

Advancing the Opportunities of Underserved Students: Lessons from Child Welfare, Education, and the Labor Market

by

Max Gross

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2020

Doctoral Committee:

Professor Brian A. Jacob, Chair
Professor Charlie Brown
Assistant Professor Michael Mueller-Smith
Associate Professor Joseph P. Ryan
Associate Professor Kevin Stange

Max Gross

maxgross@umich.edu

ORCID ID [0000-0002-8573-5270](https://orcid.org/0000-0002-8573-5270)

© Max Gross 2020

For Edie, Josh, Milton, and Rea

ACKNOWLEDGEMENTS

This dissertation could not have happened without supportive advising from my dissertation committee: Brian Jacob, Michael Mueller-Smith, Joseph Ryan, Charlie Brown, and Kevin Stange. Brian taught me how to be an applied microeconomist and showed confidence in me and my work during the many times when I did not. Mike challenged me to be a better researcher through methodological rigor and provided a crucial space to speak openly about mental health. Joe's interdisciplinary perspective grounded my work in public policy. Charlie's attention to detail forced me to wrestle with important questions that nobody else could have come up with. Kevin's contagious positive attitude encouraged me to persevere when faced with challenges.

Other faculty members also offered valuable feedback over the years, including Ana Reynoso, Ashley Craig, Christina Weiland, Jeff Smith, John Bound, Martha Bailey, Mel Stephens, Sara Heller, and Sue Dynarski. My University of Michigan colleagues were pivotal to my research and graduate school experience, including but not limited to Andrew Simon, Anirudh Jayanti, Anna Shapiro, Ariel Binder, Brittany Vasquez, Daniel Hubbard, Dhiren Patki, Fernando Furquim, Gloria Yeomans-Maldonado, Jordan Rhodes, Meghan Oster, Michael Ricks, Parag Mahajan, Shawn Martin, Silvia Robles, Stacey Brockman, Stephanie Owen, Thomas Goldring, and Xiaoyang Ye. In particular, my friendships with George Fenton and Matt Gross have kept me sane since the first day of math camp. Or at least as sane as possible.

The second and third chapters of my dissertation were coauthored with brilliant researchers and supported by fantastic colleagues. I wrote the second chapter together with Robert Fairlie and Silvia Robles. Thomas Barrios also played a critical role in the early stages of the project. Amanda Pallais, Ed Glaeser, Joshua Angrist, Larry Katz, and Lesley Turner offered helpful comments. Andrew LaManque, Bob Barr, Howard Irvin, Jerry Rosenberg, Kathleen Moberg, Lydia Hearn, Mallory Newell, Rowena Tomaneng, and Stephen Fletcher generously provided access to the student-level administrative data and detailed information on courses, minority student programs, and registration procedures. The third chapter is coauthored with Brian Jacob and Kelly Lovett. Ariella Meltzer, Hannah Zlotnick, Pieter De Vlieger, and Thomas Goldring provided excellent research assistance. Connect Detroit and

Detroit Employment Solutions Corporation were great partners in this work.

Research partnerships have fueled my dissertation work. Jasmina Camo-Biogradlija, Julie Monteiro de Castro, Kyle Kwaiser, Nicole Wagner, and Pam Soltman of the Education Policy Initiative offered guidance with both education data and financial resources. Andrew Moore, Brian Perron, Joseph Ryan, and Terri Gilbert of the Child and Adolescent Data Lab generously shared the child welfare and juvenile justice data used in the first chapter. Many child welfare employees, especially those from Jackson County, generously helped me to understand the child welfare and foster systems. I am grateful for the Michigan Department of Education (MDE) and the Center for Educational Performance and Information (CEPI) which provided access to the administrative education records used in chapters one and three. This data was structured and maintained by the Michigan Consortium for Education Research (MCER). MCER data are modified for analysis using rules governed by MCER and are not identical to data collected and maintained by MDE and CEPI. Any opinions, findings, conclusions, or recommendations expressed in this dissertation are those of the author and do not reflect the view of any other entity. I am also appreciative of funding from the Institute of Education Sciences, U.S. Department of Education through PR/Award R305B150012# and Grant R305E100008.

My experiences as an undergraduate inspired my pursuit of a doctoral degree. Ethan Kaplan gave me my first taste of economics research. John Straub offered confidence that I could succeed in a doctoral program. Jon Grunewald, Jordan Cedarleaf-Pavy, Kyle Baylor, Paul Levy, Rishi Sugla, and Ross Heise continue to be great sources of intellectual curiosity. Matt Epstein's enthusiasm for behavioral economics brought me excitement for economics broadly and enlightening conversations with Shivani Kochhar motivated me to pursue the economics of education more specifically.

A variety of outlets have provided much-needed relief from the stressors of graduate school which, in turn, has made my research stronger. Yet none was as important as *The Rights to Ricky Sanchez*. Spike Eskin and Michael Levin's podcast, and my subsequent conversations about it with Tim Parisi, have been a constant source of happiness in my life since *The Process* began several decades ago.

I could not have written this dissertation without the incredible support of my family. My Mom and Dad never miss a moment to tell me how proud they are of me; their unconditional love has made me both a better person and a more thoughtful researcher. Misha and Nate continue to educate me on issues related to social justice, which has intimately shaped the topics addressed in this dissertation. Despite falling asleep on several occasions as I practiced for academic seminars, Rachel's emotional support throughout graduate school has kept me going. She has edited the writing of my dissertation — teaching me how to use the Oxford

comma and kindly nudging me to use fewer commas otherwise. And she brought Ferguson into my life, introducing joy that I never imagined possible. I am forever grateful for each of you.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF APPENDICES	xii
ABSTRACT	xiii
CHAPTER	
I. Temporary Stays and Persistent Gains: The Causal Effects of Foster Care 1	
1.1 Introduction	1
1.2 Overview of the Child Welfare System in Michigan	5
1.2.1 Child Maltreatment Investigations	5
1.2.2 Foster Care System	7
1.3 Data Sources and Sample Construction	8
1.3.1 Administrative Data Sources	8
1.3.2 Overview of Analysis Sample	9
1.4 Empirical Strategy	10
1.4.1 Research Design	11
1.4.2 Identifying Assumptions	13
1.5 Causal Effects of Foster Care on Children’s Outcomes	14
1.5.1 Effects on Child Safety, Academics, and Crime	15
1.5.2 Mechanisms	16
1.5.3 Compliers Analysis, Subgroup Effects, and Robustness Checks	20
1.6 Potential Bias in Examiner Assignment Research Design from Censored Data	22
1.7 Conclusion	25
1.7.1 External Validity	25

1.7.2	Implications for Public Policy	26
II.	The Effect of Course Shutouts on Community College Students: Evidence from Waitlist Cutoffs	40
2.1	Introduction	40
2.1.1	Related Literature	42
2.2	Institutional Background	45
2.2.1	Data Sources	46
2.2.2	Section Enrollment	46
2.2.3	Sample Characteristics	47
2.3	Empirical Strategy	49
2.3.1	Construction of the Running Variable	49
2.3.2	Estimation	51
2.3.3	Validity Checks	53
2.4	Course Scarcity and Student Outcomes	55
2.4.1	First Stage Estimates	55
2.4.2	Reduced Form and IV Estimates	55
2.4.3	Subgroup Analysis	58
2.4.4	Sensitivity Analysis	59
2.4.5	Complier Densities	60
2.5	Conclusion	62
III.	The Effect of Summer Employment on the Educational Attainment of Under-Resourced Youth	78
3.1	Introduction	78
3.2	Prior Literature	80
3.3	Institutional Background	82
3.3.1	Junior Police and Fire Cadets	83
3.3.2	Community-Based Organizations (CBOs)	83
3.3.3	Industry Led Training and Career Pathways Programs	84
3.4	Data and Sample	84
3.5	Empirical Strategy	86
3.6	Effects of Participation in GDYT	90
3.6.1	Robustness to Omitted Variable Bias	92
3.6.2	Subgroup Analysis	93
3.7	Conclusion	95
APPENDICES		106
BIBLIOGRAPHY		175

LIST OF FIGURES

Figure

1.1	Comparison of State Foster Care Systems	28
1.2	Overview of Child Maltreatment Investigations in Michigan	29
1.3	Distribution of Investigator Removal Stringency Instrument	30
1.4	Effects of Foster Care on Likelihood of Being in Foster System Over Time .	31
1.5	Effects of Foster Care Over Time	32
2.1	A Hypothetical Registration Log	64
2.2	Density of the Running Variable	65
2.3	First Stage Effect of Missing the Waitlist Cutoff on Enrollment in Waitlisted Section and Course	66
2.4	Effect of Missing the Waitlist Cutoff on Transfers to Other Two-Year Schools, by Ethnicity	67
2.5	Effect of Missing the Waitlist Cutoff on Transfers to Four-Year Schools, by Ethnicity	68
2.6	Effect of Missing the Waitlist Cutoff on Bachelors Degree Completion, by Ethnicity	69
2.7	Reduced Form Effect of Missing a Placebo Cutoff	70
2.8	Density of Potential Outcomes for Treated and Untreated Compliers	71
3.1	Distribution of Baseline Characteristics for Participants and Non-Participants	98
A.1	Foster Care Entry Per 1000 Children	108
A.2	Assessing Arteaga (2019) Rolling Window Approach to Examiner Assignment Design with Censored Data	129
B.1	Covariate Smoothness Across the Waitlist Cutoff	153
B.1	Covariate Smoothness Across the Waitlist Cutoff	154
B.1	Covariate Smoothness Across the Waitlist Cutoff	155
B.2	Reduced Form Effect of Missing the Waitlist Cutoff on Course Load and Persistence	156
B.3	Reduced Form Effect of Missing the Waitlist Cutoff on Transfers to Another Two-Year School	157
B.4	Reduced Form Effect of Missing the Waitlist Cutoff on Transfers to a Four-Year School	158
B.5	Density of the Time Running Variable	161
B.6	First Stage Effect of Missing the Waitlist Cutoff on Enrollment in the Waitlisted Section	162

LIST OF TABLES

Table

1.1	Summary Statistics	33
1.2	First Stage Effect of Removal Stringency on Foster Placement	34
1.3	Balance Tests for the Conditional Random Assignment of Investigators	35
1.4	Effects of Foster Care on Child Outcomes	36
1.5	Effects of Foster Care Over Time	37
1.6	Effects of Adult Interventions on Child Outcomes	38
1.7	Effects of Foster Care on Child Outcomes Using Censored Data	39
2.1	Summary Statistics	72
2.2	Frandsen Manipulation Test for Discrete Running Variables	73
2.3	Test for Balance of Pre-determined Student Characteristics Across the Waitlist Cutoff	74
2.4	First Stage Effect of Missing the Waitlist Cutoff on Enrollment in Waitlisted Section and Course	75
2.5	Effect of Missing the Waitlist Cutoff on Course Load and Persistence	76
2.6	Effect of Missing the Waitlist Cutoff on Transfers and Degree Completion	77
3.1	Summary Statistics	99
3.2	Summary Statistics, by Match Group Characteristics	100
3.3	Balance Tests of Differences Between Participants and Non-Participants	101
3.4	Falsification Tests of the Effect of Participation in GDYT on 6th Grade Educational Outcomes	102
3.5	The Effect of Participation in GDYT on Educational Outcomes	103
3.6	Robustness of the Effects of Participation in GDYT to Omitted Variable Bias	104
3.7	The Effect of Participation in GDYT on Educational Outcomes, by Subgroup	105
A.1	Effects of Foster Care on Michigan Public School Enrollment	109
A.2	Full Balance Tests of the Conditional Independence of the Removal Stringency Instrument	110
A.3	Testable Implications of the Exclusion of Removal Stringency Instrument	111
A.4	Testable Implications of Monotonicity of the Removal Stringency Instrument	112
A.5	Effects of Foster Care on Taking Standardized Tests	113
A.6	Effects of Foster Care on High School Graduation and College Enrollment	114
A.7	Effects of Foster Care on Type of Foster Placement	115
A.8	Effects of Foster Care on Neighborhood and School Environment Over Time	116
A.9	Effects of Foster Care on Permanency Placements	117

A.10	Characteristics of Compliers at the Margin of Foster Placement	118
A.11	Effects of Foster Care on Children’s Experience in Foster System	119
A.12	Effects of Foster Care on Index of Child Wellbeing, by Age and Gender . .	120
A.13	Robustness Checks	121
A.14	Balance Tests Using Censored Data	122
A.15	First Stage Effect of Censored Removal Stringency on Foster Placement . .	123
A.16	Effects of Foster Care Relative to Substantiation Without Removal	124
A.17	OLS Effects of Foster Care on Child Outcomes Using Censored Data . . .	125
A.18	Assessing Arteaga (2019) Approaches to Examiner Assignment Design with Censored Data	128
A.19	OLS Effects of Foster Care on Child Outcomes, by Initial Placement Type	131
A.20	Descriptive Statistics of Households With and Without Foster Children . .	132
A.21	Sample Construction	135
A.22	Comparing Sample to School-Age Children who were Excluded from Analysis	136
A.23	Breakdown of School-Age Children Included and Excluded from Analysis Sample	137
A.24	Effects of Foster Care on Child Outcomes for Sample Comparable to Doyle (2007)	139
B.1	Student Initial Education Goal	141
B.2	Effect of Missing the Waitlist Cutoff on Transfers to Nearby Two-Year Schools within Two Years	142
B.3	Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Gender	143
B.4	Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Ethnicity	144
B.5	Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Course Popularity	145
B.6	Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Waitlisted Subject	146
B.7	Effect of Missing the Waitlist Cutoff on Transfers to Other Two-Year Schools, by Ethnicity	147
B.8	Effect of Missing the Waitlist Cutoff on Transfers to Four-Year Schools, by Ethnicity	148
B.9	Effect of Missing the Waitlist Cutoff on Bachelors Degree Completion, by Ethnicity	149
B.10	Reduced Form Effect of Missing a Placebo Cutoff	150
B.11	Robustness Checks	151
B.12	First Stage Effect of Missing the Waitlist Cutoff on Enrollment in Waitlisted Section, Using Time Running Variable	163
B.13	Effect of Missing the Waitlist Cutoff on Course Load, Using Time Running Variable	164
C.1	Balance Tests of Differences Between Junior Police and Fire Cadet Participants and Age-Eligible Non-Participants	166
C.2	Description of the Sample for Each Outcome and the Formula for Outcomes in Follow-up Years One and Two	167

C.3	The Effect of Participation in GDYT on Educational Outcomes in Follow-up Year One and Follow-up Year Two	169
C.4	Balance Tests of Differences Between Participants and Non-Participants, by Subgroup	170
C.5	Falsification Tests of the Effect of Participation in GDYT on 6th Grade Educational Outcomes, by Subgroup	172
C.6	The Effect of Participation in GDYT on Additional Educational Outcomes, by Subgroup	173

LIST OF APPENDICES

Appendix

- A. Appendix to Temporary Stays and Persistent Gains: The Causal Effects of Foster Care 107
- B. Appendix to The Effect of Course Shutouts on Community College Students: Evidence from Waitlist Cutoffs 140
- C. Appendix to The Effect of Summer Employment on the Educational Attainment of Under-Resourced Youth 165

ABSTRACT

This dissertation studies the consequences of public policy on the opportunities of underserved students. It focuses on three policy environments: child welfare, education, and the labor market. In each chapter, I use a rigorous quantitative methodology and large administrative data to understand how public institutions are—or are not—improving the wellbeing of historically disadvantaged children and youth.

The first chapter focuses on placement into foster care, which is experienced by 6% of children in the United States 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 foster system when most were reunified with their birth parents, suggesting that improvements made by their birth parents was an important mechanism. Given recent federal legislation to reduce foster placements, these findings indicate that child welfare systems must invest in more effective interventions to keep vulnerable children safe and thriving in their homes.

Chapter two, with Silvia Robles and Robert W. Fairlie, examines course availability, a frequently cited yet understudied channel through which money matters for college students. Open admissions policies, binding class size constraints, and heavy reliance on state funding may make this channel especially salient at community colleges, which enroll 47% of U.S. undergraduates in public colleges and 55% of underrepresented minority students. We use administrative course registration data from a large community college in California to test this mechanism. By exploiting discontinuities in course admissions created by waitlists, we find that students stuck on a waitlist and shut out of a course section were 25% more likely to take zero courses that term relative to a baseline of 10%. Shutouts also increased transfer rates to nearby, but potentially lower quality, two-year colleges. These results document that course availability, even through a relatively small friction, can interrupt and distort community college students educational trajectories.

The third chapter, with Brian A. Jacob and Kelly Lovett, evaluates a summer youth

employment program, a popular way for municipalities to provide adolescents with skills and experiences thought to improve labor market outcomes. While research evidence on such programs has grown in recent years, it is still limited. In particular, it is not clear how, if at all, participation influences key educational outcomes. We study the program in Detroit, Michigan using a selection on observables identification strategy. In addition to controlling for a rich set of covariates, including baseline educational measures, we match participants to their classmates of the same race and gender who applied for the program but did not participate. We find that participation is associated with a modest increase in educational attainment. Specifically, it increased the likelihood of enrolling in public school after the program by 1.5% and of graduating high school by 4%, relative to comparison means of 94.5% and 85%. Youth with the weakest academic skills benefited the most, as participation increased school enrollment by 2.2% and high school graduation by 5.5% for this group. Falsification tests of whether participation predicts pre-program characteristics, as well as robustness checks which account for omitted variable bias, as proposed in Oster (2016), suggest that our results are not driven by unobservable differences between participants and other applicants.

CHAPTER I

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

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

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

1.1 Introduction

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

This paper makes several contributions. First, it provides new estimates of the causal effects of foster care on crucial indicators of child 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 large, 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. Taken together, these estimates indicate that foster care had benefits in cases where investigators might disagree about placement, which is a critical population for child welfare policy (Berrick, 2018).

The results contrast Doyle (2007, 2008) which use the same research design but find that foster care reduced earnings and increased crime for Illinois children investigated in the 1990s and early 2000s.¹ There are several possible explanations for this discrepancy. A likely reason is that children’s experiences while in the Illinois foster system were especially harmful. For example, foster children in Illinois remained in the system longer than in any other state at the time and they changed foster homes at a higher rate than all but two (Figure 1.1). Therefore, placement in other states may have been less damaging than in Illinois and perhaps beneficial. Importantly, evidence from this study is more likely to be representative because the system in Michigan functions similarly to others across the country. A second explanation is that shifts in child welfare practice over time may have helped foster systems improve nationwide, such as increasing placements with relatives and decreasing length of stay in care (ChildTrends, 2018). A third potential reason could be that Illinois placed too many children in foster care, removing those who faced little risk in the home, while Michigan better targeted placement. This is unlikely, however, because Illinois placed children at a similar rate during the early research setting as Michigan more recently (Figure A.1).

The second contribution of this study is that it explores mechanisms by exploiting the

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

fact that foster care is a temporary intervention. In my setting, children were in the foster 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.² One explanation for this surprising pattern is that birth parents, who worked closely with social workers following child removal, improved their parenting skills. Accordingly, 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 also rule out several alternative mechanisms that could, in theory, drive impacts. For example, though by definition, children moved to new homes when they were removed, and prior work highlights the large impacts of geography on child outcomes (Chetty et al., 2016; Chyn, 2018), I find no evidence that placement caused lasting improvements to children’s neighborhoods or schools.

Third, this paper provides causal evidence on the impacts of child welfare interventions that target adults, which are an understudied channel through which foster placement can impact children. Specifically, the birth parents of foster children received community-based 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. A careful examination of mechanisms requires disentangling the role of these adult services from the dramatic changes that occurred in foster children’s own lives, yet doing so is challenging because they take place at the same time. To address this, I leverage the fact that quasi-randomly assigned investigators could also offer services to families whose children were not removed. Therefore, I separately identify the impacts of community-based services and targeted services from the combination of adult services and placement by using investigator tendencies over each as instruments. Though limited by statistical power, I find that the impact of child removal together with adult services is considerably larger than the individual effects of either adult intervention alone. Overall, these results suggest that child removal enhanced the efficacy of child welfare interventions for adults, perhaps through increased incentives to comply or temporary relief from parenting.³

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

³An important limitation of this exercise, however, is that it identifies the effects of community-based services and targeted services for children who were not candidates for foster care. For example, I estimate the effect of community-based services for cases in which investigators might disagree about referring families for these light-touch interventions, which represents a relatively low-risk group. Even so, to the extent that these services have similar impacts for struggling families, the results indicate that child removal was a crucial component of the foster care intervention.

The fourth contribution of this study is that it is the first to show that a common form of incomplete data coverage substantially biases estimates from the examiner assignment research design in practice. Specifically, this paper improves upon contemporaneous studies from Rhode Island (Bald et al., 2019) and South Carolina (Roberts, 2019) which offer quasi-experimental estimates of foster care, yet do not follow children from the start of their child welfare investigation.⁴ The data in these studies contain only the subset of substantiated allegations, those in which investigators found a preponderance of evidence to support the maltreatment allegation, which represent just 40% of the caseload in Rhode Island and 25% in South Carolina (AECF, 2017). Since the same investigator who determines foster placement also makes subjective decisions around substantiation, the set of children in censored data may not be balanced across investigators even if their cases were initially assigned at random. I replicate my primary analysis using only the sample of substantiated investigations and find estimates much smaller than the true effects. As the examiner assignment design becomes increasingly common—and similar data restrictions appear in studies of crime and education—this exercise cautions against its application with incomplete data.⁵

This study is especially relevant given the dramatic changes to child welfare policy introduced in the Family First Prevention Services Act. The legislation, which took effect in 2019, makes reducing the use of foster care a federal priority by allowing states to redirect up to eight billion federal dollars from the foster system toward services aimed to prevent foster care entry (Wiltz, 2018). My analysis, which finds that foster care in Michigan improved children’s outcomes, suggests that the effectiveness of this new federal policy hinges on whether states identify and invest in high-quality prevention services.

The rest of this paper is organized as follows. Section 3.3 details the child welfare investigation process in Michigan and describes the state’s foster system. Section 1.3 introduces the sources of administrative data and the analysis sample. Section 2.3 outlines the research design. Section 3.6 reports the main findings and explores mechanisms, and Section 1.6 describes bias from incomplete data coverage. Section 1.7 concludes and discusses implications for public policy. Appendix A provide supplemental results.

⁴Bald et al. (2019) 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. Roberts (2019) examines about 17,000 children between age two and seventeen and finds positive impacts on on-time grade progression, yet noisy estimates on daily school attendance and test scores.

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

1.2 Overview of the Child Welfare System 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 (Ryan et al., 2018). One in ten were the subject of more than one investigation and one in sixty experienced foster placement.⁶ This section reviews the maltreatment investigation process in Michigan and describes the state’s foster system.

1.2.1 Child Maltreatment Investigations

Figure 1.2 describes the maltreatment investigation process in Michigan, which is similar to most other states. It begins when someone calls an intake hotline to report child abuse (e.g., bruises, burns, or sexual abuse) or neglect (e.g., unmet medical needs, lack of supervision, or food deprivation).⁷ A hotline employee, who does not participate in the investigation process, transfers relevant reports to the child’s local child welfare office.⁸ The office assigns the report to a maltreatment investigator who has 24 hours to begin an investigation, 72 hours to establish face-to-face contact with the alleged child victim, and 30 days to complete the investigation.

Critical to 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.⁹ Since investigator assignment 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 children who could have been assigned the same investigator.¹⁰

⁶These statistics reflect my calculations using the same sample as (Ryan et al., 2018), 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 Wildeman and Emanuel (2014) 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.

⁷While 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.

⁸Reports 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.

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

¹⁰There 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.

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 (USDHHS, 2018b, Tables 3-1 and 3-3), meaning child welfare offices did not follow up with the family further.

Second, investigators decide how much risk the child faces by continuing to live in their home. They complete a 22 question risk assessment to compute a risk score, which is used to determine whether foster placement is appropriate. Many of the items require simple yes or no answers, such as “primary caretaker able to put child’s needs ahead of own” and “primary caretaker views incident less seriously than the department.” Even with guidance on how to interpret these questions, some are inherently subjective. Moreover, Bosk (2015) offers detailed qualitative evidence that investigators often manipulate their responses to ensure risk scores that match their priors. Therefore, even with a standardized system in place, 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 by child welfare. If the investigator substantiates the allegation and the risk level is high, the family also receives more intensive, targeted services, such as substance abuse treatment, parenting classes, or counseling. Local, state, and federal funding, from Title IV-E, covers the costs of these targeted services. Lastly, substantiated allegations with especially high risk not only trigger targeted and community services but also require the investigator to file a court petition for child removal.¹¹ The main analysis in this study examines the combined effects of child removal and these adult interventions on children’s outcomes, yet additional analysis explores their individual contributions.

Second, new reports involving the same child as a recent prior report are usually assigned to the original investigator since they have familiarity with the family. Anecdotally, such reports tend to re-enter the rotation after a few months. I exclude those within one year of a prior investigation from the analysis to be conservative.

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

1.2.2 Foster Care System

Foster care is a family intervention; children are temporarily removed from their homes while their birth parents receive services to improve their parenting. Removal occurs quickly, just 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.¹² 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.¹³ It is common to switch placement settings while in the foster system—60% of children in Michigan 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 (ChildTrends, 2017).

After placement, child welfare caseworkers meet with birth parents to create a reunification plan stating the conditions under which the child can return home. 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.¹⁴ Caseworkers monitor their progress and make changes to the reunification plan as needed. Family reunification only occurs if a court decides that birth parents made sufficient changes for their child to be safe in their home.

Ultimately, children in Michigan, including those outside of the analysis sample, spend seventeen months in the system on average, after which 47% were reunified with their birth parents, 34% were adopted or had legal guardianship transferred, and 9% exited the system as independent adults upon turning age eighteen. The remaining 10% fell into less common exit categories, such as informal guardianship with relatives, incarceration, or transfer to another agency. Section 1.5.3.1 offers evidence on how the foster care experience of the

¹²There is limited causal evidence on the effects of each placement type, and the instrumental variables design in this study cannot separately identify each effect. However, OLS analysis in Supplemental Appendix A.1.2 finds a larger positive association between kinship placement and children's outcomes than other placement types.

¹³There is limited data available both nationwide and in Michigan on foster families, those who takes in foster children. Estimates from the American Community Survey (ACS), which have known limitations, suggest that households with foster children tend to be larger and lower-income than other households with at least one member younger than 18 years old. Supplemental Appendix A.1.3 provides summary statistics and discusses the limitations of using ACS data to identify families with foster children.

¹⁴Though 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.

overall population of foster children compares to children at the margin of placement.

1.3 Data Sources and Sample Construction

1.3.1 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 a Michigan public school in the year of their investigation. This match rate is quite high given that many investigated children should not have matched to an enrolled public school student, e.g., private or homeschool students, high school dropouts, and those who were not permanent Michigan residents. Specifically, I estimate that if there were a common identifier, just 87.1% of investigated children would have matched to a currently enrolled student.¹⁵ Supplemental Appendix A.1.4 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 like 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 United States Census Bureau.

Juvenile justice data from SCAO includes all juvenile court petitions filed in almost every

¹⁵I estimate that the remaining 12.9% of investigated children 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%).

county in Michigan between 2008 and 2015. A court petition is an official document filed following juvenile arrest in cases where youth are not immediately diverted from the courts. Petitions can be dismissed by the court after filing and need not indicate that there was ever a formal court hearing. The SCAO data covers 75 of Michigan’s 83 counties, including Detroit and the metro-Detroit area but excluding the following five urban and three rural counties: Kent, Washtenaw, Ingham, Ottawa, Kalamazoo, Berrien, Delta and Keweenaw.¹⁶ I exclude the 19% of investigated children who lived in these eight counties from my analysis of juvenile delinquency, and the conclusions on other outcomes are similar when these children are excluded.

Using these administrative data sources, I assess the effects of foster care on child wellbeing across three dimensions: safety, schooling, and crime. Given that I study a variety of outcomes, I construct an index of child wellbeing according to Kling et al. (2007) to create a summary measure, increase statistical power, and help address multiple hypothesis testing. The index consists of unweighted means of standardized versions of seven primary outcomes, described in detail below: two measures of child safety, four academic outcomes, and one indicator of juvenile delinquency. I reverse code “bad” outcomes and impute missing values according to group means.

To measure child safety, I create indicators for whether children were the alleged victim in a subsequent maltreatment investigation and whether they were a confirmed (substantiated) victim in a subsequent investigation. Second, 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.¹⁷ Finally, I measure juvenile delinquency as the filing of a juvenile court petition.

1.3.2 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

¹⁶These 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.

¹⁷These 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 publicly available.

in grades one through eleven in the school year of their investigation to observe baseline characteristics and at least one follow-up year.¹⁸ Appendix A.1.4 describes the sample restrictions in greater detail. Overall, I focus on 242,233 investigations of 186,250 students and follow students for at most nine years after their investigation.

Table 3.1 describes the sample. Column 1 consists of all public school students in Michigan during the 2016-2017 school year, while column 2 consists of the investigations of children in the analysis sample. Black and low-income children were disproportionately involved in the child welfare system; 29% 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 also disproportionately black and low-income, had much lower daily attendance rates, and scored about a tenth of a standard deviation lower on math and reading tests. Overall, these descriptive statistics caution against a causal interpretation to mean comparisons between investigated children who were and were not removed.

1.4 Empirical Strategy

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

¹⁸The 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 the likelihood that they ever enrolled in a Michigan public school. A likely explanation for this finding is that about one-third of foster children were adopted upon exiting the foster system and may have legally changed their last name prior to enrolling in school, meaning that the administrative child welfare and education records were unlikely to match. It is also possible, however, that young children differentially moved out of state or enrolled in private schools. Importantly, I find no evidence of differential attrition out of Michigan public schools for currently enrolled students (Table A.1).

1.4.1 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 (Doyle, 2007, 2008) as well as research on incarceration (Kling, 2006; Aizer and Doyle, 2015; Mueller-Smith, 2015), disability insurance (Dahl et al., 2014), and evictions (Collinson and Reed, 2019; Humphries et al., 2019), among others. Specifically, I instrument for placement using the removal tendencies of quasi-randomly assigned investigators. By chance, children assigned to especially strict investigators—those with high propensities to remove—were more likely to enter foster care than they would have been if they happened to be assigned to a more lenient investigator.

In order to extract signal from noise in a measure of removal tendency, 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.¹⁹ This leaves 3,073 investigators assigned to 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 :

$$Z_{iw}^R = \left(\frac{1}{n_w - 1}\right) \sum_{k \neq i}^{n_w - 1} (FC_k) \quad (1.1)$$

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.²⁰ This instrument is equivalent to the investigator fixed effect from a leave-out regression where foster placement is the dependent variable.

The instrument has a mean of 0.030 and a standard deviation of 0.024, indicating considerable variation in investigator tendencies. Crucial to the research design, there is variation even among investigators who worked in the same local office. Figure 1.3 shows the distribution of the instrument net of child zipcode by investigation year effects; an investigator at the 10th percentile removed at a rate 2.1 percentage points less than others in their local team 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

¹⁹Table B.11 shows that the results are robust to both larger and smaller thresholds.

²⁰There are other reasonable ways to measure removal stringency. For example, this approach does not allow for investigator tendencies to change over time. Section 1.5.3.3 describes several alternatives and shows that the results are robust across measures.

foster care:

$$FC_{iw} = \gamma_0 + \gamma_1 Z_{iw}^R + \gamma_2 X_{iw} + \theta_r + \eta_{iw} \quad (1.2)$$

$$Y_{iw} = \beta_0 + \beta_1 \hat{C}_{iw} + \beta_2 X_{iw} + \theta_r + \epsilon_{iw} \quad (1.3)$$

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.²¹ θ_r represent child zip code by investigation year fixed effects to control for the level of investigator rotational assignment, restricting the comparison to children who could have been assigned to the same investigator.²² There are 7,534 unique rotation groups, consisting of thirteen investigators on average. Finally, I cluster standard errors at the investigator level.²³

$\hat{\beta}_1$ 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 a population that is especially relevant for child welfare policy. As Dr. Jill Duerr Berrick, Zellerbach Family Foundation Professor at the University of California Berkeley School of Social Welfare, writes in *The Impossible Imperative: Navigating the Competing Principles of Child Protection*, “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

²¹Specifically, 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 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 bachelor’s 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.

