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

By:

Max Gross

E. Jason Baron

Published in the

American Economic Journal: Applied Economics (2022)

Gross: Mathematica (email: mgross@mathematica-mpr.com); Baron: Duke University and NBER (email: jason.baron@duke.edu). David Deming was coeditor for this article. 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, Ezra Goldstein, 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 for Public Policy Analysis 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 Plevek, 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. The project received approval from the University of Michigan's Institutional Review Board: HUM00132379 and was supported by training grant award R305B170015 from the Institute of Education Sciences, U.S. Department of Education. 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.

American Economic Journal: Applied Economics 2022, 14(2): 170–199
https://doi.org/10.1257/app.20200204

Temporary Stays and Persistent Gains: The Causal Effects of Foster Care[†]

By Max Gross and E. Jason Baron*

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 were 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. (*JEL* H75, I21, J13, K42)

> *"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 and Management Presidential Address, 2019

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; Children's Bureau 2018b). 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

* Gross: Mathematica (email: mgross@mathematica-mpr.com); Baron: Duke University and NBER (email: jason.baron@duke.edu). David Deming was coeditor for this article. 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, Ezra Goldstein, 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 for Public Policy Analysis 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 Plevek, 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. The project received approval from the University of Michigan's Institutional Review Board: HUM00132379. 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.

[†] Go to https://doi.org/10.1257/app.20200204 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

(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 (Children's Bureau 2004), 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 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

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

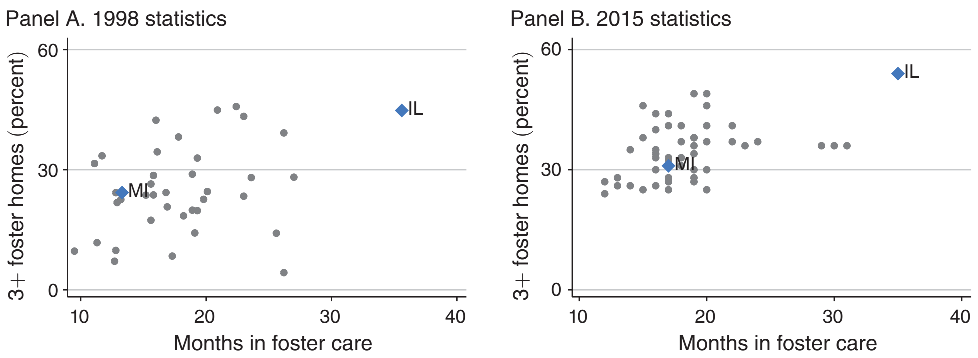

Figure 1. Comparison of State Foster Care Systems

Notes: These panels show statistics about state foster care systems from 1998, the first year of publicly available data, reported in Children's Bureau (2004), and from 2015, reported in KCDC (2015) and ChildTrends (2017b). 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, ten states did not report either of these statistics.

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.

The pattern of our results over time strongly suggests that improvements made by birth parents were the primary channel through which foster care placement improved children's outcomes. 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] A likely 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, 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, Hendren, and Katz 2016; Chyn 2018), we find no evidence that placement caused lasting improvements to children's neighborhoods or schools.

This paper improves upon Bald et al. (2019) and Roberts (2019), which also estimate the effects of foster care using a similar research design, by demonstrating that incomplete data coverage in these studies can create substantial bias.[3] Specifically, these other studies do not follow children from the start of their child welfare

[2] We 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.

[3] Bald 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

investigation. Rather, they focus exclusively on substantiated allegations (those in which investigators found a preponderance of evidence to support the maltreatment allegation). Such data restrictions may create bias because the same investigator who determines foster placement also has discretion over substantiation. Thus, even if cases are initially assigned at random, the subset of children with substantiated allegations may not be balanced across investigators. To demonstrate this bias, in online Appendix B, we replicate our primary analysis using only the sample of substantiated investigations and find estimates much smaller than the effects using the full data.

Our 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, made 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.

I. 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 third grade (Ryan et al. 2018). One in 10 was the subject of more than one investigation, and 1 in 60 was placed in foster care.[4] This section reviews the maltreatment investigation process in Michigan and describes the state's foster system.

A. *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).[5] A hotline employee, who does not participate in the investigation process, transfers relevant reports to the child's local child welfare office. The office assigns the report to a maltreatment investigator, who then 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.

children ages 2–17 and found positive impacts on on-time grade progression yet noisy estimates on daily school attendance and test scores.

[4] These 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 and 2006.

[5] The 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 percent), legal and law enforcement personnel (18.7 percent), and social service workers (10.7 percent) (Children's Bureau 2020).

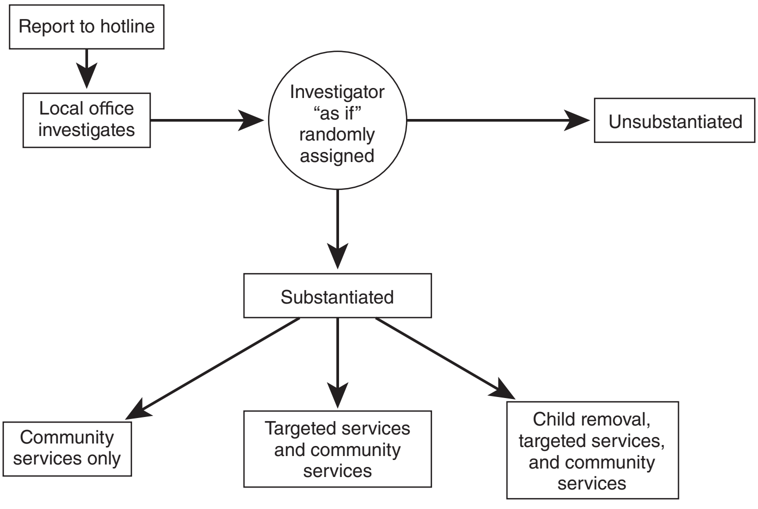

FIGURE 2. OVERVIEW OF CHILD MALTREATMENT INVESTIGATIONS IN MICHIGAN

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.

