

B.2.1 Potential Bias

To understand the source of potential bias, consider decisions made by investigators in the context of foster care. Substantiation decisions are based on the strength of the evidence, whereas placement decisions are based on the child’s risk of future harm.⁵⁶ The research design assumes that, due to random assignment, the distribution of risk is identical across investigators and, therefore, identifies impacts using exogenous variation in investigator tolerance over risk. However, if investigators also vary in their stringency over evidence, the set of substantiated cases may not be balanced across investigators. Therefore, restricted data access can create a violation of the exclusion restriction.

In addition to the usual instrumental variables assumptions of relevance, exogeneity, exclusion, and monotonicity, at least one additional assumption must be satisfied for the examiner assignment design to produce unbiased estimates from censored data (Arteaga, 2019). Either investigators must not vary over substantiation—that is, investigators always agree over evidence—or the investigator’s substantiation decision must be uncorrelated with the child’s potential outcomes. The former assumption is at odds with the motivation of the research design, given that the design hinges upon variation in investigator tendencies. Moreover, at least in Michigan, there is a large amount of variation in substantiation tendencies.⁵⁷ The latter assumption is also very strong: it would be surprising if the

⁵⁵The decision-making process for child welfare investigators in Michigan is the same as in South Carolina (Roberts, 2019) and Rhode Island (Bald et al., 2019); random assignment occurs *before* the substantiation decision is made, and substantiation is decided by the same investigator.

⁵⁶These two decisions may be correlated, yet they are distinct margins. For example, there can be clear evidence for an allegation when the child faces little risk of future harm, or less clear evidence in a higher risk scenario.

⁵⁷Investigators at the 10th percentile substantiated at a rate 8.4 percentage points less than the average

substantiation decision—which is based on how much evidence there is that the reported maltreatment actually occurred—was unrelated to children’s potential outcomes.

B.2.2 Replication with Censored Data

Although this is not the first study to describe the potential for bias from censored data, it is the first to shed light on how much it can matter in practice. Using data containing the universe of child welfare investigations in Michigan, including both unsubstantiated and substantiated allegations, we replicate the main analysis as if we only had access to substantiated cases. Using only the sample of substantiated investigations, we reconstruct the removal instrument according to Equation 1. A standard balance test reveals that a variety of baseline characteristics which are associated with foster care placement are not jointly predictive of the new instrument (Table B12). Therefore, since exogeneity appears to hold using the subset of substantiated investigations, one might expect the 2SLS results to be consistent with the full sample.⁵⁸ However, this turns out to not be the case.

Table B14 shows that the effects using the complete data (Panel A) are much larger than those found when restricted to substantiated investigations (Panel B).^{59 60} The replication exercise produces a substantively smaller impact on the index of child well-being. The effect on daily attendance rate is moderately smaller than the effect using the complete data but still statistically significant, whereas the point estimate on math test scores is much smaller and imprecise.⁶¹ The findings in Panel B of Table B14 are somewhat similar to those in Bald et al. (2019), which finds noisy estimates for school-age children, and to Roberts (2019) which reports imprecise estimates on test scores but positive effects for on-time grade progression. Although institutional differences between the child welfare systems in Michigan, Rhode Island, and South Carolina surely contribute to the different findings, this exercise documents that bias in the other studies may also play a role. Overall, this exercise cautions against applying the examiner assignment design with censored data.

investigator in their local area while investigators at the 90th percentile did so at a rate 8.9 percentage points greater.

⁵⁸When focusing on the sample of fourth grade students and older and including baseline standardized test scores, however, the censored instrument does not pass a standard balance test. In comparison, Roberts (2019) passes a balance test that includes baseline test scores, while Bald et al. (2019) rejects statistical significance at the one percent level in a joint balance test for school-age girls, but passes the balance test for school-age boys.

⁵⁹Table B13 shows that there exists a strong first stage relationship with the censored instrument. In addition, it is possible that Panel B in Table B14 represents a different LATE than Panel A. To address this potential concern, we use investigator tendencies over substantiation and removal to instrument for both foster placement and substantiation. Table B15 shows that the estimates in Panel B are also smaller than the causal effects of placement relative to substantiation from the complete data.

⁶⁰Table B16 shows that the OLS estimates are very similar from both the complete data and when restricted to substantiated investigations, however.

⁶¹Interestingly, the standard errors on point estimates from the censored sample are much smaller despite this sample containing a fewer number of observations. This is likely due to the fact that the censored sample includes only the subset of substantiated investigations, thus zooming in on the cases most likely to lead to foster placement. Even though there are considerably fewer observations, this analysis contains much less residual variation since it excludes students who contribute little to no identifying variation in the main analysis.

B.2.3 Assessing [Arteaga \(2019\)](#) Approaches to Using Examiner Assignment Design with Censored Data

What can researchers do when limited to using censored data? [Arteaga \(2019\)](#) proposes a reasonable solution in a study of the effects of parental incarceration on child outcomes. The study uses data from SISBEN, Colombia’s census of its low-income population, to link children to parents and parents to both criminal convictions and incarceration. SISBEN does not include information on parents who appeared before a court but were not convicted, however. Fortunately, anonymized records containing the universe of both conviction and incarceration decisions are publicly available for every judge in Colombia, which the study uses to create the judge instrument. Importantly though, these anonymized records can only be matched to SISBEN along the judge field and not to individual parents. Therefore, though the study accesses complete information about judge tendencies, it does not observe the full population of criminal defendants.