²²Child 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 B.11 shows that the results are robust to instead defining rotational assignment at the child county by investigation year level.

²³The 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.

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” (Berrick, 2018).

1.4.2 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 1.2 shows the first stage regression of foster placement on the removal stringency instrument. The correlation between the instrument and foster care is 0.48 (Column 1) and a one standard deviation (2.4 percentage points) increase in removal stringency 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.

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 are not jointly predictive of the instrument despite being highly predictive of placement itself (Table 3.3). As further evidence, the first stage F-statistic in Table 1.2 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 investigator experience or thoroughness (Table A.3).

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 A.3).²⁴

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

²⁴Another potential concern is that investigators vary along dimensions of the investigation process other than foster placement that may influence outcomes (Mueller-Smith, 2015). I discuss this in detail in Section 1.5.2.2, as the impacts of these other dimensions, e.g., adult interventions, are themselves of substantive interest.

that for each child, the covariance between their investigator-specific removal treatment status and investigator stringency is weakly positive.

It follows from average monotonicity that removal stringency and foster placement should be positively correlated for all child subgroups. There are two complementary ways to probe this implication. First, the first stage should be non-negative for all subgroups (Dobbie et al., 2018), which holds for gender, race/ethnicity, age, and prior child welfare involvement groups in my setting (Table A.4, Panel A).²⁵ Second, investigators who were strict for certain groups should also have been strict for others (Bhuller et al., 2018). For example, amidst serious concerns of racism in maltreatment investigations (Clifford and Silver-Greenberg, 2017), monotonicity asserts that investigators who were particularly likely to remove children of color should also have been weakly stricter than their colleagues in their investigations of white children. In support of the assumption, the first stage remains positive and statistically significant when I re-calculate the instrument as a leave-subgroup-out measure (Table A.4, Panel B).

1.5 Causal Effects of Foster Care on Children’s Outcomes

Table 1.4 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.²⁶ The OLS results suggest that removal had a near-zero impact on the index of child wellbeing. 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 findings stand out from comparing the OLS and 2SLS results on the index of child wellbeing. 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 all investigated 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

²⁵I do not create groups based on the type of maltreatment such as abuse or neglect because investigators code this information after they begin their investigation. However, to the extent that different types of maltreatment are related to observable child subgroups, the exercise offers an indirect test for non-monotonic tendencies based on these features.

²⁶Specifically, 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 at which young people are tried in the adult court system is sixteen years old in Michigan, so seventeen-year-olds are ineligible for the juvenile delinquency outcome.

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.

1.5.1 Effects on Child Safety, Academics, and Crime

Table 1.4 shows that foster children were safer than they would have been had they remained at home, indicating that the foster system achieved its primary objective. The 2SLS estimates show that removal reduced the likelihood of being an alleged victim of maltreatment in a subsequent investigation by 13.2 percentage points, a 52% 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 in theory these effects may represent a reduction in reporting behavior without a change in underlying safety, the data does not support this interpretation. For example, suppose that teachers were less likely to report minor bruises to child welfare if they knew that the bruised student was, or had been, in foster care. We would still expect them to report especially severe abuse against foster children though, since teachers and other mandated reporters are required by law to report suspected maltreatment. Therefore, if placement only reduced reporting, then the reported abuse against foster children should be more serious than the reported incidents against children who were not removed. However, I find that foster placement did not influence the likelihood of substantiation among children with a subsequent investigation. Moreover, caseworkers, who are also mandatory reporters, visited foster children regularly, both during their time in the system and after they exited, suggesting that actual maltreatment against foster children should have been reported (USDHHS, 2016b).

Consistent with an improvement in child safety, I find large gains in academic outcomes. Removal increased daily school attendance rates by 5.4 percentage points which, for the 180 day school year, is equivalent to showing up for ten additional days of school. I also find that foster children were less likely to be retained in grade, though the effect is imprecisely estimated. Furthermore, removal had a very large positive effect on standardized math test scores of about one-third of a standard deviation.²⁷ This estimate is statistically significant at the 10% level, yet I can rule out decreases greater than 6% of a standard deviation.

²⁷As a benchmark, Goodman (2014) estimates that each additional student absence reduces math achievement by 0.05 standard deviations, suggesting that the estimated math score effect is roughly in line with the increase in daily school attendance.

In addition, while the point estimate on standardized reading test scores is positive and substantively large, about half the size of the effect on math, it is not statistically significant. This is not particularly surprising because reading skills are considered less malleable than math at older ages.²⁸

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 across dimensions of safety, academics, and criminality consistently suggest that foster care improved children’s outcomes.

1.5.2 Mechanisms

I examine the channels through which 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 evaluate interventions targeted at birth parents as a potential channel through which placement influences children.

1.5.2.1 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 1.4).²⁹ I create an index of neighborhood and school characteristics according to Kling et al. (2007) in order to explore the effects of placement on childhood environment. The index consists of three neighborhood components: median household income, the fraction of adults with a bachelor’s degree, and employment rate. It also includes two school components: average math and reading test scores and the share of 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 1.5, Panel A).³⁰ Given that moving to lower-poverty areas can improve child wellbeing (Chetty et al., 2016; Kawano et al., 2017; Chyn, 2018), such exposure might lead to contemporaneous gains in children’s outcomes. However, there were no discernible differences in year one outcomes between

²⁸Removal did not influence the likelihood of taking standardized tests (Table A.5). In addition, Table A.6 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.

²⁹They spent nineteen months in foster care, on average (Table A.7).

³⁰This was driven by exposure to more highly educated neighborhoods and higher-income classmates (Table A.8).

children placed and not placed in foster care (Table 1.5, Panel A).³¹ That foster children were no more or less likely to be abused or neglected in the first year may be especially surprising since maltreatment in foster homes is extremely rare. It is possible, however, that the threat of child removal reduced the maltreatment of children who were not removed in the short-run.

Nearly all (85%) marginal foster children had exited the system after two years and the vast majority were reunified with their birth parents.³² 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 1.5, 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 school attendance rates, and standardized math test scores (Table 1.5, Panel B). Figure 1.5 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 only began to decrease after four years and continued to decrease every year for three more.³³

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. There are several institutional features which support this channel. First, after their children were removed, birth parents worked closely with social workers to address challenges in their own lives, such as confronting drug addiction, finding stable employment, securing housing, or strengthening parenting skills. Second, birth parents received fully funded services to help with these challenges, like substance abuse treatment, parenting classes, or counseling.³⁴ Lastly, a judge needs to approve that it is safe for children to return home before they can be reunified with their birth parents. In addition to these institutional reasons, I also find statistical evidence of birth parent improvement. Perpetrators of child maltreatment, almost always a birth parent, were less likely to abuse or neglect children even years later if their

³¹For ease of interpretation, Table 1.5 and all further analyses report only a set of the outcomes that were statistically significant at the 10% level from Table 1.4. Additional results are available upon request.

³²Table A.9 shows that of the remaining 15% who exited: 8% were adopted, 5% had guardianship transferred, and 2% turned eighteen years old and legally exited foster care as adults.

³³These 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.

³⁴There is no reliable measure of the cost of services for the birth parents of foster children because they are funded by many different sources. They may be covered through Medicaid, funded by the state, or contracted through individual counties. For reference, Stacie Bladen, the Deputy Director of MDHHS' Children's Services Agency, communicated via email that the state's costs associated with serving parents of children in foster care was at least \$16 million in 2019, or roughly \$1,200 per child.

initial child victim entered foster care (Figure 1.5e).

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

Second, it could be that children’s experiences while in foster care benefitted them only years later, i.e., foster care could trigger additional supports whose benefits take time to manifest. In particular, if the costs related to family separation are high in the short run yet fade over time, even the benefits from channels that have more immediate impacts may appear only years later. However, 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 1.5, 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 standardized test scores of less than one-tenth of a standard deviation, much smaller than the 0.34 standard deviation increase in math test scores found in this study.³⁵

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

1.5.2.2 Evidence from Adult Interventions

Following child removal, birth parents received two broad types of services: 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 own lives because they both happen at the same time. However, a useful comparison group exists because quasi-randomly assigned

³⁵See, for example, Carrell and Hoekstra (2014) and Mulhern (2019) for the effects of school counselors, Dee (2004) for the effects of teacher role models, Heller (2014) for the effects of summer jobs and mentors, Anderson et al. (2018) for the effects of healthier meals, and Figlio and Winicki (2005); Leos-Urbel et al. (2013); Imberman and Kugler (2014); Frisvold (2015); Schwartz and Rothbart (2017) for the effects of greater access to food.

investigators could offer these services to adults even if their children were not removed.

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 1.2, investigators placed families on one of four tracks based on the strength of evidence that maltreatment occurred and the child’s risk of future harm: (1) no services, (2) community-based services, (3) community-based and targeted services, and (4) child removal plus community-based and targeted services.³⁶ As such, I create two new instruments according to Equation 1.1: investigator propensity to recommend community-based services alone and investigator propensity to recommend both community-based and targeted services without child removal. Together with the main removal stringency measure, I use these new measures to simultaneously instrument for tracks two, three, and four.³⁷

Table 1.6 shows the three distinct local average treatment effects estimated from this exercise, which yield two key takeaways. First, I find large, 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 in Table 1.4, though less precise. This addresses a potential violation of the exclusion restriction in that by having discretion over adult services, investigators may influence children in ways other than foster care (Mueller-Smith, 2015). For example, a violation of the exclusion restriction arises if investigators who were more likely to remove children were also more likely to recommend targeted services, and tendencies over targeted services are not included in the estimation. Since the point estimates for foster care are very similar to the main specification, this exercise suggests that the removal stringency instrument operates through foster placement.

Second, though limited by statistical power, I find that the individual impacts of targeted and community services are qualitatively smaller than the combined effect of both types of adult interventions together with child removal. This offers suggestive evidence that child removal was a crucial component of the foster care intervention. Of note, however, a limitation of this exercise is that, unlike the main analysis of cases where investigators might disagree about placement, it identifies effects for children who faced lower risk and were not candidates for foster care. For example, the LATE for community-based services identifies effects for families at the margin of receiving any services. However, to the extent that these services have similar impacts for struggling families, the results indicate that child removal

³⁶While 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.

³⁷The 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 the propensity for community services alone is 0.14, between the removal instrument and tendency for both targeted and community services is 0.24, and between the two non-removal instruments is 0.60.

was necessary for adult interventions to be effective.

There are at least two explanations as to why adult interventions may have been less effective while children remained in the home. First, they might have lacked sufficient intensity. Unlike services offered to the birth parents of foster children, which lasted up to 24 months and were monitored by both the child welfare office and the courts, services lasted at most twelve months and participation was tracked only by the local child welfare office when the child was not removed. A second potential explanation is that child removal was critical to ensure adult compliance. For example, temporary relief from parenting may have provided birth parents with the time and space needed to overcome challenges in their own lives. Child removal might also have increased adult incentives to engage with these programs. Prior work highlights that services like drug rehabilitation or job training programs often have high failure rates overall (SAMHSA, 2009; Barnow and Smith, 2015), yet 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 an important mechanism to explain gains in children’s outcomes.

1.5.3 Compliers Analysis, Subgroup Effects, and Robustness Checks

1.5.3.1 Contextualizing Children at the Margin of Foster Placement

The estimates in this study represent effects for children at the margin of placement, those in which investigators might disagree over whether foster care is appropriate. In order to better understand how these children compared to the overall population of foster children, I report complier characteristics in Table A.10, estimated according to the methodology in Dahl et al. (2014). I find that 5% of investigated children in the sample were compliers. Compliers 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 looked similar in terms of demographic and baseline academic characteristics.

To contextualize the positive effects of placement, it is also useful to explore how marginal children experienced the foster system. Table A.11 compares the experiences of these children, calculated using the instrumental variables design, to the overall population of foster children, defined as the mean among all foster children in the sample. While these groups were initially placed in similar types of homes—e.g., 57.2% of children at the margin were initially placed with relatives compared to 58.2% of all placements—their experiences varied in terms of placement stability, length of time in foster care, and permanency outcomes.

Children at the margin had more stable placements; 51% experienced just one or two different placements compared to 44% of all foster children in the sample. They also spent about 38 fewer days in the system and were more likely to be reunified with their birth parents. Therefore, marginal children had more stable placements, quicker exits, and higher reunification rates than the overall population of foster children, all of which are important objectives for foster care systems.

1.5.3.2 Heterogeneity by Child Age and Gender

Previous work highlights disparities in how children respond to environmental changes by age, finding that young children benefit from moving to lower-poverty areas more than older youth (Chetty et al., 2016; Chyn, 2018). I find similar effects for foster placement. Table A.12 shows that foster care improved the index of child wellbeing for young children, those ages ten and younger at the beginning of the investigation, by 19.6% of a standard deviation, an effect statistically significant at the 5% level. This was almost twice as large as the point estimate for older youth, which was not statistically significant. Though the positive effects of placement were qualitatively driven by young children, the estimates are not statistically different from each other.

In addition, previous work shows that males are often more vulnerable than females to childhood disadvantage or disruption (Kling et al., 2005; Bertrand and Pan, 2013; Autor et al., 2019). However, Table A.12 shows that the impacts of placement were similarly positive for both groups. Further exploring this heterogeneity using age by gender groups reveals that the effects were qualitatively larger for young male children than for young female children, yet qualitatively larger for female youth than for male youth. Taken together, this exercise indicates that the intersection between age and gender may play an important role in determining the impacts of placement.³⁸

1.5.3.3 Robustness Checks

Table B.11 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 of each child. Next, I test sensitivity to the number of cases an investigator must have been assigned to be included in the sample. The main analysis excludes children assigned to investigators who worked fewer than 50 cases, so I relax this threshold to 25 and also strengthen it to 75. The results are similar to those in

³⁸Though LGBTQ youth are over-represented in foster care (HRC, 2015) and have especially traumatic experiences in the system (Sullivan et al., 2001), this study is unable to examine differences along this margin because the administrative data sources do not include information on sexual orientation or gender identity.

the main analysis across these three alternative samples. Lastly, I restrict the analysis to a balanced panel consisting of the first five follow-up years for students that could be observed in the public school system for five years after their investigation based on their grade level and year of investigation—those investigated in seventh grade or below in 2012 or earlier. The impacts of foster placement were nearly twice as large in the balanced panel as the main analysis. This is consistent with the previously discussed results. Specifically, the effects in the main analysis could not have been driven by the placement of older children or those investigated later in the sample period because Section 1.5.2.1 shows that impacts appear only several years after removal. Similarly, the subgroup analysis in Section 1.5.3.2 shows qualitatively larger effects for younger children.

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, following the measure of teacher-value added in Chetty et al. (2014).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 investigation year fixed effects.

1.6 Potential Bias in Examiner Assignment Research Design from Censored Data

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

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 stores 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.

foster care, such as juvenile incarceration (Aizer and Doyle, 2015; Eren and Mocan, 2017), adult incarceration (Kling, 2006; Mueller-Smith, 2015), disability insurance (Dahl et al., 2014), student loan repayment (Herbst, 2018), and evictions (Collinson and Reed, 2019; Humphries et al., 2019), among others. In many of these settings, treatment assignment is a two-step selection process in which individuals are assigned to treatment only after crossing an initial decision threshold. For example, in the context of foster care, children can only be removed if their maltreatment allegation is first substantiated. 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 contains only individuals that cross the initial decision threshold, e.g., only substantiated investigations or only convicted defendants. Such restrictions appear in two recent studies of foster care (Bald et al., 2019; Roberts, 2019) as well as in other contexts (Kling, 2006; Eren and Mocan, 2017; Herbst, 2018), and may introduce bias.

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 (Arteaga, 2019). Either investigators must not vary over substantiation—i.e., investigators always agree over evidence—or the investigator’s substantiation decision must be uncorrelated with the child’s potential outcomes. The former assumption is at odds with the motivation of the research design, given that the design hinges upon variation in investigator tendencies. Moreover, at least in Michigan, there is a large amount of variation in substantiation tendencies.⁴² The latter

⁴⁰These 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.

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.

⁴²Investigators at the 10th percentile substantiated at a rate 8.4 percentage points less than others in their local team while investigators at the 90th percentile did so at a rate 8.9 percentage points greater.

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 1.1 using only the sample of substantiated investigations, I find that standard balance tests are sensitive to the inclusion of baseline standardized test scores.⁴³ This offers evidence that data constraints can create a violation of the exclusion restriction.

Table 1.7 shows that the true effects using the complete data (Panel A) are larger than those found when restricted to substantiated investigations (Panel B).⁴⁴ ⁴⁵ The replication exercise produces a substantively small and statistically insignificant impact on the index of child wellbeing. 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. The findings in Panel B of Table 1.7 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. 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 in the other studies may also play a role. Overall, this exercise cautions against applying the examiner assignment design with censored data.⁴⁶

⁴³Specifically, Table A.14 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, 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. Balance tests in other studies may be underpowered, however—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 school-age sample in Rhode Island.

⁴⁴Table A.15 shows that there exists a strong first stage relationship with the censored instrument. In addition, it is possible that Panel B in Table 1.7, 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 A.16 shows that the estimates in Panel B are also smaller than the causal effects of placement relative to substantiation from the complete data.

⁴⁵Table A.17 shows that the OLS estimates are very similar from both the complete data and when restricted to substantiated investigations, however.

⁴⁶In addition, while Arteaga (2019) proposes a reasonable approach to use the examiner assignment design to recover unbiased estimates with censored data, the study cannot empirically assess how well the approach performs in practice because it only accesses censored data itself. Appendix A.1.1 shows that the proposed

1.7 Conclusion

This paper offers some of the only causal estimates of foster care on crucial indicators of child wellbeing: safety, education, and crime. To do so, I leverage the quasi-random assignment of child welfare investigators who vary in their propensity to recommend placement. Using detailed administrative data from Michigan to study over 200,000 child welfare investigations between 2008 and 2016, I find that placement improved a variety of children’s outcomes. Foster children were 50% less likely to be abused or neglected in the future, relative to a baseline mean of 25.5%. Placement also increased daily school attendance by 6%, or about ten additional days of school every year, and improved standardized math test scores by one-third of a standard deviation, the equivalent of moving from the 33rd to the 46th percentile in the state. I also estimate a substantively large, yet statistically insignificant, reduction in juvenile delinquency.

1.7.1 External Validity

As child welfare systems in the United States vary across states, it is important to consider the generalizability of these results. Overall, foster care in Michigan is similar to others in terms of how long children remain in the system and the stability of their placements (AECF, 2017). Among all children in foster care in the US in 2015, the average child had spent nineteen months in the system, ranging from just twelve months in Idaho to 35 months in Illinois. At seventeen months, Michigan, along with eight other states, ranked eighteenth in this measure. Similarly, 35% of foster children in the US had lived in at least three different foster homes, ranging from 24% in Wyoming to 54% in Illinois. Michigan ranked seventeenth in this measure at 31%. Furthermore, among all children in the US who exited the system in 2015, 51% reunified with their birth parents compared to 47% of children in Michigan.

In addition, it is also unlikely that the services for families with maltreated children who were not removed in Michigan were notably different than those in other states. Though there is no publicly available data on the quality of prevention services across states, there is little rigorous evidence of effective services from any state.⁴⁷ Overall, the child welfare system in Michigan during the sample period was fairly typical of those in other states.⁴⁸

method does little to resolve bias in the current context.

⁴⁷Specifically, there are only 32 programs, spanning infant and toddler mental health to adult substance abuse treatment, that received the highest evaluation rating on the California Evidence-Based Clearinghouse for Child Welfare (CEBC) as of October 2019. For comparison, the What Works Clearinghouse includes 76 studies of pre-kindergarten alone that met its highest standards of rigor and almost 500 of K-12 education (IES, 2017).

⁴⁸Michigan did make substantive changes to its child welfare system following a July 2008 settlement in *Dwayne B. v. Snyder*, yet this does not considerably inhibit the external validity of my analysis for several

A second important consideration for the generalizability of the results is that this paper focuses exclusively on the effects of foster care for school-age children.⁴⁹ Bald et al. (2019) and Roberts (2019) make critical advances in this regard, finding that placement can have benefits for young children, yet as previously shown in Section 1.6, data constraints in these papers may preclude a causal interpretation. Since they make up nearly half of the foster care population, more evidence is needed on how placement influences young children.

1.7.2 Implications for Public Policy

The new research findings from this paper 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 allocating up to eight billion dollars of federal Title IV-E funds for states to spend on alternatives to placement. Previously reserved for foster care and adoption budgets, except for waivers permitted in special cases, states can now use this funding stream on services to prevent foster care entry among children at-risk of placement. Therefore, the effectiveness of this policy hinges on both the population of children that states deem as candidates for foster care and the prevention services that states choose to fund. My analysis from Michigan helps to shed light on both of these issues.

First, by showing that placement improved outcomes in Michigan, which uses foster care less often than other states, this paper highlights states where the new federal policy may be most and least effective. Specifically, Michigan ranked 12th in the rate of foster care entry in 2017; just 3.1 of every 1000 children in Michigan entered foster care relative to the national median of 4.1 (USDHHS, 2017b).⁵⁰ The Children’s Bureau’s annual report to Congress notes in 2016 that, “it is unlikely that these variations can be attributed to differences in the rates of child victims” (USDHHS, 2016a).⁵¹ Therefore, my analysis suggests that foster care can

reasons. First, my sample period begins in 2008. Second, lawsuits against child welfare systems are common; sixteen other states and many cities have faced similar lawsuits between 1999 and 2018 (ChildrensRights, nd). Lastly, changes required by the settlement, such as increasing the education requirements for caseworker hiring, implementing pre-service training, and reducing worker caseload, made Michigan’s system more similar to others across the country.

⁴⁹Though I observe the child welfare investigations of younger children, it is challenging to follow them across administrative data systems over time because they are disproportionately likely to be adopted and legally change their last names.

⁵⁰Figure A.1b shows the variation in foster care entry across states in 2017, ranging from 1.5 per 1000 children in Virginia to 13.3 in West Virginia.

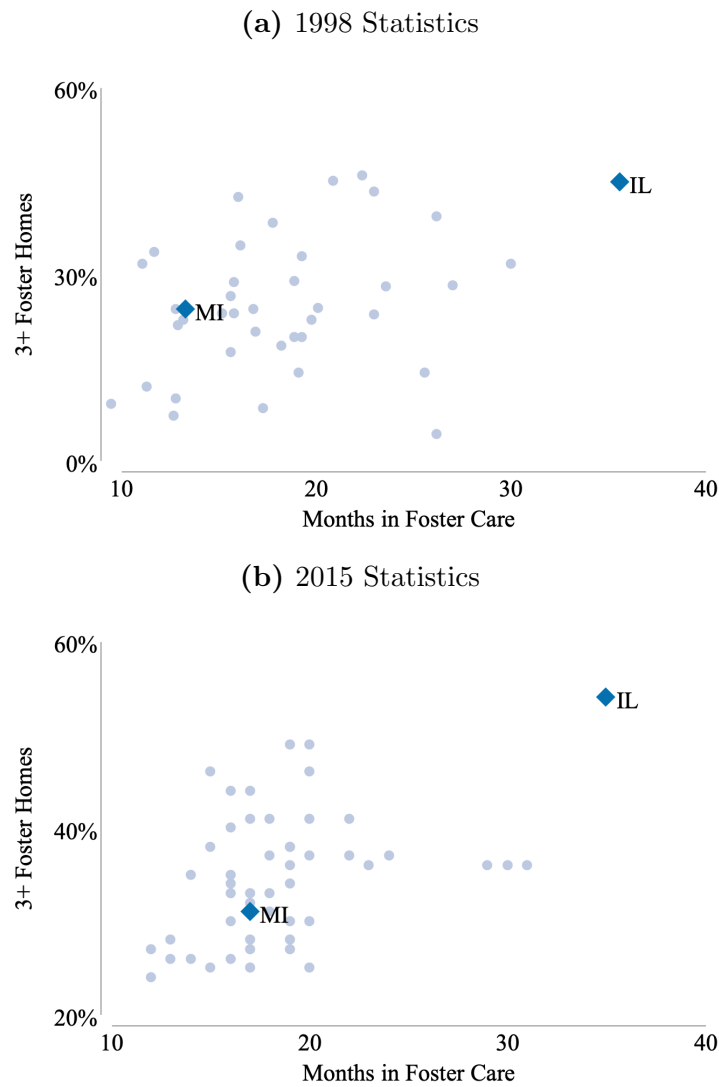
⁵¹It is difficult to measure the actual rate of foster care entry among child victims of abuse and neglect because both the definition of substantiation and the use of alternatives to formal maltreatment investigations vary dramatically across states.

be beneficial for children’s wellbeing when reserved for relatively high-risk cases.⁵² Though this paper does not study all of the outcomes that should be considered—e.g., mental health or the trauma associated with family separation—the findings suggest that, all else equal, the Family First Prevention Services Act may be helpful for children in some states, but potentially harmful for children in others.

Second, this paper underscores that states must invest in high-quality prevention services for the new federal policy to be most effective. Since this study finds that foster placement had benefits for at-risk children relative to offering services in the home, a focus on the efficacy of prevention services is critical. In particular, the evidence indicates that birth parents improve their parenting as a result of their children being placed in foster care, which suggests that states should invest in prevention programs that better engage with birth parents as they work toward transformational life changes like overcoming addiction, securing housing, or gaining employment. Learning what works to keep vulnerable children safe and thriving when they remain in their homes and understanding how to scale these interventions is a crucial next frontier for future research.

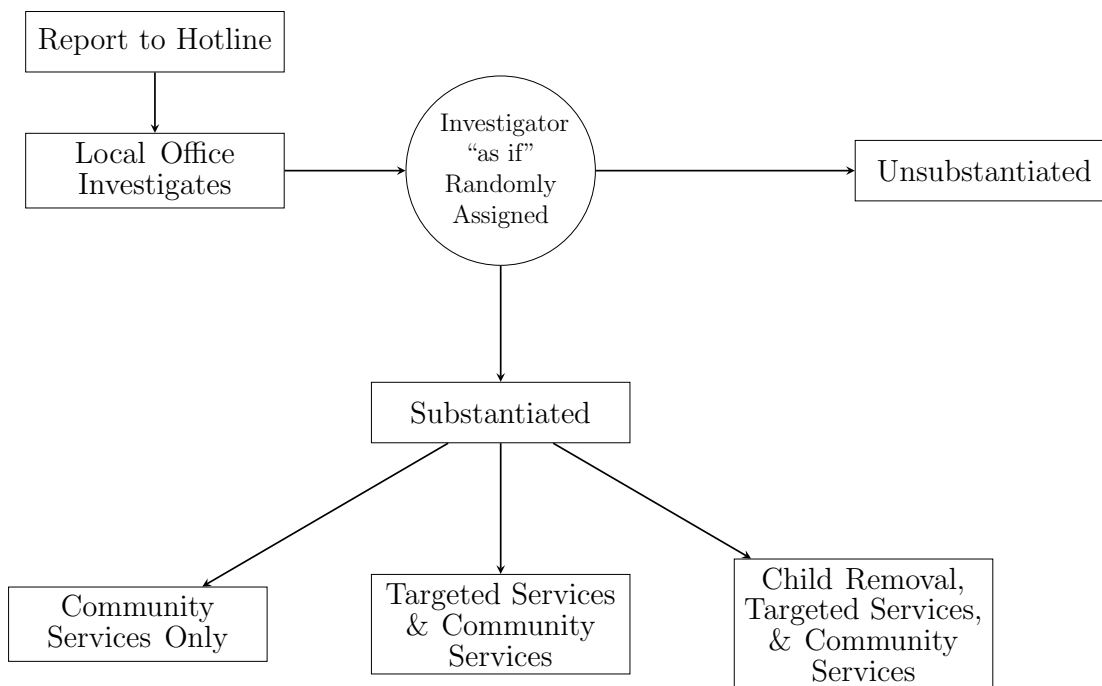
⁵²Consistent with this hypothesis, I find that the estimate of placement on the index of child wellbeing is twice as large in Michigan counties that remove at a rate below the state median than in counties that remove at a rate above it (0.26 compared to 0.13).