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. 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 to the same investigator.[6]

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 (Children's Bureau 2018a, 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

[6] There 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 they have familiarity with the family. Anecdotally, such reports tend to reenter the rotation after a few months. We exclude from the analysis those within one year of a prior investigation to be conservative.

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, 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. 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 the effect of removal above and beyond adult services.

B. *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 ten 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.[7] In many cases, though, children do not have suitable relatives available. In 2015, 41 percent of foster children in Michigan were placed 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.[8] It is common to switch placement settings while in the foster system. Sixty percent of foster children in Michigan in 2015 had lived in more than one setting, and 17 percent lived in at least 4. Michigan looks very similar to the rest of the country along these statistics (Child Trends 2017a).

After placement, child welfare caseworkers (who are different from the investigator) meet with birth parents to create a reunification plan stating the conditions under which they can regain custody of their child. 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

[7] There 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 online Appendix E finds a larger positive association between placement with relatives and children's outcomes relative to other placement types.

[8] There are limited data available on the families who take in foster children. However, as shown in online Appendix F, estimates from the American Community Survey 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 (ACS 2016).

can include substance abuse treatment, parenting classes, counseling, and job training. 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 progress 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.

Although child welfare systems vary across states, the system in Michigan is quite similar to other states around the country on key foster care indicators. Specifically, in 2015, Michigan (along with eight other states) ranked eighteenth in placement length and seventeenth in the share of foster children who experienced three or more placement settings, a measure of placement stability. Furthermore, the proportion of children in Michigan who were reunified with their birth parents after exiting foster care (47 percent) was very close to the national average (51 percent) (AECF 2017). Therefore, there is little reason to suspect that the findings from this study would not generalize to other child welfare systems across the country.

II. Data Sources and Sample Construction

A. *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.[9] Online Appendix D 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 (MDHHS 2017). They include details of each investigation, such as the allegation report date, allegation types

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

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.[10] 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.[11]

Education data from CEPI and MDE cover the universe of public school students in Michigan, including charter school students, between the 2002–2003 and 2016–2017 school years (MDE 2017; CEPI 2017). 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 US Census Bureau.

Juvenile justice data from SCAO include all juvenile court petitions filed in almost every county in Michigan between 2008 and 2015 (SCAO 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.[12] We exclude the 19 percent of investigated children who lived in these 8 counties from our analysis of juvenile delinquency; the conclusions on other outcomes are similar when these children are excluded.[13]

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 court 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.

[10] Investigators 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 ten days.

[11] 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). Online Appendix B describes how incomplete data coverage can substantially bias estimates from the examiner assignment research design.

[12] The data include Wayne County (home of Detroit) and the metro-Detroit area but exclude five urban and three rural counties: Kent, Washtenaw, Ingham, Ottawa, Kalamazoo, Berrien, Delta, and Keweenaw. Together, these 8 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.

[13] Michigan's juvenile arrest rate is similar to other states. The US Department of Justice's Office of Justice Programs reports that in 2017, Michigan ranked twentieth out of the 48 contiguous states 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 eighteenth. It is important to note, however, that our juvenile delinquency measure differs from Doyle (2007). We elaborate on this point in Table 7.

B. *Child Safety*, *Academic*, *and Crime Outcomes*

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

Following Kling, Liebman, and Katz (2007) and Deming (2009), we 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 nonmissing standardized items in the index.[14] 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.[15] Finally, we measure juvenile delinquency as the filing of a juvenile court petition.

C. *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.[16]

[14] We impute missing values in this way because our best guess of the value of a missing outcome is the average of the remaining standardized outcomes in the index.

[15] These 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.

[16] The analysis sample excludes children who were too young to have entered school at the time of their investigation. Although 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 about one-third of

Online Appendix D describes the sample restrictions in greater detail. Overall, we focus on 242,233 investigations of 186,250 students and follow students for at most 9 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 involved Black children, and 83 percent involved 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 percent of investigations that resulted in foster placement. Relative to the overall sample in column 2, foster children were 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. Taken together, these descriptive statistics indicate that children placed in foster care differ in systematic ways from children who were investigated but were not placed.

III. Empirical Strategy

A naïve 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.

A. *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, Kostøl, and Mogstad 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

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 (online Appendix Table B1).

Table 1—Summary Statistics

	All Michigan students (1)	Analysis sample All (2)	Analysis sample Foster care (3)
Child sociodemographics			
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
Prior schooling characteristics			
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
Standardized math score	0.00	−0.27	−0.36
Standardized reading score	0.00	−0.25	−0.34
Investigation characteristics			
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 sociodemographic 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.

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.[17] 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

[17] Online Appendix Table A6 shows that the results are robust to larger thresholds.

past and future, assigned to the same investigator that resulted in foster placement. Specifically, for investigation i assigned to investigator w,

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

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.[18] 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 tenth percentile removed at a rate 2.1 percentage points less than the average investigator in their local area, whereas someone at the ninetieth 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 foster care:

$$FC_{iw} = \gamma_1 Z_{iw}^{R} + \gamma_2 X_{iw} + \Theta_r + \eta_{iw}, \tag{2}$$

$$Y_{iw} = \beta_1 \hat{FC}_{iw} + \beta_2 X_{iw} + \theta_r + \epsilon_{iw}, \tag{3}$$

where Y_{iw} is a child outcome, such as daily school attendance rate or score on a standardized math test. The term X_{iw} is a vector of baseline covariates that includes a variety of sociodemographic and academic characteristics, such as gender, race/ethnicity, grade level fixed effects, baseline standardized test scores, and prior child welfare involvement.[19] 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. 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.[20]

The parameter 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

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

[19] See online Appendix Table B10 (Column 5) for the full set of baseline characteristics.

[20] There 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 IVD details several alternative approaches and reports that the results are robust to the level of clustering.

