

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

Max Gross*

October 21, 2019

Latest version available [here](#).

Abstract

6% of children in the United States enter the foster system between birth and age eighteen. Using administrative data from Michigan, I estimate the effects of foster care on children's outcomes by exploiting the quasi-random assignment of child welfare investigators. I find that foster care reduced the likelihood of being abused or neglected in the future by 50%, increased daily school attendance by 6%, and improved math test scores by 0.34 standard deviations. Gains in safety and academics emerged after children exited the system when most were reunified with their birth parents, suggesting that improvements made by birth parents were an important mechanism. I discuss implications for policy in light of the Family First Prevention Services Act of 2019, which allocates billions of federal dollars toward reducing foster placements.

*University of Michigan, Department of Economics. maxgross@umich.edu.

I thank Brian Jacob, Michael Mueller-Smith, Joseph Ryan, Charlie Brown, and Kevin Stange for their invaluable advice and guidance. I also benefitted from conversations with Ashley Craig, Susan Dynarski, Sara Heller, George Fenton, Matthew Gross, Rachel Hildrich, Anirudh Jayanti, Parag Mahajan, Stephanie Owen, and Andrew Simon as well as feedback from seminar participants at the University of Michigan and APPAM. I 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, Andrea Plevck, and Nicole Wagner Lam for coordinating data access. I also thank the many child welfare employees across Michigan for help understanding how the system works in practice and for bringing humanity to the data. This research was supported in part by the Institute of Education Sciences, U.S. Department of Education through PR/Award R305B150012# and Grant R305E100008. I 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.

“Few professionals (if any) believe that large proportions of American children should be taken from their parents, and few professionals (if any) believe that children should be kept at home in dire circumstances...The debate is not, and should not, be at the ends of the continuum. The divisions typically are animated not in the cases that are black and white, but in the cases that occupy the center, gray area of child welfare.”

— Dr. Jill Duerr Berrick, The Impossible Imperative, 2018

1 Introduction

About 250,000 children entered the foster system every year in the United States from 2000 to 2017 because they were abused or neglected at home (??). By age eighteen, 6% of children—including over 10% of black children—will have entered foster care (?). Among historically vulnerable groups, foster children experience the worst life outcomes (?). Despite this, there is very little causal evidence on the impacts of foster care. Early pathbreaking research in ?? studied placements nearly two decades ago, concluding that foster care was damaging for children, yet child welfare policy has changed over time (?). Especially given its increased use in response to the opioid epidemic (??), it is critical to understand the effectiveness of current foster care systems.

This paper makes several contributions. First, it provides new estimates of the causal effects of foster care on crucial indicators of child wellbeing: 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, I 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, which 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.

I 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% reduction relative to a baseline mean of 25.5%. In addition to improving child safety, placement had meaningful, positive impacts on academic outcomes; it increased daily school attendance by 6.0% and standardized math test scores by 0.34 standard deviations. I also find a substantial yet less precise reduction in juvenile delinquency. These estimates represent effects in cases where investigators might disagree about placement, which is a critical population for child welfare policy (?).

The results contrast ?? which find that foster care reduced earnings and increased crime

for Illinois children investigated in the 1990s and early 2000s.¹ Though the system in Illinois caused great harm, it was an outlier; placement in other states at the time may have been less damaging or even beneficial. Foster children in Illinois remained in the system for a longer amount of time than any other state in the 1990s, and they changed foster homes at a higher rate than all but two (Figure ??). Importantly, Michigan's foster system functions similarly to others across the country.

The second contribution of this study is that it explores mechanisms by exploiting the temporary nature of foster care. In my research setting, children were in the system for nineteen 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. One explanation for this surprising pattern is that birth parents, who worked closely with social workers following child removal, improved their parenting skills. Accordingly, I 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. I find little evidence for alternative explanations, such as lasting improvements to children's neighborhood or school quality, which have been associated with better outcomes (??).

The third contribution of this paper is that it examines an understudied channel through which foster placement can impact children: interventions for birth parents. Following child removal, birth parents received both community services, like referrals to local drug rehabilitation groups or food pantries, as well as more intensive, targeted services, like substance abuse treatment or parenting classes. It is challenging to disentangle the effects of adult services on foster children's outcomes from the dramatic changes occurring in children's lives because they both happen at the same time. As it turns out, though, quasi-randomly assigned investigators could recommend light-touch or intensive services to adults even if their children were not removed, creating useful comparison groups. Therefore, I study the effectiveness of adult interventions when children remain in the home by exploiting investigator tendencies over various adult programs. I find suggestive evidence that those effects were small relative to the combined impact of adult interventions together with child removal. A caveat of this exercise is that it represents the effects of adult programs for cases where investigators might disagree about recommending services, not for children at-risk of foster placement. However, to the extent that services have similar impacts for struggling

¹They also differ from a sizable correlational literature which tends to find a negative association between removal and children's outcomes (????????). Interestingly, however, they are consistent with recent evidence on parental incarceration in the United States from North Carolina (?) and Ohio (?), which is a somewhat analogous form of family separation.

families, the results indicate that child removal enhanced the efficacy of adult interventions, perhaps through increased incentives to comply or temporary relief from parenting.

The fourth contribution of this study is that it empirically documents that a common form of incomplete data coverage can substantially bias estimates from the examiner assignment design. Specifically, this paper improves on contemporaneous studies from Rhode Island (?) and South Carolina (?) which offer quasi-experimental estimates of foster care, yet do not follow children from the start of their child welfare investigation.² The data in these studies contain only the subset of substantiated allegations, which represent just 40% of the caseload in Rhode Island and 25% in South Carolina (?). Since the same investigator who determines foster placement also makes subjective decisions around substantiation, the set of children in this censored data need not be balanced across investigators even if their cases were initially assigned at random.³ I replicate the primary analysis using only the sample of substantiated investigations and find estimates much smaller than the true effects. Furthermore, I find that the method proposed in ? to identify impacts when restricted to censored data does not resolve bias in the current context.

This study is especially relevant given the dramatic changes to child welfare policy introduced in the Family First Prevention Services Act of 2019. The legislation makes reducing the use of foster care a federal priority; it allows states to redirect up to eight billion federal dollars from foster care toward alternatives to placement (?). I find that foster care can be a useful tool in some instances, however, suggesting that this policy may not have its intended effects.

The rest of this paper is organized as follows. Section ?? details the child welfare investigation process in Michigan and describes the state’s foster system. Section ?? introduces the sources of administrative data and the analysis sample. Section ?? outlines the research design. Section ?? reports the main findings and explores mechanisms, and Section ?? describes bias from incomplete data coverage. Section ?? concludes and discusses implications for public policy. Two appendices, print appendix “A” and online appendix “B,” provide supplemental results.

²? studies about 12,000 children between zero and seventeen years old and finds substantial gains for girls younger than six years old but imprecise null effects for other gender-age groups. ? examines about 17,000 children between age two and seventeen and finds noisy estimates on daily school attendance and test scores. Interestingly, however, removal increased on-time grade progression for younger children, but increased juvenile delinquency for older ones.

³Similar data constraints appear in studies of crime and education policy which use the same research design.

2 Overview of Child Welfare in Michigan

About one in five public school students in Michigan were the subject of a formal investigation of child abuse or neglect by third grade (?). One in ten were the subject of more than one investigation and one in sixty experienced foster placement.⁴ This section reviews the maltreatment investigation process in Michigan and describes the state’s foster system.

2.1 Child Maltreatment Investigations

Figure ?? describes the maltreatment investigation process in Michigan, which is similar to most other states. It begins with a call to an intake hotline to report child abuse—e.g., bruises, burns, or sexual abuse—or neglect—e.g., 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.⁶ The office assigns the report to a maltreatment investigator who has 24 hours to begin an investigation, 72 hours to establish face-to-face contact with the alleged child victim, and 30 days to complete the investigation.

Critical to my 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.⁷ Since the investigator assignment system occurs within each local office, and within local geography areas in some larger counties, all of the analysis includes zip code by investigation year fixed effects to compare only children who could have been assigned the same investigator.⁸

⁴These statistics reflect the authors’ calculations using the same sample as (?), which consists of over 700,000 third grade students born between 2000-2006. The cumulative risk of placement statistic in Michigan by third grade is smaller than the nationwide estimate in ? for a number of reasons: it reports placement by third grade rather than by age eighteen, Michigan removes children at a rate slightly lower than the national average, it represents only the population of public school students rather than the universe of children, and it follows the same students from birth to third grade rather than using a synthetic life table approach.

⁵While anyone can call the hotline to report suspected maltreatment, the most frequent reporters are people who are mandated by law to do so, such as education personnel, police officers, and social service workers. The intake process is the same regardless of the reporter.

⁶Reports are screened out if, for example, the perpetrator is younger than eighteen years old or the victim is older than eighteen. I observe only screened-in reports, which will not affect the validity of the research design since investigator assignment occurs after this initial screening.

⁷Though investigators may vary slightly in completion time, even those who take somewhat longer are assigned new reports according to the rotation. In fact, despite a legal maximum caseload size of twelve instituted in 2013, two-thirds of investigators reported having a caseload of thirteen or greater after 2014 (?).

⁸There are two exceptions to the rotational assignment of investigators, which I exclude from my analysis. First, given their sensitivity, reports of sexual abuse tend to be assigned to more experienced investigators. Second, new reports involving the same child as a recent prior report are usually assigned to the original

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. 75% of reports in 2016 went unsubstantiated (?, Tables 3-1 and 3-3), and child welfare offices do not follow up with the family further.

Second, investigators decide how much risk the child faces by continuing to live in their home. They complete a risk assessment that consists of 22 questions to compute a risk score. Many of the items require simple yes or no answers, such as “primary caretaker able to put child’s needs ahead of own” and “primary caretaker views incident less seriously than the department.” Even with guidance on how to interpret these questions, some are inherently subjective. Moreover, ? offers detailed qualitative evidence that investigators often manipulate their responses to ensure a risk score that matches their prior. Therefore, even with a standardized system in place, investigators yield immense discretion over foster placement.

Investigator judgment over both evidence and risk jointly determine the outcome of the investigation. If the investigator substantiates the allegation and the risk level is low, they must refer the family to community-based services like food pantries, support groups, or other local non-profits. These cases require no further follow-up. If the investigator substantiates the allegation and the risk level is high, the family also receives targeted services, such as substance abuse treatment, parenting classes, or counseling. Local, state, and federal fundings, from Title IV-E, cover the costs of these targeted services. Lastly, substantiated allegations with intensive risk not only trigger targeted and community services but also require the investigator to file a court petition for child removal.⁹ The main analysis in this study examines the combined effects of child removal and adult interventions on children’s outcomes, yet additional analysis explores their individual contributions.

2.2 Foster Care System

Foster care is a family intervention; children are temporarily removed from their homes while birth parents receive services to improve their parenting. Removal occurs quickly; just

investigator since they have familiarity with the family. Anecdotally, such reports tend to re-enter the rotation after a few months pass. I exclude those within one year of a prior investigation from the analysis to be conservative.

⁹Unlike investigators who no longer work with the family after completing the investigation, the same judge may preside over the family’s case and ultimately determine whether the child is reunified. Since this presents a violation of the exclusion restriction, the research design in this study leverages investigator discretion rather than judge discretion.

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.¹⁰ In many cases, though, children do not have suitable relatives available. In 2015, 41% of foster children in Michigan were living with an unrelated family, 35% lived with relatives, 9% lived in group homes or institutions, and 14% lived in other settings, such as pre-adoptive homes or supervised independent living.¹¹ It is common to switch foster homes while in the system—60% of children lived in more than one setting, and 17% lived in at least four. Michigan looks very similar to the rest of the country along these statistics (?).

After placement, child welfare caseworkers meet with birth parents to create a reunification plan stating the conditions under which the child can return home. These plans might require the parent to secure housing, overcome drug addiction, or keep enough food in the home. Birth parents receive targeted services to address the challenges in their own lives, which can include substance abuse treatment, parenting classes, counseling, and job training.¹² Caseworkers monitor their progress and make changes to the reunification plan as needed. Family reunification occurs if the parents make enough progress in convincing a court that their child is safe to return home.

Ultimately, children in Michigan, including those outside of the analysis sample, spend 17 months in the system, on average, after which 47% were reunified with their birth parents, 34% were adopted or had legal guardianship transferred, while 9% reached age eighteen and were emancipated.¹³

¹⁰There is limited causal evidence on the effects of each placement type and the instrumental variables design in this study cannot separately identify each effect. However, OLS analysis in Supplemental Appendix ?? finds a larger positive association between kinship placement and children’s outcomes than other placement types.

¹¹There is limited data available both nationwide and in Michigan on who takes in foster children. Estimates from the American Community Survey, which have known limitations, suggest that households with foster children tend to be larger and lower-income than other households with at least one member younger than 18 years old. Supplemental Appendix ?? provides summary statistics and discusses the limitations of using ACS data to identify families with foster children.

¹²Though limited local supply or high adult demand may constrain access to these services—e.g., there may be a shortage of providers or long waitlists for care—caseworkers do their best to meet the needs of their families and sometimes have priority access.

¹³The remaining 10% fell into less common permanency categories, such as informal guardianship with relatives, incarceration, or transfer to another agency.

3 Data Sources and Sample Construction

3.1 Administrative Data Sources

This study uses administrative data from the Michigan Department of Health and Human Services (MDHHS), Michigan Department of Education (MDE), Center for Educational Performance and Information (CEPI), and Michigan State Court Administrative Office (SCAO) to test the effects of foster placement on children’s outcomes. Since there is no common identifier, these files were linked using a probabilistic matching algorithm based on first name, last name, date of birth, and gender. Overall, 84% of child welfare investigations of school-age children matched to a student enrolled in public school in the year of their investigation. The match rate is quite high given that many investigated children should not have matched to an enrolled public school student, e.g., children in private school or homeschool, high school dropouts, and even children who were not permanent state residents. Specifically, I estimate that just 87.1% would be predicted to match to a currently enrolled student.¹⁴ Supplemental Appendix ?? describes the match process and match rate in greater detail.