Figure 1.1: Comparison of State Foster Care Systems



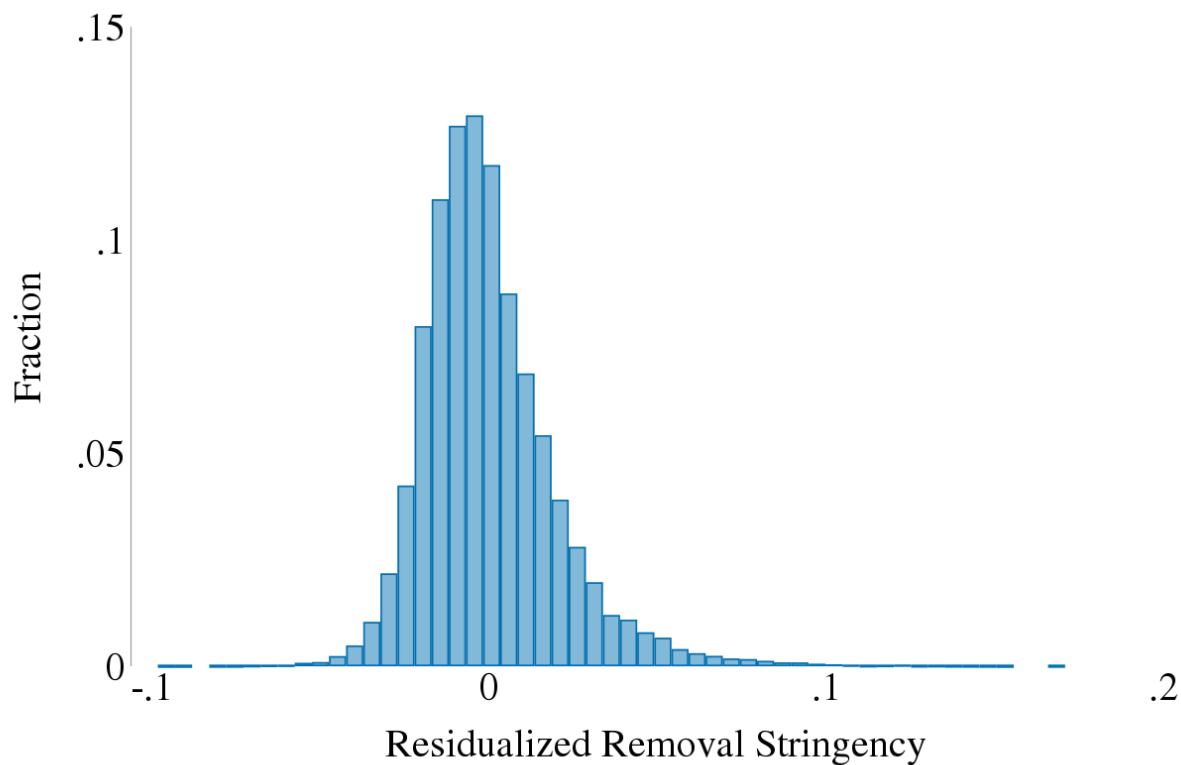
Notes. These figures show statistics about state foster care systems from 1998, the first year of publicly available data, reported in USDHHS (2003) and from 2015 reported in USDHHS (2017a). Due to a change in reporting, the horizontal axis shows the median number of months spent in foster care for each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

Figure 1.2: Overview of Child Maltreatment Investigations in Michigan



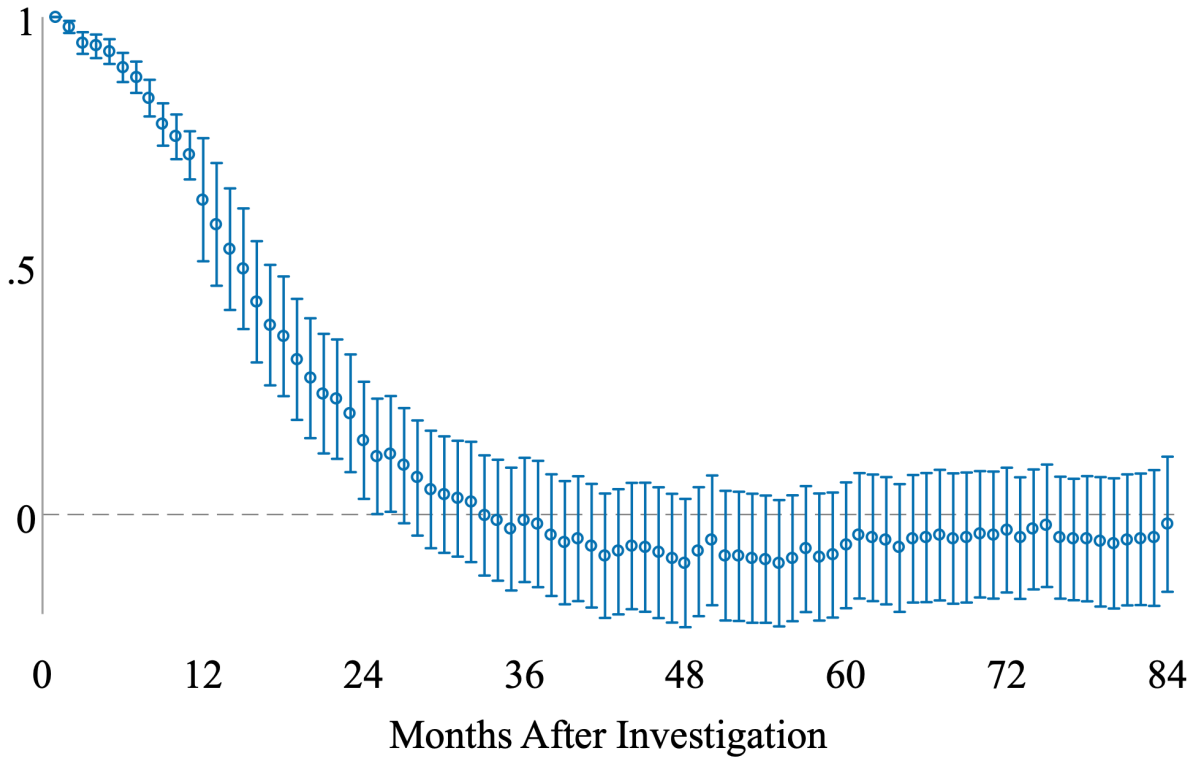
Notes. This figure describes the child maltreatment investigation process in Michigan. Substantiated means that investigators found enough evidence to support the abuse or neglect allegation. Conditional on substantiation, low-risk families received either a referral to community-based services like 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 1.3: Distribution of Investigator Removal Stringency Instrument



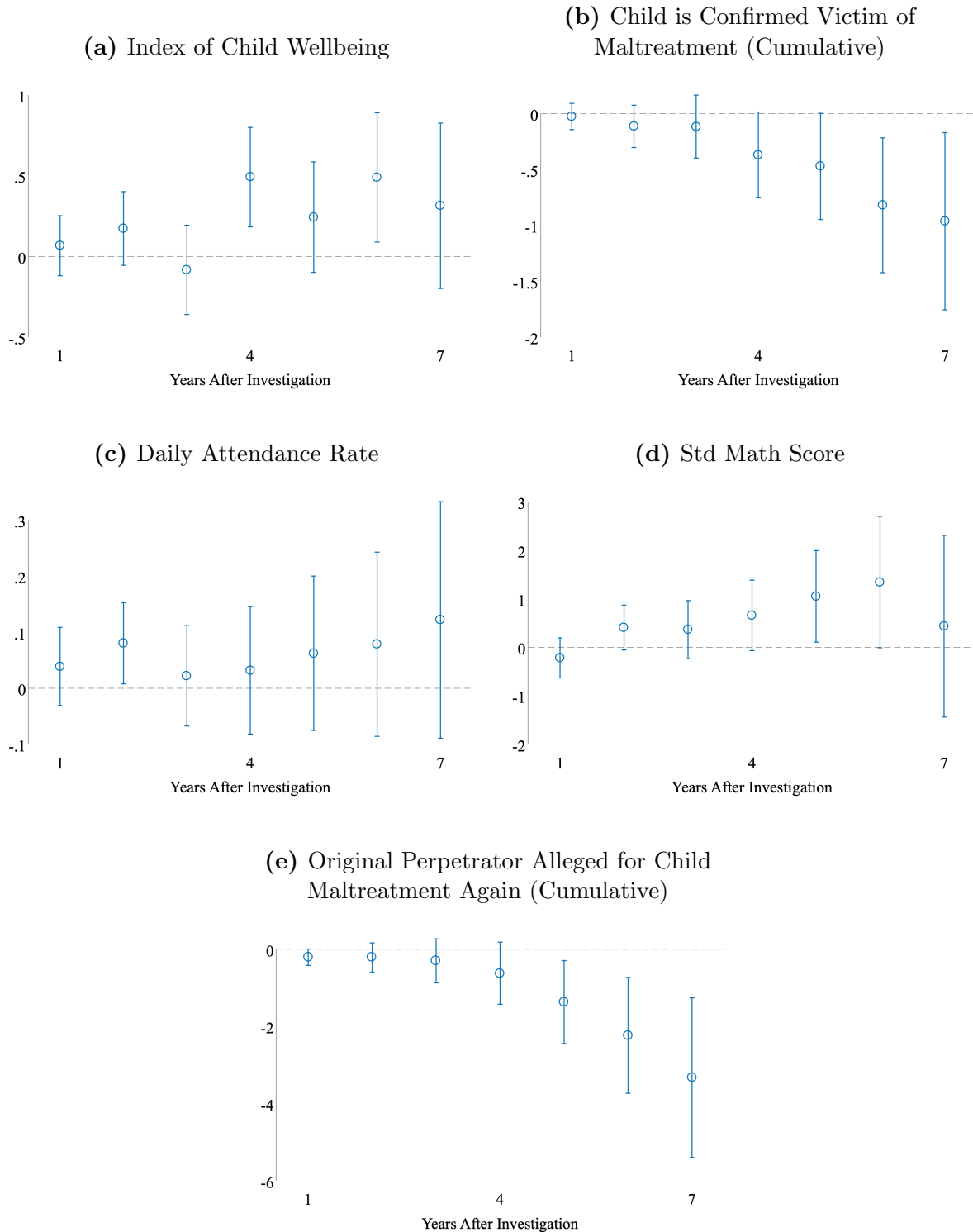
Notes. This figure shows the distribution of the removal stringency instrument residualized by the level of rotational assignment. That is, the instrument is shown net of child zipcode by investigation year fixed effects in order to show that there is variation in propensity to remove within local offices.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

Figure 1.4: Effects of Foster Care on Likelihood of Being in Foster System Over Time



Notes. This figure reports the results from 2SLS regressions of the likelihood of being in the foster system on an indicator for foster placement using removal stringency to instrument for placement. It plots both the point estimates and their 95% 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 turn eighteen years old exit from the analysis. The point estimate can be negative in the rare case that control compliers eventually entered foster care.

Figure 1.5: 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 1.5e represents the effect of child removal on the cumulative number of future allegations of child maltreatment against the original perpetrator. Since multiple perpetrators can be involved in the original case, this represents the mean effect across all perpetrators. For reference, 56% of investigations involved a single perpetrator, 97% involved one or two, and 99.4% involved three or fewer.

Table 1.1: Summary Statistics

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

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

Table 1.2: First Stage Effect of Removal Stringency on Foster Placement

	(1)	(2)	(3)	(4)
	Foster	Foster	Foster	Foster
	Care	Care	Care	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 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 retained in grade, 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 investigator. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.3: Balance Tests for the Conditional Random Assignment of Investigators

	Full Sample		4th Grade and Above	
	(1) Foster Care	(2) Removal Stringency	(3) Foster Care	(4) 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 A.2 reports the full set of results. Standard errors are clustered by investigator.

Table 1.4: Effects of Foster Care on Child Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Index of Child Wellbeing	Alleged Victim of Maltreatment	Confirmed Victim of Maltreatment	Daily Attendance Rate	Retained in Grade	Std Math Score	Std Reading Score	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 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 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 1.5: Effects of Foster Care Over Time

	(1)	(2)	(3)	(4)	(5)	(6)
	Index of Neighborhood & School Characteristics	Index of Child Wellbeing	Confirmed Victim of Maltreatment	Daily Attendance Rate	Std Math Score	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 A.8. 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 1.6: Effects of Adult Interventions 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 three 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 1.7: Effects of Foster Care on Child Outcomes Using Censored Data

	(1)	(2)	(3)	(4)
	Index of Child Wellbeing	Confirmed Victim of Maltreatment	Daily Attendance Rate	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 1.4 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 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

CHAPTER II

The Effect of Course Shutouts on Community College Students: Evidence from Waitlist Cutoffs

2.1 Introduction

State funding for higher education in the United States has dramatically declined over the past decade. Budget cuts following the Great Recession have persisted even ten years later—funding for public two and four-year colleges in 2017 was \$9 billion less than its pre-recession level in 2008 (Mitchell et al., 2016). Despite some evidence that money matters for college students, the mechanisms through which such resource effects operate are decidedly less clear. Anecdotally, overburdened college budgets are often associated with course overcrowding. When a college faces budgetary pressure, it may reduce course offerings or the number of sections per course. More students may find themselves unable to enroll in the courses they need to complete a degree. This hypothesis appears in the academic literature as well. Both Bound and Turner (2007) and Deming and Walters (2017) cite oversubscribed courses as a likely mechanism to explain the relationship between funding and college student outcomes, yet do not test it directly.

Credibly identifying the impact of limited course availability is challenging. Doing so requires detailed course registration data as a means to determine what classes students wish to take. It also requires exogenous variation in who is rationed out of a course. Using newly available data, this paper provides some of the only causal evidence on the impact of being shut out of a college course and the first estimates of its impacts among community college students. Community colleges enroll 47% of all U.S. undergraduate students in public colleges and 55% of underrepresented minority students (Snyder et al., 2018). Their open enrollment policies, binding class size constraints, lower tuition rates, and heavy reliance on state funding may make course scarcity especially salient relative to four-year colleges.

We use novel administrative course registration data from a large community college in

California to construct waitlist queues for each course. We link these to transcript data containing student course schedules, grades and degrees, as well as to the National Student Clearinghouse (NSC). The analysis measures the discontinuous impact of being stuck on a waitlist and unable to enroll in one's desired course on a student's current and future course-taking and degree completion, including transfers to other postsecondary institutions.

Our study is the first to use waitlist queues and admissions cutoffs as discontinuous breaks to determine who is able to or unable to enroll in a course. To understand the intuition behind this design, consider a section for an introductory English composition course. Suppose the section had a waitlist with two people on it, and before the end of the registration period, just one formerly enrolled student decided to drop out. This would give the first person on the waitlist the opportunity to enroll in her desired section, but not the second person. The admissions cutoff- the waitlist number below which a student does not get an opportunity to enroll- is very difficult to manipulate because waitlisted students can not reliably predict how many seats will open up. This introduces exogenous variation in who is able to take their desired courses. Our new approach leveraging waitlists for causal inference can be applied broadly in many other contexts.

The analysis primarily takes a local randomization approach to regression discontinuity analysis in order to compare students who signed up for a course-section waitlist and just missed or made the admission cutoff. Unlike a continuity-based framework, the local randomization approach explicitly treats observations in a narrow window around the cutoff like a randomized experiment. Cattaneo et al. (2018) argue that in settings with a very discrete running variable- those where there are only a few mass points around the cutoff- local randomization is the preferred estimation strategy.

Comparing students who just miss the waitlist cutoff to those who just make it, we find that students who were not able to enroll in their preferred section due to oversubscription were more likely to sit out the term altogether. Specifically, the reduced form results show that being stuck on a waitlist increased the probability of enrolling in zero courses by 1.6 percentage points, a phenomenon we call same-term drop-out. This represents a 16% increase relative to the 10.1% same-term drop-out rate among students who just got off of the waitlist.

Using the the waitlist cutoff as an instrument for being rationed out of a section, the 2SLS estimates show that being shut out of a course led to a 2.6 percentage point increase in same-term dropout. This is a 25% increase relative to the 10.4% same-term dropout rate among control compliers, those who enrolled in their desired section by the end of the registration period precisely because they got off of the waitlist. The estimated effects are robust to alternative sample definitions and different design decisions, including a continuity-based regression discontinuity analysis.

The rise in same-term dropout was driven by students waitlisted for especially popular courses, which serve as prerequisites for a variety of other courses and can be easily transferred for credit at any public four-year college in California. While students were no more or less likely to transfer to another two or four-year school within one year of being stuck on a waitlist, they were 34% more likely to transfer to other two-year schools within two years of missing the waitlist cutoff, in some cases enrolling in both two-year schools simultaneously. In particular, students tend to substitute to nearby two-year schools within 30 minutes of driving distance. These nearby schools have lower degree completion rates and their students have lower average salaries, both of which indicate a reduction in college quality.

Though we find no average effect of shutouts on completion rates for associate degrees, certificates, or bachelor's degrees within five years, there are divergent impacts by student ethnicity. Shutouts caused underrepresented minority students to transfer to another two-year school while Asian students, the largest ethnicity group at the college, responded to rationing by transferring to four-year colleges sooner than they would have otherwise. This led to a corresponding increase in bachelor's degree completion rates within five years of the waitlist for Asian students. Ethnicity is most likely a proxy for other unobservable skills and advantages- or lack thereof- in navigating the higher education system and illustrates the potential for heterogeneity in how the community college system is used. That is, the two-year system may be a direct substitute for the early years at a four-year school among students who have access to both options, a stepping-stone to eventually gain access to the four-year system, or a terminal setting itself.

Taken together, the impacts from a relatively small friction demonstrates that oversubscribed courses can meaningfully alter a student's path. These results document that limited course availability, an often cited mechanism through which funding matters for college students, can interrupt and distort student's educational trajectories

2.1.1 Related Literature

Broadly, this paper contributes to the literature on the impacts of resources in higher education. Most of the work in this space has used aggregate data, and in some cases even use variation in the incoming class sizes- which could induce course scarcity- as instruments for variation in dollars spent per student. For example, Bound and Turner (2007) uses variation in the size of graduating high school cohorts to estimate the effect of decreases in per capita funding, finding a commensurate drop in bachelor's degree attainment. Fortin (2006) uses variation in cohort sizes, state appropriations, and tuition to estimate impacts on college enrollment and ultimately the college wage premium. More recently, Deming and Walters (2017) estimates the effect of large changes in state budgets on enrollment and

degree attainment. The paper finds that budget cuts reduce the number of bachelor’s degrees, driven by a decrease in persistence among students who were already enrolled rather than decreases in matriculation rates. The authors write that, “our results are consistent with the much broader trend of informal capacity constraints in public institutions, including reduced course offerings [and] long waitlists,” yet they do not have data to test this mechanism directly (Deming and Walters, 2017).

Notwithstanding the work using aggregate data, there is limited causal evidence on micro-level pathways through which college budgets could affect degree attainment. Some studies have implied how resources could matter by evaluating resource-intensive interventions such as financial incentives (Barrow et al., 2014), tutoring, mentoring (Bettinger and Baker, 2014), or full-service wrap-around programs such as the CUNY ASAP experiment (Scrivener et al., 2015).¹ These studies generally have found positive effects.

To the best of our knowledge, there are two other papers that estimate the effects of course scarcity (Kurlaender et al., 2014; Neering, 2018). Both use administrative data from public, four-year universities in California and instrument for course shutouts using variation in the timing when students are first allowed access to the course registration systems. In Kurlaender et al. (2014), students at U.C. Davis (a moderately selective school) were put into registration priority blocks where all students in a higher priority block gained access before those in lower priority blocks, but within each block the day on which any given student could first register was randomly assigned. In Neering (2018), students in an anonymous public university were assigned to a priority sequence based on the first three letters of a student’s last name, such that students randomly assigned an earlier time in one term are intentionally assigned a later time the next term. Kurlaender et al. (2014) instruments for the average number of shutouts a student experiences over their first four years to estimate the impact of a shutout on bachelor’s degree completion and time to degree. The paper does not detect any effects of course shutouts. Neering (2018) instruments for the number of shutouts a student experiences in a given term and finds that shutouts reduce the number of credit units students attempt in that term, which seems to be offset by a rise in the rate at which students enroll over the summer. Consistent with Kurlaender et al. (2014), the author finds no downstream effects on graduation rates or time to graduation.

This paper extends the literature on course shutouts in two ways. First, we offer a new identification strategy, which directly compares students just able to get off of waitlists and into their desired courses to those who are stuck on the waitlist. Second, the granularity of the registration attempt data allow the analysis to examine heterogeneity by course

¹ASAP provided community college students with a comprehensive package of interventions, one of which was a higher course registration priority.

characteristics (such as the subject of the course, or how popular the course is). Third, this is the first paper to document the impacts of course shutouts in a community college, where there are at least four reasons why course scarcity may be more salient. On the demand side, community colleges have open enrollment policies, unlike selective four-year schools that can reject applicants in order to manage course demand. Second, tuition is much lower at community colleges, which reduces the barrier to entry and also fuels demand.

On the supply side, community colleges are particularly reliant on funding from state governments, which are affected by budgetary pressures. These first three factors make community colleges susceptible to large, unexpected swings in enrollment and funding. For example, enrollment in community colleges increased by over 8% between 2008 and 2009 during the Great Recession while enrollment in four year colleges increased by less than 1% (Dunbar et al., 2011). California's two year public schools in particular saw a sharp, per-student funding decrease of about 11% in 2009 due to the defeat of several budget proposals (IHE, 2009). Finally, section enrollment at many community colleges in California is capped at 40 students due to classroom size, while class sizes at four-year schools may be allowed to expand more readily. The potential for sectoral heterogeneity leave a gap in the current understanding of the effect of course capacity constraints.

This paper also contributes to a small literature on course registration behavior. Registration attempt data has rarely been used for descriptive analysis, let alone causal inference. Gurantz (2015) presents a review of other papers using registration attempt data and finds that they are few and far between. The paper also shows that it is not uncommon for community college students to register for classes well after their designated time, perhaps as a result of a weaker commitment to their education or a consequence of the difficulty of navigating the registration process. Understanding the reasons why students delay registration is especially important if course scarcity impacts student outcomes, as delays affect the degree to which students experience scarcity. This paper presents an innovative method for circumventing the selection bias in registration time which may prove useful in future work with similar data. Unlike other studies of course scarcity, the approach in this paper can be applied in settings where registration priority is not randomly assigned.

Finally, findings from this study can speak to documented longterm trends in the U.S., including the downward trend in bachelor's degree completion rates conditional on enrolling in college, and the upward trend in time to degree, even as there has been an overall increase in the number of students attending post-secondary institutions (Bound and Turner, 2007; Bound et al., 2010, 2012). These phenomena have been concentrated among students enrolling in non-selective two-year and four-year schools, and the literature has suggested disparities in resources per student between selective and non-selective schools as a possible

explanation.

2.2 Institutional Background

The study uses administrative data from De Anza Community College, a large two year college located in the Bay Area which is part of the California Community College system, the largest higher education system in the United States. The college has an average total enrollment of approximately 23,000 students per year and costs about \$3000 per year for a full time student. Yearly tuition is higher than the average two year school in the US (\$1,269), yet is much lower than public four year colleges (\$9,230) (Deming et al., 2012, Table 2, page 156). The college operates on a quarter system, yet enrollment is much lower during the summer term.²

De Anza offers a particularly useful setting for examining the impact of course shutouts. For one, community colleges are an important sector of the higher education landscape in California and nationally. In California, nearly half of all students attending a four year college previously attended a community college.³ Furthermore, transfers from California community colleges to the California State University (CSU) system were projected to increase by 25% from 2010 to 2020 (Wilson et al., 2010). Thus, two year schools are an increasingly vital step in the accumulation of human capital and production of labor market skills.

Most pertinent to this study, De Anza is a likely setting for observing course scarcity due to non-selective admissions, low tuition, small and capped class sizes, and the budgetary pressures of the recession. The data includes the years during the Great Recession, when California community colleges decreased the size of their staff by 8% due to budget shortfalls (Bohn et al., 2013). According to the Public Policy Institute of California, 88% of senior community college administrators surveyed in 2012 agreed that funding reductions were harmful for maintaining course offerings (Bohn et al., 2013).

Meanwhile, like all community colleges in California, De Anza has an open enrollment policy; anyone with a high school diploma or equivalent is automatically admitted. Not all open enrollment settings will automatically lead to scarcity. A college could respond to scarcity in realtime by creating additional sections if they observe excess demand during the registration period. However, both empirical evidence and anecdotal evidence from De Anza administrators offer little support for this type of dynamic course creation. There were no sections in the data where the first student enrolled a few days after a different section of

²Curious readers can see Fairlie et al. (2014) for more details about De Anza Community College.

³See U.S. Department of Education (2017); CCCCO (2012); and Sengupta and Jepsen (2006).

the same course filled up. In addition, the marginal cost of adding a section is non-trivial. According to De Anza's salary schedule, most instructors are paid between \$7,500 and \$9,000 to teach an additional section. This figure does not factor in any costs or constraints from classroom space or equipment, any increase in fringe benefit costs, or the difficulty of hiring in a part of the state with consistently lower-than-average unemployment rates. The actual marginal cost is likely more expensive.⁴ Furthermore, De Anza can not simply increase the number of students permitted into a section. Class sizes are set around the 40 student mark. Changes in class sizes are limited by available classroom configurations and need to be approved by the faculty labor union.

2.2.1 Data Sources

This study benefits from access to community college institutional records and data from the National Student Clearinghouse (NSC). Data from the college includes registration attempt logs, student demographic characteristics, and student-level transcript records. Students in the sample enrolled at the school between the fall quarter of 2002 and the spring quarter of 2010. Students are linked to their transcripts which record grades and credits for every course offered by the college during the sample period. In addition, internal data on associate degrees and certificates are available through the summer of 2010.

Especially important for the analysis, detailed logs document each registration attempt during a term's registration period. An enrollment attempt is identified by a student identifier, time- with precision to the second- and course section. For each attempt, the logs report an outcome that can take one of four values: enrolled in the section, placed on a waitlist, dropped from the section, or no change. The difficulty of obtaining data of this nature has prohibited most analyses of course scarcity on a micro level.

Students are also matched to the NSC, which records enrollment at most postsecondary institutions in the United States, through the summer of 2016. The NSC also provides data on degrees earned from these institutions, supplementing administrative records on degree completion from De Anza. This allows us to examine effects on certificate and associate degree completion from two-year colleges as well as bachelor degree completion from four-year schools many years after a students' registration attempt at De Anza.

2.2.2 Section Enrollment

The online registration process takes place one to two months before the term begins. It is governed by an automated system and students are given one of seven enrollment

⁴Larger classes also count as double or even triple teaching credit for instructors.

priority designation dates, upon which they are granted access to the registration system. Registration priority is primarily determined by credit accumulation, although some students are assigned special priority if they are an athlete, a veteran, or are involved with the Extended Opportunities Programs and Services- a service for at-risk students. The registration priority assignment rules should generate discontinuous changes in the time that students sign up for courses, independently of any waitlist effects. Therefore, we conduct all analysis within registration priority and special student categories.

When a given student searches for a desired section (eg. MWF 9-10AM) of a desired course (eg. ECON 101 Principles of Microeconomics), she is informed of the location, instructor and the available number of seats for that particular section. Students can sign up for a maximum of 21.5 credits at one time, about 7 courses. If there are no seats available, the system displays the number of other students on the waitlist.

There are a few rules governing the waitlist process. Students on a waitlist for one section of a course are not allowed to register for the waitlist of other sections of the same course and cannot register for sections of other courses that meet at the same time. According to current policies, if a seat opens up in a section during the registration period, waitlisted students are automatically enrolled in the section. While archived records of the waitlist policy are available going back to 2008, anecdotes about the policy before 2008 suggest that when students on the waitlist were notified of an opening, they were given 24 hours to enroll. If they did not enroll in 24 hours, then the next student on the waitlist could claim the spot. We check for robustness to the policy by restricting the analysis to attempts between 2008-2010 in Table B.11.

The analysis focuses on registration attempts before the term begins. After the term begins, instructors have more discretion over enrollment and often make enrollment conditional on attendance. The first stage estimates the impact of missing a waitlist cutoff on being enrolled in the waitlisted section at the end of the registration period, prior to the start of classes. Many of the outcomes concern enrollment patterns as well. For these, enrollment is defined as being enrolled after the add/drop period a few weeks into the term.

2.2.3 Sample Characteristics

Students are part of the sample if they registered for a course waitlist during the registration period between fall 2002 and spring 2010. Community colleges serve a wide variety of people, including students hoping to transfer to four year schools, those completing a vocational degree, and those taking a recreational course. Therefore, the analysis focuses on students attempting to get a two year associate degree or transfer to a four year institution, and for whom enrolling in a bachelor's program in a four-year institution could be considered a

reasonable substitute. This allows for ease of interpretation and makes a cleaner comparison to previous studies on course shutouts at four-year schools. Upon enrolling, students are asked to declare their educational goal or intention. Table B.1 lists all of the categories a student can choose from in declaring their intention. The sample includes all students who declare an intention to transfer to a four-year school, earn an associate degree, or who are undecided. The analysis is robust to including all students though. In addition, we exclude registration attempts in the optional summer term from the analysis sample.⁵

We focus on the first waitlist a student ever signed up for in order to avoid dynamic RD issues. While students may sign up for waitlists in subsequent terms, the analysis is explicitly testing the hypothesis that missing a waitlist cutoff influences whether a student appears in a subsequent semester. In addition, students may sign up for another waitlist in the same term. To the extent that the first waitlist a student signs up for represents the course that they most desire to enroll in, the analysis can be thought of as the effect of scarcity in the courses students most care about. Ultimately, the results are robust to including all waitlists and clustering standard errors at the student level.

Table 3.1 reports summary statistics at the section and student levels. Column (1) of Panel A shows that just under half of all sections were ever oversubscribed. This statistic masks differences across subject areas. 68% of all sections in science, technology, engineering, and math (STEM) courses are oversubscribed during the registration period, compared with 50% of arts & humanities sections, 60% of social science courses, and only 30% of sections for other courses. For classes that were oversubscribed, the average waitlist had about nine students still on it at the end of the registration period. Column (2) of Panel A shows the subject breakdown for all course sections included in the analysis. By definition, these sections all had waitlists. 34% of sections included in the analysis were in STEM fields, 28% were in arts and humanities, 12% were social science courses, and 26% fell into other subject areas. Average waitlist lengths at the end of the registration period for sections in the analysis were slightly lower, at 8.01 students.

Panel B shows descriptive statistics for students in the analysis compared to the California average. Column (1) reports demographics for all two-year colleges in California from the Integrated Postsecondary Education Data System (IPEDS). Column (2) contains information for all students who ever enrolled or attempted to enroll for a course at De Anza Community College during the sample period, as measured by the administrative registration records, and Column (3) reports the characteristics for students included in the analysis sample. De Anza serves slightly more women than men, though the ratio is not higher than the California

⁵The summer term lasts between 6 and 8 weeks depending on the course. The other terms are about 3 months long. Far fewer students enroll during the summer term.

average. The ethnic breakdown reflects the demographics of the Bay Area: in Column (2), 40% of students are Asian and 26% are White, while Black and Hispanic students make up only 19% of the student body. Relative to the state average, De Anza students are much less likely to be underrepresented minorities and less likely to receive financial aid.

As show in Column (3), the analysis sample contains registration attempts from 4,258 unique students. These students are more likely to receive financial aid and is younger than the De Anza student population. Students in Column (3) take an average of 1.81 courses in their first observed term relative to the population average of 1.70. Finally, in-sample students appear on 1.01 waitlists during the registration period in their first term, on average. Among all De Anza students who attempt to register during the advanced registration period, the average number of waitlists in the first observed term is just 0.42. De Anza students as a whole are thus less likely to sign up for waitlists and take fewer courses. Like the differences in age, this is consistent with the restrictions on students' educational goals, which select students with an intention to transfer or earn a two-year degree. Students who did not declare this interest are probably less attached students or students taking recreational courses.

2.3 Empirical Strategy

The analysis employs a fuzzy regression discontinuity design using waitlist queues to form a running variable. To illustrate the intuition behind the design, suppose a course section has a waitlist with two people on it. By the end of the registration period, if one formerly enrolled student decided to drop out, then the first person on the waitlist would have the opportunity to enroll in her desired section while the second person on the list would not. While the decision to sign up for a waitlist is clearly endogenous, it is difficult to anticipate how many spots will open for any given section, and therefore how deep into the queue admission offers will be extended. This makes the cutoff very difficult, if not impossible, to manipulate.

2.3.1 Construction of the Running Variable

Conceptually, the running variable represents the number of spots that would have needed to open up in order for a student to have the opportunity to enroll during the registration period, assuming she never dropped out of the queue. Figure 2.1 shows a hypothetical enrollment log to illustrate the running variable construction. The first column P_i is a student identifier that represents the chronological order in which students initially sign up for any section or section waitlist. A student who enrolls in a section without ever having

been on a waitlist also has a position P_i . However, X_i , the initial waitlist position, is only defined for students who enter a waitlist queue. In Figure 2.1, $X_{42} = 1$, as student 42 is first on the waitlist when she signs up and similarly, $X_{43} = 2$ and $X_{44} = 3$.

Importantly, the initial waitlist position is not the same as the running variable. Rather, the running variable for student i also involves D_i , the number of students who registered before student i and dropped out during the registration period after student i registered. In Figure 2.1, both student 7 and student 22 enrolled before students 42, 43, and 44, and dropped after these students entered the waitlist. Therefore, D_{42} , D_{43} and D_{44} all equal two. Although student 38 also dropped out of the queue, this occurred before students 42, 43, and 44 signed up for the waitlist and therefore student 38 has no effect on D_{42} , D_{43} or D_{44} . Essentially, D_i counts the types of drops that would move a student up on the waitlist or create a spot for her in the section.

The running variable RV_i is defined as the difference between one's initial waitlist position and the number of drops D_i ,

$$RV_i = X_i - D_i. \tag{2.1}$$

Students with a strictly positive running variable would not have had the opportunity to enroll in the section during the registration period. Students with running variables less than or equal to zero would have had an opportunity to enroll, conditional on staying in the queue. A student can only influence her own running variable by signing up, not by dropping out. For example, although student 44 eventually dropped off of the waitlist, she still received a running variable. This paper compares the outcomes of students who just made the waitlist cutoff- those with $RV_i = 0$ - to those who just missed it- students with $RV_i = 1$.

This running variable construction is preferred to other possible definitions because it preserves the order in which students sign up for the waitlist. For example, suppose student A signs up to a waitlist that already has two people on it, and student B signs up the next day, but in the interim two people have dropped out of the class. Student B would be in the second position, but student B's running variable as defined above could not be smaller than student A's. A running variable based on the time that students sign up would also have this order preserving feature, however, the construction of a cutoff time is not obvious.⁶

Of course, students continue to enroll and drop after the registration period ends. The analysis does not include these attempts because there is a larger role for instructor discretion

⁶In fact, the construction of a cutoff time fully depends on the construction of the current running variable. That is, without a cutoff waitlist position, there can be no cutoff time. Appendix B.3 tests the robustness of the results to a time-based running variable; the findings remain similar.

once the quarter begins. There is imperfect compliance since students can drop out of the queue. That is, students with $RV \leq 0$ might not actually be enrolled in the section at the end of the registration period.⁷ Thus, estimates use a fuzzy RD design as opposed to a sharp RD.