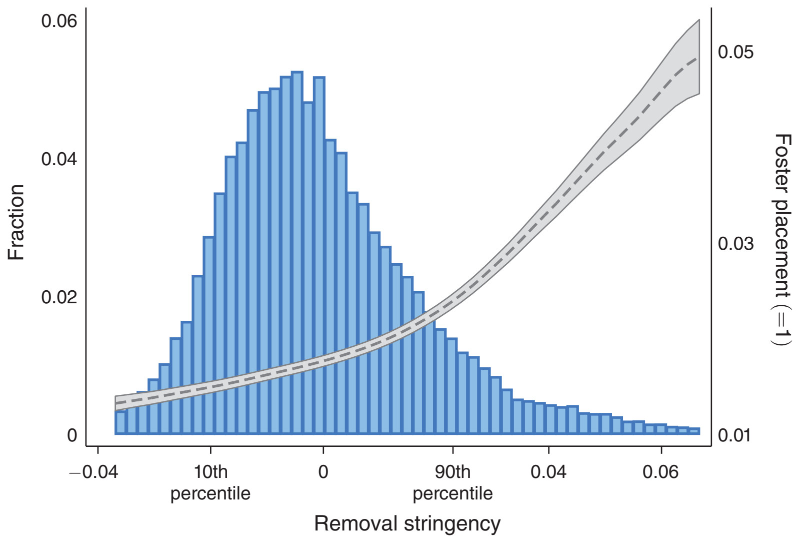

Figure 3. Distribution of Investigator Removal Stringency Instrument

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 nonparametric 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.

care influences children so clearly in danger that all investigators would remove ("always-takers") and those so clearly safe that no investigators would remove ("never-takers"). However, as Dr. Jill Duerr Berrick writes, the debate around foster care 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, 56). Therefore, compliers represent a population that is especially relevant for child welfare policy.

B. *Identifying Assumptions*

Four assumptions must be satisfied to interpret our estimates as the causal effect of foster care for children at the margin of placement: relevance, exogeneity, average monotonicity, and exclusion.

Relevance requires that investigator removal stringency predicts foster care placement ($\gamma_1 \neq 0$). Figure 3 visually depicts the strong, positive relationship between investigator removal stringency and placement, and Table 2 reports 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.

Exogeneity requires that the unobserved determinants of children's outcomes are independent of investigator removal stringency ($\text{cov}\left[Z^R, \epsilon\right] = 0$). We test

Table 2—First-Stage Effect of Removal Stringency on Foster Placement

	Foster care (1)	Foster care (2)	Foster care (3)	Foster care (4)
Removal stringency	0.480	0.451	0.450	0.449
	(0.019)	(0.021)	(0.021)	(0.021)
Observations	242,233	242,233	242,233	242,233
F-statistic	898.405	586.569	584.34	582.28
Zip code by year fixed effects		✓	✓	✓
Sociodemographic 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. Sociodemographic 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.

an implication of exogeneity—that observable child and case characteristics are uncorrelated with the removal tendencies of the assigned investigator. As expected due to the rotational assignment of child welfare investigators, a rich set of characteristics are not jointly predictive of the instrument despite being highly predictive of placement itself (Table 3).[21]

For average monotonicity to hold, the covariance between each child's investigator-specific removal treatment status and investigator stringency must be weakly positive.[22] It follows from average monotonicity that removal stringency and foster placement should be positively correlated for all child subgroups. We find that the first stage is positive and statistically significant across gender, race/ethnicity, age, and prior child welfare involvement groups (panel A of online Appendix Table B2). We also find that the first stage remains positive and statistically significant when we recalculate the instrument as a leave-subgroup-out measure (panel B of online Appendix Table B2).

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 IVD.

[21] As further evidence of exogeneity, the first-stage F-statistic in Table 2 is stable with the inclusion of covariates.

[22] Recent advances note that pairwise monotonicity—the assumption that children who were removed by a particularly lenient investigator must also have been removed by a stricter investigator—is neither realistic in most contexts nor necessary to estimate local average treatment effects (Norris 2019; Frandsen, Lefgren, and Leslie 2019). Specifically, 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, Pecenco, and Weaver 2019; Frandsen, Lefgren, and Leslie 2019).

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

	Full sample		4th grade and above	
Dependent variable:	Foster care (1)	Investigator removal stringency (2)	Foster care (3)	Investigator removal stringency (4)
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 sociodemographic 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.

IV. 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 (the mean outcomes among all investigated children who were not removed) and control complier means for 2SLS (the estimated outcomes for compliers who were not removed) are reported in curly brackets.

The OLS analysis suggests that removal had a precise but near-zero impact on the index of child well-being.[23] In contrast, the 2SLS estimate is much larger and 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. This suggests that unobserved features like the severity of maltreatment, for example, may lead OLS to understate the benefits of removal. Also, the control complier mean is less than the control mean, indicating that children at the margin 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 index components.

A. *Effects on Child Safety*, *Academics*, *and Crime*

Table 4 shows that the foster care system achieved its primary objective; placement improved children's safety. 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

[23] 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 control for lagged outcomes, and recent work by Berger et al. (2015) shows that controlling for these variables substantially reduces the negative association. We replicate the negative relationship using simple bivariate regressions in online Appendix Table B3.

TABLE 4—EFFECTS OF FOSTER CARE ON CHILD OUTCOMES

	Index of child well-being (1)	Alleged victim of maltreatment (2)	Confirmed victim of maltreatment (3)	Daily attendance rate (4)	Standardized math score (5)	Standardized reading score (6)	Juvenile delinquency (7)
Panel A. OLS							
Foster care	0.026	−0.032	−0.007	0.011	0.057	0.065	0.041
	(0.011)	(0.004)	(0.002)	(0.002)	(0.013)	(0.014)	(0.004)
	{0.002}	{0.177}	{0.046}	{0.912}	{−0.501}	{−0.479}	{0.025}
Panel B. 2SLS							
Foster care	0.392	−0.132	−0.053	0.055	0.356	0.175	−0.028
	(0.164)	(0.058)	(0.028)	(0.026)	(0.203)	(0.219)	(0.040)
	{−0.123}	{0.255}	{0.094}	{0.893}	{−0.429}	{−0.234}	{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 8 counties, available only through 2015, and relevant only for children younger than Michigan's age of majority of 17.