[Arteaga \(2019\)](#) shows how the standard examiner assignment design can not be applied in this context and derives an estimator of the causal effects of incarceration relative to conviction that can be identified using censored data.⁶² The key insight is that there is exogenous variation in incarceration among judges with identical conviction thresholds but different incarceration thresholds. In the context of this study, the variation in removal is as good as random for a given evidence threshold. More formally, the study proposes that the causal effects of removal relative to substantiation can be identified from censored data as:

$$\int_0^1 \frac{\delta \mathbb{E}[Y \cdot \mathbf{1}(T \in \{t_S, t_R\}) | P_S(Z) = p_S, P_R^*(Z) = p_R^*]}{\delta p_R^*} dp_R^* \quad (8)$$

where Y is a child outcome and T denotes treatment assignment: substantiated but not removed (t_S) or substantiated and removed (t_R).⁶³ $P_S(Z) = p_S$ represents that the evidence threshold to substantiate is held fixed at p_S and $P_R^*(Z) = p_R^*$ means that the removal threshold conditional on substantiation is equal to p_R^* . Integrating over the inside term averages the effect across all investigators.

In practice, the study derives P_S and P_R^* from the data as the leave-out measure of evidence stringency and the leave-out measure of removal conditional on substantiation respectively. Therefore, identification hinges on fixing the conviction threshold. Although [Arteaga \(2019\)](#) proposes three complementary strategies to do so, the study itself only has access to censored data and thus can not empirically assess whether these strategies actually produce unbiased estimates. Using the universe of maltreatment investigations, we compare estimates from each approach with those from the full, uncensored data.

The first, called the pooled approach, uses P_R^* to instrument for foster care while additionally controlling for linear and quadratic terms of P_S and all interactions. The second, called the tercile approach, instruments for placement with P_R^* separately for each tercile of the evidence stringency distribution. The idea is that, in addition to controlling for evidence stringency, splitting the data into terciles approximates fixing the evidence threshold. Lastly, the third approach, called the rolling window approach, mirrors the tercile approach yet estimates

⁶²This is a somewhat special context of the censoring issue given that the study has access to the universe of court records, even though they cannot be linked to parents in the SISBEN.

⁶³This is equivalent to Equation 13 in [Arteaga \(2019\)](#).

impacts more flexibly along the distribution of evidence stringency. Specifically, it sorts the sample by the evidence stringency of the assigned investigator and estimates impacts of placement for the lowest 18,000 observations of the distribution. Then it repeats this process for the lowest 500 to 18,500, and so on.

Table B17 shows the results of the first two approaches and Figure B1 shows the results from the third. As a benchmark, both the table and figure also include estimates of foster care relative to substantiation identified from the full, uncensored data. To identify this parameter, we use measures of investigator removal and substantiation stringency to simultaneously instrument for both foster placement and substantiation. The table and figure show the effects on the index of child well-being.

The approaches with censored data do not approximate the estimates from the full data especially well. With censored data, the pooled approach finds a small and statistically insignificant effect of foster care relative to substantiation, whereas the effect with full data reveals a large and statistically significant increase. Similarly, the point estimates using the full data are larger with the tercile approach, though they vary in precision. Furthermore, when using the rolling window approach, the censored data reveal a positive relationship between evidence stringency and the index of child outcomes, whereas the full data point toward the relationship being somewhat U-shaped.

Overall, estimates using these approaches are biased in the same direction as shown when using the standard examiner assignment design with censored data in Section B.2.1—they understate the benefits of foster care. Although beyond the scope of this paper, these approaches may create bias because the estimator is only valid at a given evidence threshold, yet each of these approaches uses a large window around an evidence threshold for identification. Future work may consider applying insights from recent advances in optimal bandwidth selection in the regression discontinuity context to better address the tradeoff between bias and variance when fixing the evidence threshold.

Table B12: Balance Tests Using Censored Data

	All Substantiated Investigations		4th Grade and Above	
	(1)	(2)	(3)	(4)
	Foster Care	Censored Removal Stringency	Foster Care	Censored Removal Stringency
F-Statistic from Joint Test	22.241	1.071	12.475	2.252
P-Value from Joint Test	0.000	0.369	0.000	0.000
Observations	47,469	47,469	27,036	27,036

Notes. This table reports the results from regressions of the dependent variable on a variety of socio-demographic and academic covariates as well as zip code by investigation year fixed effects. The censored removal stringency instrument is explained in detail in Section B.2.1. Columns 1 and 2 include the all substantiated investigations and exclude standardized test scores in the vector of covariates. As students in Michigan begin taking statewide standardized tests in grade 3, Columns 3 and 4 report results for students with a substantiated investigation who were enrolled in at least grade 4 during the maltreatment investigation and include standardized test scores. Full regression results are available upon request. Standard errors are clustered by child.

Table B13: First Stage Effect of Censored Removal Stringency on Foster Placement

	(1) Foster Care	(2) Foster Care	(3) Foster Care	(4) Foster Care
Censored Removal Stringency	0.592*** (0.019)	0.512*** (0.023)	0.508*** (0.023)	0.506*** (0.023)
Observations	47,469	47,469	47,469	47,469
F-Statistic	991.246	484.988	484.541	482.837
Zip code by Year FE		✓	✓	✓
Socio-Demographic Controls			✓	✓
Academic Controls				✓

Notes. This table reports the results from regressions of foster placement on the censored measure of removal stringency. The censored removal stringency instrument is explained in detail in Section B.2.1. Each column includes a different set of covariates. Socio-demographic controls include gender, race/ethnicity, indicators for grade in school, an indicator for a prior investigation, and the number of prior investigations. Academic controls include an indicator for free or reduced-price lunch eligibility, an indicator for receipt of special education services, an indicator for ever expelled, and daily attendance rate—measured in the school year prior to the investigation—as well as the most recent pre-investigation score from standardized math and reading test scores. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B14: Effects of Foster Care on Child Outcomes Using Censored Data