2.3.2 Estimation

Consider a student who placed herself on a waitlist. $NotEnroll_{ist}$ is a measure of rationing and indicates the treatment. It is one if the student does not enroll in her desired section s in term t during the registration period, and zero otherwise. Let $Y_i(NotEnroll_{ist} = 1)$ be her educational outcome if she does not enroll in her preferred section and $Y_i(NotEnroll_{ist} = 0)$ be her educational outcome if she does. The analysis estimates $\mathbb{E}[Y_i(NotEnroll_{ist} = 1) - Y_i(NotEnroll_{ist} = 0) \mid RV_{ist} = 1]$. This is interpreted as the local average treatment effect (LATE) for compliers, students who are rationed out of a section if they miss the cutoff and are induced to enroll if they make the cutoff. It is important to consider the type of student represented by a complier in this scenario. Students discouraged by a waitlist cutoff could be less motivated, less organized, or both. Furthermore, they may be less savvy navigators of institutions for reasons that reflect social inequality.

To estimate the LATE, we use a two stage least squares regression for students within one position of the waitlist cutoff. That is, for student i in section s and term t with $RV_{ist} \in [0, 1]$:

$$NotEnroll_{ist} = \alpha_0 + \alpha_1 MissWL_{ist} + \mathbf{X}_{ist}'\Gamma + \delta_t + \zeta_{ist} \quad (2.2)$$

$$Y_{ist} = \beta_0 + \beta_1 \widehat{NotEnroll}_{ist} + \mathbf{X}_{ist}'\Pi + \delta_t + \epsilon_{ist} \quad (2.3)$$

where $\widehat{NotEnroll}_{ist}$ represents the student's predicted probability of not enrolling in the section according to equation 2.2. Enrollment for the first-stage equation is measured on the last day of the advanced registration period, prior to the start of classes. RV_{ist} is the running variable, and $MissWL_{ist}$ is an indicator equal to one if $RV_{ist} = 1$ and equal to zero otherwise. \mathbf{X}_{ist} is a vector of covariates including gender, race, ethnicity, US citizenship status, age, financial aid receipt, registration priority fixed effects, special admit status, special program status, as well as indicators for missing variables. The δ_t represent a vector of term by year fixed effects and ζ_{ist} and ϵ_{ist} are error terms.

The estimates rely on local randomization assumptions to identify the causal effect of not enrolling in a desired section due to oversubscription for compliers (for a detailed description

⁷By definition, students with $RV > 0$ could not have enrolled during the registration period though. In this sense, we observe only one-sided noncompliance.

of local randomization see Cattaneo et al., 2017b, 2018). Essentially, local randomization assumes that within one position on either side of the waitlist cutoff, the running variable is unrelated to potential outcomes. That is, assignment of the running variable is “as-if random,” and there is no selection into treatment.

Local randomization is appropriate for settings with extremely discrete running variables, as opposed to the more commonly used RD assumptions involving continuity of the regression function, which require a continuous running variable.⁸ In fact, Cattaneo et al. (2018) argue that in settings with a very discrete running variable, local randomization is “possibly the only valid method for estimation and inference.” The full set of assumptions include

1. *Fixed Potential Outcomes.* Potential outcomes are non-random and fixed for students within one position the cutoff.
2. *Known randomization mechanism.* The distribution of the treatment assignment vector is known for those within one position of the cutoff.
3. *Unconfoundedness.* Whether students end up directly on the right or left of the cutoff does not depend on potential outcomes.
4. *Exclusion Restriction.* Within one position of the cutoff, the running variable influences outcomes only through treatment, not directly.
5. *SUTVA.* Locally, within one position of the cutoff, each student’s potential outcomes only depend on his or her own treatment assignment, and not anybody else’s.
6. *Monotonicity.* Within one position of the cutoff, missing the cutoff does not cause any students to be more likely to enroll than they otherwise would have been, and making the cutoff does not cause any students to be less likely to enroll.

Assumption one and two define what is meant by random. Assumption one means that a student’s potential outcomes are fixed and inherent to her.⁹ Assumption three is the key to local randomization and has some testable implications. Any manipulation of a student’s own running variable would violate this assumption. However, a student’s running variable is dependent on the number of other students who drop the section, and is out of her control.¹⁰

⁸The results are robust to using a larger bandwidth and treating the running variable as if it were continuous, however.

⁹There is a formulation of the local randomization assumptions for potential outcomes that are random variables as well, but it would not change anything in the mechanics of estimating the LATE parameter (Cattaneo et al., 2017b).

¹⁰In some sections, no students drop. In others, as many as twenty students drop. Among sections in the analysis, the 10th percentile number of drops is zero and the 90th percentile is five.

An example of a violation of the assumption is if a student is more likely to sign up for the waitlist because she knows that a friend is planning to drop. This seems unlikely, particularly for our sample which is mostly incoming students who may not know many people. Section 2.3.3 formally tests for manipulation around the cutoff.

Assumption four, the exclusion restriction, is generally not needed in RD studies that rely on continuity of the conditional regression function, and indeed, it would be unreasonable to assume that there is no direct relationship between the running variable and the potential outcomes for all values of the running variable. Clearly, somebody who signed up for a section very early in the registration period is different from somebody who signed up very late. However, it is more plausible that there is no difference, on average, between people within one waitlist position of each other.

The stable unit treatment value assumption (SUTVA) is standard in estimating LATE using an instrumental variable, though of course it's possible that there are spillovers from other students. Again, one mitigating factor for these possible spillovers is that most students are first-time enrollees and likely do not know each other well. The monotonicity assumption is also standard. Since signing up for a waitlist has a cost- students are barred from signing up for any other section at the same time or for the same course- it is implausible that being high enough on the waitlist to gain admission would cause a student to be less likely to sign up for a course than they otherwise would have been. Being more likely to sign up for a course because one missed the waitlist cutoff is also intuitively unlikely, though not testable.

Equations (2.2) and (2.3) are estimated using a two-stage least squares regression. Although Lee and Card (2008) suggest clustering standard errors by the value of the running variable when the running variable is discrete, Kolesar and Rothe (2016) point out that confidence intervals constructed in this way have poor coverage when the number of clusters is small, which is the case in this analysis. Therefore, only the usual heteroskedasticity robust standard errors are used, unless otherwise noted.

2.3.3 Validity Checks

One can test for manipulation of the running variable by checking for smoothness in the density of the running variable at the cutoff. Figure 2.2 shows the density of the running variable. Table 2.2 reports p-values from formal tests for smoothness using a McCrary-like test specifically designed for discrete running variables, introduced in Frandsen (2017). An important assumption of the Frandsen (2017) test is that the second order finite difference of the running variable's probability mass function (pmf) is bounded at zero, with the bound represented by k . Intuitively, k represents the amount of curvature or nonlinearity in the pmf of the running variable that would still be compatible with no manipulation. The choice

of k is left to the researcher, but the author notes that a natural maximum is the amount of curvature in a discretized normal distribution that is roughly as discrete as the observed distribution of the running variable- call this the “rule of thumb” maximum. If there are about twenty support points within one standard deviation of the cutoff, then the rule of thumb maximum is 0.005, whereas if there are only six support points, it is 0.047. We test for manipulation using many values of k but note that in our context there are about eight support points within one standard deviation. The density test fails to reject the null of no manipulation at the five percent level for all values of k and fails to reject it at the ten percent level at values that are much smaller than the rule of thumb maximum in our context.

Another testable implication of the FRD assumptions is that predetermined characteristics should be balanced across the waitlist cutoff. Table 3.3 reports the results of linear regressions testing for imbalance across the waitlist cutoff in student characteristics.¹¹ The regressions condition on term by year fixed effects, registration priority fixed effects, and special student categories that affect registration priority. None of the student characteristics are statistically significant at the five percent level, although age is significantly different across the threshold at the ten percent level. The difference is small in magnitude however, equal to about four months. Furthermore, the covariates are not jointly significant, with a joint F-test yielding a p-value of 0.242. In addition, although the analysis relies only on variation between students with a running variable of zero or one, Figures B.1 through B.1 show that these baseline characteristics are similar across the cutoff when looking at a wider bandwidth.

The analysis also examines two other student characteristics that should be very similar across the threshold if there is no selection into treatment. First, students across the cutoff sign up for a similar number of other waitlists during the registration period- 1.07 on average for those with running variable of zero and 1.09 for those with running variable of one. Since this behavior occurs after students sign up to the initial waitlist, the student had some information about their likely schedule, even if it wasn’t full information. Therefore, while this variable can’t be tested for balance formally, the similarity in waitlist enrollment behavior is consistent with “as-if” random assignment. In addition, there was less than a day between the registration attempts of students in the same section with running variable zero or one. Specifically, the average amount of time between registration for these students was 19 hours, just 13% of a standard deviation between any two registration attempts in a waitlisted section.

¹¹The balance tests do not include registration time because it will mechanically be earlier for those with a running variable of zero.

2.4 Course Scarcity and Student Outcomes

2.4.1 First Stage Estimates

The first stage estimates can be easily seen in discontinuities at the cutoff. Figure 2.3a shows a discontinuity at the waitlist cutoff for enrollment in the waitlisted section at the end of the registration period. 64% of students just to the left of the cutoff- the last to have the opportunity to enroll in the section during the registration period- ended up enrolled in the waitlist section. In accordance with the definition of the running variable, students who miss the waitlist cutoff are not able to enroll during the registration period. Figure 2.3b shows the enrollment rates for courses in which a student has been waitlisted for one section. Due to the rules about only being able to enroll in one waitlist per course, the first stage looks almost identical. In theory, somebody on the left of the cutoff could have switched sections within the same course. This does not appear to happen often, as 65% of students who do not miss the cutoff ultimately enroll in the waitlisted course, relative to 64% who enroll in the waitlisted section.

It is important to verify that the first stage effect of missing a waitlist cutoff is large enough to avoid a weak instruments problem. Table 2.4 examines sensitivity of the first stage to the inclusion of covariates for both enrollment in the desired section and the desired course, and reports F-statistics. The F-statistics are all greater than 3500 regardless of whether covariates are included and whether examining enrollment in the waitlisted section or course. As reported in Panel A, students who miss the waitlist are between 64.1 and 64.4 percentage points less likely to enroll in their desired section than those who just make it. The barrier to entry for a section translates into a barrier at the course level. In Panel B, students are between 64.5 and 64.8 percentage points less likely to enroll in their desired course after missing the waitlist cutoff.

Although estimates of the first stage for section enrollment and course enrollment are qualitatively similar, all further analysis uses the section enrollment as the endogenous variable of interest, as it is most directly influenced by the waitlist cutoff. The results are nearly identical regardless of whether the analysis defines treatment at the section or course level though.

2.4.2 Reduced Form and IV Estimates

The main outcomes of interest are enrollment in the concurrent term and enrollment in other two and four-year schools within one through five years of the waitlisted term. Although the estimates identify effects by comparing students immediately on either side of the cutoff, Figures B.2 through B.4 visually depict the reduced form effects using a

larger window. They plot the residuals of the main outcome variables, conditioned on the observable, pre-determined characteristics, and binned by values of the running variable. Figure B.2 is the visual representation of the reduced form effects of missing a waitlist cutoff on whether students enroll in zero, one to two, or three or more courses in the waitlisted term and whether they enrolled in any course in the following non-summer term. Enrolling in zero courses can be thought of as same-term drop-out, though the student may appear again in a later term. Enrollment in one or two courses would be like enrolling part-time, while three or more courses is roughly full-time enrollment. There is a 0.016 percentage point jump up in percentage point jump in same-term dropout, and smaller, less prominent jumps in the other enrollment outcomes.

Figure B.3 shows the reduced form impact on whether the student transfers to another two-year school. There is no noticeable rise in the share who transfer within one year, but a large 2.3 percentage point increase in the share of students who transfer within two years to another two-year school for those who missed the waitlist cutoff. Since the data only include enrollment in other two-year schools and not transcript records from those schools, the analysis cannot disentangle whether students transfer only to take their waitlisted course or for their entire course load. The difference in transfers to other two-year schools on either side of the cutoff gets smaller in later years. While reduced form effects of two percentage points may seem small, these translate to meaningfully large effects relative to the control means. For example, only 10.7% of students transfer to another two year within two years. Finally, Figure B.4 shows that there is no noticeable change in transfers to four year schools across the cutoff at any time point.

Table 2.5 presents formal estimates of the LATE of being shut out of a course on enrollment patterns in the concurrent semester. Columns (1), (2), and (3) report the effect of begin shut out on whether a student enrolls in zero, one to two, or three or more courses respectively. All results control for the full vector of covariates and use a bandwidth of one. The main results show students are 2.6 percentage points more likely to “drop out” in the waitlisted term; that is, to take no course at all that term. The estimated increase in same-term dropout is an increase of 25% relative to the same-term dropout rate of the control compliers, which is 10.4%. There are also negative, though not statistically significant, effects on course-taking for students who do take a course. The rise in same-term drop-out is accompanied by a 1.7 percentage point decrease in the probability of enrolling in three or more classes- a full course load- relative to a control complier mean of 55.9%. We also estimate a 0.8 percentage point decrease in the likelihood of enrolling in one to two courses that term relative to a control complier mean of 33.7%. These results cannot distinguish between a cascading effect- somebody who would otherwise have taken three

courses dropping down to two and somebody who would have taken two, dropping down to one, and so on- and a more dramatic shift from a plan to take a full course load to taking no courses, or some combination of these two options.

Table 2.6 shows the effect of course shutouts on transfer rates and degree completion for associate degrees, certificates, and bachelor's degrees. Students were no more or less likely to transfer to another two or four-year school within one year of being stuck on a waitlist. Taken together with the increase in same-term dropout at De Anza, this suggests that course scarcity increases the likelihood of dropout from college altogether for that term.

This dropout effect does not persist, however. There is a large positive effect of 3.6 percentage points on the transfer rate to other two-years within two years of missing the waitlist cutoff. This is relative to a control complier mean of 10.5%, which means the transfer rate increases by 34%. The point estimates for transfers to other two-year schools within three, four, and five years are also meaningfully large, 2.6, 2.6, and 2.7 percentage points respectively, but not statistically significant. It suggests the effect attenuates but might not entirely dissipate over time. There are no detectable effects on transfers to four-year schools or on the share who earn associate degrees or certificates from De Anza or any other school, or bachelor's degrees up to five years out.

In general, the three most frequent recipients of De Anza's transfer students are: Foothill College, Evergreen Valley College, and San Jose City College. These are roughly 15 minutes, 30 minutes, and 18 minutes from De Anza by car, respectively. Foothill college in particular is almost seamlessly integrated, with cross-registration between De Anza and Foothill being common and easy to do because it uses the same registration system.¹² However, as shown in Table B.2, when estimating the treatment effect on attending each of these alternative schools separately, there is a statistically significant increase in enrollment at Evergreen, San Jose City College, and all other two-year schools, but not at Foothill. It's likely that students consider classes at Foothill as part of the initial choice set when they are registering, and not as a back-up option after the fact. According to the U.S. Department of Education's College Scorecard website, De Anza Community College costs less, has a higher graduation rate, and students who attend De Anza earn higher average salaries after attending than attendees at the other two colleges. In addition, both by revealed preference and by online ranking services such as NICHE and Wallethub, which consistently rank De Anza above Evergreen and San Jose City, it is likely that students are worse off from having to substitute for the courses they need at these common alternatives.

¹²To be clear, although students often take some classes at De Anza and others at Foothill in the same term, the data consists only of course registration attempts at De Anza.

2.4.3 Subgroup Analysis

This section reports results by subgroup categories, including differential impacts by gender, ethnicity, popularity of the course, and course subject. The demographic breakdowns are proxies for student vulnerability or disadvantage. The ethnic categories in particular are not taken to have theoretical meaning in their own right, but are rather meant to serve as rough correlates of unobservable characteristics such as the human capital of a student's social network or other barriers to human capital accumulation. Course popularity and subject are meant to test the idea that not all courses are equally important to a student's educational and labor market goals.

Tables B.3 to B.6 show the differential effects on course enrollment in the concurrent term for all subgroups. There are no detectable differential impacts on enrollment patterns by demographic subgroups, either gender or ethnicity.

There is more evidence that the type of course may be important for enrollment patterns. To gauge the popularity of the course, we tallied enrollment requests for all courses across the sample period and picked the top five most requested with the rationale that more popular courses are likely to be important pre-requisites for common majors or for transfer. The top five include three introductory writing courses, a government course, and a psychology course. Indeed, course catalogs confirm that these five classes were all prerequisites for a variety of other courses at the college. Furthermore, they were part of the Intersegmental General Education Transfer Curriculum (IGETC), which allows them to be easily transferred toward a bachelor degree in the UC system. As shown in Table B.5, the point estimates for same semester drop-out are more than twice as large for the top five most popular classes, although they are not statistically different from each other. In addition, being rationed out of a top five class seems to lead students to either drop out or increase their enrollment to full time, with a significantly larger drop in part-time enrollment. Waitlists for less popular classes cause relatively larger, though not statistically significant, decreases in full-time enrollment instead. This suggests that students enrolled in the most popular classes are relatively less attached to college. Finally, as shown in Table B.6, differences in impact by subject matter such as STEM, arts and humanities, social studies, and other subjects, however, are minimal.

Interesting dynamics emerge in transfer and degree completion by ethnicity categories. Tables B.7 and B.8 report results on transfer rates to other colleges by ethnicity, where students are partitioned into three groups: Asian, White, and underrepresented minority (URM). The URM category consists of Black, Hispanic, Native American, multi-racial students, and students who do not fit into any other category. The point estimates are plotted in Figures 2.4 and 2.5.

There is a divergence in transfer responses by ethnicity. As seen in Figure 2.4, although all

students show a positive uptick in transfer rates to other two-year schools within two years of the waitlist, the point estimates are highest for URM students. For these students, transfers to two-year schools continue to increase every year through five years out. Meanwhile, other students do not transfer to two-year schools at an appreciably high rate, including near zero point estimates for Asian students and negative point estimates for white students.

In contrast, Asian students are more likely to transfer to a four-year school in response to being rationed out of a course, as shown in Figure 2.5, while URM students become increasingly less likely to transfer to a four year school as time goes on. With Asian students accelerating their transfer to a four-year school, there should be a corresponding uptick in bachelor's degree completion for Asian students. Indeed, Figure 2.6 shows a positive effect of rationing on bachelor's degree completion among Asian students, especially at the five year mark. There is no impact on bachelor's degree completion for URM students, although the control complier mean for this group is near zero for the first three years after the waitlist and still quite low at 5.8% in the fifth year out. Finally, there is evidence that bachelor's degree attainment among White students is hampered by course rationing. Being rationed out of a course reduces bachelor's degree completion within five years by 56% for White students, relative to a control complier mean of 13.4%. The estimates plotted in Figure 2.6 can be found in Table B.9.

This analysis suggests that students use the community college system differently in California. Perhaps students better prepared to navigate the college landscape, as proxied by ethnicity, strategically enroll in community college after high school because it is easier to get into a UC school as a community college transfer. For example, Berkeley and UCLA acceptance rates are almost twice as high for transfer students than for freshman admits. Anecdotally, these statistics seem well known on college discussion forums. The results highlight the many potential responses to course scarcity in a community college setting—some transfer to four-year schools thus accelerating their time to a bachelor's degree, while others transfer to lower-quality two year schools.

2.4.4 Sensitivity Analysis

The results of this paper are robust to several design decisions. First, as is standard in a regression discontinuity analysis, the analysis checks whether there are treatment effects at placebo thresholds.¹³ Figure 2.7 plots the reduced form coefficients the two main outcomes

¹³An FRD that relied on continuity assumptions might also check for sensitivity to bandwidth choices and controlling for different polynomials of the running variable. The local randomization assumptions are only valid within one position of the cutoff, however. In particular, conditional independence does not hold as the bandwidth is increased. This is not surprising because increasing the bandwidth creates a comparison between students who signed up to the waitlist at increasingly far apart in time. Given that the identification

affected by course shutouts, estimated for ten different waitlist thresholds.¹⁴ The outcomes are: took zero courses in the waitlisted term and transferred to another two-year school within two years. Table B.10 reports the corresponding point estimates and standard errors represented in the figure. The true cutoff represents the last student on the waitlist who received an offer of admission to the section. For each placebo cutoff j , students with $RV_{ist} = j$ behave as the control group and are compared to students directly to the right, with $RV_{ist} = j + 1$. The difference in outcomes at any cutoff $j \neq 0$ should not be significantly different from zero, which is the case.

Table B.11 shows the LATE of a course shutout on selected outcomes using different samples of students. Results are robust to alternative sample restrictions. Column (1) includes all students, regardless of which initial intention they declared, and all waitlists. This examines whether estimates are sensitive to conditioning on students' initial declared intentions listed in Table B.1 or to using student's first waitlist. Column (2) restricts the sample to students who declared an intention to transfer to a four year. Column (3) includes only terms after 2007, when documentation on enrollment rules is available (see section 2.2.2 for a discussion of the issue). Column (4) uses the waitlist cutoff to instrument for course enrollment, rather than course section enrollment. Finally, Column (5) uses the main analysis sample but treats the running variable as if it were continuous, performing a traditional regression discontinuity analysis using a bandwidth of ten and a linear function form.

The estimates on taking zero courses in the waitlisted term are similar in magnitude to the main results and all are statistically significant at the 10% level. In addition, the analysis is nearly identical both when instrumenting for course enrollment rather than section enrollment and when treating the running variable as if it were continuous.

2.4.5 Complier Densities

This section estimates outcome densities for treated and untreated compliers in order to better understand how enrollment patterns change. Up to this point, the analysis has looked at discrete changes in course load, examining whether students respond to course shutouts by dropping out, taking one or two courses, or three or more. This analysis may mask greater heterogeneity at different points in the course load distribution. Following Abdulkadiroglu

is only valid within one position around the cutoff, testing sensitivity to bandwidth and functions of the running variable are not relevant.

¹⁴We perform this placebo threshold exercise using the reduced form rather than the two-stage least squares estimates because the first stage is zero at placebo cutoffs by construction. That is, all students to the right of the true cutoff, where the running variable is equal to zero, were not able to enroll in the waitlisted section at the end of the registration period.

et al. (2018), the paper estimate kernel densities of the form

$$\frac{1}{h}K\left(\frac{Y_{ist} - y}{h}\right) \times NotEnroll_{ist} = \tau_y NotEnroll_{ist} + \mathbf{X}'_i \lambda_y + v_{iy} \quad (2.4)$$

where $Y_i(0)$ and $Y_i(1)$ are potential outcomes, and failing to enroll in the desired course section is the treatment. We use a Gaussian kernel for $K(u)$, and Silverman's rule of thumb for h , the bandwidth (Silverman, 1986). The instrument for treatment is missing the waitlist cutoff. The 2SLS estimate of τ_y is a consistent estimate of the density of $Y_{ist}(1)$, evaluated at y . Likewise, by substituting $Enroll_{ist} = 1 - NotEnroll_{ist}$ in equation (2.4), the equivalent of the 2SLS coefficient, τ_y , is a consistent estimate of the density of $Y_{ist}(0)$ evaluated at y . Densities are evaluated on a grid of 100 points.¹⁵

Figure 2.8a shows the complier densities for the number of courses a student is enrolled in after the add/drop date. Figure 2.8b shows the densities for the time it takes students to earn an associate degree, certificate, or bachelor's degree. For ease of interpretation, students who do not earn a degree within five years are coded as receiving a degree in six.

The orange dashed line represents the density for compliers who missed the cutoff; these students are shut out of their desired section. The blue solid line shows the estimated density for compliers who do not miss a cutoff; these students represent the counterfactual, business as usual for students who are not rationed out of the section they want. They are enrolled during the advanced registration period. There is a shift to the left in the distribution of the number of courses a student takes for students who get shut out of a course, though a small minority does seem to respond by taking even more courses, perhaps to compensate. A heterogeneous response would make it more difficult to detect an average impact on the share of students taking a full course-load, which is demonstrated by the vertical lines representing average number of courses. These are basically superimposed.

The plot for time to degree reveals that very few compliers earn any type of degree. While the average differences are too small to detect, the potential outcome densities do reveal more nuance. There is slightly less mass at four years, and slightly more mass at five and six for students shut out of a course, which means a small share of compliers may take longer to earn a degree or not earn a degree after being shut out of a course. While the magnitudes are small and not statistically detectable, this is suggestive that further investigation is necessary on long-term outcomes.

¹⁵For more examples and discussion of estimating complier densities, see Angrist et al. (2016); Walters (ming).

2.5 Conclusion

This paper studies the effect of course scarcity in a setting with open access, high enrollment and budget shortfalls. The analysis measures course scarcity by using cutoffs in waitlist queues which discontinuously change the probability of enrolling in a desired section. Comparing students who just miss the waitlist cutoff to those who just make it, the study finds that students who are not able to enroll in their preferred section due to oversubscription are 2.6 percentage points less likely to take any courses that term. At the same time, missing a waitlist cutoff causes a corresponding 3.6 percentage point increase in the share of students who transfer to other two-year schools within two years, in some cases enrolling in both schools simultaneously. This could signal substitution behavior to try to earn the credits associated with the waitlisted course. These effects are large relative to the control complier means. 10.4% of control compliers dropped out in the waitlisted semester and 10.5% transferred to another two year within two years. Therefore, the results represent a 25% increase in same-term dropout and a 34% increase in transfers to other two-year colleges.

The results of our study contrast with earlier work that suggests course scarcity in college does not have downstream effects on student outcomes (Kurlaender et al., 2014; Neering, 2018). One likely reason for this contrast is that there are important institutional differences between community colleges and public four-year universities, the settings studied in earlier work. For example, community colleges have open enrollment policies, binding class size constraints, lower tuition rates, and heavy reliance on state funding. Moreover, underfunded community colleges are not unique to California; 46 states spent less per-student in 2016 than they did before the 2008 recession (Mitchell et al., 2016). In light of sustained decreases in per-student funding for public colleges, future work should continue to explore the effects of course scarcity at the institution level.

In addition, we estimate the effect of missing a waitlist cutoff *holding availability in all other sections fixed*. This could be considered a small friction; the response to a scenario in which a large fraction of sections are eliminated at once may be very different and presumably more severe. Likewise, students often face more than one waitlist during their college careers. For example, 81% of students in our sample sign up for more than one waitlist. In this sense, we present a lower bound on the cumulative impact of missing multiple waitlists. The evidence of short-term behavior change is at least consistent with Bound et al. (2010) and Deming and Walters (2017), which find aggregate impacts of decreases in funding per student.

While we find no average impacts of course rationing on transfers to four-year schools or

bachelor's degree attainment, there is evidence of diverging impacts by ethnicity. For Asian students, facing rationing leads to an accelerated rate of transfer to a four-year college. Underrepresented minority students are more likely to continue in other two-year schools and if anything, become less likely to transfer to a four-year as time goes on. White students seem to delay their transfer to a four year. These patterns show up again in bachelor's degree completion, with Asian students reacting to rationing by earning a bachelor's degree sooner than they otherwise would have, and White students earning their degree later. URM students are earning bachelor's degrees at such a low rate within five years of the waitlist that they exhibit a floor-effect- they can't do any worse. Anecdotal evidence suggests that there are potentially two streams of students using the community college as a vehicle to access four-year schools.

The first type of student can not access a four-year initially, and uses the community college to build their skills in a stepping-stone fashion. This represents the traditional picture of how community colleges are thought to function. However, there could be a group of very positively selected students who actually could have enrolled in a four-year school initially, but instead choose to start in a two-year setting. This could be because they can complete their core courses at a lower tuition rate or because it may be less competitive to access a selective University of California campus by transferring from a two-year rather than applying directly out of high school. Whatever the case, a positively selected student who faces rationing may become frustrated with the resource constraints of a two-year setting and abandon their initial plans to start in a community college, leading them to transfer to a four-year sooner. One hypothesis is that ethnicity serves as a rough proxy for student resources and ability to navigate the higher education system. Finding differential responses is consistent with prior literature that worries about diverting students from selective four-year schools to two-year schools or less selective four-year schools by heavily subsidizing these options (Cohodes and Goodman, 2014).

In summary, this paper provides evidence of the impact of course shutouts on educational attainment, a mechanism that was previously untestable due to data limitations. It also introduces a new method for leveraging waitlist registration logs, a data resource that has been underused to perform causal inference. Finally, this study continues the work of documenting and quantifying the effects of higher education funding and specifically funding for community colleges, which disproportionately serve low-income students and students of color. In the face of unequal access to educational resources, it is extremely important to understand the exact processes through which money influences student outcomes in order to create effective solutions.

Figure 2.1: A Hypothetical Registration Log

P_i	action	date	time	X_i	D_i	RV_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
36	enroll	Aug 1, 2004	11:00:00	-	-	-
37	enroll	Aug 1, 2004	12:00:00	-	-	-
38	enroll	Aug 1, 2004	13:00:00	-	-	-
39	enroll	Aug 1, 2004	14:00:00	-	-	-
40	enroll	Aug 1, 2004	15:00:00	-	-	-
38	drop	Aug 2, 2004	8:00:00	-	-	-
41	enroll	Aug 2, 2004	10:00:00	-	-	-
42	waitlist	Aug 2, 2004	12:00:00	1	2	-1
43	waitlist	Aug 2, 2004	13:00:00	2	2	0
44	waitlist	Aug 2, 2004	14:00:00	3	2	1
7	drop	Aug 3, 2004	20:00:00	-	-	-
42	enroll	Aug 3, 2004	21:00:00	-	-	-
22	drop	Aug 4, 2004	9:00:00	-	-	-
43	enroll	Aug 4, 2004	11:00:00	-	-	-
44	drop	Aug 4, 2004	15:00:00	-	-	-
45	waitlist	Aug 4, 2004	17:00:00	1	0	1

Notes. P_i is a student identifier, X_i is the initial waitlist position, D_i counts the number of students who signed up before student i signed up for the waitlist, and dropped after student i (as long as it was during the registration period). $RV_i = X_i - D_i$ is student i 's running variable.