Child welfare data from MDHHS consists of the universe of maltreatment investigations in Michigan between August 1996 and July 2017. It includes details of each investigation, such as the allegation report date, allegation types as coded by the investigator, the child’s zip code, substantiation, and foster placement. Conditional on placement, it contains limited information on placement settings and permanency outcome—e.g., reunified with birth parents, adopted, etc. Critical to my analysis, the files also include unique investigator identifiers beginning in 2008.

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

Juvenile justice data from SCAO includes all juvenile court petitions filed in almost every county in Michigan between 2008 and 2015. A court petition is an official document filed following juvenile arrest in cases where youth are not immediately diverted from the courts.

¹⁴I estimate that the remaining 12.9% consist of private school students in Michigan (4.6%), non-Michigan residents (3.4%), homeschool students in Michigan (2.6%), and students who dropped out of high school in Michigan (2.1%).

Petitions can be dismissed by the court after filing and need not indicate that there was ever a formal court hearing. The SCAO data covers 75 of Michigan’s 83 counties, including Detroit and the metro-Detroit area yet excluding the following five urban and three rural counties: Kent, Washtenaw, Ingham, Ottawa, Kalamazoo, Berrien, Delta and Keweenaw.¹⁵ I exclude the 19% of investigated children who lived in these eight counties from my analysis of juvenile delinquency, yet the estimates on juvenile crime are similar if I include them.

Using these administrative data sources, I assess the effects of foster care on child wellbeing across three dimensions: safety, schooling, and crime. First, I capture child safety as whether they were the alleged victim in a subsequent maltreatment investigation and whether they were a confirmed (substantiated) victim in a subsequent investigation. Second, I examine schooling by studying daily attendance rates, grade retention, 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, and grade retention is a binary indicator equal to one if the student repeated the previous year’s grade level. 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.¹⁶ Finally, I measure juvenile crime as the filing of a court petition.

Given that I study a variety of outcomes, I construct an index of child wellbeing according to ? to create a summary measure, increase statistical power, and help address multiple hypothesis testing. The index consists of unweighted means of standardized versions of the seven primary outcomes. I reverse code “bad” outcomes and impute missing values according to group means.

3.2 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. I exclude cases where investigators were unlikely to have been quasi-randomly assigned: allegations of sexual abuse and allegations involving children from a recent prior report. I also restrict the sample to children enrolled in grades one through eleven in the school year of their investigation to observe baseline characteristics and at least one follow-up year.¹⁷ Overall, I focus on 242,233 investigations

¹⁵These counties include three of the state’s ten most populated cities: Grand Rapids, Lansing, and Ann Arbor, and three more of the top thirty: Kalamazoo, Wyoming, and Ypsilanti.

¹⁶These educational outcomes are included in the analysis only if they occur after a child’s investigation. That is, I 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 available.

¹⁷The analysis sample excludes children who were too young to have entered school at the time of their investigation. Though these younger children appear in the child welfare data and, years later, may appear in public school records, I find that foster placement caused a large and statistically significant reduction in

of 186,250 students and follow students for at most nine years after their investigation.

Table ?? describes the sample. Column 1 consists of all public school students in Michigan during the 2016-2017 school year, while column 2 consists of investigated children in the analysis sample. Black and low-income children were disproportionately involved in the child welfare system; 29% of investigations were of black children and 83% were of low-income children, despite making up just 21% and 49% of the population respectively. Children with child welfare involvement had noticeably lower baseline daily attendance rates and scored about a quarter of a standard deviation worse on standardized math and reading tests. Column 3 describes children involved in the 2% of investigations that resulted in foster placement. Relative to the overall sample in column 2, foster children were especially disproportionately black and low-income. They also had much lower daily attendance rates—74% compared to 81%—and scored about a tenth of a standard deviation lower on math and reading tests. Overall, these descriptive statistics caution against offering a causal interpretation to mean comparisons between investigated children who were and were not removed.

4 Empirical Strategy

A naive analysis of foster care might regress a child outcome, like daily school attendance or their score on a standardized test, 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 are likely biased because foster children differ from those who were not removed along unobservable dimensions. 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 placements and overstate the costs.

4.1 Research Design

In order to overcome omitted variable bias, I use the examiner assignment research design, which has been applied to other studies of foster care (??) as well as incarceration (???), disability insurance (?), and evictions (??). Specifically, I instrument for placement using the removal tendencies of quasi-randomly assigned investigators. By chance, children assigned

the likelihood that they ever enrolled in a Michigan public school. Perhaps they differentially moved out of state or enrolled in private schools. Importantly, I find no evidence of differential attrition out of Michigan public schools for currently enrolled students (Table ??). Table ?? describes the sample restrictions in greater detail.

to especially strict investigators—those with high propensities to remove—were more likely to enter foster care than they would have been if they happened to be assigned someone more lenient.

In order to extract signal from noise in a measure of removal tendencies, I 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.¹⁸ This leaves 3,073 investigators assigned 315 cases on average. Following the literature, I calculate the instrument as the fraction of all other investigations, both past and future, assigned to the same investigator that resulted in foster placement. Specifically, for investigation i assigned to investigator w :

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

where n_w equals the total number of cases assigned to investigator w and FC_k is an indicator equal to one if investigation k resulted in foster care.¹⁹ 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 tendencies. Crucial to the research design, there is variation even within local offices (Figure ??). An investigator at the 10th percentile removed at a rate 2.1 percentage points less than their colleagues while someone at the 90th percentile removed at a rate 2.4 percentage points greater. Relative to the average removal rate of 3%, this represents a 150% increase in the likelihood of foster placement.

I use the following instrumental variables specification to measure the causal effects of foster care:

$$FC_{iw} = \gamma_0 + \gamma_1 Z_{iw}^R + \gamma_2 X_{iw} + \theta_r + \eta_{iw} \quad (2)$$

$$Y_{iw} = \beta_0 + \beta_1 \hat{FC}_{iw} + \beta_2 X_{iw} + \theta_r + \epsilon_{iw} \quad (3)$$

where Y_{iw} is a child outcome, such as their daily school attendance rate or their score on a standardized math test, and X_{iw} is a vector of baseline covariates which includes a variety of socio-demographic and academic characteristics.²⁰ θ_r represent child zip code by

¹⁸Table ?? shows that the results are robust to both larger and smaller thresholds.

¹⁹There are other reasonable ways to measure removal stringency. For example, this approach does not allow for removal tendencies to change over time. Section ?? describes several alternatives and shows that the results are robust across measures.

²⁰Specifically, it includes controls for socio-demographic features, such as gender, grade level fixed effects, race/ethnicity, and free or reduced-price lunch receipt. It also controls for baseline academic characteristics measured in the year before the investigation, including attendance rate and receipt of special education supports, as well as an indicator for ever retained in grade. It flexibly controls for a student's most recent

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.²¹ There are 7,534 unique rotation groups, consisting of thirteen investigators on average. Finally, I cluster standard errors at the investigator level.²²

$\hat{\beta}_1$ is the local average treatment effect (LATE) of foster placement where compliers are children for whom investigators might disagree about removal. Given likely heterogeneous treatment effects, this study cannot speak to how foster care influences always takers—children so clearly in danger at home that all investigators would remove—and never takers—those so clearly safe that no investigators would remove. Even so, compliers represent the population most relevant for child welfare policy (?). Following the methodology in ?, I find that compliers made up 5% of investigated children in the sample. They were younger than the average foster child—61% were ten years old or younger at the start of their investigation relative to just 51% of foster children overall, yet otherwise look similar in terms of demographic and baseline academic characteristics (Table ??).

4.2 Identifying Assumptions

Three assumptions are necessary for the LATE to be unbiased.

1. Relevance: $\gamma_1 \neq 0$. The instrument must predict foster placement. Table ?? shows the first stage regression of foster placement on the removal stringency instrument. The correlation between the instrument and foster care is 0.48 (column 1) and a one standard deviation increase in removal stringency increased the likelihood of placement by about one percentage point (column 4). The F-statistic of 439 indicates that there is not a weak instruments problem.

baseline standardized math and reading test scores by including linear, quadratic, and cubic terms, as well as the interaction of baseline math and reading performance. It consists of some information about the maltreatment report, such as whether the allegation was for physical abuse or neglect, the child’s relation to the perpetrator, and an indicator for whether the child was previously the subject of an investigation. Furthermore, it controls for characteristics of the school that the child attended during the investigation, such as indicators for whether they were enrolled in a charter or an urban school, the fraction of white, black and Hispanic students, and the fraction who were eligible for free or reduced-price lunch. It also controls for characteristics of the child’s neighborhood, as defined by their census block group, including median household income, employment rate, the fraction of adults with at least a bachelors degree, the fraction of residents that were white, black and Hispanic, an indicator for whether the child experienced homelessness, and the number of times the child moved neighborhoods. Lastly, it includes indicators for any missing covariates.

²¹Child welfare staff from several local offices explained that some investigators only work in the northern part of the county while others only work in the south, for example. However, such geographical boundaries are neither publicly available nor observed in administrative data. Importantly, Table ?? shows that the results are robust to instead defining rotational assignment at the child county by investigation year level.

²²The results are robust to clustering standard errors at the child level or using two-way clustering at the investigator and child level. Results are available upon request.

2. Exclusion: $\mathbb{E}[Y|FC, X, \theta_r, Z_1^R] = \mathbb{E}[Y|FC, X, \theta_r, Z_2^R]$. The instrument can only influence outcomes through foster placement. Though inherently untestable, the quasi-random assignment of investigators lends credence to this assumption; a rich set of socio-demographic characteristics and baseline academic measures do not predict the instrument despite being highly predictive of placement (Table ??). As further evidence, the first stage F-statistic in Table ?? is stable with the inclusion of covariates and the instrument is unrelated to the number of cases that investigators were assigned, which is a useful proxy for experience or thoroughness (Table ??).

A potential concern is that investigators might have influenced children’s experiences in the foster system, conditional on placement. However, investigators did not work with children after the investigation; cases that required follow-up were transferred to other child welfare caseworkers. Accordingly, the instrument does not predict the initial placement setting or the number of days spent in the system (Table ??). Moreover, though in theory, stricter investigators could have argued for different or better services, in practice, the initial set of services is relatively standardized for families who face similar challenges.

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

There are two common ways to probe the average monotonicity assumption. First, the first stage should be non-negative for all subgroups (?), which holds for gender, race/ethnicity, age, and prior child welfare involvement groups (Table ??, Panel A).²³ Second, investigators who were strict for certain groups should also have been strict for others (?). For example, amidst serious concerns of racism in maltreatment investigations (?), monotonicity asserts that investigators who were particularly likely to remove children of color should also have been weakly stricter in their investigations of white children than their colleagues. In support of the assumption, the first stage remains positive and statistically significant when I re-calculate the instrument as a leave-subgroup-out measure (Table ??, Panel B).

²³I do not create groups based on the type of maltreatment such as abuse or neglect because investigators code this information after they begin their investigation. To the extent that different types of maltreatment are related to observable child subgroups, the exercise offers an indirect test for non-monotonic tendencies based on these features.

5 Causal Effects of Foster Care on Children’s Outcomes

Table ?? shows the effects of foster care on several critical indicators of child wellbeing covering the areas of safety, education, and crime. It reports the results from both the OLS and 2SLS models using panel data spanning all of the school years following a child’s investigation.²⁴ The OLS results suggest that removal had a near-zero impact on the index of child wellbeing.²⁵ In contrast, the 2SLS estimate reveals that removal improved the wellbeing index by 16.4% of a standard deviation, an effect statistically significant at the 5% level.

Two expected patterns stand out from this comparison. First, the OLS estimate is smaller than the 2SLS estimate, suggesting that unobserved features, like the severity of maltreatment, for example, lead OLS to understate the benefits of removal. Second, the control mean, the mean outcome among children who were not removed, is larger than the control complier mean, the estimated outcome for compliers who were not removed. Specifically, the control complier mean is 6% of a standard deviation less than the control mean, indicating that children at-risk of placement were worse off by remaining in the home than the average investigated child. While the index provides a useful summary, I turn to the effects on each of the seven components next in order to understand what drives the improvement as well as more easily interpret magnitudes.

5.1 Effects on Child Safety, Academics, and Crime

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

Although these effects may represent a reduction in reporting behavior without a change in underlying safety, the data does not support this interpretation. For example, perhaps teachers were less likely to report bruises to child welfare if students were in foster care. If true, it follows that the reported incidents of abuse or neglect against foster children would be more severe than those against children who were not removed. However, foster

²⁴Specifically, I construct an unbalanced panel at the investigation-school year level and restructure non-educational outcomes to follow the school year calendar. For example, I define maltreatment reports and juvenile petitions occurring between September 2010 and August 2011 as the 2010-2011 school year. Children age out of the panel for certain outcomes—e.g., the age of majority in Michigan is sixteen years old, so seventeen-year-olds are ineligible for the juvenile delinquency outcome.

²⁵As detailed in Section ??, the index is created according to ? and represents a summary measure of overall child wellbeing.

placement did not influence the likelihood of substantiation among children with a subsequent investigation. Moreover, caseworkers that were mandated by law to report suspected maltreatment visited foster children regularly, both during their time in the system and after they exited, suggesting that actual maltreatment against foster children should have been reported (?).

Consistent with the improvement in child safety, I also find large gains in academic outcomes. Removal increased daily attendance rates by 5.4 percentage points. Relative to a 180 day school year, this is equivalent to showing up for ten additional days of school. I also estimate that foster children were less likely to be retained in grade, though the effect is imprecise. Furthermore, removal had a substantial positive effect on standardized math test scores of about one-third of a standard deviation.²⁶ This estimate is statistically significant at the 10% level, yet I can rule out decreases greater than 6% of a standard deviation. In addition, while the point estimate on standardized reading scores is positive and substantively large—about half the size of the effect on math—it is not statistically significant. Since math skills are considered more malleable at older ages, this is not particularly surprising.²⁷

Lastly, I examine the effect of removal on juvenile delinquency, defined by the filing of a juvenile court petition. The point estimate suggests a large decrease in juvenile crime—a 55% drop relative to a control complier mean of 5.1%—yet the estimate is imprecise. Overall, the results consistently suggest that foster care improved children’s outcomes.

5.2 Mechanisms

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

²⁶As a benchmark, ? 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.