	(1) Index of Child Well-being	(2) Confirmed Victim of Maltreatment	(3) Daily Attendance Rate	(4) Std Math Score
<i>Panel A: Complete Data, Unsubstantiated and Substantiated</i>				
Foster Care	0.392** (0.164)	-0.053* (0.028)	0.055** (0.026)	0.356* (0.203)
Observations	242,233	242,233	224,925	177,118
<i>Panel B: Censored Data, Only Substantiated</i>				
Foster Care	0.154* (0.087)	-0.009 (0.016)	0.039*** (0.014)	0.062 (0.105)
Size of Bias	0.238	0.044	0.016	0.294
Observations	47,469	47,469	43,839	35,322

Notes. Panel A reports the 2SLS results from Table 4 while Panel B reports the results from 2SLS regressions of the outcome variable on foster care using censored removal stringency to instrument for foster care. The sample in Panel B is restricted to only substantiated investigations. The size of the bias represents the absolute value of the difference between the point estimate in Panel A (the effect using the complete data) and Panel B (the biased effect). All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B15: Effects of Foster Care Relative to Substantiation Without Removal

	(1) Index of Child Well-being	(2) Confirmed Victim of Maltreatment	(3) Daily Attendance Rate	(4) Std Math Score
Foster Care and Substantiated	0.426* (0.230)	-0.067* (0.040)	0.072** (0.036)	0.531* (0.287)
Substantiated	-0.015 (0.060)	0.006 (0.011)	-0.007 (0.009)	-0.081 (0.074)
Observations	242,233	242,233	224,925	177,118

Notes. This table reports the results from 2SLS regressions of the outcome variable on two treatment conditions: substantiation and foster care plus substantiation. It uses investigator stringency in evidence and risk levels to simultaneously instrument for the independent variables respectively. Specifically, we create an instrument for an investigator's propensity to substantiate (Z^{SUB}). Together with the main removal stringency measure (Z^{FC}), we use this new measure to simultaneously instrument for substantiation and foster care placement according to the following two first-stage equations: (1) $FC_{iw} = \gamma_1 Z_{iw}^{FC} + \gamma_2 Z_{iw}^{SUB} + \gamma_3 X_{iw} + \kappa_r + \mu_{iw}$, (2) $SUB_{iw} = \alpha_1 Z_{iw}^{FC} + \alpha_2 Z_{iw}^{SUB} + \alpha_3 X_{iw} + \chi_r + \nu_{iw}$, and one second-stage equation: $Y_{iw} = \beta_1 \hat{FC}_{iw} + \beta_2 \hat{SUB}_{iw} + \beta_3 X_{iw} + \Pi_r + \xi_{iw}$. Here, FC_{iw} is a binary variable equal to one if the child was removed. Similarly, SUB_{iw} is a binary indicator equal to one if the investigation was substantiated. By construction, FC_{iw} can only equal one whenever SUB_{iw} is equal to one, so that β_1 represents the additional impact of foster placement relative to substantiation without removal, while β_2 represents the impact of substantiation without removal. The table shows estimates of β_1 and β_2 . Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B16: OLS Effects of Foster Care on Child Outcomes Using Censored Data

	(1) Index of Child Well-being	(2) Confirmed Victim of Maltreatment	(3) Daily Attendance Rate	(4) Std Math Score
<i>Panel A: Complete Data, Unsubstantiated and Substantiated</i>				
Foster Care	0.026** (0.011)	-0.007*** (0.002)	0.011*** (0.002)	0.057*** (0.013)
Observations	242,233	242,233	224,925	177,118
<i>Panel B: Censored Data, Only Substantiated</i>				
Foster Care	0.030** (0.012)	-0.014*** (0.002)	0.010*** (0.002)	0.042*** (0.015)
Observations	47,469	47,469	43,839	35,322

Notes. Panel A reports the OLS results from Table 4 while Panel B reports the results from OLS regressions of the outcome variable on foster care using only the sample of substantiated investigations. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

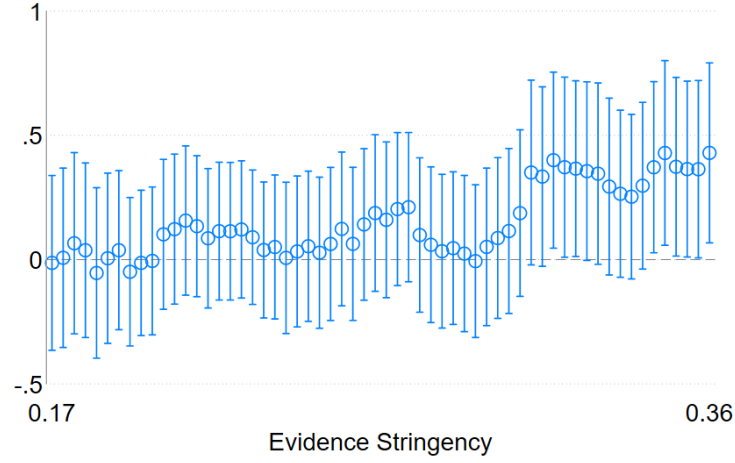
Table B17: Assessing [Arteaga \(2019\)](#) Approaches to Examiner Assignment Design with Censored Data