Figure 2.2: Density of the Running Variable

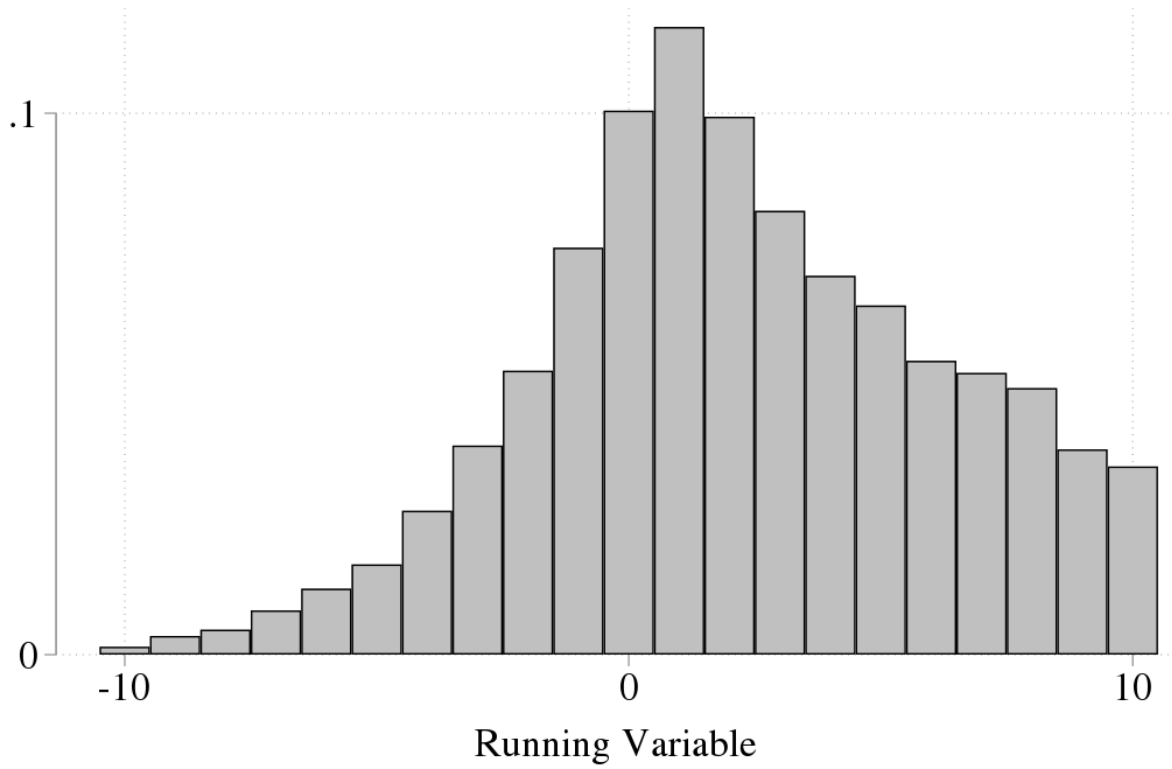
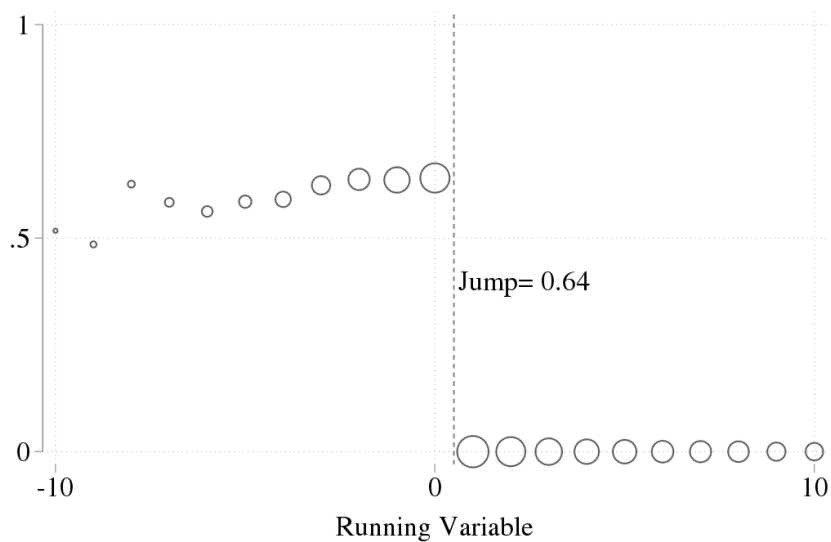
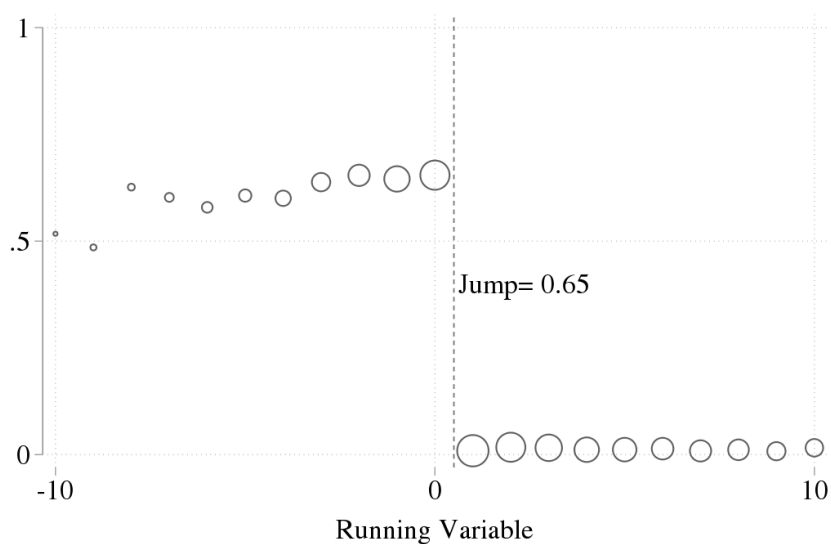


Figure 2.3: First Stage Effect of Missing the Waitlist Cutoff on Enrollment in Waitlisted Section and Course

(a) Enrolled in Waitlisted Section

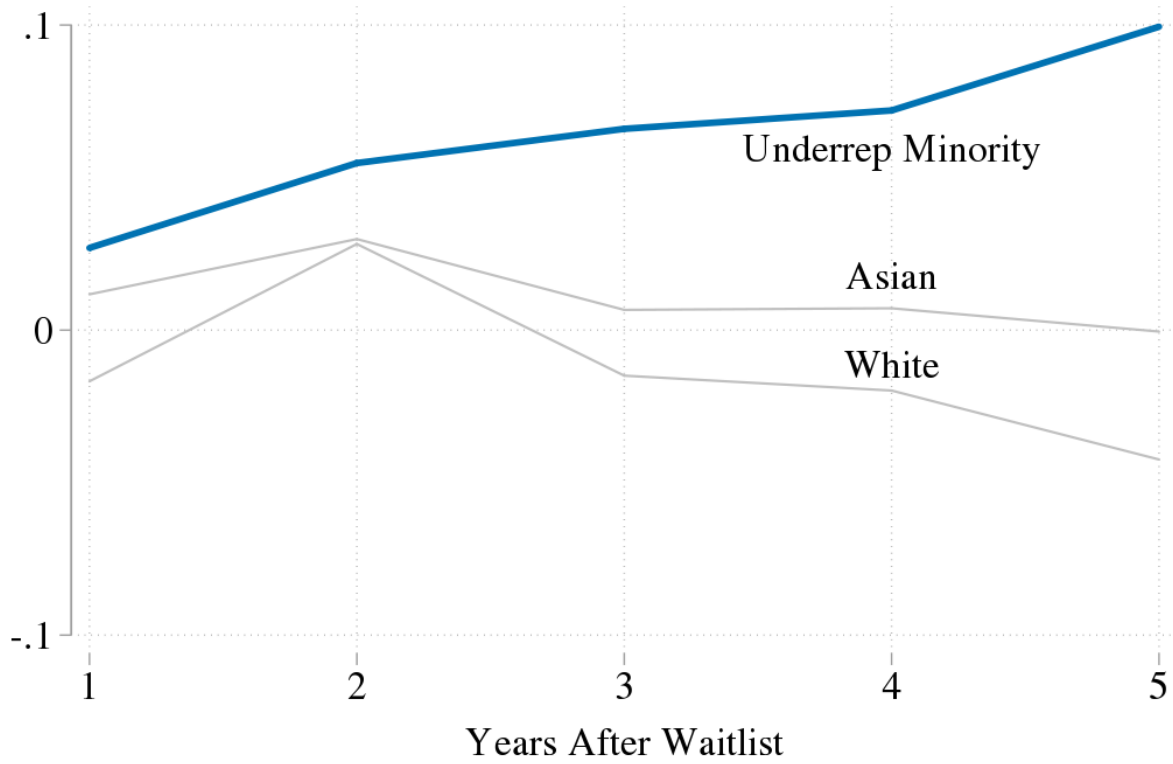


(b) Enrolled in Waitlisted Course



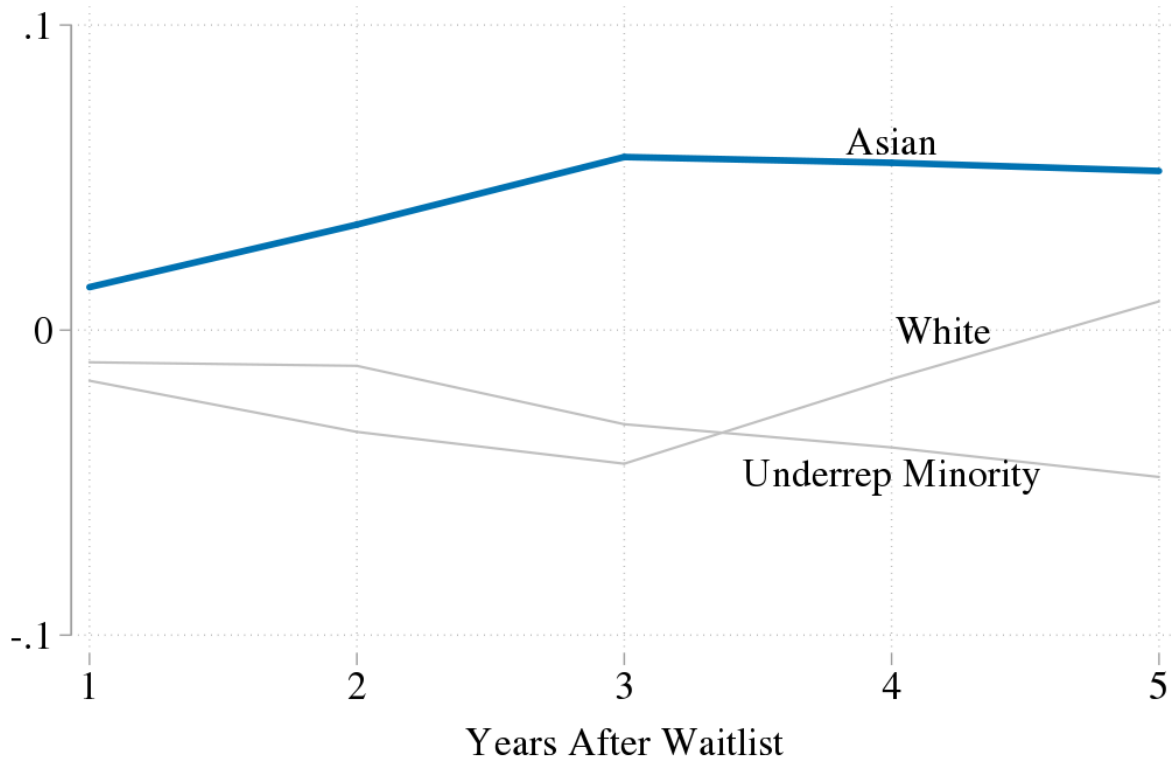
Notes. Each dot represents enrollment binned by the value of the running variable, where enrollment is equal to one if the student was enrolled in the section or course at the end of the advanced registration period. Both section and course enrollment are equal to zero for students with a running variable greater than zero by construction. The size of the dot reflects the number of observations in each bin.

Figure 2.4: Effect of Missing the Waitlist Cutoff on Transfers to Other Two-Year Schools, by Ethnicity



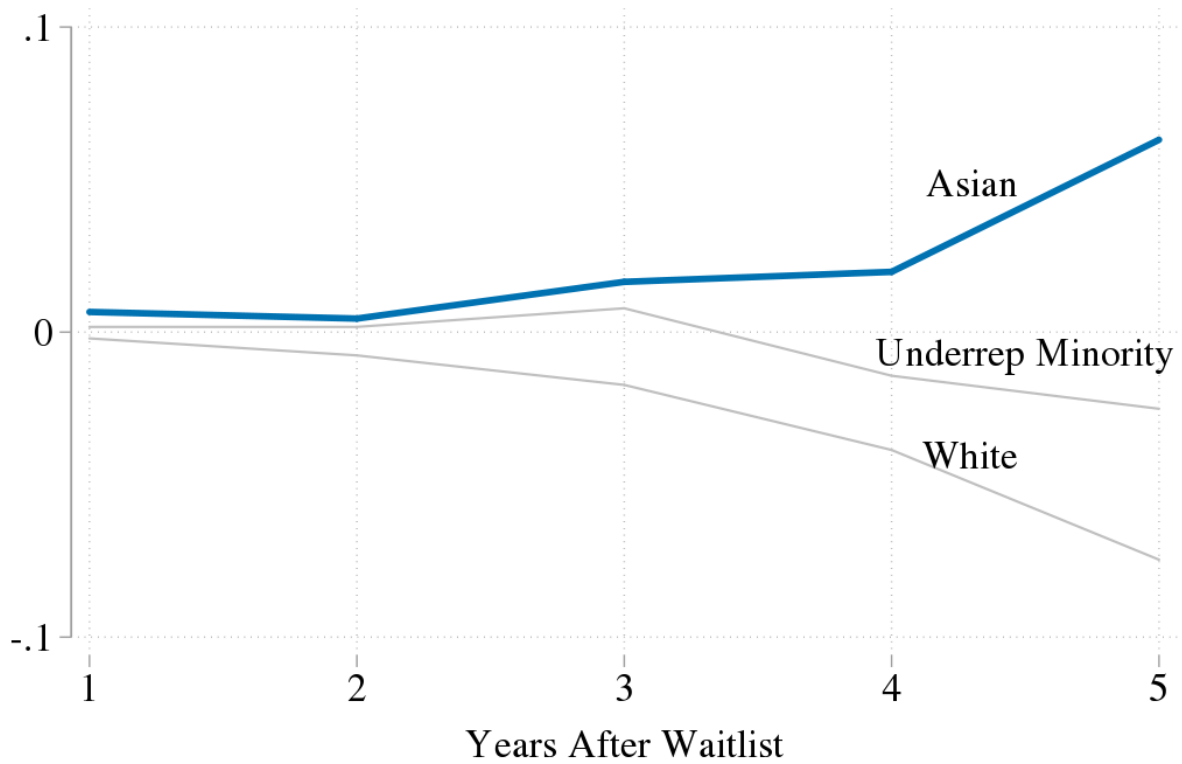
Notes. This figure shows results from a 2SLS regression as in equation 3, where effects are estimated separately by ethnicity. The outcomes are indicators for transfers to other two-year schools at different time horizons: within one through five years of the waitlisted term. All specifications include the covariates listed in Table 2.4. The exact point estimates and standard errors are reported in Table B.7.

Figure 2.5: Effect of Missing the Waitlist Cutoff on Transfers to Four-Year Schools, by Ethnicity



Notes. This figure shows results from a 2SLS regression as in equation 3, where effects are estimated separately by ethnicity. The outcomes are indicators for transfers to four-year schools at different time horizons: within one through five years of the waitlisted term. All specifications include the covariates listed in Table 2.4. The exact point estimates and standard errors are reported in Table B.8.

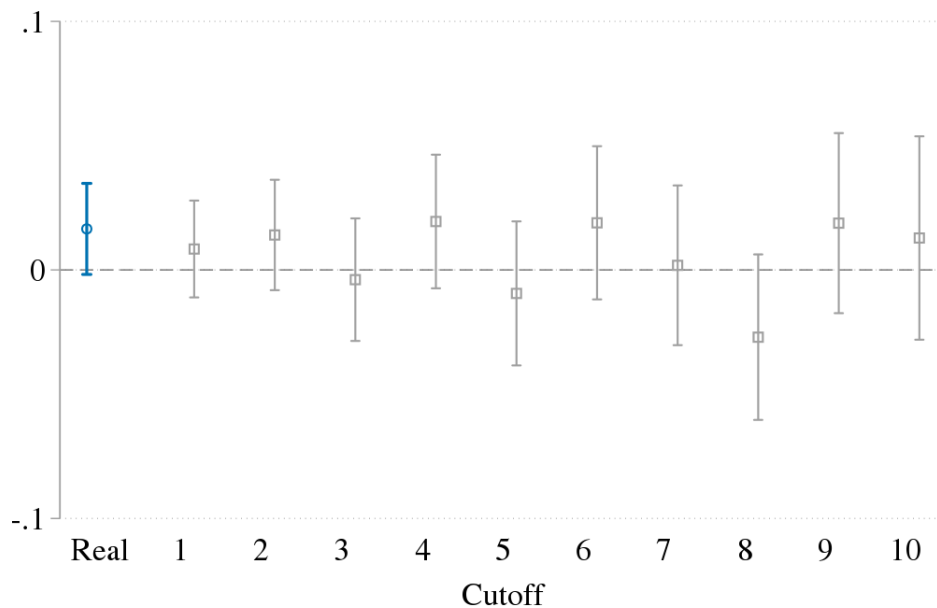
Figure 2.6: Effect of Missing the Waitlist Cutoff on Bachelors Degree Completion, by Ethnicity



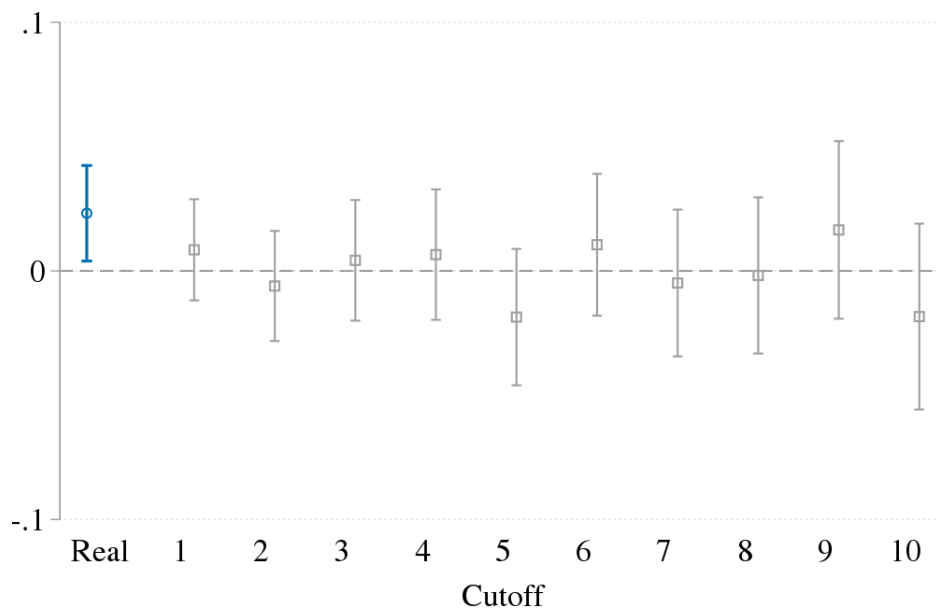
Notes. This figure shows results from a 2SLS regression as in equation 3, where effects are estimated separately by ethnicity. The outcomes are indicators for bachelors degree completion at different time horizons: within one through five years of the waitlisted term. All specifications include the covariates listed in Table 2.4. The exact point estimates and standard errors are reported in Table B.9.

Figure 2.7: Reduced Form Effect of Missing a Placebo Cutoff

(a) Enrolled in Zero Courses

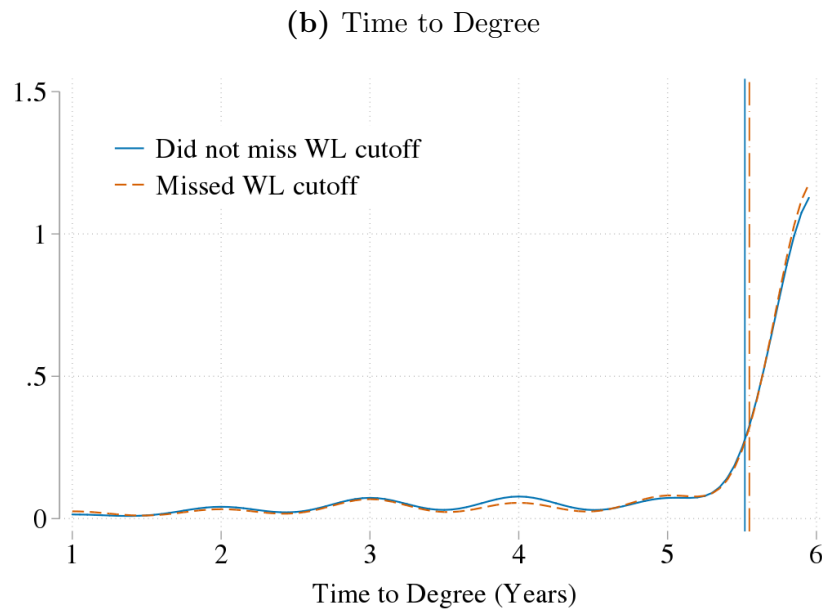
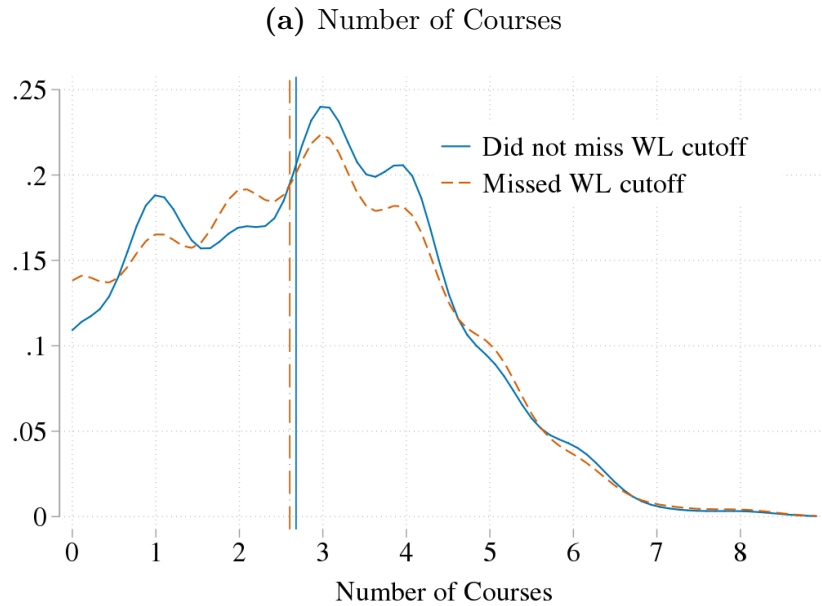


(b) Transferred to Another Two-Year School within Two Years



Notes. This figure plots point estimates and confidence intervals for the reduced form effect of missing a placebo cutoff using a bandwidth of one. For example, at the placebo cutoff of three, we show the effect of having a running variable of three relative to a running variable of four. The actual reduced form effect- where the cutoff is equal to zero- is shown in blue, whereas the other values on the x-axis represent the effects of the other placebo cutoffs.

Figure 2.8: Density of Potential Outcomes for Treated and Untreated Compliers



Notes. This figure plots estimates of the potential outcome densities for treated and untreated compliers. Treated compliers missed the waitlist cutoff and did not enroll in their desired section, and untreated compliers did not miss the cutoff and therefore enrolled in their desired section. The vertical lines represent the average outcomes for each group. Number of courses is defined as the number of courses a student was enrolled in after the add/drop date. Time to degree measures the number of years from the waitlisted term until a student earned any higher education degree, including associates, certificates, or bachelors degrees. Time to degree is equal to six for students who either take six years to complete a degree, or do not complete a degree within six years of the waitlisted term.

Table 2.1: Summary Statistics*Panel A: Section-level statistics*

	All Sections (1)	Analysis Sections (2)
% with a WL	0.49	1.00
% STEM with WL	0.68	0.34
% Arts/Humanities with WL	0.50	0.28
% Social Sciences with WL	0.60	0.12
% Other with WL	0.30	0.26
WL Length	8.98	8.01
WL Length (SD)	9.15	7.07
Observations	29,614	3,499

Panel B: Student-level statistics

	CA 2-year Public Colleges (1)	All De Anza (2)	Analysis Sample (3)
Female	0.53	0.52	0.51
Asian	0.13	0.40	0.44
White	0.27	0.26	0.24
Hispanic	0.45	0.14	0.16
Black	0.07	0.05	0.05
Ever Receives Aid	0.59	0.17	0.32
Age Under 25	0.63	0.59	0.80
Age 25 and Over	0.37	0.41	0.20
# Courses, first term		1.70	1.81
# Waitlists, first term		0.42	1.01
Observations	1,234,509	179,596	4,258

Notes: Panel A presents section-level statistics for De Anza Community College between Fall 2002 and Summer 2010. Column (1) reports the average share of sections with waitlists, by subject and before sample restrictions. For all sections in the analysis, column (2) reports the share in each subject. By definition, all sections in the analysis have a waitlist. The STEM definition follows the National Science Foundation. Waitlist length measures how many students remain on the waitlist at the end of the registration period for oversubscribed sections. In Panel B, column (1) describes student characteristics at all two-year colleges in the California, column (2) shows characteristics for De Anza students, and column (3) reports statistics for the students in the analysis (sample restrictions are detailed in Section 2.2.3). Data for all two-year public colleges in CA comes from IPEDS for Fall 2014, except for financial aid receipt which is from the 2014-2015 school year. In column (1), financial aid receipt and age represent a cross section of all undergraduates at public 2-year schools in CA. In columns (2) and (3), a student is counted as receiving aid if they received it at any time in the sample period and age represents their age in their first term in the sample period. The number of courses is the number a student was enrolled in after the drop date in the first observed term. The number of waitlists is the total that a student signed up for during the advanced registration period in the student's first observed term.

Table 2.2: Frandsen Manipulation Test for Discrete Running Variables

Nonlinearity Parameter (k) (1)	P-value (2)
0.005	0.078
0.010	0.091
0.015	0.115
0.020	0.148
0.025	0.185
0.030	0.234
0.035	0.291
0.040	0.350
0.045	0.416
0.050	0.483

Notes: This table presents results from the manipulation test proposed in (Frandsen, 2017). The parameter k , which is chosen by the researcher, represents the “maximal degree of nonlinearity in the probability mass function that is still considered to be compatible with no manipulation” (Frandsen, 2017). Column (1) reports tested values of k and Column (2) reports the p-value of a test of the null hypothesis that no manipulation occurred.

Table 2.3: Test for Balance of Pre-determined Student Characteristics Across the Waitlist Cutoff

	Coefficient (1)	Standard Error (2)	P-Value (3)
White	-0.019	0.013	0.142
Asian	0.004	0.015	0.807
Hispanic	0.018	0.011	0.103
Black	0.001	0.007	0.883
Other Race	0.007	0.007	0.290
Missing Race	-0.011	0.008	0.184
Female	-0.017	0.015	0.257
Missing Gender	0.002	0.001	0.165
Age	0.351	0.199	0.078
Missing Age	-0.001	0.001	0.315
International Student	0.004	0.014	0.792
Received Financial Aid	-0.012	0.014	0.396
Missing Financial Aid Receipt	0.001	0.002	0.775
First Time Student	0.001	0.003	0.751
Joint p-value			0.242
Observations (N_l/N_r)	1,977	2,281	

Notes: Each row reports results from a linear regression of the covariate on an indicator for missing a waitlist cutoff, term by year fixed effects, registration priority fixed effects, and indicators for special student categories. The sample includes students within one position of the waitlist cutoff. The first column shows coefficients, the second column shows the robust standard error, and the third column shows the p-value. The p-value in the last row is from an F test of whether the differences in each characteristic are jointly significant, conditional on the fixed effects and special student categories previously listed.

Table 2.4: First Stage Effect of Missing the Waitlist Cutoff on Enrollment in Waitlisted Section and Course

	(1)	(2)
<i>Panel A: Section Enrollment</i>		
Missed WL Cutoff	-0.641*** (0.011)	-0.644*** (0.011)
R-squared	0.489	0.499
F-Statistic	3526	3583
Controls	N	Y
Control Mean	0.641	0.641
<i>Panel B: Course Enrollment</i>		
Missed WL Cutoff	-0.645*** (0.011)	-0.648*** (0.011)
R-squared	0.485	0.494
F-Stat	3516	3555
Controls	N	Y
Control Mean	0.655	0.655
Observations (N_t/N_r)	1,977/2,281	

Notes: Results are from a linear regression where the dependent variable is enrollment in the waitlisted section in Panel A and enrollment in the waitlisted course in Panel B, where enrollment is equal to one if the student was enrolled at the end of the advanced registration period. All students are within one running variable position from the cutoff. The first column does not include controls while the second controls for race/ethnicity, gender, age, citizenship, financial aid receipt, first time students, special student status, special program status, registration priority fixed effects, term by year fixed effects, and indicators for missing variables. The control mean is the mean of the dependent variable for students with a running variable of zero. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table 2.5: Effect of Missing the Waitlist Cutoff on Course Load and Persistence

	# Courses Enrolled in Concurrent Term			Enrolled
	Zero (1)	One or Two (2)	Three or More (3)	Next Term (4)
2SLS	0.026* (0.014)	-0.008 (0.021)	-0.017 (0.022)	-0.019 (0.021)
Reduced Form	0.016* (0.009)	-0.005 (0.014)	-0.011 (0.014)	-0.012 (0.013)
CCM	0.104	0.337	0.559	0.688
Observations (N_l/N_r)	1,977	2,281		

Notes: This table shows results from a 2SLS regression as in equation 3. The outcome is an indicator for whether the student took no courses in the concurrent term in Column (1), took one or two courses in Column (2), or took three or more courses in Column (3). A course is counted if the student is enrolled after the add/drop date. The outcome in column (4) is an indicator for whether the student enrolls in any classes the following major term. The standard errors are in parentheses, with the control complier means (CCM) and the reduced form displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table 2.6: Effect of Missing the Waitlist Cutoff on Transfers and Degree Completion

Outcome	Within 1 Year (1)	Within 2 Years (2)	Within 3 Years (3)	Within 4 Years (4)	Within 5 Years (5)
Transfer Other Two-Year	0.009 (0.011)	0.036** (0.015)	0.026 (0.017)	0.026 (0.019)	0.027 (0.020)
CCM	[0.056]	[0.105]	[0.152]	[0.189]	[0.222]
Reduced Form	0.006 (0.007)	0.023** (0.010)	0.017 (0.011)	0.017 (0.012)	0.017 (0.013)
Transfer Four-Year	0.000 (0.008)	0.009 (0.012)	0.004 (0.017)	0.013 (0.019)	0.019 (0.020)
CCM	[0.030]	[0.062]	[0.141]	[0.190]	[0.219]
Reduced Form	0.000 (0.005)	0.006 (0.008)	0.003 (0.011)	0.008 (0.012)	0.013 (0.013)
Certificate/ Associate	0.003 (0.005)	-0.002 (0.009)	-0.010 (0.013)	-0.016 (0.015)	-0.011 (0.015)
CCM	[0.008]	[0.033]	[0.077]	[0.106]	[0.117]
Reduced Form	0.002 (0.003)	-0.002 (0.006)	-0.007 (0.008)	-0.011 (0.009)	-0.007 (0.010)
Bachelors	0.004 (0.003)	0.002 (0.003)	0.006 (0.006)	-0.003 (0.010)	-0.002 (0.014)
CCM	[0.002]	[0.006]	[0.014]	[0.043]	[0.094]
Reduced Form	0.003 (0.002)	0.001 (0.002)	0.004 (0.004)	-0.002 (0.007)	-0.001 (0.009)
Observations (N_l/N_r)	1,977	2,281			

Notes: This table shows results from a 2SLS regression as in equation 3. The outcomes are indicators for transferring and degree completion at different time horizons: within one through five years of the waitlisted term. Associate and certificate completion data comes from both De Anza administrative records and the National Student Clearinghouse. The standard errors are in parentheses, with the control complier means (CCM) and the reduced form displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

CHAPTER III

The Effect of Summer Employment on the Educational Attainment of Under-Resourced Youth

3.1 Introduction

Even though the U.S. economy has climbed out of the Great Recession, the labor market opportunities for many people remain a concern. The labor force participation rate among all civilians age 16 or older was 5% lower in 2016 than its pre-recession level of 66.2% in 2006, according to data from the U.S. Bureau of Labor Statistics. The recovery for young adults age 16 to 24, particularly for young men, has been slower than for the population overall; the labor force participation rate for these groups declined by 9% and 11% respectively over the same timeframe. As policymakers seek to improve work opportunities for youth, there has been a growing interest in alternative pathways for them to obtain career skills.

Summer youth employment programs are a popular way for municipalities to provide adolescents with skills and experiences thought to improve labor market outcomes. The first such program began in the United States over 50 years ago in 1964, and as of 2016, at least 42 cities across the country offered summer jobs to over 115,000 youths.¹ New York City and Chicago operate the largest programs, providing jobs to about 50,000 and 24,000 youths respectively, though many smaller cities operate them as well, including Tuscaloosa, AL, Charlottesville, VA and Madison, WI (Dollarwise, 2016).

These programs are arguably more important in today's economy than ever before, as recent work has highlighted the increasing returns to non-cognitive skills in the labor force (Deming, 2017). Working during the summer offers youth the opportunity to interact with professionals who can teach them valuable interpersonal skills. Also, many summer youth employment programs include an explicit curriculum designed to teach a variety of work

¹There is no exact count of how many local governments organize summer youth employment programs as these programs are decentralized (Belotti et al., 2010).

readiness and life skills, ranging from how to prepare a resume to the benefits of establishing a bank account. Finally, some have a particular focus on career and technical education (CTE), which is receiving increased attention from many cities and states (Alfeld, 2016; Jacob, 2017). In this sense, summer youth employment programs are similar in spirit to school-based CTE programs, in which schools offer students internships or connections to employers, in that both are part of the education for career readiness.