²⁷Removal did not influence the likelihood of taking standardized tests (Table ??). In addition, Table ?? shows the effects on high school graduation and college enrollment. However, the sample of children old enough to be eligible for these outcomes is small and the analysis can not rule out considerable positive or negative effects.

5.2.1 Evidence from the Timing of Impacts

40% of children who were removed had exited the foster system after one year and nearly all had exited after two years (Figure ??).²⁸ I construct an index of neighborhood and school characteristics to explore the effects of placement on childhood environment. The index, constructed according to ?, consists of three neighborhood components: median household income, the fraction of adults with a bachelor’s degree, and employment rate, and two school components: average math and reading test scores and share of free or reduced-price lunch eligible students. There was a large and statistically significant increase in the index during the first year after placement (Table ??, Panel A).²⁹ Given that moving to lower-poverty areas can improve child wellbeing (???), such exposure might lead to contemporaneous gains in children’s outcomes. However, there were no discernible differences in year one outcomes between children placed and not placed in foster care (Table ??, Panel A).³⁰

Nearly all marginal foster children had exited the system after two years and the vast majority, 85%, were reunified with their birth parents.³¹ Upon exiting, foster children returned to similar neighborhoods and schools as untreated compliers; I do not detect differences in the characteristics of their neighborhoods or schools after the first year (Table ??, Panel B). Despite this, gains in safety and academic outcomes emerged several years after removal. Specifically, the index of child wellbeing increased by 19% of a standard deviation across all years after the first, driven by gains in safety, daily attendance rates, and standardized math scores (Table ??, Panel B). Figure ?? shows the effects separately by year, revealing that there were steady improvements in most outcomes that persist for several years. For example, the likelihood of being the victim of maltreatment began to decrease only after four years and continued to decrease every year for three more.³²

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

²⁸They spent nineteen months in foster care on average, consisting of eleven months living with relatives, six months living with an unrelated family, and two months living in group homes or institutions (Table ??).

²⁹This was driven by exposure to more highly educated neighborhoods and higher-income classmates (Table ??).

³⁰For ease of interpretation, I report only a set of the outcomes that were statistically significant at the 10% level from Table ?? in Table ?? and in all further analyses. Additional results are available upon request.

³¹Table ?? shows that of the remaining 15% who exited: 8% were adopted, 5% had guardianship transferred, and 2% turned eighteen years old and were emancipated.

³²These estimates represent time-since-treatment effects rather than age-of-treatment effects because all specifications include fixed effects for student grade level at the time of the investigation.

or strengthening parenting skills. They received fully funded services and met regularly with caseworkers and court officials to track their progress. I find evidence of such improvement. Perpetrators of child maltreatment, almost always a birth parent, were less likely to abuse or neglect children even years later if their initial child victim entered foster care (Figure ??).

Though I can not definitively rule out them out, I find little evidence for two alternative explanations of the pattern of impacts. First, though it is possible that moving to lower-poverty areas during placement improved child outcomes, well-identified studies of mobility find that effects increase with duration (??), yet exposure in this context was only temporary. Furthermore, the long-run benefits of moving do not run through schooling channels (??), yet foster care had large impacts on academics.

Second, it could be that children’s experiences while in foster care benefitted them only years later, i.e., foster care could trigger additional supports whose benefits take time to manifest themselves. In particular, if the costs related to family separation are high in the short run yet fade over time, even the benefits from channels that have more immediate impacts may appear only years later. However, I find no evidence that foster care increased supports in school either during placement or after exiting, as proxied by receipt of special education services (Table ??, Column 6). Moreover, while children may have benefitted from placement in other ways, perhaps through access to better counseling, new role models, or more nutritious meals, credible estimates of these channels for school-age children consistently find effects on test scores of less than one-tenth of a standard deviation, much smaller than the 0.34 standard deviation increase in math scores found in this study.³³

5.2.2 Evidence from Adult Interventions

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

³³See, for example, ? for the effects of counselors, ? for the effects of teacher role models, ? for the effects of summer jobs and mentors, ? for the effects of healthier meals, and ????? for the effects of greater meal availability.

To study the role of adult interventions, I exploit the fact that quasi-randomly assigned investigators had discretion over adult services in addition to child removal. As shown in Figure ??, investigators placed families on one of four tracks based on the strength of evidence that maltreatment occurred and the child’s risk of future harm: (1) no services, (2) community-based services, (3) community and targeted services, and (4) child removal plus community and targeted services.³⁴ As such, I create two new instruments according to Equation ??: investigator propensity to recommend community services alone and investigator propensity to recommend community and targeted services without child removal. I use these new measures, along with the main removal stringency measure, to simultaneously instrument for the three tracks that include adult services.³⁵

Table ?? shows the three distinct local average treatment effects estimated from this exercise. Although less precise, I find positive effects of child removal plus targeted and community services, relative to both types of services without child removal, that are nearly identical to the main analysis. The robustness of the main result confirms that, as expected, the removal stringency instrument nudged families from track three to track four. It also addresses a potential violation of exclusion in the form of omitted treatment bias (?), since the main effects appear not to be driven entirely by adult services. Also, though limited by statistical power, the individual impacts of targeted and community services are qualitatively smaller than the combined effect of both types of adult interventions with child removal. Importantly, while the main analysis studied cases where investigators might disagree about placement, this new exercise represents effects for children who faced lower risk and were not candidates for foster care.³⁶ However, to the extent that services have similar impacts for struggling families, this offers suggestive evidence that child removal was a crucial component of the foster care intervention.

There are at least two explanations as to why child removal may have been necessary for adult services to be effective. First, they may have lacked the required intensity. Unlike services offered to the birth parents of foster children, which lasted up to 24 months and were monitored by both the child welfare office and the courts, they lasted at most twelve months and participation was tracked only by the local child welfare office when the child was

³⁴While I observe track assignment in the child welfare records, I do not observe the specific types of services received, e.g., substance abuse treatment or parenting classes

³⁵The three instruments are positively, but not perfectly, correlated with each other, indicating that there is independent identifying variation from each. Within local office teams, the correlation between the removal instrument and propensity for community services alone is 0.14, between the removal instrument and tendency for targeted and community services is 0.24, and between the two non-removal instruments is 0.60.

³⁶For example, the LATE for community services identifies effects for families at the margin of receiving any services.

not removed. A second potential explanation, however, is that child removal was critical to ensure adult compliance. For example, temporary relief from parenting may have provided birth parents with the time and space needed to overcome challenges in their own lives. Child removal might also have increased adult incentive to engage with service programs. Though prior work highlights that services like drug rehabilitation or job training programs often have high failure rates overall (??), birth parents of foster children may have put in more effort than the average participant.

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

5.3 Further Empirical Exercises

5.3.1 Heterogeneity by Student and County Characteristics

I test whether the effects of foster placement varied across four subgroups: age, gender, prior child welfare involvement, and county urbanicity. Previous work highlights disparities in how children respond to environmental changes by gender and age, finding that male children are more vulnerable than females (?) and young children benefit from moving to lower-poverty areas more than older youth (??). In addition to studying effects by gender and age, I explore differences by whether the child was ever previously the subject of a maltreatment investigation in Michigan because it is a factor in determining removal. Lastly, I explore differences based on county urbanicity, defined according to the National Center for Health Statistics, because of the unique challenges to child welfare in rural communities, such as the shortage of medical service providers (?) and limited supply of foster homes (?).³⁷

Panel A of Table ?? shows that foster care improved outcomes for both male and female children; there were no statistically significant differences in outcomes by gender. There were notable differences by age, however (Panel B). Improvements in safety and attendance rates were driven by children ages ten and younger, while those ages eleven and older experienced more substantial test score gains. In terms of prior child welfare involvement, removal improved safety only for children who were the subject of a previous investigation, yet improved school daily attendance for those with no prior involvement (Panel C). Lastly, the

³⁷It is also worth noting that though LGBTQ youth are over-represented in foster care (?) and have especially traumatic experiences in the system (?), this study is unable to examine differences along this margin because the administrative data sources do not include information on sexual orientation or gender identity.

gains were qualitatively driven entirely by children in urban/suburban counties rather than rural areas, though the differences between these groups were not statistically significant (Panel D).³⁸

5.3.2 Robustness Checks

Table ?? shows that the main results are robust in both sign and magnitude to a variety of design decisions. I conduct the analysis using alternative samples (Panel A). First, I limit the sample to only the first investigation for every child. Next, I restrict the analysis to a balanced panel consisting of the first five follow-up years for students whom I could observe in the public school system for at least five years based on their grade level and year of investigation. Lastly, I test sensitivity to the number of cases an investigator must have been assigned to show up in the sample. The main analysis excludes children assigned to investigators who worked fewer than 50 cases, so I relax this threshold to 25 and also strengthen it to 75. The results are similar to those in the main analysis across all alternative samples.

I also check for robustness using other reasonable ways to measure investigator removal tendencies (Panel B). First, I randomly split the sample in half and define the instrument as the investigator’s removal rate from the other half of the sample. Second, I allow tendencies to vary over time by creating a leave-out-other-years measure. Third, I address concerns that removal decisions occurring around the same time may be correlated by constructing a leave-out-same-year measure. Lastly, I use an empirical Bayes shrinkage procedure, similar to measures of teacher-value added in ?, to more precisely estimate investigator tendencies.³⁹ Though they vary in precision, I find large, positive effects of foster care across all of the alternative instruments.

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

³⁸Overall, there were no clear patterns between subgroup outcomes and the amount of time that children spent in foster care or the types of foster homes that they lived in (Table ??).

³⁹Specifically, I randomly split the sample in half and create a shrunken measure using investigations from the other half of the sample. The procedure first regresses foster placement on investigation year fixed effects and investigator fixed effects and store the investigator fixed effect plus the residual term. I collapse the data to the investigator by year level, keeping the mean of this stored value for every cell. Then, I regress this stored value in year t for each investigator on their stored value in years $t - 2, t - 1, t + 1$, and $t + 2$, along with missing indicators where necessary. The shrunken stringency measure is the predicted value of this final regression.

investigation year fixed effects.

6 Potential Bias in Examiner Assignment Research Design from Censored Data

The examiner assignment design used in this study is increasingly popular as access to large administrative datasets allows researchers to exploit discretionary decision-making. It has been used to study a variety of interventions other than foster care, such as juvenile incarceration (??), adult incarceration (??), disability insurance (?), student loan repayment (?), evictions (??), among others. In many of these settings, treatment assignment is a two-step selection process in which individuals are assigned to treatment only after crossing an initial decision threshold. For example, in the context of foster care, children can only be removed if their maltreatment allegation is first substantiated. Similarly, in the criminal justice setting, defendants can only be incarcerated conditional on being convicted. Whether due to restrictions from data partners or privacy considerations, some studies apply this design using partially censored data containing only individuals that cross the initial decision threshold, e.g., only substantiated investigations or only convicted defendants. Such restrictions appear in two contemporaneous studies of foster care (??) as well as in other contexts (???), and may introduce bias.

To understand the source of potential bias, consider decisions made by investigators in the context of foster care. Substantiation decisions are based on the strength of the evidence, while placement decisions are based on the child’s risk of future harm.⁴⁰ The research design assumes that, due to random assignment, the distribution of risk is identical across investigators and therefore identifies impacts using exogenous variation in investigator tolerance over risk. However, if investigators also vary in their stringency over evidence, the set of substantiated cases may not be balanced across investigators. Therefore, restricted data access can create a violation of the exclusion restriction.⁴¹

In addition to the usual instrumental variables assumptions of relevance, exclusion, and monotonicity, at least one additional assumption must be satisfied for the examiner assignment design to produce unbiased estimates from censored data. Either investigators must not vary over substantiation—i.e., investigators always agree over evidence—or the

⁴⁰These two decisions may be correlated, yet they represent distinct margins. For example, there can be clear evidence for an allegation when the child faces little risk of future harm or less clear evidence in a higher risk scenario.

⁴¹The exclusion restriction is inherently untestable. Though standard balance tests offer one way to probe whether it holds, they may be underpowered or not fully capture unobservable differences. Balance tests are most reliable when backed with institutional evidence about the randomization process.

investigator’s substantiation decision must be uncorrelated with the child’s potential outcomes. The former assumption is somewhat at odds with the motivation of the research design, given that it hinges upon variation in investigator tendencies. Moreover, at least in Michigan, there is a large amount of variation in substantiation tendencies.⁴² The latter assumption is also very strong; it would be surprising if the substantiation decision—which is based on how much evidence there is that the reported maltreatment actually occurred—was unrelated to children’s potential outcomes.

Though this is not the first study to describe the potential for bias from censored data, it is the first to shed light on how much it can matter in practice. Using data containing the universe of child welfare investigations in Michigan, including both unsubstantiated and substantiated allegations, I replicate the main analysis as if I only had access to substantiated cases. After reconstructing the removal instrument according to Equation ?? using only the sample of substantiated investigations, I find that standard balance tests are sensitive to the inclusion of baseline test scores.⁴³ This offers evidence that, at least in the context of foster care in Michigan, data constraints can create a violation of the exclusion restriction.

Table ?? shows that the true effects using the complete data (Panel A) are larger than those found when restricted to substantiated investigations (Panel B).⁴⁴ The replication exercise produces substantively small and statistically insignificant impacts on the index of child wellbeing and child safety. The effect on daily attendance rate is moderately smaller than the true effect yet still statistically significant, while the point estimate on math test scores is over 0.28 standard deviations smaller and is imprecise.⁴⁵ In addition, while ? proposes a reasonable approach to use the examiner assignment design to recover unbiased estimates with censored data, the study can not empirically assess how well the approach

⁴²An investigator at the 10th percentile substantiated at a rate 8.4 percentage points less than their colleagues while someone at the 90th percentile did so at a rate 8.9 percentage points greater.

⁴³Specifically, Table ?? shows that the censored instrument is unrelated to many observable child characteristics, yet does not pass a standard balance test when prior test scores are included as covariates. In comparison, ? passes a balance test that includes baseline test scores, while ? does not report balance tests with prior test scores for the sample of school-age children. For reference, even the sample of substantiated investigations in Michigan with available baseline test scores in Michigan is 1.6 times larger than the sample in South Carolina and 2.3 times larger than the sample in Rhode Island.