	Tercile Approach			
	(1) Pooled Approach	(2) Lenient in Evidence	(3) Middle in Evidence	(4) Strict in Evidence
<i>Panel A: Censored Data</i>				
Foster Care	0.087 (0.091)	-0.138 (0.198)	0.098 (0.183)	0.365* (0.202)
Observations	47,470	15,823	15,823	15,824
<i>Panel B: Full Data</i>				
Foster Care	0.408** (0.187)	0.583 (0.682)	0.159 (0.382)	0.793*** (0.290)
Observations	242,233	80,744	80,744	80,745

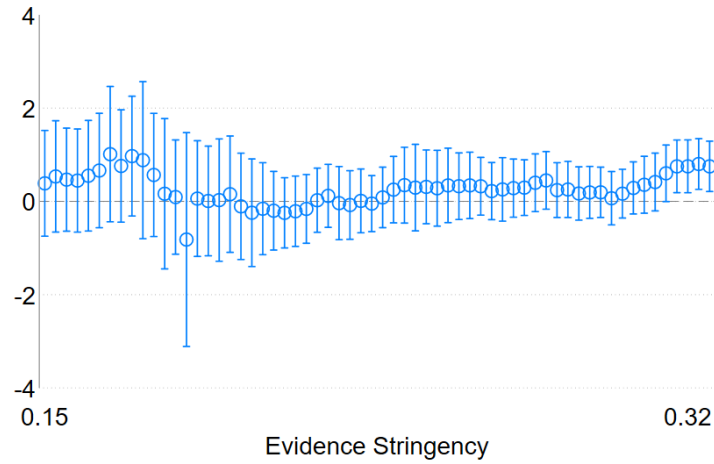
Notes. This table compares the estimates of foster care relative to substantiation on the index of child well-being using approaches proposed in [Arteaga \(2019\)](#). All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. Panel A applies the approaches to censored data, restricted to only children with substantiated maltreatment reports. In Panel A, investigators who were lenient in evidence substantiated between 0–21 percent of reports, whereas those in the middle and strict categories substantiated between 21–28 percent and 28–67 percent, respectively. Panel B applies the approaches to the full, uncensored data. We use removal stringency to instrument for foster care and evidence stringency to instrument for substantiation. In Panel B, investigators who were lenient in evidence substantiated between 0–18 percent of reports, whereas those in the middle and strict categories substantiated between 18–25 percent and 25–69 percent, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure B1: Assessing [Arteaga \(2019\)](#) Rolling Window Approach to Examiner Assignment Design with Censored Data

(a) Censored Data



(b) Full Data



Notes. This figure compares the estimates of foster care relative to substantiation on the index of child well-being using the rolling window approach proposed in [Arteaga \(2019\)](#) with both the censored and full data. The graphs plot both the point estimates and their 95 percent confidence intervals. All specifications include the covariates as listed in the text as well as zip code by investigation year fixed effects. Standard errors are clustered by child. Figure [B1a](#) sorts the censored data based on evidence stringency and estimates separate regressions of the index of child outcomes on foster care using removal stringency conditional on substantiation to instrument for foster care and including evidence stringency as a covariate. Since the sample size is similar to our study, we follow [Arteaga \(2019\)](#) in using a rolling window of 18,000 observations and adjust the window by 500 observations each time along the evidence threshold. Figure [B1b](#) applies the same approach to the full, uncensored data. We use removal stringency to instrument for foster care and evidence stringency to instrument for substantiation to estimate the effect of foster care relative to substantiation. Since the sample size is about five times larger with the full data, we use a rolling window of 90,000 observations and adjust the window by 2,500 observations each time.

B.3 Comparison to Doyle (2007, 2008)

Pathbreaking research in Doyle (2007) using administrative data from Illinois found that foster placement greatly harmed children’s outcomes. It reduced quarterly earnings as an adult (ages 18 to 28) by about \$1,300, increased teenage pregnancy by two times, and increased juvenile delinquency by three times. Follow-up work in Doyle (2008) also found that placement increased adult criminality by three times.⁶⁴ Using the same research design, we find that placement had a protective effect, improving children’s safety and educational outcomes. As discussed in Section 6, we can statistically reject that foster placement in Michigan caused the large harmful impacts found in the early work. Moreover, using a one-sided hypothesis test, we can rule out altogether that placement reduced the index of child well-being.

There are several reasons why the results in this study starkly contrast the findings in Doyle (2007, 2008), which broadly fit into four categories: (1) State-level differences in foster care placements, (2) national changes to foster care over time, (3) differences in sample definition, and (4) differences in the marginal placement. The rest of this section describes each in detail.

1. State-level differences in foster care placements. Foster placements were considerably longer and less stable in Illinois during the Doyle (2007, 2008) sample period than in Michigan more recently. For example, the median duration of foster care in Illinois during the early period was 40 months, compared to just 15.8 months in Michigan in 2008 (the first year of our sample period) and 12.8 months in 2017 (the final year of our panel) (Wulczyn et al., 2000; USDHHS, 2003a, 2017b).⁶⁵ Similarly, 44.8 percent of foster children in Illinois in 1998 had lived in three or more different foster homes compared to an average of just 31 percent across our 10-year panel in Michigan (USDHHS, 2003a; AECF, 2017). Similar trends hold among marginal placements as well; Illinois children at the margin of placement spent an average of four to five years in foster care, relative to 19 months in our context. Thus, the difference in findings across settings may largely be explained by these tremendous institutional differences.

