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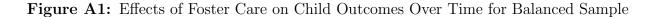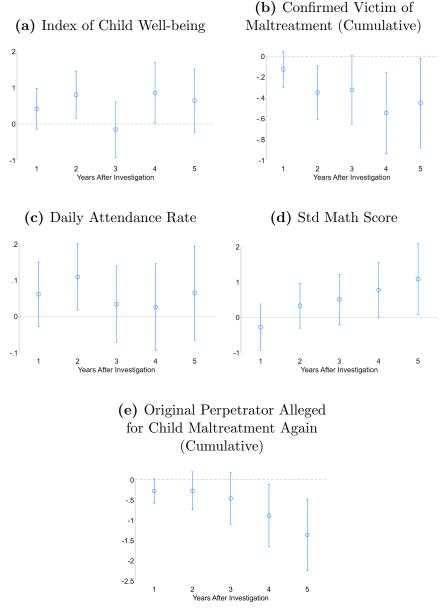
Temporary Stays and Persistent Gains: The Causal Effects of Foster Care

Max Gross and E. Jason Baron

A   Online Figures and Tables

A.1   Primary Online Figures and Tables

Figure A1: Effects of Foster Care on Child Outcomes Over Time for Balanced Sample

(a) Index of Child Well-being

(b) Confirmed Victim of Maltreatment (Cumulative)

(c) Daily Attendance Rate

(d) Std Math Score

(e) Original Perpetrator Alleged for Child Maltreatment Again (Cumulative)

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

1

Table A1: Effects of Foster Care on Permanency Placements

	(1) Reunified	(2) Adopted	(3) Guardianship	(4) Emancipated	(5) Still in FC in Sep 2017
Foster Care	0.703 (0.020)	0.064 (0.011)	0.040 (0.009)	0.017 (0.006)	0.176 (0.016)
% Conditional on Exiting	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 zip code by investigation year fixed effects. Standard errors are clustered by child. Each permanency outcome is mutually exclusive. Some students were still in the foster system at the end of the sample period in September 2017; these students are coded as such for their permanency outcome.

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

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

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

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

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

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

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

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

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

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

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

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

Table A6: Robustness Checks

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

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

A.2 Additional Tables

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

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

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

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

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

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

Table B3: Relationship Between Foster Care and Child Outcomes

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

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

Table B4: Effects of Foster Care on Taking Standardized Tests

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table B10: Robustness of the Main Results to Control Selection

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

	(1)	(2)	(3)	(4)	(5)
Std Math Score X Std Reading Score			0.007	0.007	0.007
			(0.004)	(0.004)	(0.004)
Std Math Score Squared			0.001	0.001	0.001
			(0.003)	(0.003)	(0.003)
Std Reading Score Squared			-0.004	-0.004	-0.004
			(0.002)	(0.002)	(0.002)
Std Math Score Cubed			-0.007	-0.007	-0.007
			(0.001)	(0.001)	(0.001)
Std Reading Score Cubed			-0.003	-0.003	-0.003
			(0.001)	(0.001)	(0.001)
School Controls					
Urban				-0.028	-0.027
				(0.005)	(0.005)
Charter				0.098	0.098
				(0.006)	(0.006)
% White				-0.010	0.006
				(0.020)	(0.021)
% Black				-0.064	-0.057
				(0.019)	(0.020)
%Free or Reduced Price Lunch				-0.131	-0.089
				(0.012)	(0.012)
Neighborhood Controls					
# Neighborhoods Lived in Before Investigation					-0.009
					(0.001)
Household Median Income					0.001
					(0.000)
Employment Rate					0.016
					(0.019)
% Bachelor's Degree or Higher					0.094
					(0.018)
% White					-0.003
					(0.029)
% Black					0.005
					(0.030)
Homeless in SY Before Investigation					-7.625
					(0.738)
Observations	242,233	242,233	242,233	242,233	242,233
Rotation Group FE	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓
Investigation Controls			✓	✓	✓	✓
Academic Controls			✓	✓	✓	✓
School Controls				✓	✓
Neighborhood Controls					✓