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 (without improving safety), then the reported incidents involving foster children should, on average, be more severe than those involving children who were not placed. However, we find no clear evidence that foster placement influenced the likelihood of substantiation among children with a subsequent investigation.[24] Moreover, caseworkers, who are also mandatory reporters, visited foster children regularly, both during their time in the foster system and after they exited, suggesting that actual maltreatment against foster children would have been reported (Child Welfare Information Gateway 2016).

Consistent with an improvement in child safety, we find large gains in academic outcomes. Removal increased daily school attendance rates by 5.5 percentage

[24] Specifically, we estimate equations (2) and (3) using an indicator for substantiation as an outcome and limit the analysis to children with a subsequent investigation. We obtain a point estimate on foster placement of −0.14 and a standard error of 0.12.

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 of a standard deviation.[25] 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.[26] Lastly, we find 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 use a one-sided hypothesis test to assess whether foster care worsened children's 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.

B. *Mechanisms: Evidence from the Timing of Impacts*

Forty percent of children who were placed had exited the foster system after one year, and nearly all had exited after two years (Figure 4).[27] We create an index of neighborhood and school characteristics according to Kling, Liebman, and Katz (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 (panel A of Table 5).[28] Given that moving to areas with lower poverty levels can improve child well-being (Chetty, Hendren, and Katz 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 (panel A of Table 5). That foster children were no more or less likely to be abused or neglected in the first year may be especially

[25] As 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.

[26] Removal did not influence the likelihood of taking standardized tests (online Appendix 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 (online Appendix Table B5).

[27] They spent 19 months in foster care, on average (online Appendix Table B6).

[28] This was driven by exposure to more highly educated neighborhoods and higher-income classmates (online Appendix Table B7).

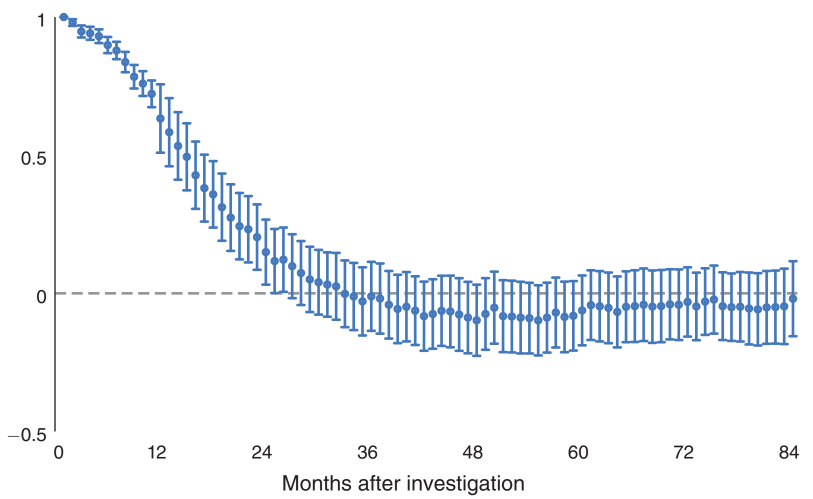

FIGURE 4. EFFECTS OF FOSTER CARE ON LIKELIHOOD OF BEING IN FOSTER SYSTEM OVER TIME

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.

surprising since maltreatment in foster homes is extremely rare.[29] It is possible, however, that the threat of child removal reduced the maltreatment of children who were not placed in the short run.

Nearly all (85 percent) foster children at the margin of placement had exited the system after two years and reunified with their birth parents.[30] Upon exiting, foster children returned to neighborhoods and schools similar to those of children at the margin who were not placed; we do not detect differences in the characteristics of their neighborhoods or schools after the first year (panel B of Table 5). 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 (panel B of Table 5). 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.

[29] In 2018, 0.71 percent of foster children in Michigan were maltreated while in the foster system (Children's Bureau 2018c).

[30] Online Appendix Table A1 shows that of the remaining 15 percent who exited, 8 percent were adopted, 5 percent had guardianship transferred, and 2 percent exited foster care as legal adults.

Table 5—Effects of Foster Care over Time

	Index of neighborhood and school characteristics (1)	Index of child well-being (2)	Confirmed victim of maltreatment (3)	Daily attendance rate (4)	Standardized math score (5)	Received special education services (6)	Retained in grade (7)
Panel A. One year after investigation							
Foster care	0.257	0.231	−0.024	0.040	−0.207	−0.013	−0.035
	(0.100)	(0.211)	(0.053)	(0.034)	(0.217)	(0.063)	(0.049)
	{−0.147}	{0.028}	{0.068}	{0.912}	{0.062}	{0.099}	{0.065}
Panel B. Two+ years after investigation							
Foster care	0.066	0.446	−0.064	0.060	0.579	0.014	−0.008
	(0.125)	(0.196)	(0.032)	(0.031)	(0.239)	(0.105)	(0.036)
	{−0.011}	{−0.159}	{0.102}	{0.885}	{−0.624}	{0.035}	{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 are shown in online Appendix Table B7. All regressions include the covariates as listed in the text and zip code by investigation year fixed effects.

Because most foster children were reunified with their birth parents upon exiting the foster care system, the pattern of impacts over time strongly suggests that improvements made by birth parents were the primary mechanism at work. There are two institutional features that support this channel. First, 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. Birth parents received fully funded services to help with these challenges, such as substance abuse treatment, parenting classes, or counseling. Second, a judge needs to approve that it is safe for children to return home before they can be reunified with their birth parents. Moreover, 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 (panel E of Figure 5).