⁴⁴Table ?? shows that there exists a strong first stage relationship with the censored instrument. In addition, it is possible that Panel B, which compares placement to substantiated cases, represents a different LATE than Panel A. To address this potential concern, I use investigator tendencies over substantiation and removal to instrument for both foster placement and substantiation. Table ?? shows that the estimates in Panel B are also smaller than the causal effects of placement relative to substantiation from the complete data.

⁴⁵The findings in Table ?? are somewhat similar to those in ?, which finds noisy estimates for the set of school-age children, and to ? which reports imprecise estimates on test scores but positive effects for on-time grade progression. While institutional differences between the child welfare systems in Michigan, Rhode Island, and South Carolina surely contribute to the different findings, this exercise documents that bias may also play a role.

performs in practice because it only accesses censored data itself. Appendix ?? shows that the proposed method does little to resolve bias in the current context. Overall, these exercises caution against applying the examiner assignment design with censored data.

7 Discussion and Policy Takeaways

Among groups easily identifiable in administrative data, foster children are the most vulnerable (?). For example, in Michigan, they graduated high school at less than half the state average and at lower rates than even students who ever experienced homelessness and those who were eligible for free or reduced-price lunch during every year of school.⁴⁶ Despite this, there is very little evidence of how placement influences child wellbeing. This paper offers new causal estimates of foster care by leveraging the quasi-random assignment of child welfare investigators. I find that foster placement improved child safety and schooling. Specifically, it reduced the likelihood of being abused or neglected in the future by 50%, increased daily school attendance by 6%, and improved math test scores by one-third of a standard deviation.

These results contrast earlier evidence from ?? which studied placements in Illinois from nearly two decades ago and found that foster care was very harmful to children.⁴⁷ There are at least two reasons why the findings in this study might contrast earlier work. First, the findings from Illinois may not have generalized to systems in other states at the time (Figure ??). Illinois was an outlier in terms of how long children stayed in the foster system and how often they changed foster homes.⁴⁸ As such, placement in the rest of the country during the early sample period may have been less damaging than in Illinois or even beneficial. Importantly, Michigan’s foster system more recently functions similarly to most other states.

A second reason why the results might differ from ?? is that nationwide efforts since the early 2000s to reduce foster placements led to fewer unnecessary removals. Specifically, the last two decades has seen a sizable decline in the share of investigated children who were removed, from 15% in 1998 to just 11% in 2015 (??). Investigators during the early period may have removed some children who would have been better off without placement. More

⁴⁶From author’s calculations based on the four-year high school graduation rate of the 2012 through 2014 cohorts of first-time ninth-grade public school students.

⁴⁷Appendix ?? offers evidence that neither differences in sample composition nor observed outcomes explain the difference in results.

⁴⁸For example, the median length of stay in Illinois in 1998 was almost two years longer than the median nationwide, 40 months relative to just 18 months. Also, foster children switched foster homes nearly twice as often as the rest of the country, 48% experienced three or more placement settings compared to the national average of 25% (?).

recently, however, some of the children who were not removed may have stood to benefit from foster care. In this sense, even if the damages from placement found in ?? generalized to other states during the early period, the significant shift in child welfare practices can help explain the gains from removal found in this study.

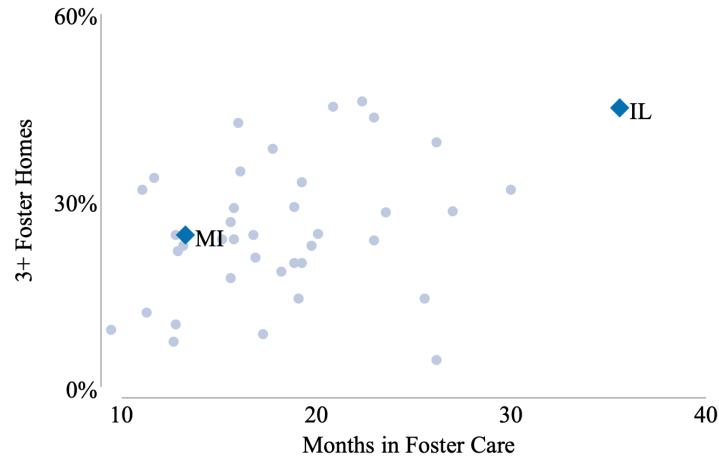
In addition to studying the impacts of placement on children’s outcomes, this paper also explores mechanisms. In particular, it highlights an understudied channel through which foster care influences children: improvements made by birth parents. I find a surprising pattern of impacts over time that is consistent with this pathway; the gains emerged after children exited the system when most were reunified with their birth parents. Birth parents worked closely with social workers following child removal and received intensive services like substance abuse treatment, parenting classes, and counseling. This paper offers suggestive evidence that child removal was crucial to ensure adult compliance with these programs.

These findings have important implications for public policy, especially in light of the Family First Prevention Services Act, which 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 allowing states to use up to eight billion dollars of federal Title IV-E funds on alternatives to placement. Previously reserved for foster care and adoption budgets, states can spend the new funding stream on programs proven to keep safe children who remain in the home. However, the evidence base around successful interventions is limited.⁴⁹ Taken together with new evidence from this study, which finds that foster care can be a useful tool in certain instances, this policy may not have its intended effects. Therefore, this paper underscores that learning how to keep vulnerable children safe without foster placement is a crucial next frontier for future research.

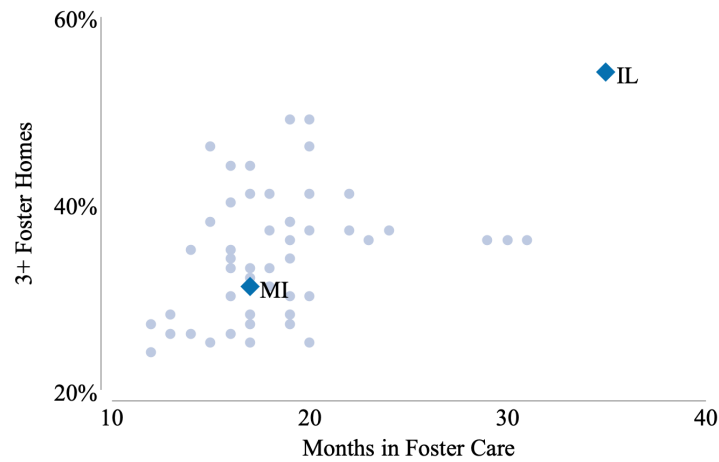
⁴⁹There are only 32 programs that received the highest evaluation rating on the California Evidence-Based Clearinghouse for Child Welfare, which the federal government used as a model for the types of programs that states can choose. Moreover, these programs span a wide range of topic areas, from infant and toddler mental health to adult substance abuse treatment (?). For comparison, the What Works Clearinghouse includes 76 studies that met its highest standards of rigor of pre-kindergarten alone and almost 500 of K-12 education (?).

Figure 1: Comparison of State Foster Care Systems

(a) 1998 Statistics

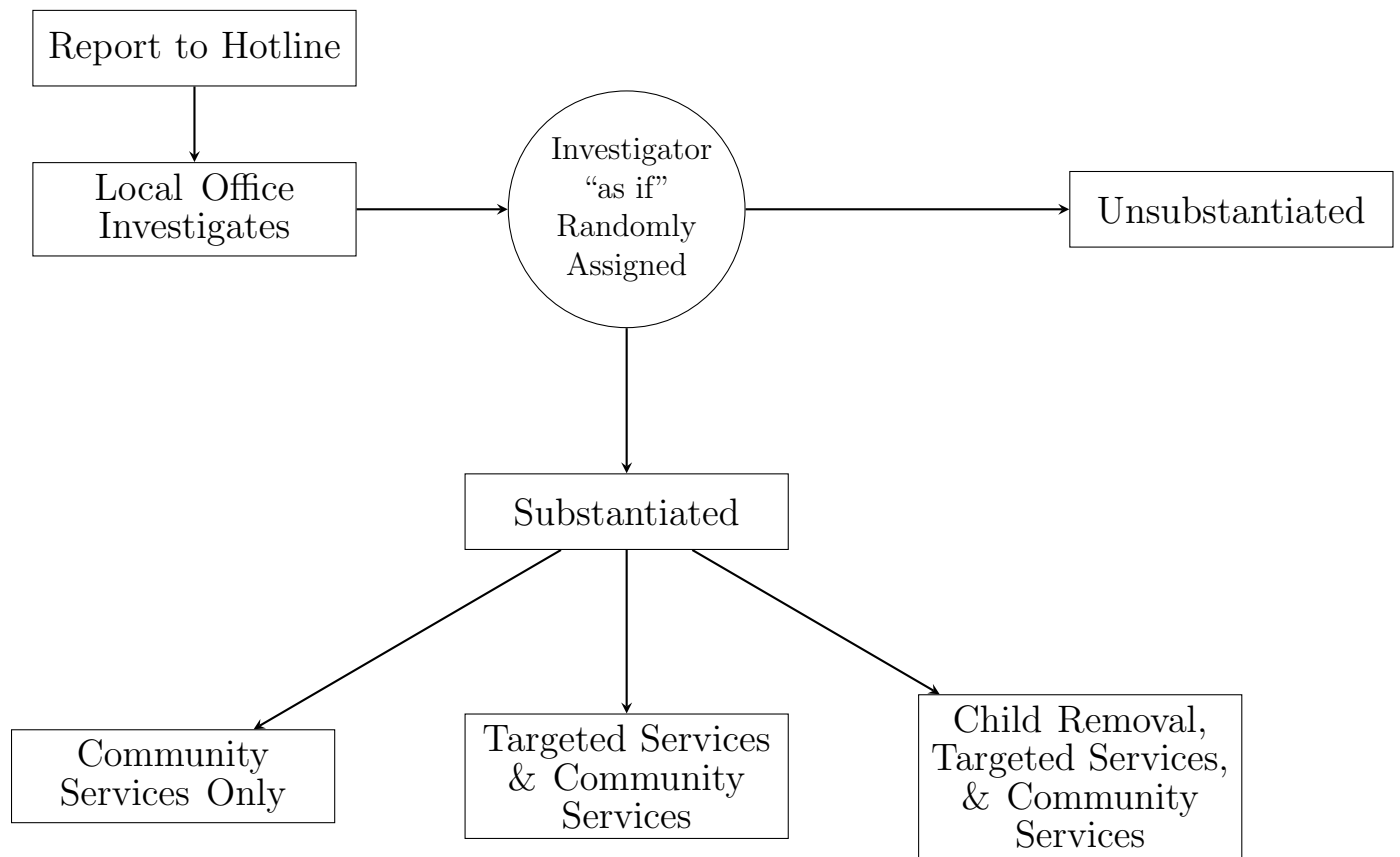


(b) 2015 Statistics



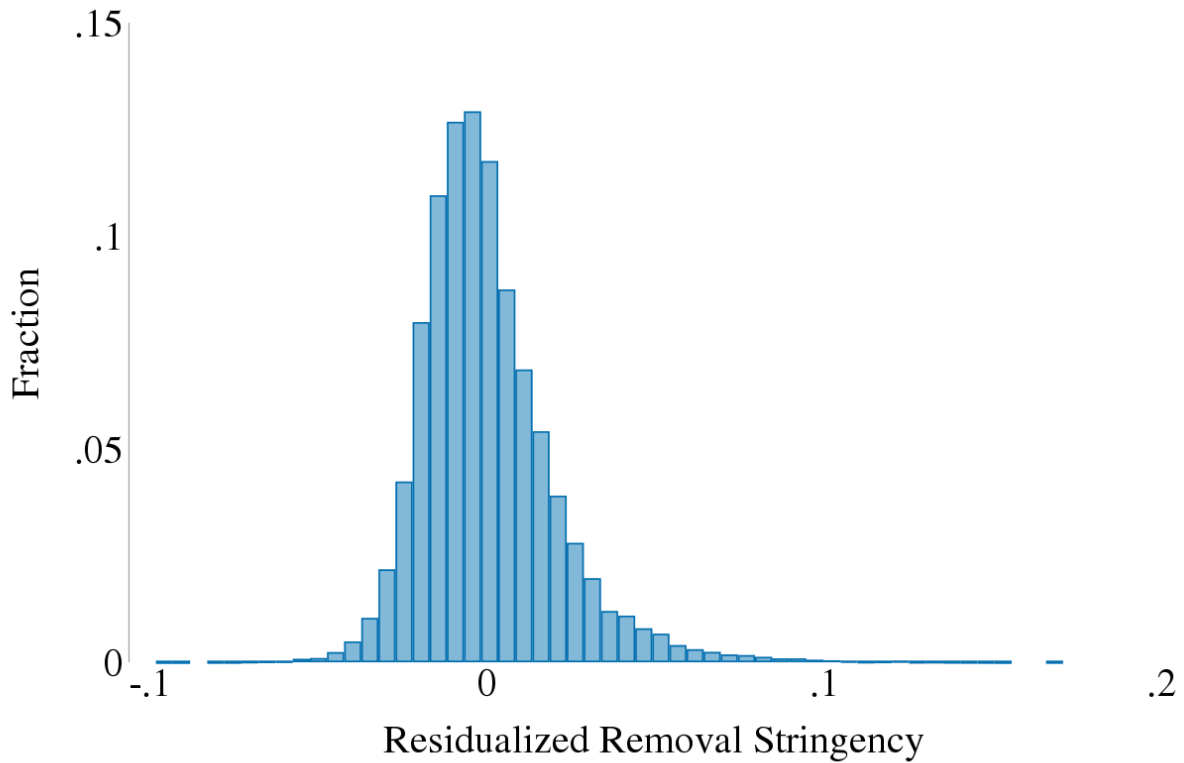
Notes. These figures show statistics about state foster care systems from 1998, the first year of publicly available data, reported in ? and from 2015 reported in ?. Months in foster care represents the median in 1998 and the average in 2015 due to a change in reporting definition. The share of foster children who experienced at least three different foster homes is reported as the average in both periods. In 1998, ten states did not report both length of stay and the share of foster children who experienced at least three different placement homes.