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

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

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

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

B Censored Data and the Examiner Assignment Research Design

The examiner assignment research design used in this study has been widely applied recently as increased access to large administrative datasets allows researchers to exploit discretionary decision-making. It has been used to study a variety of interventions other than foster care, such as juvenile incarceration (Aizer and Doyle, 2015; Eren and Mocan, 2017), adult incarceration (Kling, 2006; Mueller-Smith, 2015), disability insurance (Dahl, Kostøl and Mogstad, 2014), student loan repayment (Herbst, 2018), and evictions (Collinson and Reed, 2019; Humphries et al., 2019), among others. In many of these settings, treatment assignment is a two-step selection process in which individuals are assigned to treatment only after crossing an initial decision threshold. For example, in the context of foster care, children can only be removed if their maltreatment allegation is first substantiated.[1] Similarly, in the criminal justice setting, defendants can only be incarcerated conditional on being convicted. Whether due to restrictions from data partners or privacy considerations, some studies apply this design using partially censored data that contain only individuals who cross the initial decision threshold—for example, only substantiated investigations or only convicted defendants. Such restrictions appear in two recent studies of foster care (Bald et al., 2019; Roberts, 2019) as well as in other contexts (Eren and Mocan, 2017; Herbst, 2018; Kling, 2006), and may introduce bias.

Potential Bias

To understand the source of potential bias, consider decisions made by investigators in the context of foster care. Substantiation decisions are based on the strength of the evidence, whereas placement decisions are based on the child's risk of future harm.[2] The research design assumes that, due to random assignment, the distribution of risk is identical across investigators and, therefore, identifies impacts using exogenous variation in investigator tolerance over risk. However, if investigators also vary in their stringency over evidence, the set of substantiated cases may not be balanced across investigators. Therefore, restricted data access can create a violation of the exclusion restriction.

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

---

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

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

tendencies.[3] 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.

Replication with Censored Data

Although this is not the first study to describe the potential for bias from censored data, it is the first to shed light on how much it can matter in practice. Using data containing the universe of child welfare investigations in Michigan, including both unsubstantiated and substantiated allegations, we replicate the main analysis as if we only had access to substantiated cases. Using only the sample of substantiated investigations, we reconstruct the removal instrument according to Equation 1.[4] A standard balance test reveals that a variety of baseline characteristics which are associated with foster care placement are not jointly predictive of the new instrument (Table B12). Therefore, since exogeneity appears to hold using the subset of substantiated investigations, one might expect the 2SLS results to be consistent with the full sample.[5] However, this turns out to not be the case.

Table B14 shows that the effects using the complete data (Panel A) are much larger than those found when restricted to substantiated investigations (Panel B).[6] [7] The replication exercise produces a substantively smaller impact on the index of child well-being. The effect on daily attendance rate is moderately smaller than the effect using the complete data but still statistically significant, whereas the point estimate on math test scores is much smaller and imprecise.[8] The findings in Panel B of Table B14 are somewhat similar to those in Bald et al. (2019), which finds noisy estimates for school-age children, and to Roberts (2019) which reports imprecise estimates on test scores but positive effects for on-time grade progression. Although institutional differences between the child welfare systems in Michigan, Rhode Island, and South Carolina surely contribute to the different findings, this exercise documents that bias in the other studies may also play a role. Overall, this exercise cautions against

---

[3]Investigators at the 10th percentile substantiated at a rate 8.4 percentage points less than the average investigator in their local area while investigators at the 90th percentile did so at a rate 8.9 percentage points greater.

[4]Table B13 shows that there exists a strong first stage relationship with the censored instrument.

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

[6]It is possible that Panel B in Table B14 represents a different LATE than Panel A. To address this potential concern, we use investigator tendencies over substantiation and removal to instrument for both foster placement and substantiation. Table B15 shows that the estimates in Panel B are also smaller than the causal effects of placement relative to substantiation from the complete data.

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

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

applying the examiner assignment design with censored data.

Assessing Arteaga (2019) Approaches to Using Examiner Assignment Design with Censored Data

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

Arteaga (2019) shows how the standard examiner assignment design can not be applied in this context and derives an estimator of the causal effects of incarceration relative to conviction that can be identified using censored data.[9] 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:

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

where Y is a child outcome and T denotes treatment assignment: substantiated but not removed (t_S) or substantiated and removed (t_R).[10] $P_S(Z) = p_S$ represents that the evidence threshold to substantiate is held fixed at p_S and $P_R^*(Z) = p_R^*$ means that the removal threshold conditional on substantiation is equal to p_R^*. Integrating over the inside term averages the effect across all investigators.