Though we cannot 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, Hendren, and Katz 2016; Chyn 2018), whereas exposure in our context was only temporary. Studies also find that the long-run benefits of moving do not run through schooling channels (Jacob 2004; Sanbonmatsu et al. 2006), yet foster care had large impacts on educational outcomes. Second, it could be that foster care triggers additional supports whose benefits take time to manifest. However, we find no evidence that foster care increased supports in school either during placement or

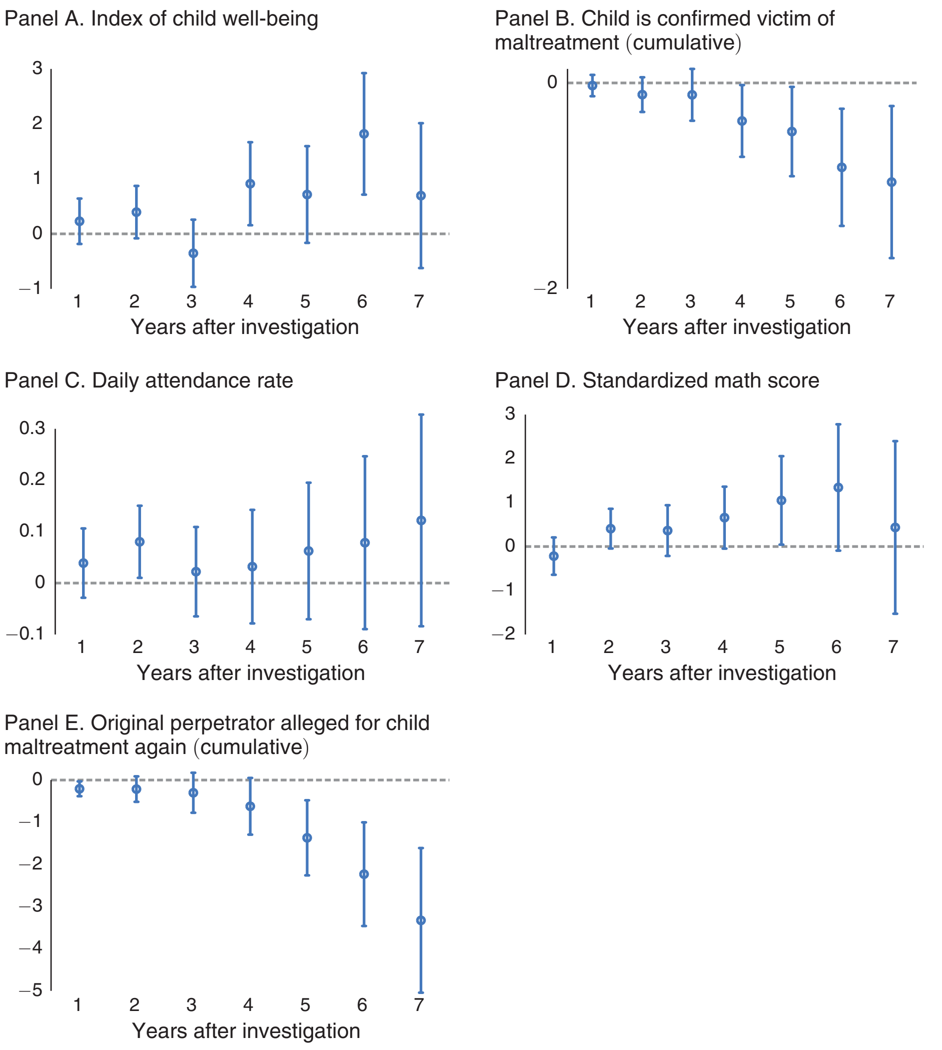

Figure 5. Effects of Foster Care over Time

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 nonschooling outcomes. Panel E of Figure 5 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.

after exiting, as proxied by receipt of special education services and grade retention (Table 5, columns 6 and 7).

Therefore, evidence from the timing of impacts suggests that positive changes made by birth parents were the main channel through which foster placement improved children's safety and schooling.

C. *Compliers Analysis and Subgroup Effects*

Contextualizing Children at the Margin of Foster Placement.—The estimates in this study represent effects for children at the margin of placement, those for whom investigators might disagree over whether foster care is appropriate. 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. Yet these groups otherwise looked similar in terms of demographic and baseline academic characteristics (online Appendix Table A2). In terms of experiences in the foster care system, compliers had shorter and more stable placements and were more likely to be reunified with their birth parents than the overall population of foster children (online Appendix Table A3).

Heterogeneity by Child Age and Gender.—Prior research 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, Hendren, and Katz 2016; Chyn 2018). Similarly, we find large and statistically significant impacts of foster care on the index of child well-being for children age ten and younger, yet smaller and statistically insignificant effects for older youth (online Appendix Table A4). In addition, although previous work shows that male children are often more vulnerable than female children to disadvantage or disruption (Kling, Ludwig, and Katz 2005; Bertrand and Pan 2013; Autor et al. 2019), we find that the impacts of placement were positive for both groups.[31]

D. *Threats to the Research Design and Robustness Checks*

Exclusion and the Multidimensionality of Treatment.—The exclusion restriction requires that the removal stringency instrument can only influence outcomes through foster placement. A potential concern is that investigators might have influenced children's experiences in foster care. However, investigators did not work with children after the investigation; cases that required follow-up were transferred to other child welfare caseworkers. That is, investigators were not involved in determining where children were placed, how long they stay in foster care, or the stability of their placements.[32] Accordingly, the instrument does not predict these indicators of children's experiences in foster care, nor is it jointly related to these factors (online Appendix Table A5).

[31] Previous studies of foster care using a similar research design 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.

[32] For example, the length of placement depends on a variety of factors, none of which involve the initial investigator. First, 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 caseworker, 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 or guardian homes.

Although 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, however, we would not be able to address all of the potential channels through which investigators could impact children's outcomes.

However, we can empirically account for perhaps the most important way in which investigators might impact children's outcomes other than removal: investigators influence whether families are referred to prevention-focused services. As shown in Figure 2 and discussed in Section IA, investigators place 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) both community-based and targeted services, and (4) child removal plus community-based and targeted services. The exclusion restriction would be violated if investigators who were more likely to remove children were also more likely to recommend prevention services, and tendencies over prevention services are not included in the estimation (Mueller-Smith 2015).