There are three main purposes of summer youth employment programs. First, they intend to give young adults the opportunity to earn an income during the summer. Second, they aim to reduce crime during the summer by keeping youth ‘off the streets.’ Finally, they serve to improve labor market outcomes beyond the summer.

While research evidence on such programs has grown in recent years, it is still limited. In particular, it is not clear how, if at all, participation influences key educational outcomes such as high school graduation. This is a critical gap as education is one of the strongest predictors of labor market success. Participation might increase educational attainment by increasing the opportunity cost of dropping out of high school or by increasing academic engagement as a result of improved soft skills. On the other hand, it may decrease educational attainment if youth have an unsatisfying work experience. Using detailed administrative data, this paper analyzes a more extensive set of educational outcomes than previous studies, including continued enrollment in school, attendance rates, test scores, high school graduation, and college enrollment.

In this paper, we study how participation in a summer youth employment program is associated with educational outcomes. Specifically, we study the program in Detroit, Michigan using a selection on observables identification strategy, comparing youth who participated in the program to those who applied but did not participate. The main threat to our identification is that participants differ from applicants who did not participate along unobservable dimensions that are related to educational outcomes. To address this, we first match participants to applicants who did not participate that were of the same race and gender and were in the same grade in the same school in the same year. Participants look quite similar to their matched peers along a variety of other observable dimensions.

We also control for information related to the selection process in our analysis. While details vary across job placements, program administrators select participants using information from the application, from brief conversations at a career fair, or, if applicable, from an existing relationship with the applicant. In our analysis, we control for all of the information from the application, including gender, race and a second-order polynomial of age. We also control for some things that administrators may be able to infer from the application, such as characteristics of the neighborhood where the applicant lives. Lastly, we control for

information that was unobservable to program staff, such as prior attendance records and test scores. Specifically, we control for prior school attendance, special education status, limited English proficiency, eligibility for free or reduced-price lunch, ever retained in grade as well as prior standardized math and reading test scores, and the interaction of the two.

We perform falsification exercises and do not find that participation predicts pre-program characteristics. For the 99.9% of youth who completed 6th grade before applying, we test whether participation predicted educational performance in 6th grade. These measures could not have been caused by participation in the summer program. We condition on all of the covariates listed above, including baseline attendance and test scores, and do not find that participation predicts sixth-grade attendance or test scores.

We find that participation is associated with a modest increase in educational attainment. Specifically, it increased the likelihood of enrolling in public school after the program by 1.5% and of graduating from high school by 4%. We perform the robustness checks proposed in Oster (2016) and find that, consistent with the falsification exercises, these results are not driven by selection bias from unobservable characteristics that are correlated with observables. Specifically, we can bound the effect on public school enrollment between 0.8% and 1.5% and on high school graduation between 3.4% and 4%, assuming that selection on unobservables was not larger than selection on observables.

Youth with the weakest academic skills benefited the most, as participation increased school enrollment and high school graduation by 2.2% and 5.5% respectively for this group. Our findings suggest that summer youth employment is an important complement to in-school work-based learning programs, such as career technical education and school-employer partnerships. A youth's first experience in the labor force may have a significant impact on future education and career aspirations.

3.2 Prior Literature

Until recently, research on summer youth employment programs was limited to descriptive analyses of program implementation or short-run participant outcomes. To the best of our knowledge, there are only seven studies which credibly estimate causal effects of these programs — four that study the same New York City program (Leos-Urbel, 2014; Schwartz et al., 2015; Gelber et al., 2016; Valentine et al., 2017), two that study the same Chicago program (Heller, 2014; Davis and Heller, 2017) and one that focuses on the program in Boston (Modestino, 2017). The NYC and Boston studies utilize a quasi-experimental lottery design that compares youth who were provided an opportunity to participate via a random admissions lottery to those who applied but did not receive an offer to participate. The

Chicago study analyzes two randomized controlled trials, comparing young adults who were randomly assigned to participate in the program to those who were assigned to a control group.

Several consistent findings stand out. First, participation in summer employment programs seems to be associated with a reduction in crime. In New York City, Gelber et al. (2016) find that participation reduced the likelihood of incarceration up to 9 years after the program by 10% relative to a comparison mean of 1%. In Chicago, the program reduced violent-crime arrests by 35% during the summer and the following school year (Heller, 2014; Davis and Heller, 2017). In Boston, participation reduced the number of arraignments for violent and property crime in the following year by 35% and 57% respectively, with the largest impacts for African-American and Hispanic males (Modestino, 2017).

Second, participation in the programs does not seem to have a meaningful impact on employment or earnings. Researchers followed NYC participants for up to 7 years and do not find that participation had a long-run effect on labor market outcomes. They do find a small increase in the likelihood of having a job in the first two follow-up years, accompanied by a small decline in wages, which they associate with a greater likelihood of working in the public sector (Gelber et al., 2016; Valentine et al., 2017). However, the studies of the Boston and Chicago programs follow youth for one and two years respectively and did not find an effect on employment or earnings over this timeframe. (Davis and Heller, 2017; Modestino, 2017).

It is decidedly less clear how, if at all, participation in a summer youth employment program influences educational outcomes. An early study in NYC of the 2007 cohort found that participation increased attendance in the following school year by 1.7%, driven by youth age 16 and older who had low baseline attendance rates (Leos-Urbel, 2014). However, a study that included a broader sample of five cohorts of the program from 2006 to 2010 found a precise zero effect on attendance (Valentine et al., 2017). Moreover, studies following NYC participants over a longer time period find precise null effects on high school graduation, college enrollment or degree completion (Gelber et al., 2016; Valentine et al., 2017).²

Modestino (2017) finds that participation in Boston's summer youth employment program increased attendance rates in the following school year by 3% relative to a baseline of 87%. This was driven by larger increases for Hispanic youths, males and those older than 16. Studying the Chicago program, however, Davis and Heller (2017) find that participation did not affect daily attendance or grade point average the following school year. It also did

²Schwartz et al. (2015) studies the test-taking and performance of four cohorts of applicants between 2005 and 2008 and find that participation did not increase either the number of Regents exams or the number of exams passed among first-time applicants. Youth who participated in the program for more than one summer do benefit, though.

not affect persistence in school, defined as continued enrollment or high school graduation, within 3 years. However, the authors do find that relatively more advantaged youth, those with the highest pre-program school attendance and grades, do benefit from participation. Specifically, participation increased school persistence by 13% for those in the top quartile of predicted employment impact relative to a baseline mean of 60%, which was statistically significant at the 10% level.

Our paper makes several contributions to the literature. First, we use more detailed education data than previously available, allowing us to look at a broader set of educational outcomes altogether than any other single study. Second, in an effort to understand contradictory findings in the prior literature, we provide some additional evidence as to which subgroups benefit the most from summer employment programs. Finally, our study is the first to evaluate the summer youth employment program in Detroit, Michigan. While it is similar to those in NYC and Boston in terms of implementation, the program in Detroit serves a less-resourced population than these other cities.³ For example, according to the 2016 American Community Survey, 51% of children in Detroit live in poverty, compared to 27% in NYC, 31% in Boston and 28% in Chicago. Furthermore, 6th graders in Detroit score 2.3 grade levels below the national average on standardized tests, relative to 0.3 grade levels in NYC and Boston and one grade level in Chicago (Reardon et al., 2017). Therefore, youth in Detroit may benefit from having a summer job more than those in other cities.

3.3 Institutional Background

The current iteration of Detroit’s summer youth employment program began in 2009 with federal stimulus funding provided at the onset of the Great Recession. It has provided summer jobs to over 15,000 youths between 2015 and 2017. The program, commonly known as Grow Detroit’s Young Talent or GDYT, employs young adults for 20 hours per week for six weeks from July through August, at between \$8 and \$9.50 per hour depending on age and job type. Youth selected for the program receive 24 hours of work readiness training before and during their employment.

All Detroit residents between the ages of 14 to 24 are eligible to apply for the program. The application period begins in February and is open for five weeks. It is widely advertised in the city as the Mayor’s office holds a kickoff event and works with schools, religious leaders and community organizations to recruit a pool of applicants. About 10% of the eligible

³The program in Chicago has more of an emphasis on a social-emotional learning curriculum than programs in other cities. Participants in Chicago spent up to 40% of their hours engaged with the curriculum. In contrast, participants in NYC, Boston and Detroit spent about 10-20% of their hours in programming about work readiness.

population applies for the program.⁴ The application is very simple; in it, youth provide basic demographic information, indicate whether they have any past work experience and specify their interest in different industries. The application does not include a resume or personal statement and does not ask for details about past job responsibilities or school achievement.

GDYT consists of four sub-programs, each of which has a distinct selection process and involves a different summer experience.⁵ Although there may be meaningful differences in how each influences educational outcomes, we can not explore heterogeneity across sub-programs in our analysis. Except for the Junior Police and Fire Cadets program, we do not have data on which particular sub-program youth participated in.

3.3.1 Junior Police and Fire Cadets

The Junior Police and Fire Cadets program (JPC) is reserved for 14 to 15-year-olds and is structured to provide a first work experience for youth. Working with JPC includes a variety of community service activities, including providing support and companionship to seniors and cleaning parks and other neighborhood commons. College students serve as day-to-day supervisors for the young adults, and police officers interact with the youth in various capacities throughout the summer. JPC employees make up 20 percent of GDYT participants.

According to our conversations with program staff, youth are not purposely chosen for this program. Instead, staff pick a (somewhat) random set of 14- and 15-year-old applicants to GDYT and invite them to participate. When we compare pre-program characteristics of JPC participants to age-eligible youth who applied but did not participate in GDYT, we find they are quite similar, though not identical. As reported in Table C.1, there were no statistically significant differences in academic performance, including attendance and test scores, between these groups, yet JPC participants lived in lower-income neighborhoods.

3.3.2 Community-Based Organizations (CBOs)

The largest group of GDYT youth — roughly 60 percent — work for one of many non-profit community-based organizations (CBO) in Detroit, similar to the employment models in NYC and Chicago. The work covers a wide range of activities, from camp counselor for younger children to clerical office support to community beautification. These organizations recruit applicants on their own, frequently selecting youth with whom they

⁴According to authors' calculations from the 2015 ACS population estimates.

⁵Youth apply to participate in GDYT broadly, and not to a specific sub-program.

have a long-term relationship (e.g., those who participated in programming during the academic year with the CBO, or were known to staff in the neighborhood).⁶ CBO participants typically range in age from 15-18 years old.

3.3.3 Industry Led Training and Career Pathways Programs

The final two opportunities for employment are through Industry Led Training (ILT) and Career Pathways internships. The ILT program is a new and relatively smaller component for applicants at least 16 years old, with about 10% of GDYT youths participating each summer. It consists of work-based training programs in high-growth sectors (e.g., hospitality, child care, IT, advanced manufacturing, healthcare), which are typically run by non-profit organizations in Detroit. The Career Pathways program provides an internship for older applicants, usually at least 19 years of age, at a variety of major private sector employers in the city, including Detroit Manufacturing Systems, Touchpoint Support Services, and Wayne State University. To be eligible for an ILT program or Career Pathways internship, applicants must have had some prior work experience, and be referred by some other organization. Eligible youths are invited to attend a career fair to meet the employers, and then employers select who they want to hire.⁷ Given this selection process, we assume that applicants who participate in the ILT and Career Pathways programs are likely to be quite different than those who do not.

3.4 Data and Sample

This study uses administrative records from both GDYT, the Michigan Department of Education (MDE) and the Center for Educational Performance and Information (CEPI). From GDYT, we have application and participation data from the program for the summer of 2015. The application data consists of all applications which were started during the submission window between February and March of 2015, regardless of whether they were completed.⁸ We define applicants as those who completed the entire application during the submission window or, in some rare cases, as those who did not complete an application during the window yet still participated in the program.⁹ In total, there were 12,255

⁶Youth who are recruited to participate in a CBO summer program must still submit a GDYT application. Therefore, we have complete data on these applicants.

⁷GDYT staff assigns any remaining spots to other eligible youths.

⁸Only 2.44% of applications were incomplete. 97% of these stalled after the first step of the application, which only asked for first name, last name, and email address. These incomplete applications do not contain sufficient information to match youth to the education data, so we drop them from our analysis.

⁹Youth who are recruited to participate in a CBO summer program can apply after the submission window ends and still work with GDYT in the summer. In total, 0.79% of completed applications were

applicants in 2015.¹⁰

We matched the application data to payroll data using exact matches on first and last name. The payroll data was maintained by multiple organizations that managed records for the youth who worked during the summer. Some of the payroll systems were unreliable and crashed repeatedly over the course of the summer. Therefore, we are missing important information about job placement, hours worked and earnings. We consider any youth who appeared in a payroll system to have participated with GDYT.¹¹ In total, 2,807 youths worked in 2015.¹²

We matched the GDYT data to administrative records from MDE and CEPI consisting of the universe of K-12 public school students in the state from 2003 to 2017. We used a quasi-probabilistic matching algorithm based on first name, middle initial, last name, suffix, date of birth and gender. We successfully matched 94% of applicants to the education data, and exclude youth who did not match from our main analysis.¹³ The education records contain information on school enrollment, attendance, test scores, and graduation. They are linked to data from the National Student Clearinghouse, which provides information on college enrollment. We analyze the effect of participation in GDYT on a variety of educational outcomes, including K-12 enrollment, attendance, expulsions, taking the SAT, SAT score, high school graduation and college enrollment.

Table 3.1 shows summary statistics for four groups. Column 1 shows characteristics for all Detroit residents in 8th to 12th grade who were enrolled in a public school during the 2014-2015 school year, while column 2 focuses only on the youth who applied for GDYT. Comparing column 1 to column 2 provides insight about who applied for the summer program.¹⁴ Applicants were disproportionately black, as 95% of those who applied for

submitted after the submission window.

¹⁰While we do not have information on whether applicants in 2015 applied or participated in prior summers, data from a different summer indicates that about 30% of all applicants, and 42% of participants, participated in the program during the previous summer.

¹¹Although we do not have information on hours worked for the 2015 cohort, data from a different summer indicates that 12% of youth in the payroll system did not ever show up to work. If youth appeared in the payroll data but did not actually work, then this would attenuate our estimates of the effects of program participation toward zero.

¹²The number of participants reported in official GDYT reports is greater than what we report here because the GDYT counts include youth who are employed by affiliate organizations, a separate employment model that isn't captured in the available application or payroll records.

¹³Applicants may not have matched to the education data because (1) they only attended private K-12 schools in Michigan, (2) they moved to Michigan after high school, or (3) the combination of their name, birthdate and gender did not provide enough information to identify a match. Of the 6% of applicants who did not match to the education data, about 80% were 18 years old or younger when they applied, suggesting that the non-matches were not driven by (2). We cannot disentangle explanations (1) and (3) with our data.

¹⁴The distribution of age and educational status between these groups differs by construction, as column 1 is restricted to middle and high school students while column 2 includes all applicants, regardless of their age or school enrollment status.

the program were black compared to 86% of Detroit youths. Females were also more likely to apply. To draw comparisons between student needs in school as well as the neighborhoods where they live, we measure applicant characteristics in the most recent pre-program year that they were enrolled in a public K-12 school. Although applicants and Detroit residents overall were similar in terms of eligibility for free or reduced-price lunch, there are notable differences in terms of where they live and how they performed on standardized tests. Applicants lived in relatively more advantaged neighborhoods; 13% of adults had a Bachelors's degree and 33% of households earned below the poverty line, compared to 12% and 35% respectively for all Detroit youths. Similarly, applicants were more likely to take standardized tests in 8th grade, and those who took them scored higher. Overall, this comparison shows that applicants were a somewhat positively selected group of Detroit youths.

Since youth who applied for the program were different than those who did not along a variety of observable dimensions, our empirical strategy compares youth who participated in GDYT to those who applied but did not participate. Comparing columns 3 and 4 of Table 3.1 shows that these two groups look nearly identical in terms of age and academic enrollment during the 2014-2015 school year, and quite similar in terms of race/ethnicity, gender and needs in school. Participants and non-participants lived in very similar neighborhoods, where about one-third of households live below the poverty line. While participants were slightly less likely to take standardized tests in 8th grade, those who took the tests were more likely to be proficient, suggesting that they were similar in terms of academic performance overall.

3.5 Empirical Strategy

The key empirical challenge in our analysis is to control for pre-program differences between participants and other applicants that directly influence educational outcomes, and would bias our impact estimates. We use a combination of exact matching and regression adjustment to control for such differences.

We create match groups consisting of all applicants who were the same race and gender and also were in the same grade in the same school in the most recent pre-program year that they attended a public K-12 school in Michigan.¹⁵ For example, if a student was in 11th

¹⁵The most recent pre-program year an applicant attended a public K-12 school in Michigan (match year) was almost always (1) the 2014-2015 school year, or (2) the year that they graduated or dropped out of high school. 81% of applicants have match years of 2014-2015, 11% have match years of 2012-2013 or 2013-2014 and 8% have match years before the 2012-2013 school year. In some rare cases, though, the match year represents when an applicant moved out of state or began attending private school. As a result, 1% of applicants have match groups that consist of their classmates when they were in 6th grade or below, even though they applied for GDYT when they were much older. In addition, ten applicants first enrolled in a

grade during the 2014-2015 school year, then his or her match group consists of all of the other 11th grade applicants from their school who were of the same race and gender. Using this example, we define the match year as the 2014-2015 school year and the match school as the school they attended during the match year. Similarly, if a student was in 12th grade in 2012-2013 and graduated high school that year, then their match group consists of all of the other applicants who were also in 12th grade in their school in 2012-2013 of the same race and gender.

Of all participants, 81% are in match groups with at least one participant and one non-participating applicant, and thus provide variation with which we can identify effects of the program. As shown in Table 3.2, these youth differ in several ways from other applicants, which influences the generalizability, or external validity, of our analysis. Relative to the participants in a degenerate match group, those in a match group with at least one participant and non-participant were substantially younger, more likely to be black and lived in a somewhat higher poverty neighborhood.¹⁶

Because our identification is driven by within match group differences between participants and other applicants, it is useful to compare these two groups in terms of pre-program characteristics. To do so, we estimate the following model

$$X_{ij} = \alpha_0 + \alpha_1 \textit{Participated}_{ij} + \gamma_j + \nu_{ij} \tag{3.1}$$

where X_{ij} represents a baseline characteristic of youth i in match group j , and $\textit{Participated}_{ij}$ is a binary indicator for participation. Lastly, γ_j is the full set of match group fixed effects. We give participants a weight of one and other applicants a weight equal to the ratio of participants to non-participants in their match group.¹⁷ This weighting scheme accounts for the fact that match groups vary in terms of the proportion of participants to non-participants. We cluster standard errors by match school.¹⁸

Table 3.3 shows the results of the balance tests estimated using equation 3.1. Column 1 shows the comparison means, the weighted average of the baseline characteristic for

Michigan public school after the 2014-2015 school year. We exclude them from the analysis since we cannot construct match groups for them without pre-program information.

¹⁶For example, 10% of youths in a non-degenerate match group were older than 18, 98% were black and 34% of households in their neighborhoods lived below the poverty line, compared with 30%, 89%, and 30% respectively for youth in a match group with only participants or non-participants.

¹⁷That is, if a match group contains two participants and six non-participants then we give each participant a weight of one and each non-participant a weight of $\frac{1}{3}$. Similarly, if a match group contains three participants and two non-participants then we give each participant a weight of one and each non-participant a weight of $\frac{3}{2}$.

¹⁸All of our results are robust to using an unweighted regression model as well as to clustering standard errors by the zip code where youth resided in the match year or by the school they attended during the 2015-2016 school year.

non-participants, and column 2 shows α_1 , the coefficient on the indicator for participation. Participants look quite similar to non-participants along most observable dimensions.¹⁹ For example, 14.5% of participants and 13.2% of non-participants received special education services in the match year and 82.9% of participants were eligible for free or reduced-price lunch compared to 82.7% of non-participants. The differences in these baseline characteristics, as well as in neighborhood poverty rate and those related to 8th-grade standardized tests, are not statistically significant. However, there are meaningful differences between participants and non-participants in terms of baseline attendance. Participants had higher attendance rates by 1.5 percentage points and were 4.4 percentage points less likely to be chronically absent.²⁰ To address these differences, we control for a rich set of baseline controls, including baseline attendance and chronic absenteeism, in our main analysis.

In our context, testing for mean differences in baseline characteristics may not be very informative as to how similar participants and non-participants were before the program. Some CBO's employ youth who are particularly 'at risk' while others target those who are especially high-achieving. Therefore, it is possible that participants and non-participants look similar on average, even though they are actually quite different. To address this concern, Figure 3.1 shows kernel density plots of the entire distribution of four baseline characteristics, separately for participants and non-participants.²¹ They are residualized by match group and are restricted to the sample of youth who are in a match group with at least one participant and one non-participant. The figures show that there exists a common support between participants and non-participants; participants are not drawn from the two extremes of the distribution.

Having established that participants and non-participants within match group are quite similar on most observable dimensions, we estimate the impact of participation in the program with the following regression model

$$Y_{ij} = \beta_0 + \beta_1 \text{Participated}_{ij} + \beta_2 \mathbf{X}_{ij} + \gamma_j + \epsilon_{ij} \quad (3.2)$$

where outcome Y for youth i in match group j is a function of a binary indicator for participation (i.e., worked in the GDYT program during the 2015 summer), and other covariates. We include a vector of match group fixed effects, γ_j , which ensures we are comparing outcomes within *race * gender * school * grade* cells. In addition, we control for a rich set of covariates, \mathbf{X}_{ij} , consisting of all of the information from the application, some

¹⁹We do not test for differences along race/ethnicity, gender or age because the match group fixed effect ensures that participants and non-participants are comparable along these dimensions.

²⁰We define chronic absenteeism as having an attendance rate less than 90%.

²¹The kernel density plots for the binary characteristics in Table 3.3 look similar to those in Figure 1 but are much less smooth. They are available upon request.

information not from the application but likely observable to program administrators, as well as information that was not available to program staff during the selection process. The information from the application that we control for includes an indicator of whether the applicant graduated from high school before the program as well as linear and quadratic terms of age as of July 1, 2015. Program administrators could infer neighborhood characteristics from the application, so we use controls, measured at the census block group level, for the fraction of adults with at least a Bachelor’s degree, the fraction of households living in poverty, the fraction of households that are owner-occupied and the employment rate.²² Finally, program staff did not have access to past academic records, yet we use academic controls, including attendance rate, binary indicators for chronic absenteeism, special education status, limited English proficiency, eligibility for free or reduced-price lunch, ever retained in grade, as well as standardized math and reading test scores and an interaction of math and reading scores.²³ We include indicators for missing control variables. As before, we give participants a weight of one and other applicants a weight equal to the ratio of participants to non-participants in their match group. We cluster standard errors by match school.

The coefficient of interest, β_1 , represents the difference in average outcomes between participants and those who applied but did not participate. β_1 represents the true causal effect of program participation if the conditional independence assumption holds. There cannot be any unobservable differences between participants and non-participants that directly influence educational outcomes. While we cannot explicitly test this assumption, we can probe it. There are two ways in which we examine the robustness of our results to omitted variable bias. First, we implement the tests proposed in Oster (2016) to determine the extent to which selection on unobservables explains our results. We discuss these tests in detail in section 3.6.1.

Second, if participants differ from non-participants along unobservable dimensions, even after controlling for match group fixed effects and the rich set of covariates listed above, we might expect participation to predict youth outcomes prior to the program. Therefore, we estimate variants of equation 3.2 where the dependent variable is a pre-program educational outcome. We focus on 6th-grade outcomes, as 99.9% of youths applied after completing the sixth grade. Weighting and standard errors are handled as before.

Table 3.4 shows the results of these falsification tests. Column 1 reports the weighted average of the 6th-grade outcome for non-participants and column 2 shows the coefficient on the indicator for participation. Participation does not predict prior student outcomes,

²²Unless otherwise specified, all of the pre-program characteristics are measured in the match year.

²³We use 8th-grade test scores whenever possible. If the applicant did not reach 8th grade before they applied or did not take the tests in 8th grade, we use 7th-grade scores. If neither 7th nor 8th-grade scores are available, we use 6th-grade scores.

including attendance and performance on standardized tests. This exercise suggests that it is unlikely that there are important omitted variables that confound our impact estimates.

3.6 Effects of Participation in GDYT

We assess the effect of participation in GDYT on a variety of short-run educational outcomes. We observe outcomes during two follow-up school years, 2015-2016 and 2016-2017. Overall, we find consistently positive effects of program participation on enrollment, test-taking rates, and high school graduation, with no evidence of subsequent decreases in attendance or test scores.

Table 3.5 reports the effect of participating in GDYT on educational outcomes, showing the estimates from equation 3.2. Since applicants range in age from 14-24, each outcome is defined only for a certain set of youth.²⁴ We describe the sample who were eligible for each outcome in column 1 and the number of youths in the sample in column 2.²⁵ Column 3 shows the weighted average of the outcome for non-participants and column 4 displays the cumulative effect of participation in GDYT in the two years following the program. While we focus our discussion on the effects after two years, we show estimates of the effect of participation after each follow-up year separately in Table C.3.

Participation in GDYT increased the likelihood of being enrolled in a Michigan public school by 1.4 percentage points. In the two years after the program, 95.9% of participants remained enrolled compared to 94.5% of non-participants, a difference which is statistically significant at the 1% level. Although we cannot directly test it, this is likely driven by a reduction in dropping out of school. It is unlikely that this is a result of differential mobility out of state or enrollment in private school given that we find even larger positive effects on high school graduation from a Michigan public school, which will be discussed in further detail below.

We interpret our results as the effect of working with GDYT compared to a ‘business as usual’ control group. We do not observe the employment status of non-participants during the 2015 summer, so we do not know whether they worked outside of GDYT or did not have a job at all.²⁶ However, all of the prior studies of summer youth employment programs

²⁴For example, we do not analyze the effect of participation on subsequent enrollment in K-12 for a 19-year-old participant who already graduated high school.

²⁵Table C.2 further describes the sample of eligible youths for the outcomes in each of the first two follow-up years and clarifies the formula used to calculate the cumulative measure of the outcomes after the first two follow-up years.

²⁶Some non-participants were offered a summer job with GDYT but declined, although we also do not observe offers to participate with GDYT. There are a few reasons why someone who received an offer would not participate: 1) they did not complete registration paperwork, 2) they were selected for the ILT or Career Pathways Internship programs but did not show up to the career fairs, 3) they needed to attend summer

suggest that participation increases the likelihood of having any job during the summer and we have little reason to suspect that this result would not generalize to Detroit (Gelber et al., 2016; Davis and Heller, 2017; Modestino, 2017).

Participation in GDYT did not influence daily attendance rates or chronic absenteeism; we find precise zero point estimates for these outcomes. Students who remained in public school were no more or less likely to show up throughout the school year. If the marginal student were any less qualified, prepared or resourced to remain in school, then they might have had lower attendance rates or higher chronic absenteeism, conditional on enrollment. This is not what we find, however. In addition, we do not find evidence that participation influenced the likelihood of expulsion.

Michigan offers the SAT for free and during school hours, as part of its 11th grade standardized tests. Some students still do not take the test, though, due to exemptions as a result of special education plans or absences on test days. GDYT participants were 3.4 percentage points more likely to take the SAT than non-participants, representing a 4.7% increase from a comparison mean of 72.9%. This estimate is significant at the 10% level. Some, but not all, of the increase in test-taking is driven by increased enrollment in school, yet the magnitude of the effect on enrollment is less than half the size of the effect on taking the SAT. Importantly, there was no statistically significant effect on SAT scores among those who took the SAT.

Participation in GDYT had a large and positive effect on high school graduation. Youth who participated were 3.4 percentage points more likely to graduate than non-participants, representing a 4% increase relative to a comparison mean of 85%. This finding stands in contrast to Valentine et al. (2017) which finds that summer youth employment did not influence high school graduation in NYC. As shown in Table C.3, this is mostly driven by an increase in the second follow-up year.²⁷ Finally, we find meaningfully large, yet statistically insignificant, estimates of participation on college enrollment for applicants who had graduated high school before applying. They suggest that participation increased the likelihood of enrolling in college by 8%. This is driven by an increase in four-year college enrollment.

school, or 4) they did not wish to participate.

²⁷This is probably because 11th grade applicants were likely to graduate regardless of participation in the program, whereas the program had a greater influence for 10th-grade applicants. However, an alternative explanation could be that there is a cumulative effect of participation, whereby working with GDYT for more than one summer has a larger effect on graduation than working for a single summer. This would be consistent with findings from the NYC program in Schwartz et al. (2015). We do find that 10th grade participants in 2015 were more likely to work with GDYT during the 2016 summer. Descriptively, though, there was not a larger increase in high school graduation among youth who participated in both the 2015 and 2016 summers. Results are available upon request.

3.6.1 Robustness to Omitted Variable Bias

Although we use both matching and regression adjustment to control for many observable differences between participants and other applicants, our estimates may still suffer from omitted variable bias. Unobservable differences between these groups could come both from the labor supply side, if the most motivated or committed youth who were offered jobs chose to participate, and from the labor demand side, if employers chose to hire the best applicants based on unobservable features like personality. We expect either to lead us to overstate the positive effects of the program.

In order to test the extent to which omitted variable bias may be driving our results, we perform two exercises proposed by Oster (2016). Building on earlier work by Altonji et al. (2005), Oster proposes that, under some reasonable assumptions, one can identify a consistent estimator of the bias. In particular, Oster suggests two complementary methods to assess the robustness of results to omitted variable bias. The first is to generate a bias-adjusted treatment effect, which represents the value of the treatment effect assuming a given degree of selection on unobservables. Researchers can bound the true treatment effect using the bias-adjusted treatment effects. The second is to examine the amount of selection on unobservables, relative to selection on observables, that would need to exist for the true treatment effect to be equal to zero. If a large amount of selection on unobservables is needed, then the treatment effect can be considered robust to omitted variables bias.

Both of these exercises require a proportional selection assumption, whereby selection on unobservables is proportional to selection on observables. The former exercise also assumes that the omitted variable bias does not change the direction of the covariance between the observables and the treatment. This holds as long as the bias not too large. An important caveat with these exercises is that they only address omitted variable bias created by unobservables that are related to observable characteristics. In the context of this study, it seems unlikely that there are unobservables that are orthogonal to prior academic records or socio-demographic characteristics.

Table 3.6 reports the results of these two exercises.²⁸ Column 1 shows our estimate of the effect of program participation, as previously reported in Table 3.5. This is equivalent to the estimate of the treatment effect if there were no omitted variable bias. Column 2 shows the bias-adjusted treatment effect, assuming that the amount of selection on unobservables is equal to the amount of selection on observables.²⁹ Both Oster (2016) and Altonji et al. (2005)

²⁸We used the STATA package *psacalc* for these calculations.

²⁹We also assume that the maximum R^2 from a regression of the outcome on the observable characteristics and all unobservables would be equal to one. Assuming a smaller R^2 , perhaps due to measurement error in the outcome variable, would only move the bias-adjusted treatment effect away from zero, suggesting this may be a conservative approach.

suggest this as an upper bound on the amount of omitted variable bias. Together, columns 1 and column 2 report bounds for the true treatment effect. Finally, column 3 shows the amount of proportional selection needed such that the treatment effect equals zero. Values smaller in magnitude than one mean that selection on unobservables would not need to be as large as selection on observables, whereas values larger than one mean that selection on unobservables would need to be larger than the amount of selection on observables.³⁰

Our impact estimates do not appear to be driven by omitted variable bias. We can bound the effect on public school enrollment between 0.8 and 1.4 percentage points. Selection on unobservables would need to be almost twice as large as the selection on unobservables in order for the true effect to equal zero. Consistent with earlier results, the amount of bias needed to generate an effect of zero is much smaller for attendance and expulsion outcomes.