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

The first, called the pooled approach, uses P_R^* to instrument for foster care while additionally controlling for linear and quadratic terms of P_S and all interactions. The second, called the tercile approach, instruments for placement with P_R^* separately for each tercile of the evidence stringency distribution. The idea is that, in addition to controlling for evidence stringency,

---

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

[10]This is equivalent to Equation 13 in Arteaga (2019).

splitting the data into terciles approximates fixing the evidence threshold. Lastly, the third approach, called the rolling window approach, mirrors the tercile approach yet estimates impacts more flexibly along the distribution of evidence stringency. Specifically, it sorts the sample by the evidence stringency of the assigned investigator and estimates impacts of placement for the lowest 18,000 observations of the distribution. Then it repeats this process for the lowest 500 to 18,500, and so on.

Table B17 shows the results of the first two approaches and Figure B2 shows the results from the third. As a benchmark, both the table and figure also include estimates of foster care relative to substantiation identified from the full, uncensored data. To identify this parameter, we use measures of investigator removal and substantiation stringency to simultaneously instrument for both foster placement and substantiation. The table and figure show the effects on the index of child well-being.

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

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

Table B12: Balance Tests Using Censored Data

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

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

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

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

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

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

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

 Notes. Panel A reports the 2SLS results from Table 4 while Panel B reports the results from 2SLS regressions of the outcome variable on foster care using censored removal stringency to instrument for foster care. The sample in Panel B is restricted to only substantiated investigations. The size of the bias represents the absolute value of the difference between the point estimate in Panel A (the effect using the complete data) and Panel B (the biased effect). All regressions include the covariates as listed in the text and zip code by investigation year fixed effects. Standard errors are clustered by child.
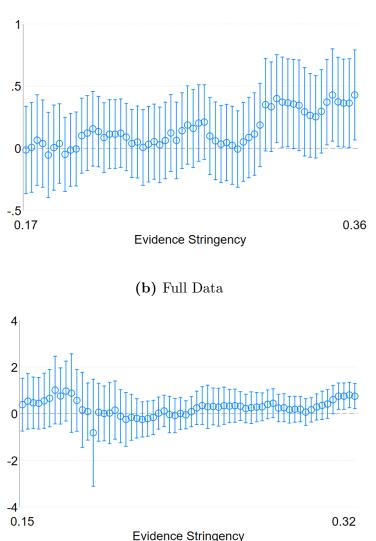
Table B15: Effects of Foster Care Relative to Substantiation Without Removal

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

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

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

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

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

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

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

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

Figure B2: Assessing Arteaga (2019) Rolling Window Approach to Examiner Assignment Design with Censored Data

(a) Censored Data

(b) Full Data

Notes. This figure compares the estimates of foster care relative to substantiation on the index of child well-being using the rolling window approach proposed in Arteaga (2019) with both the censored and full data. The graphs plot both the point estimates and their 95 percent confidence intervals. All specifications include the covariates as listed in the text as well as zip code by investigation year fixed effects. Standard errors are clustered by child. Figure B2a sorts the censored data based on evidence stringency and estimates separate regressions of the index of child outcomes on foster care using removal stringency conditional on substantiation to instrument for foster care and including evidence stringency as a covariate. Since the sample size is similar to our study, we follow Arteaga (2019) in using a rolling window of 18,000 observations and adjust the window by 500 observations each time along the evidence threshold. Figure B2b applies the same approach to the full, uncensored data. We use removal stringency to instrument for foster care and evidence stringency to instrument for substantiation to estimate the effect of foster care relative to substantiation. Since the sample size is about five times larger with the full data, we use a rolling window of 90,000 observations and adjust the window by 2,500 observations each time.