In terms of external validity, it is also worth noting that foster care in Illinois during the early period was dramatically longer and less stable than in other states at the time (Figure 1). For example, among children’s first spell, the median duration of placement over the decade from 1988 to 1998 was nearly four-times longer than the average across 11 other states that had high-quality administrative data (39.4 versus 10.0 months) (Wulczyn et al. (2000), Figure 4.2). This was not driven by a few children with especially long stays; the median duration in Illinois was considerably longer than the 11 other states at every quartile

⁶⁴These studies examine slightly different samples. Specifically, analysis of juvenile delinquency in Doyle (2007) is limited to Cook County (home to Chicago), whereas Doyle (2008) notes that the data outside of Cook County are of higher quality for the analysis of adult criminality. They also cover slightly different years; Doyle (2007) examines children investigated between 1990 and 2001, whereas Doyle (2008) includes investigations through 2003. Lastly, Doyle (2007) includes children ages 5–15, whereas Doyle (2008) examines children ages 4–16. Both focus exclusively on children who had received Medicaid before their investigation.

⁶⁵Due to changes in reporting over time, the statistics for Illinois include all children who first entered foster care between 1988 and 1998, whereas those for Michigan include the average among children in foster care at the end of each fiscal year.

of the length distribution (Wulczyn et al. (2000), Figure 4.1).⁶⁶ To offer more evidence, just over one in four children who entered foster care in Illinois between 1988 and 1995 were still in the foster care system as of December 1998, compared to an average of less than one in 10 across 10 other states with reliable data (Wulczyn et al. (2000), Figure 5.1). Placement in Illinois was also less stable than in other states; in 1998, Illinois had the third-highest share of foster children who lived in three or more different foster homes among 41 states with quality data (USDHHS, 2003a).

In contrast, foster care in Michigan looked much more similar to other states during the years studied in this paper. For example, in 2015, median duration in foster care was 13.6 months compared to a national median of about 12 months (USDHHS, 2016b). Similarly, 31 percent of foster children lived in three or more different foster homes compared to the national average of 35 percent. For these reasons, our analysis is likely more generalizable to the rest of the country than the findings in Doyle (2007, 2008) were at the time they were published.

2. National changes to foster care over time. There have been substantive changes to child welfare practice since the period studied in Doyle (2007, 2008) that may have improved foster care across the country. One such legislative change is the Adoption and Safe Families Act of 1997 which sought to reduce the length of foster placements by requiring that states terminate parental rights for children who had been in the system for 15 out of 22 consecutive months (with some exceptions, such as children placed in kinship care). Accordingly, the proportion of children in foster care with short stays (between one and two years) increased from 18 percent to 30 percent from 1998 to 2017 (ChildTrends, 2018). There has also been a cultural push toward kinship placements since the end of the early sample period. For example, 28 percent of foster children were placed with relatives in 1998; this declined to 24 percent between 2001 and 2003, the final years of the Doyle (2007, 2008) sample period. This proportion had risen to 32 percent in 2017, the final year of the panel in this study.

These shifts over time reflect changes in what the field believes is best for abused and neglected children, though there is little credible research on the efficacy of reducing placement length and/or placing children with relatives. To the extent that child welfare practice has improved over time, these national trends might contribute to the differences in findings between Doyle (2007, 2008) and this study.

3. Differences in the sample definition. The sample in Doyle (2007) included children ages 5–15 who had received Medicaid before their investigation. To assess whether these sample restrictions could have driven the differences in findings, we restrict our analysis to children ages 5–15 who were eligible for free or reduced-price lunch in any school year prior to the investigation. We find estimates of foster care placement very similar to our main analysis (Table B18).⁶⁷ Moreover, using a one-sided hypothesis test, we can statistically reject that placement worsened the index of child well-being. Therefore, differences in sample

⁶⁶Specifically, the median duration was 4.5 times longer at the 25th percentile than the average of the 11 other states, 4.0 times longer at the 50th percentile, and 2.5 times longer at the 75th percentile (Wulczyn et al. (2000), Figure 4.1).

⁶⁷The sample in Doyle (2008) includes children ages 4–16. The results do not substantively change when we add in 16-year-olds, though we do not include 4-year-olds because of differential enrollment in public schools discussed in Section 3.3.

definition do not appear to contribute to the differences in findings.

4. Differences in marginal placements. The examiner assignment research design identifies the impact of foster care for children at the margin of placement. That is, children for whom investigators might disagree over whether placement is appropriate. To address whether there were substantive differences in marginal placements across settings, we first compare the observable characteristics of the complier populations.

Compliers in [Doyle \(2008\)](#) were older than in our study. Specifically, they were 45 percent more likely to be ages 11 to 13 than the overall sample. In contrast, compliers in our setting were 11 percent more likely to be age 10 or below.⁶⁸ For this difference in the complier population to translate into differences in findings, there must also be heterogeneous impacts by age. We find that the benefits of foster care were largest for younger children (Table A4), whereas [Doyle \(2007\)](#) finds the harm was greatest for older youth. This pattern is consistent with the differences in results. However, [Doyle \(2008\)](#) finds similar results for children older and younger than age 10, so it is unclear whether heterogeneity by age drives the divergent findings.

We find less evidence that other observable complier characteristics contribute to the differences in findings. For example, compliers were more likely to be female in [Doyle \(2008\)](#) than in our setting (66 percent versus 52 percent). Although [Doyle \(2007, 2008\)](#) found that the impact of placement on juvenile and adult crime was more negative for female children, we find similar benefits of placement for male and female children. There are smaller differences between compliers along race/ethnicity (40 percent of compliers were African American in [Doyle \(2008\)](#) versus 47 percent who are students of color in our setting, comprising mostly African American students but also Latinx, Native American and other underrepresented minority students), and [Doyle \(2008\)](#) finds that placement had similar impacts for White and African American children.