To account for investigator discretion over prevention services, 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:

$$(4) \qquad RTC_{iw} = \gamma_1 Z^{RTC}_{iw} + \gamma_2 Z^{TC}_{iw} + \gamma_3 Z^{C}_{iw} + \gamma_4 X_{iw} + \kappa_r + \mu_{iw},$$

$$(5) \qquad TC_{iw} = \alpha_1 Z^{RTC}_{iw} + \alpha_2 Z^{TC}_{iw} + \alpha_3 Z^{C}_{iw} + \alpha_4 X_{iw} + \chi_r + \nu_{iw},$$

$$(6) \qquad C_{iw} = \delta_1 Z^{RTC}_{iw} + \delta_2 Z^{TC}_{iw} + \delta_3 Z^{C}_{iw} + \delta_4 X_{iw} + \pi_r + \zeta_{iw},$$

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

where RTC_{iw} is a binary variable equal to one if the child was removed, 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 when 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.[33]

[33]The first-stage relationships in equations (4) through (6) are strong, with F-statistics above 200 (online Appendix Table B8), and balance tests indicate that each of the three instruments is unrelated to a rich set of baseline child characteristics (online Appendix Table B9). The 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, corr$(Z^{RTC}, Z^{C}) = 0.14$, corr$(Z^{RTC}, Z^{TC}) = 0.24$, and corr$(Z^{TC}, Z^{C}) = 0.60$.

Table 6—Effects of Adult Interventions on Child Outcomes

	Index of child well-being (1)	Alleged victim of maltreatment (2)	Confirmed victim of maltreatment (3)	Daily attendance rate (4)	Std math score (5)	Std reading score (6)	Juvenile delinquency (7)
Panel A. One year after investigation							
Child removal, targeted, and community services	0.033	0.031	−0.038	0.015	−0.337	−0.288	−0.005
	(0.294)	(0.128)	(0.074)	(0.046)	(0.296)	(0.318)	(0.068)
Targeted and community services	0.075	−0.108	−0.005	−0.000	0.157	0.040	−0.007
	(0.114)	(0.049)	(0.028)	(0.014)	(0.105)	(0.112)	(0.027)
Community services	0.009	0.046	0.011	0.009	−0.110	0.050	0.008
	(0.073)	(0.032)	(0.018)	(0.009)	(0.069)	(0.075)	(0.017)
Panel B. Two+ years after investigation							
Child removal, targeted, and community services	0.350	−0.069	−0.037	0.075	0.564	0.188	−0.013
	(0.260)	(0.087)	(0.043)	(0.040)	(0.313)	(0.334)	(0.064)
Targeted and community services	0.128	−0.045	−0.024	0.003	0.081	0.031	−0.006
	(0.093)	(0.032)	(0.016)	(0.013)	(0.113)	(0.119)	(0.024)
Community services	−0.090	0.018	0.012	−0.010	−0.086	−0.005	−0.006
	(0.060)	(0.020)	(0.010)	(0.009)	(0.074)	(0.079)	(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.

Table 6 provides evidence that the removal stringency instrument operates through foster care placement. Specifically, the impacts of prevention services without child removal (both community-based services alone and community-based and targeted services) are mostly small and statistically insignificant. This indicates that investigator stringency over prevention services was largely unrelated to children's outcomes. In contrast, we find that the impacts of child removal above and beyond prevention services are very similar in magnitude to the main estimates in Table 5. Although the 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. Taken together, this exercise suggests that any bias from investigators' involvement in referring prevention services is likely to be small.[34] Overall, the evidence presented in this section suggests that there is little cause for concern regarding the exclusion restriction in our context.

Beyond a test of the exclusion restriction, the analysis in Table 6 is one of the first to test the causal impacts of prevention-focused child welfare interventions. We find

[34] The exclusion restriction could also be violated if the decision of whether to substantiate influences children's outcomes regardless of the decision to place children in foster care. However, online Appendix Table B15 shows that the additional impact of foster placement relative to substantiation without removal is nearly identical to the main results shown in Table 4. Finally, a remaining concern is whether investigator placement tendencies are related to the match quality of families and services. This is unlikely in our context because investigators choose from a standard set of community service options and do not make specific referrals to targeted services.

little evidence that prevention services without removal influence children's outcomes.[35] It is critical for future research to more deeply explore the implementation and impacts of such prevention-focused services.

Robustness Checks.—Online Appendix 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 the number of cases an investigator must have been assigned to be included in the sample.[36] 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 (that is, children investigated in seventh grade or below in 2012 or earlier). The results from these first three alternative samples are similar to those in the main analysis. Although the impacts of foster placement are substantially larger in the balanced panel than in the main analysis, the time pattern of impacts is nearly identical (online Appendix Figure A1).[37]

We also check for robustness using other reasonable ways to measure investigator removal tendencies (panel B of online Appendix Table A6). 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.[38] Lastly, since including many controls (for example, over 7,500 rotation fixed effects) can induce bias in leave-out instrument approaches, we implement the unbiased jackknife instrumental variables (UJIVE) approach of Kolesár (2013), which is robust

[35] This is consistent with Grimon (2020), who finds that having a child welfare case opened without removal has no impacts on mothers' criminal justice involvement or use of social services up to six years after the investigation, despite increasing the take-up of mental health and substance abuse services in the short run.

[36] The main analysis excludes children assigned to investigators who worked fewer than 50 cases, so we strengthen this threshold to 75.

[37] The larger effects for this balanced sample are 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 IVB shows that impacts appear only several years after removal. Similarly, the subgroup analysis in Section IVC shows considerably larger effects for younger children.

[38] Specifically, 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) race/ethnicity (child of color or White); (3) allegation type (physical abuse or not); (4) perpetrator type (parent or not); (5) age at investigation (younger than ten 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 example, 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 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.

to this issue.[39] 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. Yet the results are very similar when we instead include county by investigation year fixed effects.

In addition, 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. However, estimates of the effects 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 (online Appendix Table B10).

Finally, we test the sensitivity to alternative levels of clustering standard errors. We cluster standard errors at the child level in our main specification in order to account for the correlation in outcomes that arises mechanically by including the same child more than once in the panel. However, the results are robust to clustering at the investigator level, the zip code–year level, two-way clustering at the child and investigator level, and two-way clustering at the child and zip code–year level (online Appendix Table B11).[40]

V. 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

[39] The UJIVE approach of Kolesár (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).