C   Comparison to Doyle (2007, 2008)

Pathbreaking research in Doyle (2007) using administrative data from Illinois found that foster placement greatly harmed children's outcomes. It reduced quarterly earnings as an adult (ages 18 to 28) by about $1,300, increased teenage pregnancy by two times, and increased juvenile delinquency by three times. Follow-up work in Doyle (2008) also found that placement increased adult criminality by three times.[11] Using the same research design, we find that placement had a protective effect, improving children's safety and educational outcomes. As discussed in Section V, we can statistically reject that foster placement in Michigan caused the large harmful impacts found in the early work. Moreover, using a one-sided hypothesis test, we can rule out altogether that placement reduced the index of child well-being.

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

1. State-level differences in foster care placements. Foster placements were considerably longer and less stable in Illinois during the Doyle (2007, 2008) sample period than in Michigan more recently. For example, the median duration of foster care in Illinois during the early period was 40 months, compared to just 15.8 months in Michigan in 2008 (the first year of our sample period) and 12.8 months in 2017 (the final year of our panel) (USDHHS, 2003, 2017; Wulczyn, Hislop and Goerge, 2000).[12] Similarly, 44.8 percent of foster children in Illinois in 1998 had lived in three or more different foster homes compared to an average of just 31 percent across our 10-year panel in Michigan (AECF, 2017; USDHHS, 2003). Similar trends hold among marginal placements as well; Illinois children at the margin of placement spent an average of four to five years in foster care, relative to 19 months in our context. Thus, the difference in findings across settings may largely be explained by these tremendous institutional differences.

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

---

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

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

every quartile of the length distribution (Wulczyn, Hislop and Goerge (2000), Figure 4.1).[13] To offer more evidence, just over one in four children who entered foster care in Illinois between 1988 and 1995 were still in the foster care system as of December 1998, compared to an average of less than one in 10 across 10 other states with reliable data (Wulczyn, Hislop and Goerge (2000), Figure 5.1). Placement in Illinois was also less stable than in other states; in 1998, Illinois had the third-highest share of foster children who lived in three or more different foster homes among 41 states with quality data (USDHHS, 2003).

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

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

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

3. Differences in the sample definition. The sample in Doyle (2007) included children ages 5–15 who had received Medicaid before their investigation. To assess whether these sample restrictions could have driven the differences in findings, we restrict our analysis to children ages 5–15 who were eligible for free or reduced-price lunch in any school year prior to the investigation. We find estimates of foster care placement very similar to our main analysis (Table B18).[14] Moreover, using a one-sided hypothesis test, we can statistically reject that placement worsened the index of child well-being. Therefore, differences in sample

---

[13]Specifically, the median duration was 4.5 times longer at the 25th percentile than the average of the 11 other states, 4.0 times longer at the 50th percentile, and 2.5 times longer at the 75th percentile (Wulczyn, Hislop and Goerge (2000), Figure 4.1).

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

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

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

Compliers in Doyle (2008) were older than in our study. Specifically, they were 45 percent more likely to be ages 11 to 13 than the overall sample. In contrast, compliers in our setting were 11 percent more likely to be age 10 or below.[15] For this difference in the complier population to translate into differences in findings, there must also be heterogeneous impacts by age. We find that the benefits of foster care were largest for younger children (Table A4), whereas Doyle (2007) finds the harm was greatest for older youth. This pattern is consistent with the differences in results. However, Doyle (2008) finds similar results for children older and younger than age 10, so it is unclear whether heterogeneity by age drives the divergent findings.

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

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

About 2.5 per 1000 children in Illinois entered foster care in 1990, the first year of the early studies (Wulczyn, Hislop and Goerge, 2000). This rose by 76 percent over the next four years such that in 1994, 4.4 per 1000 children entered foster care, and declined to around 2 per 1000 in 2001, the final year of the Doyle (2007) sample period (USDHHS, 2003).[16] In comparison, the placement rate in Michigan remained around 3 per 1000 children during the

---

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

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

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

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

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

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

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

D   Data Appendix

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

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

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

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

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

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

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

Table B19: Sample Construction

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

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

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

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

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

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

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

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

E OLS Effects of Foster Care Placement Types

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

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

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

$$Y_{iw} = \beta_0 + \beta_1 KINSHIP_{iw} + \beta_2 UNRELATED_{iw} + \beta_3 GROUP_{iw} + \beta_4 X_{iw} + \theta_r + \epsilon_{iw} \quad \text{(E.1)}$$

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

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

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

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

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

F   Who Takes in Foster Children?