Although examining complier characteristics permits a direct comparison of children at the margin of placement along some dimensions, it may be less informative about the underlying risk that marginal children face across studies. To address this, we turn to the overall placement rates in each setting. The intuition is that, all else being equal, we would expect foster placement to be more beneficial for children at the margin in places with lower overall placement rates since they face more risk in the home. Similarly, we would expect placement to be more harmful for children at the margin in places with higher overall placement rates since they face less risk in the home. Therefore, to the extent that the share of children who are at-risk is similar across settings, comparing the overall placement rate in Illinois during the early sample period to that of Michigan more recently informs us about the risk of marginal children.

About 2.5 per 1000 children in Illinois entered foster care in 1990, the first year of the early studies ([Wulczyn et al., 2000](#)). This rose by 76 percent over the next four years such that in 1994, 4.4 per 1000 children entered foster care, and declined to around 2 per 1000 in 2001, the final year of the [Doyle \(2007\)](#) sample period ([USDHHS, 2003a](#)).⁶⁹ In comparison, the placement rate in Michigan remained around 3 per 1000 children during the sample

⁶⁸Section V.C in [Doyle \(2008\)](#) describes the complier population, whereas [Doyle \(2007\)](#) does not include complier characteristics. Table A2 reports the characteristics of compliers for our study.

⁶⁹Illinois placed 1.79 per 1000 children in 2003, the final year of the [Doyle \(2008\)](#) sample period.

period in this study, from 3.4 per 1000 in 2008 to 3.0 per 1000 in 2016 ([USDHHS, 2008, 2016b](#)). Therefore, at its peak in 1994, the placement rate in Illinois was 25 percent higher than the highest rate in Michigan during our sample period (3.51 in 2010). If marginal children in [Doyle \(2007, 2008\)](#) were primarily investigated in the mid-1990s, compliers who were placed in foster care may have faced considerably less risk in the home than those in this study, which could explain the contrast in findings. However, this may not hold if a large share of marginal children were investigated earlier or later in the sample period. Therefore, it is unclear whether the differences in findings across studies can be attributed to differences in the risk that marginal children faced in the home.

Overall, the most likely reason for the contrast in findings between our study and [Doyle \(2007, 2008\)](#) appears to be the tremendous difference in what foster care placement looked like for children during the two study periods. Children at the margin of placement in Illinois spent nearly 2.5 years longer in foster care than in Michigan—the “treatment conditions” were fundamentally different across studies. It is also possible that both national legislative and cultural changes to child welfare over time, like shorter stays and increased placement with relatives, improved foster systems across the country. We find less evidence that differences in the marginal placement across studies—and no evidence that differences in sample composition—play a role in explaining the stark contrast in findings.

Table B18: Effects of Foster Care on Child Outcomes for Sample Comparable to [Doyle \(2007\)](#)

	(1) Index of Child Well-being	(2) Alleged Victim of Maltreatment	(3) Confirmed Victim of Maltreatment	(4) Daily Attendance Rate	(5) Std Math Score	(6) Std Reading Score	(7) Juvenile Delinquency
Foster Care	0.428** (0.167)	-0.161*** (0.059)	-0.071** (0.029)	0.057** (0.026)	0.373* (0.198)	0.146 (0.216)	-0.020 (0.042)
One-Sided P-Value	0.005	0.003	0.007	0.015	0.030	0.250	0.319
Observations	204,909	204,909	204,909	190,620	156,834	156,802	117,270

Notes. This table reports the results from 2SLS regressions of foster care on the dependent variable, using removal stringency to instrument for foster care. The analysis sample is restricted to children between the ages of 5 and 15 during their investigation who were ever eligible for free or reduced-price lunch prior to the investigation. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.4 Data Appendix

We use administrative data from the Michigan Department of Health and Human Services (MDHHS), Michigan Department of Education (MDE), Center for Educational Performance and Information (CEPI), and Michigan Courts State Court Administrative Office (SCAO) to test the effects of foster placement on a variety of child outcomes. There is no common identifier between these administrative data sources, so the files were linked using a probabilistic matching algorithm. The linkage procedure was identical between the three sources, so we describe only the match between the child welfare and education data here.

As described in [Ryan et al. \(2018\)](#), the child welfare data were matched to education records based on first name, last name, date of birth, and gender, and was implemented using the Link King program. Race/ethnicity was not included in the match because the categories were different across data systems. The match was restricted to children born between 1989 and 2012 and compared 846,870 individuals of any age who had a child maltreatment investigation against approximately 5.1 million public school students. 742,269 children (87.6%) with an investigation matched to a public school record. For each of these matched records, the Link King software rates the certainty level of the match on a seven-point scale, ranging from one, a “definite match,” to seven, a “probabilistic maybe.” Overall, 92% of the matches were rated with a certainty-level of one or two and were kept for analysis.

For our analysis, we restrict the sample to include maltreatment reports that entered the investigator rotational assignment system and involved children enrolled in public school. Table B19 describes each sample restriction, step by step. The first restriction ensures the maltreatment report entered the rotation assignment system. The second ensures that nobody in the sample had already been treated. Restrictions three and four limit the sample to children included in the record linkage. The fifth restriction, like the first, drops cases unlikely to have been quasi-randomly assigned. The sixth drops a small fraction of investigations missing pertinent information to construct rotation groups. Restriction seven makes sure that investigators were assigned enough cases to reliably measure their tendencies, yet the results are similar if we relax this. The eighth restriction drops a large fraction of investigations but allows me to observe at least one year of public school records both before and after the investigation for nearly all investigations. Finally, restriction nine ensures that we can observe at least one follow-up school year after the investigation and restriction ten ensures that there were enough children to make within-rotation group comparisons.