Figure 2: Overview of Child Maltreatment Investigations in Michigan



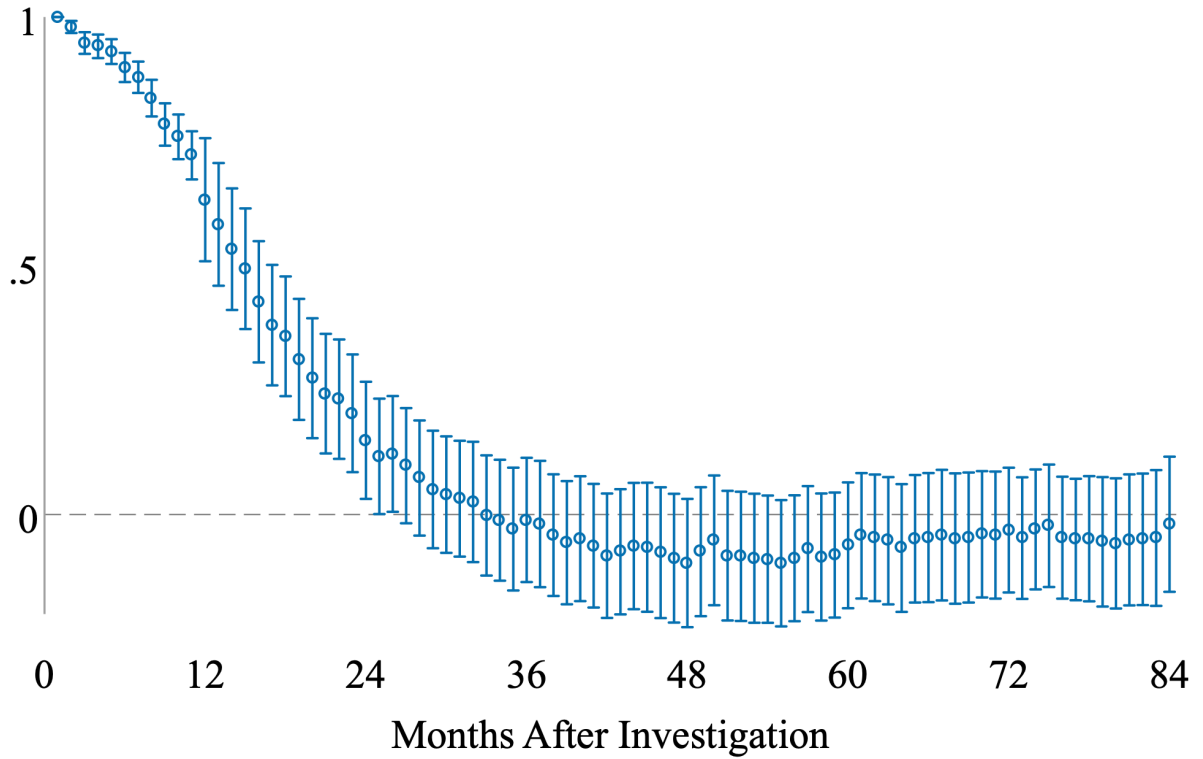
Notes. This figure describes the child maltreatment investigation process in Michigan. Substantiation means that investigators find enough evidence to support the abuse or neglect allegation. Conditional on substantiation, low-risk families receive either a referral to community-based services like a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

Figure 3: Distribution of Investigator Removal Stringency Instrument



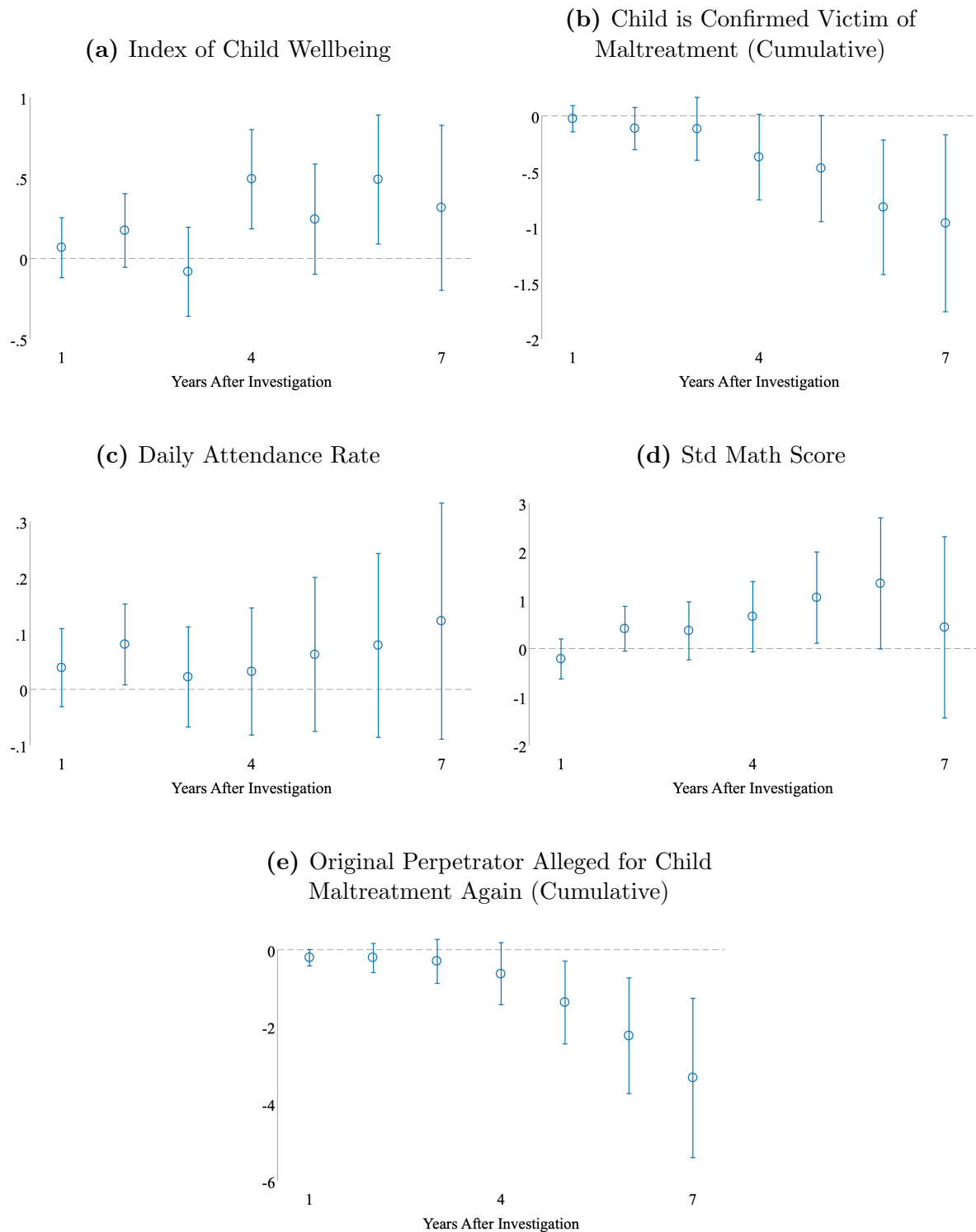
Notes. This figure shows the distribution of the removal stringency instrument residualized by the level of rotational assignment. That is, the instrument is shown net of zipcode by investigation year fixed effects in order to show variation in propensity to remove even within local offices. The instrument is calculated as the fraction of all other investigations- both past and future- assigned to the same investigator that resulted in foster placement.

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 outcome variable on foster care using removal stringency to instrument for foster care. It plots both the point estimates and their 95% confidence intervals. All specifications include the covariates as listed in the text, as well as zipcode by investigation year fixed effects. Standard errors are clustered by investigator. 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 are ineligible for being in foster care (eg. age-out of eligibility) exit from the analysis. The point estimate can be negative because control compliers may eventually enter foster care.

Figure 5: Effects of Foster Care Over Time



Notes. These figures report the results from 2SLS regressions of the outcome variable on foster care using removal stringency to instrument for foster care. They plot both the point estimates and their 95% confidence intervals. All specifications include the covariates as listed in the text, as well as zipcode by investigation year fixed effects. Standard errors are clustered by investigator. Follow-up years after the investigation are defined by school years even for non-schooling outcomes. Figure ?? 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% of investigations involve a single perpetrator, 97% involve one or two, and 99.4% involve three or fewer.

Table 1: Summary Statistics

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

Notes. This table reports summary statistics. Column one consists of the cross-section of Michigan public school students during the 2016-2017 academic year enrolled in grades one through eleven. 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 while column 3 contains the subset of investigations that resulted in foster placement. The socio-demographic variables in columns 2 and 3 are measured in the school year of the investigation. Low income is measured by free or reduced price lunch eligibility. The prior schooling characteristics in columns 2 and 3 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 while neglect is equal to one for all investigations that did not involve physical abuse.

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

| | (1) Foster Care | (2) Foster Care | (3) Foster Care | (4) Foster Care |
|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Removal Stringency | 0.480*** (0.019) | 0.451*** (0.021) | 0.450*** (0.021) | 0.449*** (0.021) |
| Observations | 242,233 | 242,233 | 242,233 | 242,233 |
| F-Statistic | 658.980 | 441.260 | 440.570 | 438.750 |
| Zipcode by Year FE | | ✓ | ✓ | ✓ |
| Socio-Demographic Controls | | | ✓ | ✓ |
| Academic Controls | | | | ✓ |

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

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

| | Full Sample | | 4th Grade and Above | |
|-----------------------------|----------------|-----------------------|---------------------|-----------------------|
| | (1) | (2) | (3) | (4) |
| | Foster Care | Removal Stringency | Foster Care | Removal Stringency |
| F-Statistic from Joint Test | 18.119 | 0.953 | 12.505 | 1.001 |
| P-Value from Joint Test | 0.000 | 0.530 | 0.000 | 0.463 |
| Observations | 242,233 | 242,233 | 144,032 | 144,032 |

Notes. This table reports the results from regressions of the dependent variable on a variety of socio-demographic and academic covariates as well as zipcode 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 three, columns 3 and 4 report results for students enrolled in at least grade four during the maltreatment investigation and include standardized test scores. Table ?? reports the full set of results. Standard errors are clustered by investigator.

Table 4: Effects of Foster Care on Child Outcomes

| | (1) Index of Child Wellbeing | (2) Alleged Victim of Maltreatment | (3) Confirmed Victim of Maltreatment | (4) Daily Attendance Rate | (5) Retained in Grade | (6) Std Math Score | (7) Std Reading Score | (8) Juvenile Delinquency |
|----------------------|---------------------------------------|---|---|------------------------------------|-------------------------------|---------------------------------|---------------------------------|--------------------------------|
| <i>Panel A: OLS</i> | | | | | | | | |
| Foster Care | -0.005 (0.005) {0.000} | -0.032*** (0.004) {0.177} | -0.007*** (0.002) {0.046} | 0.011*** (0.002) {0.912} | 0.005** (0.002) {0.051} | 0.056*** (0.013) {-0.501} | 0.064*** (0.015) {-0.479} | 0.041*** (0.004) {0.025} |
| <i>Panel B: 2SLS</i> | | | | | | | | |
| Foster Care | 0.164** (0.075) {-0.063} | -0.132** (0.066) {0.255} | -0.053* (0.031) {0.094} | 0.054** (0.027) {0.893} | -0.019 (0.029) {0.063} | 0.339* (0.203) {-0.429} | 0.162 (0.216) {-0.234} | -0.028 (0.038) {0.051} |
| Observations | 242,233 | 242,233 | 242,233 | 224,925 | 242,204 | 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. 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 zipcode by investigation year fixed effects. Standard errors are clustered by investigator. The education and crime outcomes do not include all of the observations in the sample. Specifically, some grade level and daily attendance rate 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 is missing for eight counties, is available only through 2015, and is relevant only for children younger than Michigan's age of majority of sixteen. *p < 0.10, **p < 0.05, *** p < 0.01.

Table 5: Effects of Foster Care Over Time

| | (1) Index of Neighborhood & School Characteristics | (2) Index of Child Wellbeing | (3) Confirmed Victim of Maltreatment | (4) Daily Attendance Rate | (5) Std Math Score | (6) Received Special Education Services |
|--|--|---------------------------------------|---|------------------------------------|--------------------------------|---|
| <i>Panel A: One Year After Investigation</i> | | | | | | |
| Foster Care | 0.257** (0.117) {-0.147} | 0.067 (0.095) {0.028} | -0.024 (0.060) {0.068} | 0.039 (0.036) {0.912} | -0.218 (0.211) {0.062} | -0.012 (0.065) {0.099} |
| <i>Panel B: Two+ Years After Investigation</i> | | | | | | |
| Foster Care | 0.066 (0.141) {-0.011} | 0.194** (0.089) {-0.092} | -0.065* (0.035) {0.102} | 0.060* (0.032) {0.885} | 0.558** (0.239) {-0.624} | 0.016 (0.107) {0.035} |
| Observations | 242,233 | 242,233 | 242,233 | 224,925 | 177,118 | 242,233 |

Notes. This table reports the results from 2SLS regressions of the outcome variable on foster care using removal stringency to instrument for foster care. 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. The curly brackets below the standard error represents the control complier mean. The index of neighborhood and school characteristics is made up of neighborhood median income, educational attainment, and employment rate, as well as school average test scores and income level. The effects on each component of the index of neighborhood and school characteristics is shown in Table ?? . The p-value reports whether the subgroup estimates are statistically different from each other. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. *p< 0.10, ** p< 0.05, *** p< 0.01.