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

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

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

| | USA | | Michigan | |
	(1) At Least One Child Under 18	(2) At Least One Foster Child	(3) At Least One Child Under 18	(4) At Least One Foster Child
# Adults	2.14	2.25	2.08	2.06
# Children Under Age 18	1.88	2.61	1.89	2.97
Pre-Tax Income	$141,431	$69,948	$131,038	$62,067
Head of Household				
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.

References

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

Aizer, Anna, and Joseph J Doyle. 2015. "Juvenile incarceration, human capital, and future crime:
Evidence from randomly assigned judges." *The Quarterly Journal of Economics*, 130(2): 759–803.

Arteaga, Carolina. 2019. "The Cost of Bad Parents: Evidence from the Effects of Parental Incarceration
on Children's Education." Working paper.

Bald, Anthony, Eric Chyn, Justine S. Hastings, and Margarita Machelett. 2019. "The causal
impact of removing children from abusive and neglectful homes." National Bureau of Economic Research
Working Paper 25419.

ChildTrends. 2018. "Foster Care." Child Trends Databank.

Collinson, Robert, and Davin Reed. 2019. "The effects of evictions on low-income households."
Working paper.

CRHE. 2017. "Homeschooling by the numbers." Coalition for Responsible Home Education.
https://www.responsiblehomeschooling.org/homeschooling-101/homeschooling-numbers/.

Dahl, Gordon B, Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family welfare cultures."
The Quarterly Journal of Economics, 129(4): 1711–1752.

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

Doyle, Joseph J. 2008. "Child protection and adult crime: Using investigator assignment to estimate
causal effects of foster care." *Journal of political Economy*, 116(4): 746–770.

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

Hayduk, Iryana. 2017. "The Effect of Kinship Placement Laws on Foster Children's Well-Being." *The
B.E. Journal of Economic Analysis & Policy*, 17(1): 1–23.

Herbst, Daniel. 2018. "Liquidity and Insurance in Student Loan Contracts: Estimating the Effects of
Income-Driven Repayment on Default and Consumption." Working paper.

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

Kling, Jeffrey R. 2006. "Incarceration length, employment, and earnings." *American Economic Review*,
96(3): 863–876.

Lovett, Nicholas, and Yuhan Xue. 2018. "Family First or the Kindness of Stangers? Foster Care
Placements and Adult Outcomes." Working paper.

Mack, Julie. 2017. "Where Michigan children attended school in 2016-2017 – public and private." MLive.
https://www.mlive.com/news/2017/09/where_michigan_children_attend.html.

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

OCA. 2018. "Examining Connecticut's Safety Net for Children Withdrawn from School for the Purpose of Homeschooling." Office of the Child Advocate, State of Connecticut.

O'Hare, William P. 2007. "Census Bureau Plans to Eliminate 'Foster Child' Category." Population Reference Bureau, https://www.prb.org/censusbureaufosterchildcategory/.

Roberts, Kelsey V. 2019. "Foster Care and Child Welfare." Working paper.

Ryan, Joseph P, Brian A Jacob, Max Gross, Brian E Perron, Andrew Moore, and Sharlyn Ferguson. 2018. "Early exposure to child maltreatment and academic outcomes." *Child maltreatment*, 23(4): 365–375.

Stagner, Matthew. 2019. "Getting Closer: Embracing the Emotional Aspects of Our Craft to Help Policy Research Matter More." https://www.mathematica.org/commentary/getting-closer-embracing-the-emotional-aspects-of-our-craft-to-help-policy-research-matter-more.

USDHHS. 2003. "Child Welfare Outcomes." Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau.

USDHHS. 2008. "Child Welfare Outcomes." Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau.

USDHHS. 2016. "Child Welfare Outcomes." Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau.

USDHHS. 2017. "Child Welfare Outcomes Report Data." Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services. https://cwoutcomes.acf.hhs.gov/cwodatasite/byState.

White, Rachel, and Ben DeGrow. 2016. "A Survey of Michigan's Private Education Sector." Mackinac Center for Public Policy.

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