In addition, the estimates for taking the SAT and graduating high school are robust. Selection on unobservables would need to be over twenty times as large as selection on observables for the effect on test-taking to equal zero, and almost four times as large for the effect on graduating high school. We can bound our estimate for graduating from high school between 2.9 and 3.4 percentage points, which still represents a large increase of at least 3.4% relative to the comparison mean of 85%. Although estimates for the effect of participation on college enrollment were not statistically significant, these exercises suggest that the substantively large estimate on college enrollment for applicants who graduated high school was unlikely to be driven by omitted variable bias.

3.6.2 Subgroup Analysis

While we find positive effects of the summer youth employment program overall, we also seek to identify for whom the program was most beneficial. We explore heterogeneity along four dimensions: grade in school, gender, prior academic achievement, and prior chronic absenteeism. First, we compare the effect of participation for those in 9th grade or below when they applied to those who were in 10th or 11th grade because previous studies find differential effects by age.³¹ Second, we compare males to females because of growing gender gaps in non-cognitive skills and educational attainment (Jacob, 2002; Goldin et al., 2006). Finally, we examine differences based on both prior academic achievement and prior chronic

³⁰Negative values indicate that selection on unobservables must operate in the opposite direction as selection on observables for the treatment effect to equal zero.

³¹Leos-Urbel (2014) finds that, conditional on low baseline attendance, youth age 16 and older benefited more from the program than younger participants. Modestino (2017) finds that youth older than 16 benefited more from participation as well. Davis and Heller (2017) also find that those who benefited most from the program were 16-17 year olds, although their analysis primarily compares this group to participants age 18 and older. In Detroit, 16-17 year olds tend to be in grades 10-12, so our analysis is closer in spirit to Leos-Urbel (2014) and Modestino (2017), in that we compare these youths to younger participants.

absenteeism because of past findings of heterogeneity by levels of school engagement.³² Since test scores and attendance offer different, albeit correlated, measures of engagement, we study these two groups separately.

We tested for balance in pre-program characteristics for each subgroup and show the results in Table C.4. Consistent with our main analysis, participants and other applicants look quite similar along most observable dimensions, yet participants had higher attendance rates and were less likely to be chronically absent. As before, we control for a rich set of covariates, including baseline attendance rate and chronic absenteeism, to address these differences. We also conduct the same falsification tests as in our main analysis to check whether participation predicts past academic outcomes. We report the results in Table C.5. While we expect to find some statistically significant differences as we are testing many hypotheses, participation is generally not predictive of prior outcomes. Therefore, it is unlikely that there are important unobserved factors that bias our subgroup impact estimates.

Table 3.7 shows the effect of participation in GDYT on a handful of key educational outcomes for each subgroup in the two years after the program.³³ The program was beneficial for students who were in 9th grade and below when they applied (column 1) as well as for those who were in 10th and 11th grade (column 2). The magnitude of the effect on subsequent enrollment was similar for both groups, although the estimate for younger participants is not statistically significant. Younger participants were 1.0 percentage point more likely to remain enrolled in school, compared to a 1.7 percentage point increase for their older counterparts. Older students who enrolled were more likely to be chronically absent than their younger peers, however. This was likely driven by the enrollment of older students who otherwise would have dropped out; although they remained in school, they were less likely to show up. Despite higher rates of chronic absenteeism, we find that participation still increased the educational attainment of older youths. They were 3.4 percentage points more likely to graduate high school. We conclude that having a summer job with GDYT improved the educational outcomes for middle and high school students.

Male and female participants benefited from the program in different ways. While both males (column 3) and females (column 4) enrolled in school at similar rates, participation increased chronic absenteeism for males and reduced it for females. Although neither is

³²While Leos-Urbel (2014) finds that students with low baseline attendance benefit most, Davis and Heller (2017) conclude that the program was most beneficial for youth with higher GPAs and attendance. These studies use very different definitions of attendance, though. Leos-Urbel (2014) focuses on youth with attendance rates less than 95% while Davis and Heller (2017) find that youth who had average pre-program attendance rates of 77% (attended 139 days of school out of 180) benefited more than those who had average attendance rates of 63% (attended 114 days).

³³Table C.6 reports similar results for the full set of outcomes.

statistically different from zero, they are close to being statistically different from each other, with a p-value of 0.119. Surprisingly, participation may have increased the gender gap in high school graduation by a modest amount, yet reduced the gap in college enrollment by even more.³⁴ The main effect of GDYT on high school graduation was driven by an increase for females, who were already more likely to graduate. Although our estimates on college enrollment are not statistically significant, they are consequential. If taken literally, they suggest that participation reduced the gender gap in college enrollment by 60%, from 11.5 to 4.6 percentage points.

To examine differences based on prior academic achievement, we compare youth who scored above the median on their 8th-grade math test (column 5) to those who scored below it (column 6).³⁵ The benefits of participation were largest for those who entered high school with the weakest academic skills. Participation increased school enrollment by 2.1 percentage points and high school graduation by 4.4 percentage points among lower-achieving youths. In contrast, it did not have a meaningful effect on educational outcomes for higher-achieving participants. The differences between these groups in terms of school enrollment and high school graduation are statistically significant at the 10% level. Consistent with these results, we also find that those who were not chronically absent before the program (column 7) benefited less from participation than those who were (column 8). Our estimates of the effects on school enrollment and high school graduation were larger for youth who were chronically absent. Similar to our previous findings, though, despite positive effects on other outcomes, participation did increase chronic absenteeism among youth who were chronically absent before the program. Overall, we conclude that the program was particularly beneficial for those with the weakest academic achievement and lowest attendance.

3.7 Conclusion

The labor market continues to present challenges for young, low-income, and less-educated workers. This has spurred a growing interest in finding new ways for young adults to obtain career skills, both inside and outside of school. Summer youth employment programs

³⁴This appears to be because these outcomes are defined for different samples. That is, the high school graduation outcome is defined for 10th and 11th-grade applicants while college enrollment is defined for 11th and 12th-grade applicants. When we limit the analysis to 11th-grade applicants in order to hold the sample constant, we find that the point estimate for high school graduation is slightly larger in magnitude for males than females, while the estimates for college enrollment for each group are similar to those reported in Table 3.7.

³⁵This analysis only includes youth who reached 8th grade by the 2014-2015 school year. We define the median here as the median among all 8th-grade math scores in our sample. Test scores were normalized within each year so they are comparable across years. Our results are robust to using reading scores instead of math scores.

represent one such intervention. In Detroit, the Grow Detroit’s Young Talent program has provided over 15,000 summer work opportunities to youth between 2015 and 2017.

This study is complementary to prior evaluations of similar programs. Despite the prevalence of summer youth employment programs across the United States, there is surprisingly limited evidence of their effectiveness, particularly in terms of whether participation influences educational outcomes. Despite findings from New York City, Chicago, and Boston which suggest that there were small, if any, impacts on education, we find that the program in Detroit increased education attainment. While the implementation of GDYT is similar to the programs in NYC and Boston, the cities themselves are very different. Detroit is a less-resourced city than these other sites. The poverty rate is higher, the schools are of lower quality in terms of standardized test scores and the city has uniquely struggled with population decline over the past several decades (Infoplease, 2017). Our analysis suggests that in this context, providing jobs during the summer can change the educational trajectories of young adults.

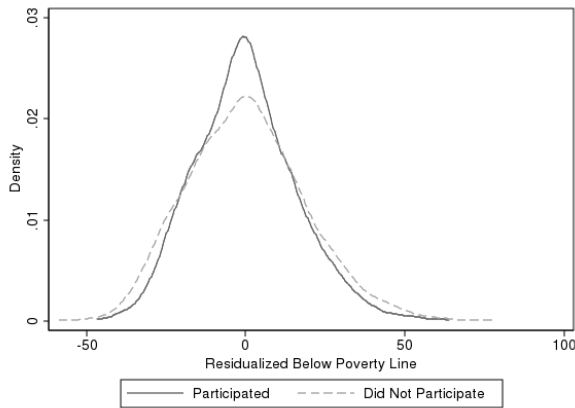
We find that youth who participate in GDYT experience consistently better educational outcomes than their classmates who applied but did not participate. GDYT increased subsequent enrollment in a Michigan public school by 1.5%, likely driven by a reduction in dropping out. Even with increased enrollment, participants were no more or less likely to be chronically absent. Most importantly, participation increased high school graduation by 4% relative to a comparison mean of 85%. Following two exercises proposed in Oster (2016), these estimates are not driven by selection bias from unobservables characteristics that are related to observables. In addition, those who entered high school with the weakest academic skills benefited the most, as participation increased school enrollment and high school graduation by 2.2% and 5.5% respectively for this group. Overall, the benefits of participation in GDYT continued long after youth received their last paycheck.

Our study has several limitations. First, we focus on only a single cohort of applicants and follow them for only two years after the program. As we continue our partnership with GDYT, we plan to study more cohorts for a longer follow-up period in future work. Second, our study relies on a selection on observables identification strategy, which is considered a less credible design than those used in evaluations of other summer youth employment programs. We are currently working with GDYT to randomly assign a subset of summer jobs to applicants, which will allow for a more refined evaluation in the future. However, after controlling for a rich set of covariates and using matching methods, our analysis shows that it is unlikely that there are important omitted variables that bias our estimates. Finally, we do not address whether participation influenced outcomes other than education. In order to conduct a cost-benefit analysis of the program, it will be important to examine its effect

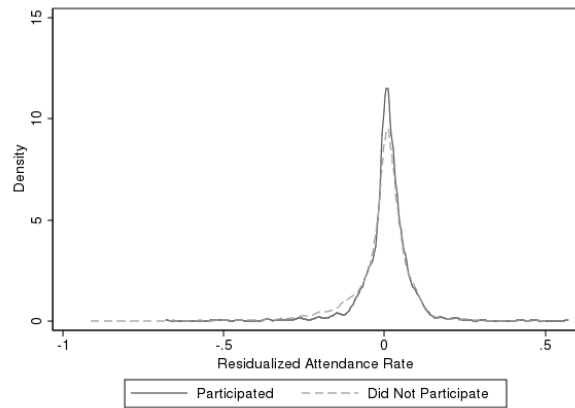
on employment, crime, and health in future work. We are working to establish partnerships to bring in data about these other indicators of wellbeing for future work.

As policymakers from all levels of governance look for ways to improve the job prospects for many people, and particularly for young people, it is important to understand whether summer youth employment programs offer a pathway to success in the labor market.

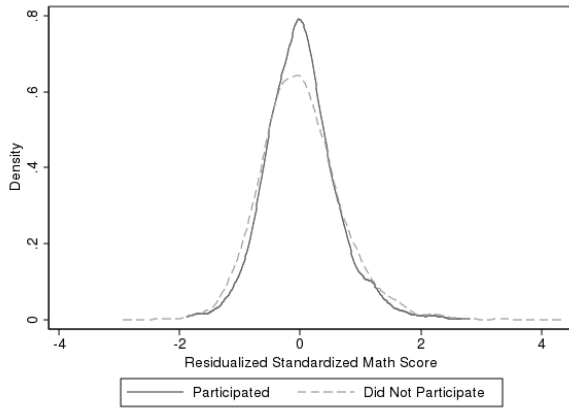
Figure 3.1: Distribution of Baseline Characteristics for Participants and Non-Participants



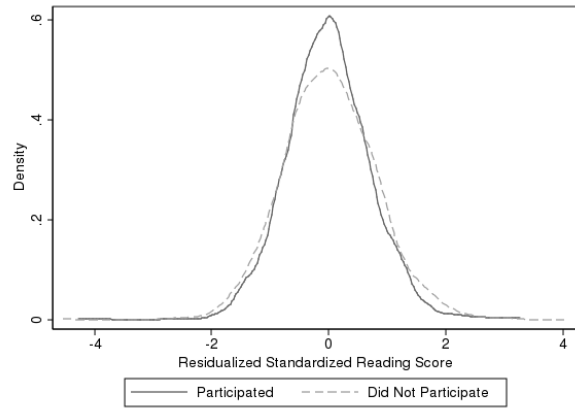
(a) Below Poverty Line in Neighborhood



(b) Attendance Rate



(c) Standardized 8th Grade Math Score



(d) Standardized 8th Grade Reading Score

Notes. Each figure shows a kernel density plot of the distribution of a baseline characteristic for participants and non-participants, residualized by match group. The sample is restricted to youth in match groups with at least one participant and non-participant. All baseline characteristics are measured in the match year, except for 8th grade test scores. Youth who did not reach 8th grade before the program are not included in the analysis of 8th grade test scores.

Table 3.1: Summary Statistics

	(1) Detroit Grades 8-12	(2) All Applicants	(3) Participated	(4) Did Not Participate
Total	44394	12255	2807	9448
Matched to Education Data	1.00	0.94	0.90	0.95
Age as of July 1, 2015				
15 and Younger	0.41	0.38	0.38	0.37
16-18	0.52	0.46	0.45	0.46
19 and Older	0.07	0.16	0.16	0.16
Educational Status				
Grade 9 and Below	0.43	0.34	0.33	0.34
Grades 10-11	0.39	0.35	0.36	0.35
Grade 12	0.18	0.12	0.12	0.12
Enrolled In College	0.00	0.08	0.09	0.08
Not in School	0.00	0.10	0.10	0.11
Demographics				
Black	0.86	0.95	0.90	0.96
Hispanic	0.08	0.03	0.07	0.02
White	0.04	0.02	0.03	0.01
Asian	0.01	0.01	0.00	0.01
Female	0.51	0.58	0.55	0.60
Needs in School				
Limited English Proficient	0.08	0.03	0.05	0.02
Special Education	0.15	0.12	0.15	0.12
Free or Reduced Priced Lunch	0.82	0.81	0.82	0.81
Neighborhood Characteristics				
BA Degree or Higher	11.94	13.30	12.69	13.47
Below Poverty Line	35.15	32.82	33.02	32.77
Owner Occupied Housing	40.05	41.31	40.84	41.45
Employed (16 and over)	75.10	75.10	75.15	75.09
8th Grade Test Scores				
Took Math Test	0.84	0.88	0.86	0.89
Proficient Math	0.09	0.10	0.11	0.09
Math Score	-0.67	-0.62	-0.60	-0.62
Took Reading Test	0.84	0.87	0.86	0.88
Proficient Reading	0.37	0.41	0.42	0.41
Reading Score	-0.55	-0.46	-0.46	-0.46

Notes. Column 1 shows summary statistics for all 8th to 12th graders who lived in Detroit and attended a Michigan public school during the 2014-2015 school year, regardless of whether they applied to GDYT. Column 2 describes GDYT applicants for the 2015 summer while columns 3 and 4 describes GDYT participants and GDYT applicants who did not participate, respectively. We report educational status in the 2014-2015 school, where not enrolled indicates that the youth was neither enrolled in public school in Michigan nor in college. Needs in school and neighborhood characteristics are measured in the most recent pre-program that a youth was enrolled in a public K-12 school in Michigan. Youth who did not reach 8th grade before the program are not included in the 8th grade test scores summary statistics.

Table 3.2: Summary Statistics, by Match Group Characteristics

	(1) Match Group with at Least 1 Participant and 1 Non-Participant	(2) Match Group with Only Participants or Non-Participants
Total	7960	3587
Age as of July 1, 2015		
15 and Younger	0.40	0.33
16-18	0.50	0.37
19 and Older	0.10	0.30
Educational Status		
Grade 9 and Below	0.36	0.29
Grades 10-11	0.41	0.21
Grade 12	0.13	0.09
Enrolled In College	0.06	0.15
Not in School	0.04	0.25
Demographics		
Black	0.98	0.89
Hispanic	0.02	0.05
White	0.00	0.04
Asian American	0.00	0.02
Female	0.59	0.56
Needs in School		
Limited English Proficient	0.02	0.04
Special Education	0.12	0.13
Free or Reduced Price Lunch	0.82	0.80
Neighborhood Characteristics		
BA Degree or Higher	12.69	14.65
Below Poverty Line	34.00	30.21
Owner Occupied Housing	39.98	44.28
Employed (16 and over)	74.39	76.67
8th Grade Test Scores		
Non-missing math score	0.89	0.86
Proficient Math	0.10	0.09
Math Score	-0.61	-0.63
Non-missing reading score	0.89	0.84
Proficient Reading	0.43	0.35
Reading Score	-0.45	-0.49

Notes. Column 1 shows summary statistics for youth who are in match groups with at least one participant and one non-participating applicant while column 2 describes youth in match groups with only participants or non-participating applicants. The sample is restricted to applicants who matched to the education data. We report educational status in the 2014-2015 school, where not enrolled indicates that the youth was neither enrolled in public school in Michigan nor in college. Needs in school and neighborhood characteristics are measured in the most recent pre-program that a youth was enrolled in a public K-12 school in Michigan. We define a neighborhood as a census block group. Youth who did not reach 8th grade before the program are not included in the 8th grade test scores summary statistics.

Table 3.3: Balance Tests of Differences Between Participants and Non-Participants

	(1) Comparison Mean	(2) Participated
Special Education	0.132	0.013 (0.011)
Low Income	0.827	0.002 (0.010)
% Below Poverty Line in Neighborhood	34.258	-0.053 (0.556)
Attendance Rate	0.895	0.015*** (0.004)
Chronically Absent	0.351	-0.044*** (0.017)
Took 8th Grade Math Test	0.873	-0.006 (0.010)
Proficient on 8th Grade Math Test	0.091	0.013 (0.011)
Standardized 8th Grade Math Score	-0.631	0.019 (0.025)
Took 8th Grade Reading Test	0.875	-0.007 (0.009)
Proficient on 8th Grade Reading Test	0.406	0.016 (0.017)
Standardized 8th Grade Reading Score	-0.476	-0.010 (0.034)
N	11,547	

Notes. This table reports the results from a weighted regression of a baseline characteristic on an indicator for participated and match group fixed effects, where participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. Column 1 shows the weighted mean of the baseline characteristic for non-participants and column 2 shows the coefficient on the indicator for participation. All baseline characteristics are measured in the match year, except for 8th grade test scores. Youth who did not reach 8th grade before the program are not included in the analysis of 8th grade test scores. Standard errors are clustered by match school. Stars indicate: *p<0.1, ** p<0.05, *** p<0.01.

Table 3.4: Falsification Tests of the Effect of Participation in GDYT on 6th Grade Educational Outcomes

	(1) Comparison Mean	(2) Participated
Attendance Rate	0.909	0.005 (0.003)
Chronically Absent	0.337	-0.013 (0.013)
Took Math Exam	0.937	0.003 (0.006)
Proficient on Math	0.144	-0.011 (0.011)
Std Math Score	-0.543	-0.039 (0.024)
Took Reading Exam	0.941	-0.002 (0.006)
Proficient on Reading	0.405	-0.002 (0.014)
Std Reading Score	-0.525	-0.013 (0.027)
N	11,531	

Notes. This table reports the results from a weighted regression of a 6th grade educational outcome on an indicator for participated, a vector of control variables as listed in the text and match group fixed effects, where participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. Column 1 shows the weighted mean of the baseline characteristic for non-participants and column 2 shows the coefficient on the indicator for participation. The sample is restricted to the 99.9% of youths who reached 7th grade before the program so that the control variables, most of which are measured in the match year, are not measured in 6th grade. Standard errors are clustered by match school. Stars indicate: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: The Effect of Participation in GDYT on Educational Outcomes

	(1) Description of Sample	(2) N	(3) Comparison Mean	(4) Participated
Enrolled in K12	11th grade or below in 2015	6,340	0.945	0.014*** (0.006)
Attendance Rate	Enrolled in K-12 in 2016 or 2017	5,637	0.876	0.002 (0.004)
Chronically Absent	Enrolled in K-12 in 2016 or 2017	5,637	0.409	-0.000 (0.013)
Expelled	Enrolled in K-12 in 2016 or 2017	6,173	0.002	-0.001 (0.001)
Took SAT	9th or 10th graders in 2015	4,080	0.729	0.034* (0.018)
SAT Composite Score	Took the SAT in 2016 or 2017	3,012	860.307	-0.164 (3.791)
Graduated HS	10th or 11th grade in 2015	3,131	0.850	0.034** (0.014)
College from HS	11th or 12th grade in 2015	2,484	0.435	0.006 (0.027)
2 Year College from HS	11th or 12th grade in 2015	2,484	0.210	0.012 (0.026)
4 Year College from HS	11th or 12th grade in 2015	2,484	0.230	-0.008 (0.018)
College as a HS Grad	Graduated HS before 2015	1,701	0.449	0.038 (0.040)
2 Year College as a HS Grad	Graduated HS before 2015	1,701	0.216	0.015 (0.032)
4 Year College as a HS Grad	Graduated HS before 2015	1,701	0.244	0.024 (0.029)

Notes. This table reports the results from a weighted regression of an educational outcome on an indicator for participated, a vector of control variables as listed in the text and match group fixed effects, where participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. Column 1 describes the sample for whom the outcome is defined, column 2 shows the number of observations in the sample, column 3 shows the weighted mean of the baseline characteristic for non-participants and column 4 shows the coefficient on the indicator for participation. Standard errors are clustered by match school. Stars indicate: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.6: Robustness of the Effects of Participation in GDYT to Omitted Variable Bias

	(1)	(2)	(3)
	Observed Effect	Bias-Adjusted Effect	Proportional Degree of Selection for Effect of Zero
Enrolled in K12	0.014	0.008	1.892
Attendance Rate	0.002	-0.006	0.265
Chronically Absent	-0.000	0.028	0.012
Expelled	-0.001	0.002	0.536
Took SAT	0.034	0.036	22.633
SAT Composite Score	-0.164	0.754	0.183
Graduated HS	0.034	0.029	3.809
College from HS	0.006	-0.038	0.152
2 Year College from HS	0.012	0.059	-0.495
4 Year College from HS	-0.008	-0.053	-0.200
College as a HS Grad	0.038	0.15	1.404
2 Year College as a HS Grad	0.015	0.088	-0.490
4 Year College as a HS Grad	0.024	-0.012	0.720

Notes. This table reports the results of robustness checks proposed in Oster (2016). Column 1 reports the observed estimate of the effect of program participation on the outcome, without accounting for omitted variable bias. Column 2 reports the bias-adjusted treatment effect, assuming that selection on unobservables is as large as selection on observables. Column 3 shows the amount of the selection on unobservables, relative to selection on observables, necessary for the treatment effect to equal zero. Columns 2 and 3 were estimated using the STATA package *psacalc*. Stars indicate: *p<0.1, ** p<0.05, *** p<0.01.

Table 3.7: The Effect of Participation in GDYT on Educational Outcomes, by Subgroup

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	9th Grade and Below	10th and 11 Grade	Male	Female	Above Median Math Score	Below Median Math Score	Not Chronically Absent	Chronically Absent
Enrolled in K12	0.010 (0.007) [0.957]	0.017** (0.007) [0.934] {0.438}	0.016* (0.008) [0.939]	0.011* (0.007) [0.949] {0.634}	0.004 (0.006) [0.965]	0.021** (0.008) [0.940] {0.077}	0.006 (0.005) [0.975]	0.024 (0.016) [0.907] {0.172}
Chronically Absent	-0.030 (0.021) [0.387]	0.027* (0.015) [0.428] {0.016}	0.026 (0.023) [0.377]	-0.023 (0.021) [0.435] {0.119}	0.003 (0.020) [0.333]	-0.002 (0.024) [0.455] {0.876}	0.004 (0.020) [0.243]	0.051* (0.029) [0.743] {0.103}
Graduated HS		0.034** (0.014) [0.850]	0.008 (0.025) [0.833]	0.051** (0.021) [0.862] {0.210}	-0.006 (0.020) [0.928]	0.044* (0.024) [0.791] {0.082}	0.018 (0.014) [0.933]	0.035 (0.047) [0.787] {0.693}
College from HS		0.004 (0.035) [0.453]	0.046 (0.033) [0.372]	-0.023 (0.036) [0.487] {0.114}	0.016 (0.041) [0.547]	0.006 (0.054) [0.347] {0.861}	-0.031 (0.041) [0.547]	0.008 (0.053) [0.329] {0.422}

Notes. This table reports the results from a weighted regression of an educational outcome on an indicator for participated, a vector of control variables as listed in the text and match group fixed effects, separately for each subgroup. Participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. Standard errors, which are clustered by match school, are reported in parentheses and the weighted mean of the baseline characteristic for non-participants is shown in brackets. In the even columns, we report the p-value from a test of whether the point estimates are equal across subgroups in curly braces. Columns 5 and 6 are restricted to youth who reached 8th grade before the program since we define prior achievement by 8th grade test scores. Stars indicate: *p<0.1, ** p<0.05, *** p<0.01.

APPENDICES

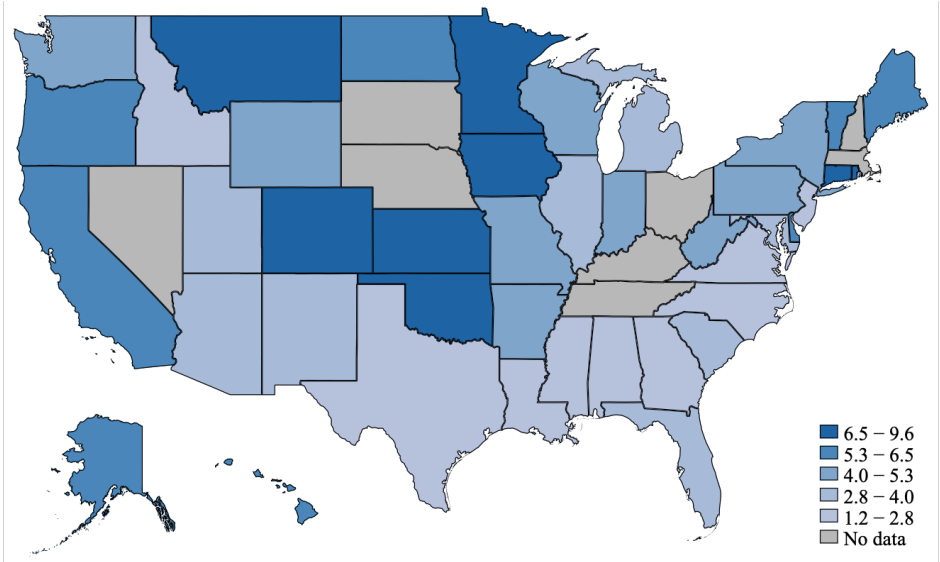
APPENDIX A

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

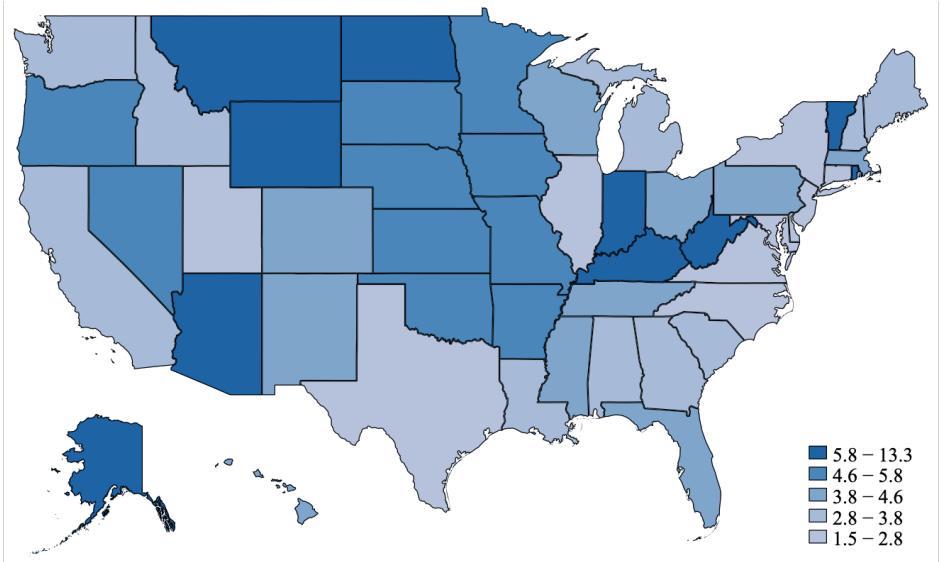
A.1 Supplemental Figures and Tables

Figure A.1: Foster Care Entry Per 1000 Children

(a) 1998 Statistics



(b) 2017 Statistics



Notes. These figures show the rate of foster care entry per 1000 children for each state in the first and last year of available data as of March 2019. There are five different shades of blue which represent the quantile of the foster care entry rate for each state, with darker shading indicating higher rates of entry. Eight states do not report the number of children who entered foster care in 1998 and are shaded in gray in Figure A.1a. The 1998 information comes from USDHHS (2003) and the 2017 information comes from USDHHS (2017b).

Table A.1: Effects of Foster Care on Michigan Public School Enrollment

	(1)	(2)	(3)	(4)	(5)	(6)
	Ever	Enrolled	Enrolled	Enrolled	Enrolled	Enrolled
	Enrolled	One Year	Two Years	Three Years	Four Years	Five Years
	After	After	After	After	After	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 Grades One to Eleven During Investigation</i>						
Foster Care	-0.033	-0.018	0.001	-0.126	-0.001	0.039
	(0.039)	(0.049)	(0.070)	(0.090)	(0.110)	(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 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 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 A.2: 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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Coefficient	Std Error	Coefficient	Std Error	Coefficient	Std Error	Coefficient	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 baseline 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 include indicators for the following variables which contain some missing information: female, low income, and each of the prior schooling characteristics. Standard errors are clustered by investigator. *p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.3: Testable Implications of the Exclusion of Removal Stringency Instrument

	(1)	(2)	(3)	(4)	(5)	(6)
	Investigator's Number of Investigations	Days in Foster Care	# Foster Homes	First Placed with Relatives	First Placed with Unrelated Family	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 A.4: Testable Implications of Monotonicity of the Removal Stringency Instrument

	(1) Female	(2) Male	(3) White	(4) Student of Color	(5) Age ≤ 10	(6) Age > 10	(7) Had Prior Inv	(8) No Prior Inv
<i>Panel A: Main Leave-One-Out Instrument</i>								
Removal Stringency	0.481*** (0.030)	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. 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 investigator. *p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.5: Effects of Foster Care on Taking Standardized Tests

	(1)	(2)
	Took Std Math Test	Took Std Reading Test
<i>Panel A: OLS</i>		
Foster Care	0.007 (0.004)	0.008* (0.004)
<i>Panel B: 2SLS</i>		
Foster Care	0.019 (0.063)	-0.029 (0.065)
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 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 also excluded from this analysis.*p< 0.10, ** p< 0.05, *** p< 0.01.

Table A.6: Effects of Foster Care on High School Graduation and College Enrollment

	(1)	(2)	(3)	(4)
	Graduated High School	Ever Enrolled in College	Ever Enrolled in a Two-Year College	Ever Enrolled in a Four-Year College
<i>Panel A: OLS</i>				
Foster Care	-0.024* (0.014)	0.001 (0.017)	-0.008 (0.016)	0.013 (0.012)
<i>Panel B: 2SLS</i>				
Foster Care	0.085 (0.274)	0.192 (0.368)	-0.001 (0.343)	0.029 (0.286)
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 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. The analysis of college enrollment is similarly restricted to students expected to be in 12th grade by 2016. Some colleges are missing information on their type, so the two and four-year college enrollment estimates need not add up to the overall college enrollment estimate. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Effects of Foster Care on Type of Foster Placement

	(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 A.8: 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}
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. 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 A.9: 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 Observations	85.3% 242,233	7.8% 242,233	4.9% 242,233	2.1% 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 such for their permanency outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Characteristics of Compliers at the Margin of Foster Placement

	(1)	(2)	(3)
	All	Foster Care	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. I follow Dahl et al. (2014) 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 standardized math and reading tests.

Table A.11: Effects of Foster Care on Children’s Experience in Foster System

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

Notes. This table compares the experiences of the average foster placement and the marginal foster placement while in the foster system. Column one reports the mean outcome among all foster placements while column two 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. 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, so these students are coded as such for their permanency outcome.