Table 6: Effects of Child Welfare Services and Foster Care on Child Outcomes

| | (1) Index of Child Wellbeing | (2) Confirmed Victim of Maltreatment | (3) Daily Attendance Rate | (4) Std Math Score |
|--|---------------------------------------|---|------------------------------------|--------------------------|
| Child Removal, Targeted Services, and Community Services | 0.123 (0.098) | -0.038 (0.041) | 0.060* (0.037) | 0.310 (0.266) |
| Targeted Services and Community Services | 0.049 (0.037) | -0.018 (0.016) | 0.002 (0.012) | 0.105 (0.096) |
| Community Services | -0.032 (0.024) | 0.012 (0.010) | -0.005 (0.008) | -0.095 (0.064) |
| Observations | 242,233 | 242,233 | 224,925 | 177,118 |

Notes. This table reports the results from 2SLS regressions of the outcome variable on the treatment conditions: community services, targeted and community services, and foster care plus targeted and community services. It uses investigator stringency in evidence and risk levels to simultaneously instrument for the independent variables respectively. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Effects of Foster Care on Child Outcomes Using Censored Data

| | (1) Index of Child Wellbeing | (2) Confirmed Victim of Maltreatment | (3) Daily Attendance Rate | (4) Std Math Score |
|--|---------------------------------------|---|------------------------------------|--------------------------|
| <i>Panel A: Complete Data, Unsubstantiated and Substantiated</i> | | | | |
| Foster Care | 0.164** (0.075) | -0.053* (0.031) | 0.054** (0.027) | 0.339* (0.203) |
| Observations | 242,233 | 242,233 | 224,925 | 177,118 |
| <i>Panel B: Censored Data, Only Substantiated</i> | | | | |
| Foster Care | 0.026 (0.040) | -0.009 (0.018) | 0.039*** (0.014) | 0.062 (0.107) |
| Size of Bias | 0.138 | 0.044 | 0.015 | 0.277 |
| Observations | 47,469 | 47,469 | 43,839 | 35,322 |

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

A Print Appendix

A.1 Supplemental Print Tables

Table A1: Effects of Foster Care on Michigan Public School Enrollment

| | (1) Ever Enrolled After | (2) Enrolled One Year After | (3) Enrolled Two Years After | (4) Enrolled Three Years After | (5) Enrolled Four Years After | (6) Enrolled Five Years After |
|---|----------------------------------|--------------------------------------|---------------------------------------|---|--|--|
| <i>Panel A- Children Six Years Old and Younger During Investigation</i> | | | | | | |
| Foster Care | -0.191*** (0.057) | | | | | |
| Observations | 236,925 | | | | | |
| <i>Panel B- Analysis Sample, Enrolled in Grade 1 to 11 During Investigation</i> | | | | | | |
| Foster Care | -0.033 (0.039) | -0.018 (0.049) | 0.001 (0.070) | -0.126 (0.090) | -0.001 (0.110) | 0.039 (0.130) |
| Observations | 248,730 | 248,730 | 212,718 | 168,711 | 133,268 | 99,014 |

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

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

| | (1) All | (2) Foster Care | (3) Compliers |
|----------------------------|------------|--------------------|------------------|
| Female | 0.49 | 0.47 | 0.52 |
| White | 0.62 | 0.52 | 0.52 |
| Student of Color | 0.38 | 0.48 | 0.47 |
| 10 Years Old & Younger | 0.55 | 0.51 | 0.61 |
| 11 Years Old & Older | 0.45 | 0.49 | 0.39 |
| Urban/Suburban County | 0.64 | 0.63 | 0.63 |
| Rural County | 0.36 | 0.37 | 0.33 |
| Low Income | 0.83 | 0.87 | 0.89 |
| Ever Retained in Grade | 0.36 | 0.39 | 0.38 |
| Above Median Math Score | 0.50 | 0.41 | 0.39 |
| Above Median Reading Score | 0.50 | 0.42 | 0.38 |
| Share of Sample | 1.00 | 0.02 | 0.05 |

Notes. I follow ? to calculate the share and characteristics of compliers. Specifically, I compute the share of compliers as the difference in the first stage effect between children assigned to an investigator with removal stringency at the 99th and the 1st percentiles. Then, I calculate the characteristics of compliers as the fraction of compliers across each characteristic subgroup. Above median math and reading scores are indicators for scoring higher than the median child in the sample on baseline math and reading tests.

Table A3: Testable Implications of the Exclusion of Removal Stringency Instrument

| | (1) Investigator's Number of Investigations | (2) Days in Foster Care | (3) # Foster Homes | (4) First Placed with Relatives | (5) First Placed with Unrelated Family | (6) First Placed in Group Home |
|--------------------|--|-------------------------------|--------------------------|---------------------------------------|---|---|
| Removal Stringency | 179.734 (197.998) | 25.797 (647.431) | 0.241 (3.534) | 0.166 (0.541) | -0.021 (0.495) | -0.145 (0.300) |
| Observations | 242,233 | 4,809 | 4,809 | 4,809 | 4,809 | 4,809 |

Notes. This table reports the results from a regression of the dependent variable on the removal stringency instrument. The dependent variable in Columns 2 through 6 are conditional on foster placement. Standard errors are clustered by investigator. *p< 0.10, ** p< 0.05, *** p< 0.01.

Table A4: Effects of Foster Care on Permanency Placements

| | (1) Reunified | (2) Adopted | (3) Guardianship | (4) Emancipated | (5) Still in FC in Sep 2017 |
|--------------------------|---------------------|---------------------|---------------------|---------------------|-----------------------------------|
| Foster Care | 0.703*** (0.024) | 0.064*** (0.012) | 0.040*** (0.010) | 0.017*** (0.006) | 0.176*** (0.020) |
| % Conditional on Exiting | 85.3% | 7.8% | 4.9% | 2.1% | |
| Observations | 242,233 | 242,233 | 242,233 | 242,233 | 242,233 |

Notes. This table reports the results from 2SLS regressions of the outcome variable on foster care, using removal stringency to instrument for foster care. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. Each permanency outcome is mutually exclusive. Some students were still in the foster system at the end of the sample period in September 2017, so these students are coded as zeroes for all permanency outcomes and instead are included in the group in column 5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Effects of Foster Care on Child Outcomes, by Subgroup

| | (1) Index of Child Wellbeing | (2) Confirmed Victim of Maltreatment | (3) Daily Attendance Rate | (4) Std Math Score |
|-----------------------------|---------------------------------------|---|------------------------------------|---------------------------------|
| <i>Panel A: Gender</i> | | | | |
| Male (N=123,715) | 0.191** (0.099) {-0.074} | -0.036 (0.039) {0.090} | 0.061 (0.037) {0.890} | 0.435 (0.285) {-0.427} |
| Female (N=118,436) | 0.145* (0.086) {-0.054} | -0.068* (0.036) {0.097} | 0.051* (0.030) {0.894} | 0.273 (0.226) {-0.420} |
| P-value | 0.669 | 0.457 | 0.794 | 0.596 |
| <i>Panel B: Age</i> | | | | |
| 10 & Younger (N=133,476) | 0.196** (0.085) {-0.031} | -0.074** (0.034) {0.098} | 0.080*** (0.028) {0.908} | 0.161 (0.223) {-0.418} |
| 11 & Older (N=108,757) | 0.108 (0.115) {-0.106} | -0.002 (0.043) {0.087} | 0.004 (0.051) {0.874} | 1.356*** (0.332) {-0.480} |
| P-value | 0.483 | 0.107 | 0.144 | 0.001 |

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 by gender and Panel B reports results by age during the investigation. The curly brackets below the standard error represents the control complier mean. The p-value reports whether the subgroup estimates are statistically different from each other. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. *p< 0.10, ** p< 0.05, *** p< 0.01.

Table A5: Effects of Foster Care on Child Outcomes, by Subgroup (continued)

| | (1) Index of Child Wellbeing | (2) Confirmed Victim of Maltreatment | (3) Daily Attendance Rate | (4) Std Math Score |
|---|---------------------------------------|---|------------------------------------|--------------------------------|
| <i>Panel C: Prior Child Welfare Involvement</i> | | | | |
| Involved (N=142,034) | 0.163** (0.082) {-0.110} | -0.093*** (0.036) {0.103} | 0.003 (0.029) {0.883} | 0.486** (0.219) {-0.503} |
| Not Involved (N=100,199) | 0.156 (0.116) {-0.001} | 0.030 (0.046) {0.081} | 0.142*** (0.043) {0.903} | 0.123 (0.313) {-0.354} |
| P-value | 0.950 | 0.015 | 0.002 | 0.265 |
| <i>Panel D: County Urbanicity</i> | | | | |
| Urban or Suburban (N=154,697) | 0.229** (0.093) {-0.060} | -0.073** (0.037) {0.092} | 0.076** (0.035) {0.892} | 0.530** (0.264) {-0.420} |
| Rural (N=87,536) | 0.029 (0.120) {-0.073} | -0.011 (0.054) {0.099} | 0.009 (0.042) {0.894} | 0.018 (0.303) {-0.452} |
| P-value | 0.180 | 0.341 | 0.207 | 0.193 |

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 C reports results by prior child welfare involvement where involved is equal to one if the child had a prior maltreatment investigation and Panel D reports results by county urbanicity as defined by the National Center for Health Statistics 2006 urban-rural classification scheme. The curly brackets below the standard error represents the control complier mean. The p-value reports whether the subgroup estimates are statistically different from each other. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. *p < 0.10, ** p < 0.05, *** p < 0.01.

Table A6: Robustness Checks

| | (1) Index of Child Wellbeing |
|---|---------------------------------------|
| <i>Panel A: Alternative Samples</i> | |
| Child's First Investigation (N=180,859) | 0.163* (0.086) |
| Balanced Panel (N=96,156) | 0.310*** (0.106) |
| Investigator Assigned ≥ 25 Investigations (N=249,228) | 0.137* (0.072) |
| Investigator Assigned ≥ 75 Investigations (N=232,818) | 0.146* (0.077) |
| <i>Panel B: Alternative Removal Stringency Instruments</i> | |
| Split Sample (N=242,233) | 0.141 (0.089) |
| Leave-out Other Years (N=242,233) | 0.078* (0.046) |
| Leave-out Same Year (N=242,233) | 0.318* (0.166) |
| Empirical Bayes Shrinkage (N=242,233) | 0.240 (0.229) |
| <i>Panel C: Alternative Level of Rotational Assignment</i> | |
| County by Year (N=242,233) | 0.204*** (0.079) |

Notes. Panel A reports the results from 2SLS regressions using alternative sample definitions, Panel A uses alternative measures of removal stringency to instrument for foster care, and Panel C reports the results using the main stringency instrument yet replaced zipcode by investigation year fixed effects with county by investigation year fixed effects. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. In Panel A, the balanced panel sample is restricted to the first five follow-up years for children investigated in seventh grade or below in 2012 or earlier. In Panel B, the split sample measure is the removal rate of the assigned investigator from a random half of the sample. The leave-out other years measure is the leave-out removal rate of the assigned investigator from other children who had investigations in the same calendar year. The leave-out same year measure is the leave-out removal rate of the assigned investigator from other children who had investigations in different calendar years. The empirical bayes shrinkage measure allows stringency to vary across years and shrinks the main removal stringency instrument toward its mean. *p< 0.10, ** p< 0.05, *** p< 0.01.

A.2 Assessing ? Approaches to Using Examiner Assignment Design with Censored Data

When treatment is a two-step selection process, the standard examiner assignment design may lead to biased estimates if researchers use data that is censored to include only individuals who cross an initial decision threshold. In the context of foster care, children are only removed if their maltreatment report is first substantiated. Similarly, in the criminal justice context, people are incarcerated conditional on conviction. As the design becomes increasingly popular, so does its use with censored data (?????). Section ?? describes the source of this bias in greater detail and shows that—at least when applied to the context of foster care in Michigan—censored data does create bias.

What can researchers do when limited to using censored data? ? proposes a reasonable solution in a study of the effects of parental incarceration on child outcomes. The study uses data from SISBEN, Colombia’s census of its low-income population, to link children to parents and parents to both criminal convictions and incarceration. SISBEN does not include information on parents who appeared before a court but were not convicted however, so the data is censored to include only parents who cross the conviction threshold. Fortunately, anonymized records containing the universe of conviction and incarceration decisions are publicly available for every judge in Colombia, which the author uses to create leniency instruments. These anonymized records can only be matched to SISBEN on the judge field and, importantly, can not be matched to individual parents.

? describes how the standard examiner assignment design can not be applied in this context and derives an estimator of the causal effects of incarceration relative to conviction that can be identified using censored data.⁵⁰ The key insight is that there is exogenous variation in incarceration among judges with identical conviction thresholds but different incarceration thresholds. In the context of this study, the variation in removal is as good as random for a given evidence threshold. More formally, the study proposes that the causal effects of removal relative to substantiation can be identified from censored data as

$$\int_0^1 \frac{\delta \mathbb{E}[Y \cdot \mathbb{1}(T \in \{t_S, t_R\}) | P_S(Z) = p_S, P_R^*(Z) = p_R^*]}{\delta p_R^*} dp_R^* \quad (4)$$

where Y is an outcome and T denotes treatment assignment- either substantiated but not removed (t_S) or substantiated and removed (t_R).⁵¹ $P_S(Z) = p_S$ represents that the evidence threshold to substantiate is held fixed at p_S and $P_R^*(Z) = p_R^*$ means that the removal threshold conditional on substantiation is equal to p_R^* . Integrating over the inside term averages the effects across all investigators.

In practice, the study derives P_S and P_R^* from the data as the leave-out measure of evidence stringency- equivalent to Z^E of Equation ?? in the main text- and the leave-out measure of removal conditional on substantiation respectively. Identifying this parameter in practice hinges on being able to fix the conviction threshold. While ? proposes three

⁵⁰This is a somewhat special context of the censoring issue given that the study has access to the universe of court records, even though they cannot be linked to SISBEN. The estimator derived in the study can only be identified in contexts where such records are available and may not generally applicable to all studies restricted by censored data.

⁵¹This is equivalent to Equation 13 in ?.

complementary strategies to do so, the study itself only has access to censored data and therefore can not empirically assess whether these strategies produce unbiased estimates. Using the universe of maltreatment investigations, I compare estimates from each approach with those from the full, uncensored data.

The first, called the pooled approach, uses P_R^* to instrument for home removal while additionally controlling for linear and quadratic terms of P_S and all interactions. The second, called the tercile approach, instruments for removal with P_R^* separately for each tercile of the evidence stringency distribution. The idea is that- in addition to controlling for evidence stringency- splitting the data into terciles approximates fixing the evidence threshold.

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

None of the approaches with censored data approximate the estimates from the full data well. With censored data, the pooled approach finds a small and not statistically significant effect of foster care relative to substantiation yet the true effect with full data represents a large and statistically significant increase. The results are similar when using the tercile approach- the point estimates are larger in the full data, though they vary in precision. Furthermore, the pattern of estimates across the evidence threshold distribution is not qualitatively similar when using the rolling window approach. The censored data reveals a positive relationship between evidence stringency and the index of child outcomes while the full data points toward the true relationship being somewhat U-shaped.