This leaves 248,730 investigations of 190,980 children. Some of these children never enrolled in a Michigan public school after their investigation which, as reported in the eleventh restriction, are later dropped from the analysis since we do not observe their outcomes. However, there were 295,892 investigations of children old enough to be enrolled in grades one through eleven, meaning only 84.1% matched to a public school student record. The remaining 47,162 investigations, or 15.9%, are excluded from our analysis. These investigated children may not have been enrolled in public school for any of the following five reasons: (1) they were enrolled in private school, (2) they were homeschooled, (3) they had dropped out of school, (4) they went to school in a different state, or (5) they actually were enrolled in public school but did not match to a public school record with high certainty. While excluding these investigations should not influence the internal validity of our results, they may affect the external validity. To explore this, we compare

the investigations included in our analysis sample to those of school-age children that were excluded, along the observable characteristics included in the child welfare files.

Table B20 shows that the investigations excluded from our analysis look relatively similar to those included. However, they were slightly more likely to be black, a bit older, and more likely to have occurred during the summer. The increased likelihood of occurring in the summer suggests that some of the investigations that did not match to public school student records involved children who lived out-of-state during the school year but were in Michigan in the summer.

Using this information, as well as publicly available statistics about private school enrollment, homeschool enrollment, and high school dropout rates, we estimate the relative share of children that were excluded from our analysis for each of the five reasons listed above. Table B21 shows these estimates. This allows me to assess the quality of the match between the education and child welfare files. Back of the envelope calculations suggest that private school students make up 4.6% of investigations, homeschool students make up 2.6%, dropouts make up 2.1%, and children who live in another state make up 3.4%. Therefore, we estimate that only 3.2% of investigations were of children who were truly enrolled in a Michigan public school, but did not match to a student record with high enough certainty. These estimates suggest that the education and child welfare link performed very well.

Table B19: Sample Construction

	(1) # Investigations	(2) # Children
0. Start with all maltreatment investigations between 2008-2017	1,366,742	657,196
<i>Drop if...</i>		
1. Investigation was within one year of a prior case involving the same child	926,407	651,534
2. Investigation occurred after child was placed in foster care	891,883	637,207
3. Child was born before August 1, 1996	818,008	537,371
4. Child was born after December 31, 2012	707,500	476,143
5. Maltreatment report was for sexual abuse	673,349	458,390
6. Investigation records were missing zip code	663,379	450,338
7. Investigator was assigned fewer than 50 cases	627,580	433,662
8. Child was not enrolled in grades 1 to 11 in a Michigan public school in year of investigation	272,153	202,183
9. Investigation occurred during the 2017 or 2018 school year	250,095	191,872
10. Degenerate zip code by year group	248,730	190,980
11. Never enrolled in Michigan public school after investigation	242,233	186,250

Notes. The final analysis sample contains all child maltreatment investigations in Michigan that entered the rotational assignment system during the 2008–2016 school year of children enrolled in a public school in grades 1 through 11 and that were assigned to investigators who worked at least 50 cases. We check for differential attrition out of the public school system using the sample reported in step 10 consisting of 248,730 investigations (shown in Table B1); since there is no evidence of differential attrition, the final analysis sample consists of students who ever enrolled in a Michigan public school after their investigation.

Table B20: Comparing Sample to School-Age Children who were Excluded from Analysis

	(1)	(2)
	In Sample	Not in Sample
<i>Child Socio-Demographics</i>		
Female	0.49	0.49
White	0.67	0.61
Black	0.24	0.29
Multiracial	0.08	0.09
Other Race	0.01	0.01
Age	10.37	11.63
Had a Prior Investigation	0.58	0.50
Investigated in Summer (June–Aug)	0.22	0.29
Observations	248,730	47,162

Notes. Column 1 consists of investigations in the analysis sample and those who would have been included in the analysis sample had they enrolled in a Michigan public school after their investigation (step 10 in Table B19). Column 2 consists of investigations that would have been included in the analysis sample had the child been enrolled in a Michigan public school in grades 1 through 11 during the investigation. That is, the investigation entered the rotational assignment system, was assigned to an investigator who was assigned at least 50 investigations, and the child was old enough to have been enrolled in 1st grade—at least 7 years old.

Table B21: Breakdown of School-Age Children Included and Excluded from Analysis Sample

	(1) Notes	(2) Estimated Share of Investigations
0. Enrolled in Public School	- Included in analysis sample	84.1%
1. Enrolled in Private School	- Private schools enroll 10% of students in MI (Mack, 2017) - 10% of private school students were low income (White and DeGrow, 2016)	4.6%
2. Homeschooled	- About 3% of students in MI are home-schooled (CRHE, 2017) - $\frac{1}{3}$ of home-schooled children in CT had an investigation (OCA, 2018) - We assume that 20% of homeschooled children in MI did	2.6%
3. Dropped out of School	- 10% of investigated children not enrolled were ≥ 16 years old - Of these, 21% were enrolled in a MI public school before investigation	2.1%
4. Went to School in Other State	- Children could have investigation in MI while visiting family - Most likely to be investigated in the summer - 7.7pp increase in summer investigations among children not in sample - We assume that half of this increase is from out-of-state children	3.4%
5. Enrolled in Public School, But Did not Match	- 96.8% investigations fall into categories 0-4 - The rest were likely to have been enrolled, but did not match	3.2%
Total		100.0%

Notes. To estimate the share of children with an investigation who fall into each category, we use Baye's Theorem to calculate, for example, the probability that a child was enrolled in private school conditional on having a maltreatment investigation. In doing so, we use the following statistics, derived from the data: $P(\text{inv}) = 0.23$, $P(\text{inv}|\text{low income}) = 0.38$, $P(\text{inv}|\text{high income}) = 0.08$ and we assume that the probability of being investigated conditional on income level is the same across public and private schools.