[40] Because the same investigator is assigned to multiple investigations throughout the panel, clustering at the investigator level accounts for potential correlations in the error term within investigator but across investigations and time. Clustering at the zip code–year level accounts for potential neighborhood- or child welfare office–level shocks. Lastly, two-way clustering by child and investigator as well as two-way clustering by child and zip code–year level account for all of the potential correlations in outcomes described above.

Table 7—Comparison to Doyle (2007)

Main estimate Gross and Baron (2022) (1)	Comparable estimate Doyle (2007) (2)	*p*-value (1) = (2) (3)	Table 6, panel B Gross and Baron (2022) (4)	*p*-value (4) = (2) (5)
Panel A. Standardized math test scores				
0.356	−0.83	0.000	0.564	0.000
(0.203)			(0.313)	
Panel B. Juvenile delinquency				
−0.028	0.10	0.002	−0.013	0.077
(0.040)			(0.064)	

Notes: This table compares the main results of this paper to those in Doyle's (2007) 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's (2007) 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" (p. 1589). 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's (2007) 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's (2007) 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's (2007) comparable estimates.

explicitly account for investigator propensity to refer families to services. Similarly, 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. Taken together, the evidence strongly suggests that foster care in Michigan during our sample period did not have the same harmful 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 for in online Appendix C. 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, Hislop, and Goerge 2000; Children's Bureau 2003, 2017). These long placements were also less stable; 45 percent of foster children in Illinois lived in 3 or more different homes in 1998 compared to an average of 31 percent across our 10-year panel in Michigan

(Children's Bureau 2003; 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—which took effect in the middle of Doyle's (2007, 2008) sample period—sought to reduce the length of foster placements. There has also been a considerable push to increase placements with relatives over 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. For example, there is mixed evidence that observable characteristics of compliers like age, gender, and race/ethnicity drive the differences. Furthermore, it is unlikely that the counterfactual for children placed (that is, 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 (Child Welfare Information Gateway 2017). 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.

VI. 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 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 thirty-third to the forty-sixth percentile in the state. We also estimate a substantively large yet imprecise reduction in juvenile delinquency. Furthermore, 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, policymakers 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

ACS. 2016. "American Community Survey 5-Year Public Use Microdata Sample, 2012–2016." IPUMS. https://ipums.org/ (accessed November 20, 2017).

AECF. 2017. "KIDS COUNT Data Center." Annie E. Casey Foundation. https://datacenter.kidscount.org.

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

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2019. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *American Economic Journal: Applied Economics* 11 (3): 338–81.

Bald, Anthony, Eric Chyn, Justine S. Hastings, and Margarita Machelett. 2019. "The Causal Impact of Removing Children from Abusive and Neglectful Homes." NBER Working Paper 25419.

Barrat, Vanessa X., and BethAnn Berliner. 2013. *The Invisible Achievement Gap, Part 1: Education Outcomes of Students in Foster Care in California's Public Schools.* San Francisco, CA: WestEd.

Berger, Lawrence M., Maria Cancian, Eunhee Han, Jennifer Noyes, and Vanessa Rios-Salas. 2015. "Children's Academic Achievement and Foster Care." *Pediatrics* 135 (1): 109–16.

Berrick, Jill Duerr. 2018. *The Impossible Imperative: Navigating the Competing Principles of Child Protection*. New York, NY: Oxford University Press.

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

Berzin, Stephanie Cosner. 2010. "Understanding Foster Youth Outcomes: Is Propensity Scoring Better Than Traditional Methods?" *Research on Social Work Practice* 20 (1): 100–111.

Billings, Stephen B. 2019. "Parental Arrest and Incarceration: How Does It Impact the Children?" Unpublished.

Bosk, Emily Adlin. 2015. "All Unhappy Families: Standardization and Child Welfare Decision-Making." PhD diss. University of Michigan.

CEPI. 2017. "Individual K-12 Student Records." Michigan Consortium for Education Research. (accessed September 20, 2017).

Chetty, Raj, Nathaniel Hendren, and Lawrence F. 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.

Child Trends. 2017a. "Michigan Foster Care, Federal Fiscal Year 2015." https://www.childtrends.org/wp-content/uploads/2017/01/Michigan-Foster-Care-Factsheet_2015.pdf.

Child Trends. 2017b. "State-Level Data for Understanding Child Welfare in the United States." Child Trends. https://www.childtrends.org/publications/state-level-data-for-understanding-child-welfare-in-the-united-states (accessed August 25, 2018).

Child Trends. 2018. "Child Trends Databank." Child Trends. https://www.childtrends.org/indicators/foster-care.

Child Welfare Information Gateway. 2016. *Reunification: Bringing Your Children Home from Foster Care*. Washington, DC: United States Department of Health and Human Services.

Child Welfare Information Gateway. 2017. *Child Maltreatment Prevention: Past, Present, and Future.* Washington, DC: United States Department of Health and Human Services.

Children's Bureau. 2003. *Child Welfare Outcomes*. Washington, DC: United States Department of Health and Human Services.

Children's Bureau. 2004. *Child Welfare Outcomes 2001: Annual Report*. Washington, DC: United States Department of Health and Human Services.

Children's Bureau. 2017. "Child Welfare Outcomes Report Data." United States Department of Health and Human Services. https://cwoutcomes.acf.hhs.gov/cwodatasite/byState.

Children's Bureau. 2018a. *Child Maltreatment 2016*. Washington, DC: United States Department of Health and Human Services.

Children's Bureau. 2018b. *The AFCARS Report: Preliminary FY 2017 Estimates*. Washington, DC: United States Department of Health and Human Services.

Children's Bureau. 2018c. "Child Welfare Outcomes Report Data." United States Department of Health and Human Services. https://cwoutcomes.acf.hhs.gov/cwodatasite/recurrence/index.