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

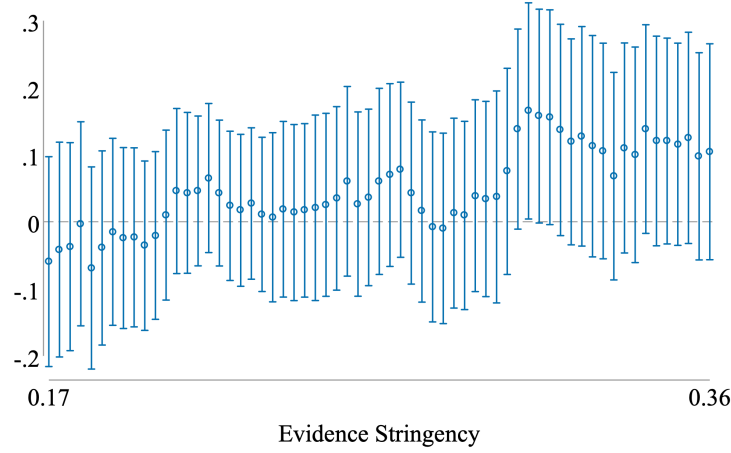
Table A7: Assessing ? Approaches to Using Judge Fixed Effects Design with Censored Data

| | Tercile Approach | | | |
|-------------------------------|---------------------------|-------------------------------|------------------------------|------------------------------|
| | (1) Pooled Approach | (2) Lenient in Evidence | (3) Middle in Evidence | (4) Strict in Evidence |
| <i>Panel A: Censored Data</i> | | | | |
| Foster Care | 0.028 (0.040) | -0.101 (0.091) | 0.039 (0.077) | 0.112 (0.086) |
| Observations | 47,470 | 15,823 | 15,823 | 15,824 |
| <i>Panel B: Full Data</i> | | | | |
| Foster Care | 0.181*** (0.085) | 0.311 (0.327) | 0.087 (0.173) | 0.346*** (0.128) |
| Observations | 242,233 | 80,744 | 80,744 | 80,745 |

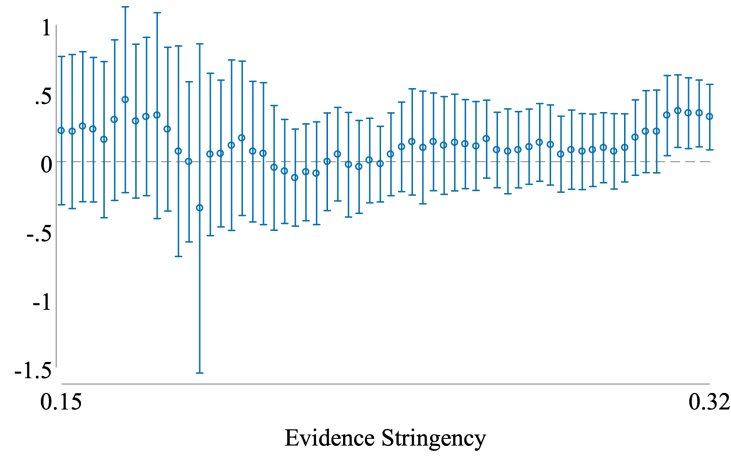
Notes. This table compares the estimates of foster care relative to substantiation on the index of child wellbeing using approaches proposed in ?. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. Panel A applies the approaches to censored data, restricted to only children with substantiated maltreatment reports. In Panel A, investigators who were lenient in evidence substantiated between 0-21% of reports, while those in the middle and strict substantiated between 21-28% and 28-67% respectively. Panel B applies the approaches to the full, uncensored data. I use removal stringency to instrument for foster care and evidence stringency to instrument for substantiation. In Panel B, investigators who were lenient in evidence substantiated between 0-18% of reports, while those in the middle and strict substantiated between 18-25% and 25-69% respectively. *p< 0.10, ** p< 0.05, *** p< 0.01.

Figure A1: Assessing ? Rolling Window Approach to Using Judge Fixed Effects Design with Censored Data

(a) Censored Data



(b) Full Data



Notes. This figure compares the estimates of foster care relative to substantiation on the index of child wellbeing using the rolling window approach proposed in ?. They plot both the point estimates and their 95% confidence intervals. All specifications include the covariates as listed in the text, as well as zipcode by investigation year fixed effects. Standard errors are clustered by investigator. Figure ?? sorts the censored data based on evidence stringency and estimating separate regressions of the index of child outcomes on foster care using removal stringency conditional on substantiation to instrument for foster care and including evidence stringency as a covariate. Since the sample sizes are similar, I follow ? by using a rolling window of 18,000 observations and adjust the window by 500 observations each time along the evidence threshold. Figure ?? applies the same approach to the full, uncensored data. I use removal stringency to instrument for foster care and evidence stringency to instrument for substantiation to estimate the effect of foster care relative to substantiation. Since the sample size is about five times larger, I use a rolling window of 90,000 observations and adjust the window by 2,500 observations each time.

B Online Appendix

B.1 Supplemental Online Tables

Table B1: Full Balance Tests of the Conditional Independence of the Removal Stringency Instrument

| Dependent Variable: | Full Sample | | | | 4th Grade and Above | | | |
|--|--------------------|------------------|--------------------|------------------|---------------------|------------------|--------------------|------------------|
| | Foster Care | | Removal Stringency | | Foster Care | | Removal Stringency | |
| | (1) Coefficient | (2) Std Error | (3) Coefficient | (4) Std Error | (5) Coefficient | (6) Std Error | (7) Coefficient | (8) Std Error |
| <i>Socio-Demographic Characteristics</i> | | | | | | | | |
| Female | -0.002*** | (0.001) | 0.000 | (0.000) | -0.004*** | (0.001) | -0.000 | (0.000) |
| White | -0.001 | (0.002) | -0.000 | (0.000) | -0.000 | (0.002) | -0.000 | (0.000) |
| Black | 0.006*** | (0.003) | 0.000 | (0.000) | 0.009*** | (0.002) | 0.000 | (0.000) |
| Hispanic | 0.003 | (0.003) | -0.000 | (0.000) | 0.004 | (0.002) | -0.000 | (0.000) |
| Low Income | 0.004*** | (0.001) | -0.000 | (0.000) | 0.004*** | (0.001) | -0.000 | (0.000) |
| Grade 2 | 0.000 | (0.001) | -0.000 | (0.000) | | | | |
| Grade 3 | 0.000 | (0.001) | -0.000 | (0.000) | | | | |
| Grade 4 | -0.000 | (0.001) | -0.000 | (0.000) | | | | |
| Grade 5 | 0.001 | (0.016) | -0.000 | (0.005) | 0.028* | (0.001) | -0.004 | (0.000) |
| Grade 6 | 0.001 | (0.016) | 0.000 | (0.005) | 0.028*** | (0.001) | -0.004 | (0.000) |
| Grade 7 | 0.003** | (0.016) | -0.000 | (0.005) | 0.030* | (0.002) | -0.004 | (0.000) |
| Grade 8 | 0.006*** | (0.016) | -0.000 | (0.005) | 0.032* | (0.002) | -0.004 | (0.000) |
| Grade 9 | 0.007*** | (0.016) | -0.000 | (0.005) | 0.034** | (0.002) | -0.004 | (0.000) |
| Grade 10 | 0.007*** | (0.017) | -0.000 | (0.005) | 0.033** | (0.002) | -0.004 | (0.000) |
| Grade 11 | 0.001 | (0.017) | -0.000 | (0.005) | 0.028* | (0.002) | -0.004 | (0.000) |
| Had a Prior Investigation | 0.001 | (0.001) | 0.000 | (0.000) | 0.001 | (0.001) | -0.000 | (0.000) |
| # Prior Investigations | 0.002*** | (0.000) | 0.000 | (0.000) | 0.002*** | (0.000) | 0.000 | (0.000) |
| <i>Prior Academic Characteristics</i> | | | | | | | | |
| Attendance Rate | -0.033*** | (0.006) | -0.001* | (0.001) | -0.035*** | (0.005) | -0.001 | (0.001) |
| Special Education | -0.000 | (0.001) | 0.000* | (0.000) | 0.002 | (0.001) | 0.000 | (0.000) |
| Ever Repeated Grade | 0.000 | (0.001) | -0.000 | (0.000) | 0.000 | (0.001) | 0.000 | (0.000) |
| Std Math Score | | | | | -0.001** | (0.001) | -0.000 | (0.000) |
| Std Reading Score | | | | | 0.001 | (0.001) | -0.000 | (0.000) |
| Observations | 242,233 | | 242,233 | | 144,032 | | 144,032 | |
| F Stat from Joint Test | 18.119 | | 0.953 | | 12.505 | | 1.001 | |
| P-Value from Joint Test | 0.000 | | 0.530 | | 0.000 | | 0.463 | |

Notes. This table reports the results from a regression of the dependent variable on a variety of socio-demographic and prior academic characteristics as well as zipcode by investigation year fixed effects. Students in Michigan begin taking statewide standardized tests in third grade so I include prior standardized tests scores only for the sample of students enrolled in at least fourth grade. Not shown here to save space- but included in the joint test- the regressions also includes indicators for missing variables: female, low income, and each of the prior schooling characteristics. Results with the missing indicators are available upon request. Standard errors are clustered by investigator. *p< 0.10, ** p< 0.05, *** p< 0.01.

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

| | (1) Female | (2) Male | (3) White | (4) Student of Color | (5) Age ≤ 10 | (6) Age > 10 | (7) Had Prior Inv | (8) No Prior Inv |
|---|---------------------|---------------------|---------------------|----------------------------|---------------------|---------------------|-------------------------|------------------------|
| <i>Panel A: Main Leave-One-Out Instrument</i> | | | | | | | | |
| Removal Stringency | 0.481*** (0.030) | 0.422*** (0.027) | 0.399*** (0.027) | 0.515*** (0.038) | 0.480*** (0.027) | 0.411*** (0.030) | 0.544*** (0.031) | 0.323*** (0.027) |
| <i>Panel B: Leave-Subgroup-Out Instrument</i> | | | | | | | | |
| Removal Stringency | 0.365*** (0.027) | 0.305*** (0.023) | 0.161*** (0.018) | 0.226*** (0.029) | 0.195*** (0.022) | 0.317*** (0.027) | 0.269*** (0.027) | 0.160*** (0.020) |
| Observations | 118,436 | 123,715 | 149,527 | 92,706 | 133,476 | 108,757 | 142,034 | 100,199 |

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

Table B3: Effects of Foster Care on Taking Standardized Tests

| | (1) | (2) |
|----------------------|-----------------------------|------------------------------|
| | Took Std Math Test | Took Std Reading Test |
| <i>Panel A: OLS</i> | | |
| Foster Care | 0.007 (0.004) {0.905} | 0.008* (0.004) {0.904} |
| <i>Panel B: 2SLS</i> | | |
| Foster Care | 0.019 (0.063) {0.967} | -0.029 (0.065) {1.00} |
| Observations | 189,084 | 189,084 |

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

Table B4: Effects of Foster Care on High School Graduation and College Enrollment

| | (1) Graduated High School | (2) Ever Enrolled in College | (3) Ever Enrolled in a Two-Year College | (4) Ever Enrolled in a Four-Year College |
|----------------------|---------------------------------|------------------------------------|--|---|
| <i>Panel A: OLS</i> | | | | |
| Foster Care | -0.024* (0.014) {0.532} | 0.001 (0.017) {0.348} | -0.008 (0.016) {0.241} | 0.013 (0.012) {0.151} |
| <i>Panel B: 2SLS</i> | | | | |
| Foster Care | 0.085 (0.274) {0.723} | 0.192 (0.368) {0.502} | -0.001 (0.343) {0.466} | 0.029 (0.286) {0.321} |
| Observations | 60,776 | 36,661 | 36,661 | 36,661 |

Notes. Panel A reports the results from OLS regressions of the outcome variable on foster care while Panel B reports the results from 2SLS regressions using removal stringency to instrument for foster care. 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 zipcode by investigation year fixed effects. Standard errors are clustered by investigator. Only students expected to be in 12th grade by 2017 based on an on-time grade progression from the school year of their investigation are included in the analysis of high school graduation while the analysis of college enrolled is restricted to students expected to be in 12th grade by 2016. Some colleges are missing information on whether they were a two or four-year school so the two and four-year college enrollment estimates need not add up to the overall college enrollment estimate. *p< 0.10, ** p< 0.05, *** p< 0.01.

Table B5: Effects of Foster Care on Removal Experience

| | (1) | (2) | (3) | (4) |
|--------------|---------------------------|----------------------------|----------------------------------|--------------------------|
| | Days in Foster Care | Days in Kinship Care | Days with Unrelated Family | Days in Group Home |
| Foster Care | 581*** (44) | 345*** (28) | 185*** (24) | 50*** (16) |
| Observations | 242,233 | 242,233 | 242,233 | 242,233 |

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

Table B6: Effects of Foster Care on Neighborhood and School Environment Over Time

| | | Neighborhood | | | School | |
|--|--|--|--------------------------------|------------------------------|-------------------------------|--------------------------------|
| | (1) Index of Neighborhood & School Characteristics | (2) Median Income (\$100,000) | (3) BA Degree or Higher | (4) Employment Rate | (5) Test Scores | (6) Low Income |
| <i>Panel A: One Year After Investigation</i> | | | | | | |
| Foster Care | 0.257** (0.117) {-0.147} | 0.071 (0.044) {0.406} | 0.084*** (0.028) {0.121} | 0.021 (0.027) {0.848} | -0.003 (0.096) {-0.119} | -0.100** (0.044) {0.649} |
| <i>Panel B: Two+ Years After Investigation</i> | | | | | | |
| Foster Care | 0.066 (0.141) {-0.011} | 0.055 (0.054) {0.411} | 0.034 (0.038) {0.157} | -0.011 (0.033) {0.875} | 0.086 (0.109) {-0.239} | -0.021 (0.051) {0.538} |
| Obserservations | 242,233 | 209,446 | 209,446 | 209,446 | 217,956 | 241,267 |

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

Table B7: Effects of Foster Care on Removal Experience, by Subgroup

| | (1) Days in Foster Care | (2) Days in Kinship Care | (3) Days with Unrelated Family | (4) Days in Group Home |
|-----------------------------|----------------------------------|-----------------------------------|---|---------------------------------|
| <i>Panel A: Gender</i> | | | | |
| Male (N=123,715) | 539*** (57) | 348*** (35) | 144*** (31) | 47** (21.876) |
| Female (N=118,436) | 619*** (57) | 343*** (36) | 223*** (31) | 53** (21) |
| P-value | 0.277 | 0.918 | 0.047 | 0.832 |
| <i>Panel B: Age</i> | | | | |
| 10 & Younger (N=133,476) | 518*** (51) | 359*** (34) | 127*** (26) | 32* (18) |
| 11 & Older (N=108,757) | 680*** (66) | 324*** (36) | 277*** (40) | 78*** (27) |
| P-value | 0.033 | 0.404 | 0.001 | 0.124 |

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 by gender and Panel B reports results by age during the investigation. The p-value reports whether the subgroup estimates are statistically different from each other. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. *p< 0.10, ** p< 0.05, *** p< 0.01.