B.5 OLS Effects of Foster Care Placement Types

In recent years, states have prioritized placing foster children with relatives, known as kinship care, whenever possible. Kinship care is thought to be less disruptive to children’s lives because it allows them to live with someone they know and who shares their culture. These placements also exhaust fewer state resources as it is difficult to recruit unrelated families to take in foster children. Despite this trend, there is mixed research evidence on the effectiveness of kinship care relative to other placement types.

[Lovett and Xue \(2018\)](#) exploit changes in monthly compensation rates and note that although low compensation rates to unrelated foster families are predictive of increased placements in kinship care, previous studies have found that they are not associated with children’s outcomes. The study finds that children who were placed in kinship care were more likely to be employed or in school, less likely to be incarcerated, and less likely to receive public assistance relative to children placed with an unrelated foster family. In contrast, [Hayduk \(2017\)](#) exploits state and time variation in the adoption of laws that prioritize kinship placements and does not detect evidence that they improved children’s physical or mental health.

We add to this evidence by testing the effects of various types of foster placement. We cannot perform this analysis using the examiner assignment research design because placement type is endogenous to unobservable characteristics of the child, such as having support from nearby family members. Therefore, we use OLS to describe how the effects of removal vary based on initial placement type. Specifically, we estimate the following model:

$$Y_{iw} = \beta_0 + \beta_1 KINSHIP_{iw} + \beta_2 UNRELATED_{iw} + \beta_3 GROUP_{iw} + \beta_4 X_{iw} + \theta_r + \epsilon_{iw} \quad (9)$$

where β_1 represents the association between initial kinship placement and the outcome relative to children who were not placed into foster care. Similarly, β_2 and β_3 report this relationship for initial placement with an unrelated foster family and in a group home respectively.

Table [B22](#) shows the results. Overall, placement with relatives is associated with greater improvements than placement with an unrelated foster family or in a group home. Notably, the OLS estimates in the main analysis understate the benefits of removal and overstate the costs relative to the 2SLS estimates. To the extent that this analysis suffers from similar selection bias, this analysis might offer a lower bound for the effects of each placement type.

Table B22: OLS Effects of Foster Care on Child Outcomes, by Initial Placement Type

	(1) Index of Child Well-being	(2) Confirmed Victim of Maltreatment	(3) Daily Attendance Rate	(4) Std Math Score
Kinship	0.116*** (0.014)	-0.007*** (0.002)	0.018*** (0.002)	0.093*** (0.017)
Unrelated	0.080*** (0.019)	-0.003 (0.004)	0.017*** (0.003)	0.050** (0.024)
Group Home or Institution	0.028 (0.042)	0.008 (0.006)	0.005 (0.007)	-0.046 (0.052)
Comparison Mean	0.002	0.046	0.912	-0.501
Kinship vs Unrelated	0.120	0.341	0.823	0.141
Kinship vs Group	0.048	0.019	0.085	0.011
Unrelated vs Group	0.264	0.110	0.120	0.093
Observations	242,264	242,264	224,925	177,118

Notes. This table reports results from OLS regressions of the outcome variable on mutually exclusive indicators for initial foster placement types. The mean outcome for children who were not removed as well as the p-values testing whether the point estimates for each placement type are statistically different from each other are shown below the regression results. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.6 Who Takes in Foster Children?

The administrative records in this study do not contain individual level information about foster parents. Moreover, there are limited public data about who takes in foster children. The best information comes from the American Community Survey (ACS), administered by the Census Bureau, which includes “foster children” as a category in a question about the members of a household. However, the ACS is known to understate the number of foster children in the country by almost half relative to administrative records and is not thought to be representative. The leading explanations for why the ACS fails to account for so many foster children are that unrelated families who care for a foster child for only a short amount of time may not list them as a member of their household and that households who take in a relative may list them as relatives instead of as foster children (O’Hare, 2007).

With these limitations in mind, Table B23 describes households with foster children and compares them to other households with members younger than 18 years old, using the 2012–2016 five-year sample of the ACS. Nationwide, households with foster children were larger and much lower income. The head of households were older, less likely to be employed, and more likely to be Black. The comparison looks similar when restricted to households in Michigan.

Table B23: Descriptive Statistics of Households With and Without Foster Children

	USA		Michigan	
	(1) At Least One Child Under 18	(2) At Least One Foster Child	(3) At Least One Child Under 18	(4) At Least One Foster Child
# Adults	2.14	2.25	2.08	2.06
# Children Under Age 18	1.88	2.61	1.89	2.97
Pre-Tax Income	\$141,431	\$69,948	\$131,038	\$62,067
<i>Head of Household</i>				
Married	0.66	0.63	0.64	0.56
White	0.71	0.68	0.77	0.67
Black	0.14	0.22	0.15	0.25
Observations	37,489,148	143,580	1,136,414	5,533

Notes. This table reports descriptive statistics comparing households with and without foster children for the United States overall and for Michigan. All statistics are weighted estimates from the American Community Survey 2012-2016 five year sample.