Children's Bureau. 2020. *Child Maltreatment 2020*. Washington, DC: United States Department of Health and Human Services.

Chyn, Eric. 2018. "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children." *American Economic Review* 108 (10): 3028–56.

Collinson, Robert, and Davin Reed. 2019. "The Effects of Evictions on Low-Income Households." Unpublished.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Quarterly Journal of Economics* 129 (4): 1711–52.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1 (3): 111–34.

Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. 2016. "School Accountability, Postsecondary Attainment, and Earnings." *Review of Economics and Statistics* 98 (5): 848–62.

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

Doyle, Joseph J., Jr. 2008. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." *Journal of Political Economy* 116 (4): 746–70.

Frandsen, Brigham R., Lars J. Lefgren, and Emily C. Leslie. 2019. "Judging Judge Fixed Effects." NBER Working Paper 25528.

Goodman, Joshua. 2014. "Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time." NBER Working Paper 20221.

Grimon, Marie-Pascale. 2020. "Effects of the Child Protection System on Parents." Unpublished.

Gross, Max, and E. Jason Baron. 2022. "Replication data for: Temporary Stays and Persistent Gains: The Causal Effects of Foster Care." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E127441V1.

Humphries, John Eric, Nicholas S. Mader, Daniel I. Tannenbaum, and Winnie L. van Dijk. 2019. "Does Eviction Cause Poverty? Quasi-Experimental Evidence from Cook County, IL." NBER Working Paper 26139.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Jacob, Brian A. 2004. "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago." *American Economic Review* 94 (1): 233–58.

Kawano, Laura, Bruce Sacerdote, William Skimmyhorn, and Michael Stevens. 2017. "On the Determinants of Young Adult Outcomes: An Examination of Random Shocks to Children in Military Families." Unpublished.

KCDC. 2015. "Children in Foster Care with More than Two Placements in the United States." Kids Count Data Center. https://datacenter.kidscount.org/data/tables/8822-children-in-foster-care-with-more-than-two-placements#detailed/1/any/false/1729,37,871,870,573,869,36,868,867,133/any/17680,17681 (accessed August 25, 2018).

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96 (3): 863–76.

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

Kling, Jeffrey R., Jens Ludwig, and Lawrence 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.

Kolesár, Michal. 2013. "Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity." Unpublished.

MDE. 2017. "Individual K-12 Student Records." Michigan Consortium for Education Research. (accessed September 20, 2017).

MDHHS. 2017. "Individual Child Maltreatment Complaints and Placement Records." Child and Adolescent Data Lab. (accessed September 20, 2017).

Mueller-Smith, Michael. 2015. "The Criminal and Labor Market Impacts of Incarceration." Unpublished.

Neilson, Susie. 2019. "More Kids Are Getting Placed in Foster Care Because Of Parents' Drug Use." *NPR*, July 15. 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, Samuel. 2019. "Examiner Inconsistency: Evidence from Refugee Appeals." Unpublished.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver. 2019. "The Effects of Parental and Sibling Incarceration: Evidence from Ohio." Unpublished.

O'Brien, Peter C. 1984. "Procedures for Comparing Samples with Multiple Endpoints." *Biometrics* 40 (4): 1079–87.

Pears, Katherine, and Philip A. Fisher. 2005. "Developmental, Cognitive, and Neuropsychological Functioning in Preschool-Aged Foster Children: Associations with Prior Maltreatment and Placement History." *Journal of Developmental and Behavioral Pediatrics* 26 (2): 112–22.

Pecora, Peter J., Ronald C. Kessler, Kirk O'Brien, Catherine Roller White, Jason Williams, Eva Hiripi, Diana English, James White, and Mary Anne 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): 1459–81.

Roberts, Kelsey V. 2019. "Foster Care and Child Welfare." PhD diss. Clemson University.

Ryan, Joseph P., and Mark F. Testa. 2005. "Child Maltreatment and Juvenile Delinquency: Investigating the Role of Placement and Placement Instability." *Children and Youth Services Review* 27 (3): 227–49.

Ryan, Joseph P., Brian A. Jacob, Max Gross, Brian E. Perron, Andrew Moore, and Sharlyn Ferguson. 2018. "Early Exposure to Child Maltreatment and Academic Outcomes." *Child Maltreatment* 23 (4): 365–75.

Sanbonmatsu, Lisa, Jeffrey R. Kling, Greg J. Duncan, and Jeanne Brooks-Gunn. 2006. "Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment." *Journal of Human Resources* 41 (4): 649–91.

SCAO. 2015. "Individual Juvenile Delinquency Petitions." Child and Adolescent Data Lab. (accessed August 24, 2019).

Scherr, Tracey G. 2007. "Educational Experiences of Children in Foster Care: Meta-Analyses of Special Education, Retention and Discipline rates." *School Psychology International* 28 (4): 419–36.

Talbot, Margaret. 2017. "The Addicts Next Door." *New Yorker,* May 29. https://www.newyorker.com/magazine/2017/06/05/the-addicts-next-door.

Trout, Alexandra L., Jessica Hagaman, Kathryn Casey, Robert Reid, and Michael 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): 979–94.

Warburton, William P., Rebecca N. Warburton, Arthur Sweetman, and Clyde 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.

Wildeman, Christopher, and Natalia Emanuel. 2014. "Cumulative Risks of Foster Care Placement by Age 18 for U.S. Children, 2000–2011." *PLoS ONE* 9 (3): e92785.

Wiltz, Teresa. 2018. "This New Federal Law Will Change Foster Care as We Know It." *PEW,* May 2. 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.

Wulczyn, Fred, Cheryl Smithgall, and Lijun Chen. 2009. "Child Well-Being: The Intersection of Schools and Child Welfare." *Review of Research in Education* 33 (1): 35–62.

Wulczyn, Fred H., Kristen Brunner Hislop, and Robert M. Goerge. 2000. Foster Care Dynamics 1983–1998. Chicago, IL: Chapin Hall Center for Children at the University of Chicago.

Zlotnick, Cheryl, Tammy W. Tam, and Laurie 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): 534–40.