B.2 Data Appendix

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

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

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

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

Table B7: Effects of Foster Care on Removal Experience, by Subgroup (Continued)

| | (1) Days in Foster Care | (2) Days in Kinship Care | (3) Days with Unrelated Family | (4) Days in Group Home |
|---|----------------------------------|-----------------------------------|---|---------------------------------|
| <i>Panel C: Prior Child Welfare Involvement</i> | | | | |
| Involved (N=142,034) | 614*** (47) | 386*** (30) | 187*** (26) | 41** (17) |
| Not Involved (N=100,199) | 508*** (85) | 255*** (50) | 181*** (44) | 71** (33) |
| P-value | 0.251 | 0.015 | 0.901 | 0.377 |
| <i>Panel D: County Urbanicity</i> | | | | |
| Urban or Suburban (N=154,697) | 612*** (58) | 351*** (38) | 192*** (31) | 69*** (22) |
| Rural (N=87,536) | 528*** (63) | 338*** (38) | 174*** (36) | 16 (21) |
| P-value | 0.318 | 0.802 | 0.697 | 0.079 |

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 C reports results by prior child welfare involvement where involved is equal to one if the child had a prior maltreatment investigation and Panel D reports results by county urbanicity as defined by the National Center for Health Statistics 2006 urban-rural classification scheme. The curly brackets below the standard error represents the control complier mean. The p-value reports whether the subgroup estimates are statistically different from each other. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. *p< 0.10, ** p< 0.05, *** p< 0.01.

Table B8: Balance Tests Using Censored Data

| | Full Sample | | 4th Grade and Above | |
|-----------------------------|----------------|-----------------------------------|---------------------|-----------------------------------|
| | (1) | (2) | (3) | (4) |
| | Foster Care | Censored Removal Stringency | Foster Care | Censored Removal Stringency |
| F-Statistic from Joint Test | 16.826 | 0.932 | 10.385 | 1.770 |
| P-Value from Joint Test | 0.000 | 0.557 | 0.000 | 0.008 |
| Observations | 47,469 | 47,469 | 27,036 | 27,036 |

Notes. This table reports the results from regressions of the dependent variable on a variety of socio-demographic and academic covariates as well as zipcode by investigation year fixed effects. The censored removal stringency instrument is explained in detail in Section ???. 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 three, columns 3 and 4 report results for students enrolled in at least grade four during the maltreatment investigation and include standardized test scores. Full regression results available upon request. Standard errors are clustered by investigator.

Table B9: First Stage Effect of Censored Removal Stringency on Foster Placement

| | (1) | (2) | (3) | (4) |
|-----------------------------|---------------------|---------------------|---------------------|---------------------|
| | Foster Care | Foster Care | Foster Care | Foster Care |
| Censored Removal Stringency | 0.592*** (0.022) | 0.512*** (0.026) | 0.508*** (0.026) | 0.506*** (0.026) |
| Observations | 47,469 | 47,469 | 47,469 | 47,469 |
| F-Statistic | 699.783 | 388.389 | 389.063 | 387.127 |
| Zipcode by Year FE | | ✓ | ✓ | ✓ |
| Socio-Demographic Controls | | | ✓ | ✓ |
| Academic Controls | | | | ✓ |

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

Table B10: Effects of Foster Care Relative to Substantiation Without Removal

| | (1) Index of Child Wellbeing | (2) Confirmed Victim of Maltreatment | (3) Daily Attendance Rate | (4) Std Math Score |
|----------------------------------|---------------------------------------|---|------------------------------------|--------------------------|
| Foster Care and Substantiated | 0.181** (0.085) | -0.060* (0.035) | 0.063** (0.032) | 0.435* (0.229) |
| Substantiated | -0.007 (0.014) | 0.003 (0.006) | -0.004 (0.005) | -0.041 (0.036) |
| Observations | 242,233 | 242,233 | 224,925 | 177,118 |

Notes. This table reports the results from 2SLS regressions of the outcome variable on two treatment conditions: substantiation and foster care plus substantiation. It uses investigator stringency in evidence and risk levels to simultaneously instrument for the independent variables respectively. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. *p< 0.10, ** p< 0.05, *** p< 0.01. *p< 0.10, ** p< 0.05, *** p< 0.01.

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

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

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

Table B11: Sample Construction

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

Notes. Notes. The final analysis sample contains all child maltreatment investigations in Michigan that entered the rotational assignment system during the 2008-2016 school year of children enrolled in a public school in grades one through eleven old, that was assigned to investigators who worked at least 50 cases. I check for differential attrition out of the public school system using the sample reported in step 10 consisting of 248,730 investigations (shown in Table ?? and, since there is no differential attrition, the final analysis sample consists of students who ever enroll in a Michigan public school after their investigation.

Table B12: Comparing Sample to School Age Children who were Excluded from Analysis

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

Notes. Column 1 consists of investigations of children enrolled in a Michigan public school in grades one through eleven during their investigation. This corresponds to the sample detailed in step ten of Table ?? and is inclusive of students eventually dropped from the analysis sample because they did not enroll in a Michigan public school after their investigation. Column 2 consists of investigations that would have been included in the analysis sample had the child been enrolled in a Michigan public school in grades one through eleven. That is, the investigation entered the rotational assignment system, was assigned to an investigator who was assigned at least 50 investigations, and the child was old enough to have been enrolled in first grade—at least seven years old. These school age children were excluded from my analysis sample for at least five reasons: they were enrolled in private school, were homeschooled, dropped out of school altogether, went to school in a different state, or actually were enrolled in public school, but did not match to a public school record with high certainty.

Table B13: Breakdown of School Age Children Included and Excluded from Analysis Sample

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

Notes. To estimate the share of children with an investigation who fall into each category, I use Baye's Theorem to calculate, for example, the probability that a child was enrolled in private school, conditional on having a maltreatment investigation. In doing so, I use the following statistics, derived from the data: $P(\text{inv}) = 0.23$, $P(\text{inv}|\text{low income}) = 0.038$, $P(\text{inv}|\text{high income}) = 0.08$ and assume that the probability of being investigated conditional on income level is the same across public and private schools.

B.3 OLS Effects of Foster Care Placement Types

In recent years, many child welfare systems have prioritized placing children with relatives whenever possible. This is thought to limit disruption because it allows children to live with someone that they know and who shares their culture. Kinship placements also exhaust fewer state resources as it can be difficult to recruit other families to take in foster children. However, there is mixed research evidence on the effectiveness of kinship care relative to other placement types.

? exploit changes in monthly compensation rates and note that while low compensation rates are predictive of increased placements in kinship care, previous studies have found no relationship between compensation rates and child outcomes. They find that children who were placed in kinship care were more likely to be employed or in school, less likely to be incarcerated, and less likely to receive public assistance relative to children placed with an unrelated foster family. In contrast, ? exploits state and time variation in the adoption of laws that prioritize kinship placements and while they find that these laws shortened the amount of time that children spent in foster care, they do not detect evidence that it improves their children’s physical or mental health.

I add to this evidence by testing the effects of various types of foster placement. I can not perform this analysis in the IV framework because placement type is endogenous to unobservable characteristics of the child, like if they have support from nearby family members. However, I can use OLS to describe how the effects of removal vary based on placement type. Specifically, I estimate the following model:

$$Y_{icw} = \beta_0 + \beta_1 KINSHIP_{icw} + \beta_2 UNRELATED_{icw} + \beta_3 GROUP_{icw} + \beta_4 \mathbf{X}_{icw} + \theta_r + \epsilon_{icw} \quad (5)$$

where β_1 represents the association between kinship placement and the outcome relative to children who were not removed. Similarly, β_2 and β_3 report this relationship for unrelated foster homes and group homes respectively. The OLS estimates in the main analysis were biased such that they understated the benefits of removal and overstated the costs. To the extent that this analysis suffers from the same selection bias, this analysis might offer a lower bound for the effects of each placement type.

Table ?? shows the results. Overall, placement with relatives was associated with greater improvements than placement with an unrelated foster family or in a group home.

Table B14: OLS Effects of Foster Care on Child Outcomes, by Initial Placement Type

| | (1) Index of Child Wellbeing | (2) Confirmed Victim of Maltreatment | (3) Daily Attendance Rate | (4) Std Math Score |
|------------------------------|---------------------------------------|---|------------------------------------|--------------------------|
| Kinship | 0.027*** (0.007) | -0.007** (0.003) | 0.018*** (0.002) | 0.093*** (0.018) |
| Unrelated | -0.001 (0.009) | -0.003 (0.004) | 0.017*** (0.003) | 0.050** (0.025) |
| Group Home or Institution | -0.027 (0.017) | 0.008 (0.007) | 0.005 (0.008) | -0.046 (0.051) |
| Comparison Mean | 0.000 | 0.046 | 0.912 | -0.501 |
| Kinship vs Unrelated | 0.011 | 0.410 | 0.836 | 0.168 |
| Kinship vs Group | 0.002 | 0.044 | 0.104 | 0.010 |
| Unrelated vs Group | 0.158 | 0.165 | 0.142 | 0.088 |
| Observations | 242,264 | 242,264 | 224,925 | 177,118 |

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

B.4 Who Takes in Foster Children?

The administrative records that I use offer no individual-level information about foster parents. Moreover, there is limited public data about who takes in foster children. The best information comes from the American Community Survey (ACS), administered by the Census Bureau, which includes “foster children” as a category in a question about the members of a household. The ACS, however, is known to understate the number of foster children in the country by almost half relative to administrative records and is not thought to be representative. The leading explanations for why the ACS fails to account for so many foster children are that unrelated families who care for a foster child for only a short amount of time may not list them as a member of their household and that households who foster a relative may list them as relatives instead of as foster children (?).

Table ?? describes households with foster children and compares them to other households with members younger than 18 years old, using the 2012-2016 five year sample of the ACS. Nationwide, households with foster children were larger and much lower income. The head of households were older, less likely to be employed, and more likely to be black. The comparison looks similar when focusing only on households in Michigan.

Table B15: Descriptive Statistics of Households With and Without Foster Children

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

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

B.5 Reconciling Differences Between the Sample and Outcomes in ?

? finds that foster placement reduced earnings and increased criminality for children ages five to fifteen who were investigated in Illinois between 1990 and 2001. Though the study accesses both unsubstantiated and substantiated reports, it is limited to children who had received public assistance before their investigation. To best compare results, I restrict my analysis of the foster system in Michigan more recently to children in the same age window as ? who were eligible for free or reduced-price lunch in any school year before their investigation. I find positive effects of foster care for this comparable sample (Table ??).

Although the studies focus on different outcomes, there are two promising avenues to compare the earlier results to this study. First, both study juvenile justice involvement. ? finds that removal increased the likelihood of appearance before a juvenile court by about 300%. Although my estimate is somewhat imprecise, I can rule out increases in juvenile petition filings filed by more than 99%. Assuming foster children were no more or less likely to have petitions dismissed, then these two outcomes are comparable.

Second, ? finds that removal reduced annual earnings by \$1300 for adults between 18 and 28 years old. A back of the envelope calculation using estimates from ? suggests that an increase in standardized math scores of 0.47 standard deviations increases earnings at age 25 by about \$700. I use information from ?, Table 2 since it is one of the few studies linking test scores to adult earnings. The study reports that a school accountability program increased 10th grade math scores for students who had failed their 8th grade exam by 0.19 standard deviations (1.3 scale score points) and earnings at age 25 by \$298. I use this subgroup of students to mirror the low average baseline performance of children with a report of abuse or neglect. Therefore, although the outcomes are not entirely comparable, the evidence strongly suggests that foster care in Michigan between 2008 and 2016 did not have the same large and lasting negative effects as it did for those who were removed in Illinois between 1990 and 2002.

Table B16: Effects of Foster Care on Child Outcomes for Sample Comparable to ?

| | (1) Index of Child Wellbeing | (2) Alleged Victim of Maltreatment | (3) Confirmed Victim of Maltreatment | (4) Daily Attendance Rate | (5) Retained in Grade | (6) Std Math Score | (7) Std Reading Score | (8) Juvenile Delinquency |
|--------------|---------------------------------------|---|---|------------------------------------|------------------------------|--------------------------------|------------------------------|--------------------------------|
| Foster Care | 0.223*** (0.076) {-0.134} | -0.162** (0.068) {0.296} | -0.067** (0.032) {0.108} | 0.061** (0.028) {0.885} | -0.025 (0.031) {0.064} | 0.467** (0.194) {-0.683} | 0.203 (0.208) {-0.381} | -0.026 (0.040) {0.053} |
| Observations | 204,909 | 204,909 | 204,909 | 190,620 | 204,903 | 156,834 | 156,802 | 117,270 |

Notes. This table reports the results from 2SLS regressions of foster care on the dependent variable, using removal stringency to instrument for foster care. The analysis sample is restricted to children between the ages of five and fifteen during their investigation who were ever eligible for free or reduced price lunch prior to the investigation. The curly brackets below the standard error represent the control complier mean. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. The education and crime outcomes do not include all of the observations in the sample. Specifically, some grade level and daily attendance rate 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 is missing for eight counties, is available only through 2015, and is relevant only for children younger than Michigan's age of majority of sixteen. *p< 0.10, ** p< 0.05, *** p< 0.01.