Table A.12: Effects of Foster Care on Index of Child Wellbeing, 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	0.196** (0.085)	0.108 (0.115)	0.191** (0.099)	0.145* (0.086)	0.320*** (0.119)	0.135 (0.102)	-0.028 (0.164)	0.157 (0.143)
P-value		0.483		0.669		0.147		0.316
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 wellbeing on foster care for a variety of subgroups, using removal stringency to instrument for foster care. The young subgroup includes children ages ten and younger at the start of the child welfare investigation while the old subgroup includes children age eleven 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 zipcode by investigation year fixed effects. Standard errors are clustered by investigator. *p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.13: Robustness Checks

	(1) Index of Child Wellbeing
<i>Panel A: Alternative Samples</i>	
Child's First Investigation (N=180,859)	0.163* (0.086)
Investigator Assigned \geq 25 Investigations (N=249,228)	0.137* (0.072)
Investigator Assigned \geq 75 Investigations (N=232,818)	0.146* (0.077)
Balanced Panel (N=96,156)	0.310*** (0.106)
<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)	0.266 (0.239)
<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 B uses alternative measures of removal stringency to instrument for foster care, and Panel C reports the results using the main stringency instrument yet replaces zipcode 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, 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.

Table A.14: 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	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 1.6. Columns one and two 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 three, columns 3 and 4 report results for students with a substantiated investigation who were enrolled in at least grade four during the maltreatment investigation and include standardized test scores. Full regression results are available upon request. Standard errors are clustered by investigator.

Table A.15: First Stage Effect of Censored Removal Stringency on Foster Placement

	(1)	(2)	(3)	(4)
	Foster	Foster	Foster	Foster
	Care	Care	Care	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 1.6. 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-price 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 pre-investigation score from standardized math and reading test scores. Standard errors are clustered by investigator. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.16: 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.

Table A.17: OLS Effects of Foster Care on Child Outcomes Using Censored Data

	(1)	(2)	(3)	(4)
	Index of Child Wellbeing	Confirmed Victim of Maltreatment	Daily Attendance Rate	Std Math Score
<i>Panel A: Complete Data, Unsubstantiated and Substantiated</i>				
Foster Care	-0.005 (0.005)	-0.007*** (0.046)	0.011*** (0.002)	0.056*** (0.013)
Observations	242,233	242,233	224,925	177,118
<i>Panel B: Censored Data, Only Substantiated</i>				
Foster Care	-0.006 (0.005)	-0.014*** (0.002)	0.010*** (0.002)	0.042*** (0.015)
Observations	47,469	47,469	43,839	35,322

Notes. Panel A reports the OLS results from Table 1.4 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 investigator. *p< 0.10, ** p< 0.05, *** p< 0.01.

A.1.1 Assessing Arteaga (2019) Approaches to Using Examiner Assignment Design with Censored Data

Section 1.6 documents that censored data can create bias in the examiner assignment research design. 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.¹ 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:

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

where Y is a child outcome and T denotes treatment assignment: substantiated but not removed (t_S) or substantiated and removed (t_R).² $P_S(Z) = p_S$ 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 fixing the conviction threshold. While Arteaga (2019) proposes three complementary strategies to do so, the study itself only has access to

¹This 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.

²This is equivalent to Equation 13 in Arteaga (2019).

censored data and thus can not empirically assess whether these strategies actually 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 foster care while additionally controlling for linear and quadratic terms of P_S and all interactions. The second, called the tercile approach, instruments for placement with P_R^* separately for each tercile of the evidence stringency distribution. The idea is that, in addition to controlling for evidence stringency, splitting the data into terciles approximates fixing the evidence threshold. Lastly, the third approach, called the rolling window approach, mirrors the tercile approach yet estimates 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 A.18 shows the results of the first two approaches and Figure A.2 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 both foster placement and substantiation. The table and figure show the effects on the index of child wellbeing.

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 yet the true 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 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 shown when using the standard examiner assignment design with censored data in Section 1.6—they understate the benefits of foster care. While 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 A.18: Assessing Arteaga (2019) Approaches to Examiner Assignment Design with Censored Data

	Tercile Approach			
	(1) Pooled Approach	(2) Lenient in Evidence	(3) Middle in Evidence	(4) Strict in Evidence
<i>Panel A: Censored Data</i>				
Foster Care	0.028 (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

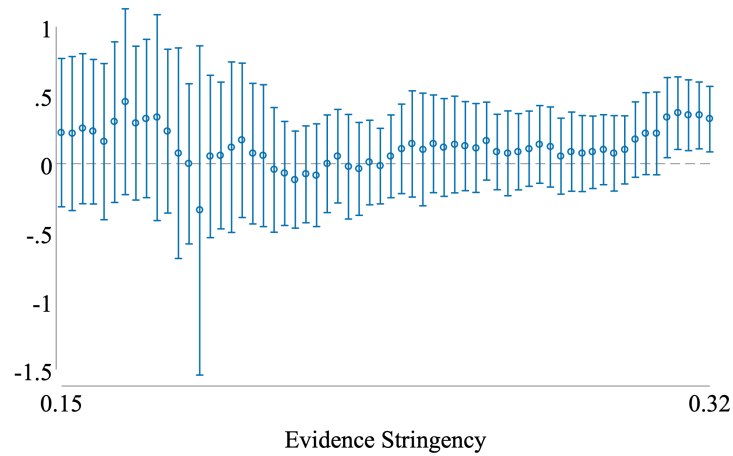
Notes. This table compares the estimates of foster care relative to substantiation on the index of child wellbeing using approaches proposed in Arteaga (2019). All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by 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 A.2: 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 wellbeing using the rolling window approach proposed in Arteaga (2019) with both the censored and full data. 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 A.2a 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 my study, I 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 A.2b 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 with the full data, I use a rolling window of 90,000 observations and adjust the window by 2,500 observations each time.

A.1.2 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 that 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 toward kinship care, 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 while 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.

I add to this evidence by testing the effects of various types of foster placement. I can not 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, I use OLS to describe how the effects of removal vary based on initial placement type. Specifically, I estimate the following model:

$$Y_{iw} = \beta_0 + \beta_1 KINSHIP_{iw} + \beta_2 UNRELATED_{iw} + \beta_3 GROUP_{iw} + \beta_4 X_{iw} + \theta_r + \epsilon_{iw} \quad (\text{A.2})$$

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 A.19 shows the results. Overall, placement with relatives was associated with greater improvements than placement with an unrelated foster family or in a group home. Notably, the OLS estimates in the main analysis understated the benefits of removal and overstated 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 A.19: OLS Effects of Foster Care on Child Outcomes, by Initial Placement Type

	(1)	(2)	(3)	(4)
	Index of Child Wellbeing	Confirmed Victim of Maltreatment	Daily Attendance Rate	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.

A.1.3 Who Takes in Foster Children?

The administrative records in this study do not contain 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. 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 A.20 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 A.20: 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.

A.1.4 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 probabilistic 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 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 my analysis, I restrict the sample to include maltreatment reports that entered the investigator rotational assignment system and involved children enrolled in public school. Table A.21 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 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 nearly all investigations. Finally, restriction nine ensures that I 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 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 a public school student record. 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 my analysis sample to those of school-age children that were excluded, along the observable characteristics included in the child welfare files.

Table A.22 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 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, I estimate the relative share of children that were excluded from my analysis for each of the five reasons listed above. Table A.23 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 A.21: 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. 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, 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 A.1) and, 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 A.22: 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 one 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 ten in Table A.21). Column two 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 during their 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 first grade—at least seven years old.

Table A.23: 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) - 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: $P(\text{inv}) = 0.23$, $P(\text{inv}|\text{low income}) = 0.38$, $P(\text{inv}|\text{high income}) = 0.08$ and I assume that the probability of being investigated conditional on income level is the same across public and private schools.

A.1.5 Reconciling Differences Between the Sample and Outcomes in Doyle (2007)

Doyle (2007) 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 to children in the same age window as Doyle (2007) 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 A.24).

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. Doyle (2007) 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, Doyle (2007) finds that removal reduced annual earnings by \$1300 for adults at ages 18 to 28 years old. A back of the envelope calculation using estimates from Deming et al. (2016) suggests that an increase in standardized math test scores of 0.34 standard deviations increases earnings at age 25 by about \$500.³ 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 foster children in Illinois between 1990 and 2002.

³I use information from Deming et al. (2016, 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.

Table A.24: Effects of Foster Care on Child Outcomes for Sample Comparable to Doyle (2007)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Index of Child Wellbeing	Alleged Victim of Maltreatment	Confirmed Victim of Maltreatment	Daily Attendance Rate	Retained in Grade	Std Math Score	Std Reading Score	Juvenile Delinquency
Foster Care	0.223*** (0.076)	-0.162** (0.068)	-0.067** (0.032)	0.061** (0.028)	-0.025 (0.031)	0.467** (0.194)	0.203 (0.208)	-0.026 (0.040)
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. All regressions include the covariates as listed in the text and zipcode by investigation year fixed effects. Standard errors are clustered by investigator. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

APPENDIX B

Appendix to The Effect of Course Shutouts on Community College Students: Evidence from Waitlist Cutoffs

B.1 Supplemental Tables

Table B.1: Student Initial Education Goal

Included in Sample	Code	Description
Yes	A	Obtain an associate degree and transfer to a 4-year institution
Yes	B	Transfer to a 4-year institution without an associate degree
Yes	C	Obtain a two year associate degree without transfer
	D	Obtain a two year vocational degree without transfer
	E	Earn a vocational certificate without transfer
	F	Discover/formulate career interests, plans, goals
	G	Prepare for a new career (acquire job skills)
	H	Advance in current job/career (update job skills)
	I	Maintain certificate or license (e.g., Nursing, Real Estate)
	J	Educational development (intellectual, cultural); often recreational course-takers
	K	Improve basic skills in English, reading, or math
	L	Complete credits for high school diploma or GED; often high school students
Yes	M	Undecided on goal
	N	To move from noncredit coursework to credit course work
	O	4 year college student taking courses to meet 4 year college requirement
	X	Uncollected/unreported
	Y	Not Applicable

Notes: At application, students are asked to indicate their initial educational goal from the above list. The sample is restricted to community college students who might consider a bachelors degree at a four-year institution a reasonable substitute to their current program.

Table B.2: Effect of Missing the Waitlist Cutoff on Transfers to Nearby Two-Year Schools within Two Years

Outcome	Foothill (1)	Evergreen Valley (2)	San Jose City (3)	Other Two-Year (4)
2SLS	-0.008 (0.010)	0.013** (0.006)	0.015** (0.006)	0.017** (0.008)
Reduced Form	-0.005 (0.007)	0.008* (0.004)	0.010** (0.004)	0.011** (0.005)
CCM	0.051	0.013	0.013	0.027
Observations (N_l/N_r)	1,977	2,281		

Notes: This table shows results from a 2SLS regression as in equation 3. The outcomes are indicators for whether the student transferred to Foothill College within two years of the waitlist in Column (1), Evergreen Valley College in Column (2), San Jose City College in Column (3), and any other two-year college in Column (4). The standard errors are in parentheses, with the control complier means (CCM) and the reduced form displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.3: Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Gender

	# Courses Enrolled in Concurrent Term			Enrolled
	Zero (1)	One or Two (2)	Three or More (3)	Next Term (4)
Male	0.016 (0.020)	0.014 (0.031)	-0.030 (0.032)	-0.004 (0.030)
CCM Male N Male (N_l/N_r)	0.108 963/1,142	0.310	0.581	0.689
Female	0.034* (0.021)	-0.028 (0.029)	-0.006 (0.029)	-0.031 (0.029)
CCM Female N Female (N_l/N_r)	0.100 1,012/ 1,131	0.362	0.537	0.687
P-value Male=Female	0.530	0.313	0.575	0.525

Notes: This table shows results from a 2SLS regression as in equation 3, where effects are estimated separately by gender. The outcome is an indicator for whether the student took no courses in the concurrent term in Column (1), took one or two courses in Column (2), or took three or more courses in Column (3). A course is counted if the student is enrolled after the add/drop date. The outcome in column (4) is an indicator for whether the student enrolls in any classes the following major term. The standard errors are in parentheses, with the control complier means (CCM) and p-value from a test for the difference in point estimates between groups displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.4: Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Ethnicity

	# Courses Enrolled in Concurrent Term			Enrolled
	Zero (1)	One or Two (2)	Three or More (3)	Next Term (4)
Asian	0.027 (0.022)	-0.006 (0.032)	-0.022 (0.033)	-0.023 (0.031)
CCM Asian N Asian (N_l/N_r)	0.094 860/988	0.287	0.619	0.758
White	0.022 (0.032)	0.013 (0.046)	-0.035 (0.047)	-0.047 (0.046)
CCM White N White (N_l/N_r)	0.121 484/ 517	0.362	0.518	0.654
URM	-0.000 (0.027)	-0.016 (0.041)	0.017 (0.042)	0.034 (0.042)
CCM URM N URM (N_l/N_r)	0.101 478/617	0.390	0.509	0.623
P-value White=Asian	0.885	0.737	0.820	0.659
P-value URM=Asian	0.425	0.841	0.477	0.274
P-value URM=White	0.598	0.635	0.416	0.190

Notes: This table shows results from a 2SLS regression as in equation 3, where effects are estimated separately by ethnicity. The outcome is an indicator for whether the student took no courses in the concurrent term in Column (1), took one or two courses in Column (2), or took three or more courses in Column (3). A course is counted if the student is enrolled after the add/drop date. The outcome in column (4) is an indicator for whether the student enrolls in any classes the following major term. The standard errors are in parentheses, with the control complier means (CCM) and p-value from a test for the difference in point estimates between groups displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.5: Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Course Popularity

	# Courses Enrolled in Concurrent Term			Enrolled
	Zero (1)	One or Two (2)	Three or More (3)	Next Term (4)
Top 5	0.057 (0.037)	-0.107* (0.064)	0.050 (0.066)	-0.008 (0.066)
CCM Top 5 N Top 5 (N_l/N_r)	0.071 170/209	0.304	0.625	0.705
Other Courses	0.024 (0.015)	0.003 (0.022)	-0.027 (0.023)	-0.025 (0.022)
CCM Other N Other (N_l/N_r)	0.107 1,807/2,072	0.340	0.552	0.687
P-value Top 5= Other	0.401	0.106	0.274	0.803

Notes: This table shows results from a 2SLS regression as in equation 3, where effects are estimated separately by popularity of the course. Top five courses are those that are the most frequently requested. The outcome is an indicator for whether the student took no courses in the concurrent term in Column (1), took one or two courses in Column (2), or took three or more courses in Column (3). A course is counted if the student is enrolled after the add/drop date. The outcome in column (4) is an indicator for whether the student enrolls in any classes the following major term. The standard errors are in parentheses, with the control complier means (CCM) and p-value from a test for the difference in point estimates between groups displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.6: Effect of Missing the Waitlist Cutoff on Course Load and Persistence, by Waitlisted Subject

	# Courses Enrolled in Concurrent Term			Enrolled
	Zero (1)	One or Two (2)	Three or More (3)	Next Term (4)
STEM	0.007 (0.024)	0.018 (0.035)	-0.025 (0.036)	-0.011 (0.034)
CCM STEM N STEM (N_l/N_r)	0.098 678/771	0.294	0.607	0.738
Arts/Humanities	0.053** (0.027)	-0.051 (0.038)	-0.002 (0.039)	0.021 (0.039)
CCM Arts/Hum. N Arts/Hum. (N_l/N_r)	0.099 542/663	0.347	0.554	0.649
Social Studies	-0.015 (0.041)	-0.007 (0.062)	0.023 (0.062)	-0.092 (0.062)
CCM Soc. Stud. N Soc. Stud. (N_l/N_r)	0.129 242/261	0.355	0.516	0.658
Other	0.050* (0.030)	0.013 (0.043)	-0.063 (0.044)	-0.033 (0.042)
CCM Other N Other (N_l/N_r)	0.105 515/586	0.373	0.521	0.680
P-value STEM=Arts/Hum	0.194	0.178	0.663	0.532
P-value STEM=Soc. Stud.	0.646	0.717	0.509	0.255
P-value STEM=Other	0.261	0.928	0.500	0.689
P-value Arts/Hum = Soc. Stud	0.164	0.545	0.742	0.124
P-value Arts/Hum= Other	0.940	0.264	0.295	0.348
P-value Soc. Stud.=Other	0.203	0.785	0.259	0.433

Notes: This table shows results from a 2SLS regression as in equation 3, where effects are estimated separately by the subject of the waitlisted course. Top five courses are those that are the most frequently requested. The outcome is an indicator for whether the student took no courses in the concurrent term in Column (1), took one or two courses in Column (2), or took three or more courses in Column (3). A course is counted if the student is enrolled after the add/drop date. The outcome in column (4) is an indicator for whether the student enrolls in any classes the following major term. The standard errors are in parentheses, with the control complier means (CCM) and p-value from a test for the difference in point estimates between groups displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.7: Effect of Missing the Waitlist Cutoff on Transfers to Other Two-Year Schools, by Ethnicity

	Within 1 Year (1)	Within 2 Years (2)	Within 3 Years (3)	Within 4 Years (4)	Within 5 Years (5)
Asian	0.012 (0.016)	0.030 (0.022)	0.007 (0.027)	0.007 (0.029)	-0.000 (0.031)
CCM Asian N Asian (N_l/N_r)	0.046 860/988	0.090	0.148	0.185	0.217
White	-0.017 (0.026)	0.028 (0.033)	-0.015 (0.038)	-0.020 (0.041)	-0.042 (0.043)
CCM White N White (N_l/N_r)	0.072 484/ 517	0.117	0.176	0.215	0.257
URM	0.027 (0.023)	0.055* (0.030)	0.066* (0.034)	0.072** (0.037)	0.099** (0.039)
CCM URM N URM (N_l/N_r)	0.049 478/617	0.110	0.114	0.175	0.206
P-value Asian=White	0.352	0.968	0.641	0.590	0.426
P-value Asian=URM	0.587	0.508	0.171	0.165	0.043
P-value White=URM	0.208	0.556	0.113	0.094	0.014

Notes: This table shows results from a 2SLS regression as in equation 3, where effects are estimated separately by ethnicity. The outcomes are indicators for transfers to other two-year schools at different time horizons: within one through five years of the waitlisted term. The standard errors are in parentheses, with the control complier means (CCM) and p-value from a test for the difference in point estimates between groups displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.8: Effect of Missing the Waitlist Cutoff on Transfers to Four-Year Schools, by Ethnicity

	Within 1 Year (1)	Within 2 Years (2)	Within 3 Years (3)	Within 4 Years (4)	Within 5 Years (5)
Asian	0.014 (0.012)	0.035* (0.020)	0.057** (0.028)	0.055* (0.032)	0.052 (0.033)
CCM Asian N Asian (N_l/N_r)	0.025 860/988	0.063	0.142	0.206	0.237
White	-0.017 (0.020)	-0.033 (0.027)	-0.044 (0.036)	-0.016 (0.040)	0.009 (0.043)
CCM White N White (N_l/N_r)	0.046 484/ 517	0.085	0.173	0.212	0.251
URM	-0.011 (0.014)	-0.012 (0.019)	-0.031 (0.027)	-0.038 (0.032)	-0.048 (0.034)
CCM URM N URM (N_l/N_r)	0.021 478/617	0.034	0.092	0.135	0.156
P-value Asian=White	0.182	0.042	0.028	0.169	0.431
P-value Asian=URM	0.170	0.091	0.024	0.040	0.036
P-value White=URM	0.800	0.512	0.772	0.664	0.295

Notes: This table shows results from a 2SLS regression as in equation 3, where effects are estimated separately by ethnicity. The outcomes are indicators for transfers to four-year schools at different time horizons: within one through five years of the waitlisted term. The standard errors are in parentheses, with the control complier means (CCM) and p-value from a test for the difference in point estimates between groups displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.9: Effect of Missing the Waitlist Cutoff on Bachelors Degree Completion, by Ethnicity

	Within 1 Year (1)	Within 2 Years (2)	Within 3 Years (3)	Within 4 Years (4)	Within 5 Years (5)
Asian	0.007 (0.004)	0.004 (0.006)	0.016* (0.009)	0.020 (0.018)	0.063** (0.025)
CCM N Asian (N_l/N_r)	0.002 860/988	0.006	0.012	0.040	0.094
White	-0.002 (0.007)	-0.008 (0.009)	-0.017 (0.015)	-0.039 (0.024)	-0.075** (0.031)
CCM N White (N_l/N_r)	0.007 484/ 517	0.013	0.029	0.068	0.134
URM	0.002 (0.002)	0.002 (0.002)	0.008 (0.006)	-0.014 (0.016)	-0.025 (0.021)
CCM N URM (N_l/N_r)	0.000 478/617	0.000	0.003	0.025	0.058
P-value Asian=White	0.279	0.246	0.057	0.050	0.001
P-value Asian=URM	0.276	0.639	0.437	0.160	0.007
P-value White=URM	0.593	0.298	0.121	0.392	0.188

Notes: This table shows results from a 2SLS regression as in equation 3, where effects are estimated separately by ethnicity. The outcomes are indicators for bachelors degree completion at different time horizons: within one through five years of the waitlisted term. The standard errors are in parentheses, with the control complier means (CCM) and p-value from a test for the difference in point estimates between groups displayed below. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.10: Reduced Form Effect of Missing a Placebo Cutoff

Cutoff	Enrolled in Zero Courses, Concurrent Term (1)	Transfer to Other 2 Year, Within 2 Years (2)	Observations (N_l/N_r) (3)
0 (Real)	0.016* (0.009)	0.023** (0.010)	1,977/2,281
1	0.008 (0.010)	0.009 (0.010)	2,281/1,955
2	0.014 (0.011)	-0.006 (0.011)	1,955/1,613
3	-0.004 (0.013)	0.004 (0.012)	1,613/1,377
4	0.019 (0.014)	0.007 (0.013)	1,377/1,269
5	-0.009 (0.015)	-0.019 (0.014)	1,269/1,068
6	0.019 (0.016)	0.011 (0.015)	1,068/1,024
7	0.002 (0.016)	-0.005 (0.015)	1,024/969
8	-0.027 (0.017)	-0.002 (0.016)	969/746
9	0.019 (0.018)	0.016 (0.018)	746/684
10	0.013 (0.021)	-0.018 (0.019)	684/579

Notes: This table shows the coefficient from a regression of the outcome on an indicator for missing the placebo cutoff- equal to one if the student has the running variable of the cutoff plus one. For each row, the sample includes only students with running variable equal to the cutoff value and one plus the cutoff. The outcome in column (1) is an indicator for being enrolled in zero courses after drop date in the waitlisted term. The outcome in column (2) is an indicator for being enrolled in another two-year school within two years. All columns include the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.11: Robustness Checks

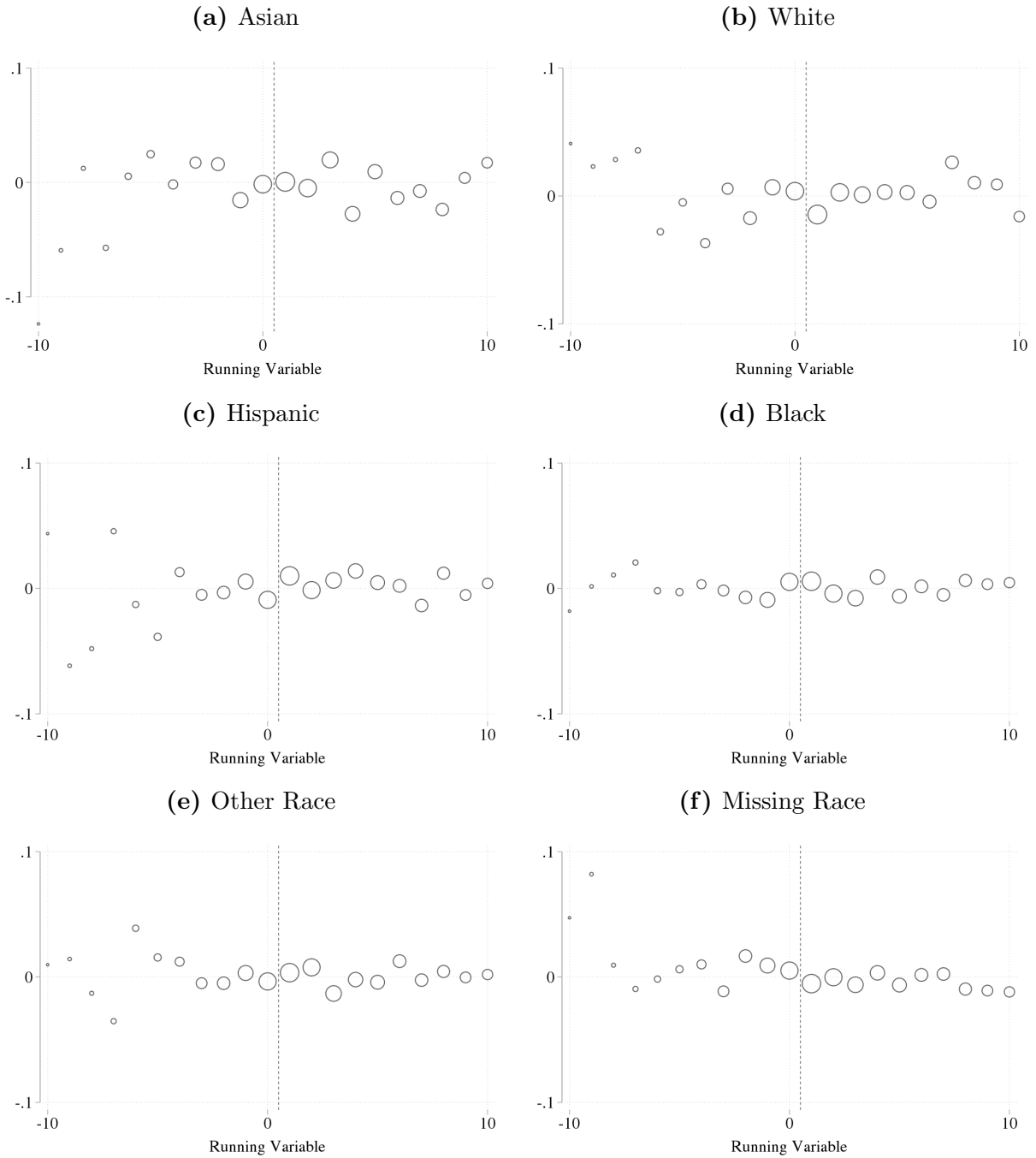
Outcome	No Sample Restrictions (1)	Intend to Transfer (2)	Post 2007 (3)	Course Enrollment (4)	Continuous RDD (5)
Enrolled in 0 Courses	0.012*** (0.003)	0.012*** (0.004)	0.019*** (0.005)	0.025* (0.014)	0.023** (0.012)
Enrolled in 1-2 Courses	0.003 (0.005)	0.005 (0.006)	-0.007 (0.008)	-0.008 (0.021)	-0.003 (0.016)
Enrolled in 3+ Courses	-0.015*** (0.005)	-0.017*** (0.007)	-0.012 (0.008)	-0.017 (0.021)	-0.020 (0.017)
Observations (N_l/N_r)	30,329/37,103	17,338/20,873	11,932/14,533	1,977/2,281	6,659/12,986

Notes: This table shows the coefficient from a 2SLS regression as in equation 3. Column (1) includes all students, and all waitlists. Column (2) includes only students who declared an intention to transfer to a four year school upon enrolling at De Anza, and all waitlists. Column (3) restricts the sample to observations after 2007 and, to increase statistical power, includes all students and all waitlists. Column (4) uses the sample in the main analysis but uses the waitlist cutoff to instrument for course enrollment, rather than course section enrollment. Column (5) uses the sample in the main analysis with a continuous regression discontinuous design with a bandwidth of ten and a linear function form. The standard errors are in parentheses. All columns include the covariates listed in Table 2.4. Standard errors are clustered at the student level when more than one observation per student is used, and are robust to heteroskedasticity otherwise. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

B.2 Covariate Smoothness and Reduced Form Figures Further from the Cutoff

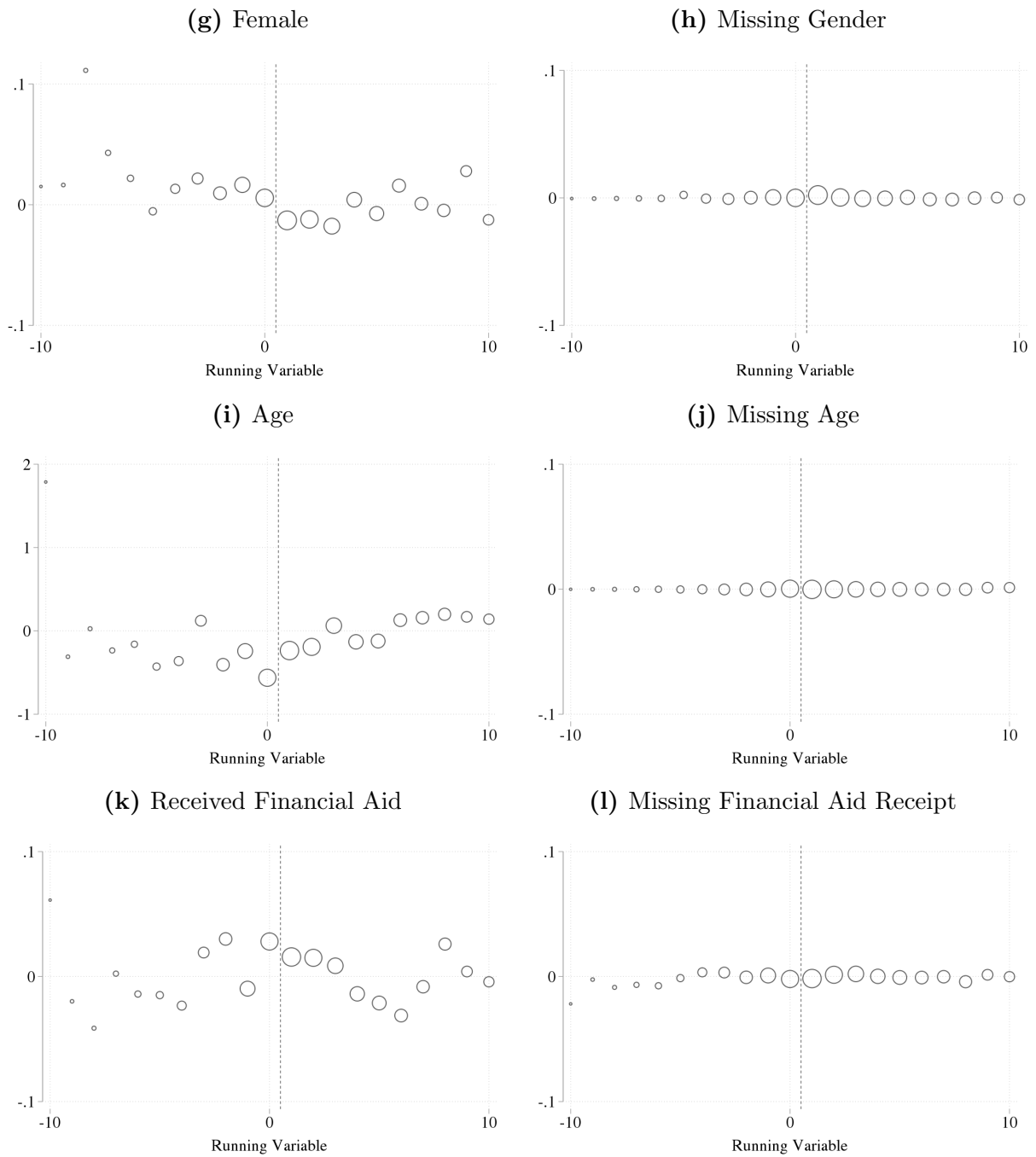
The main analysis relies on a local randomization approach to identify the effects of course shutouts on student outcomes. We use local randomization because the running variable is discrete. As such, the estimates are identified only from variation between students assigned a running variable of zero and one. Although we do not identify effects off of variation from values of the running variable further from the cutoff, it may still be useful to see the larger picture. This section shows a variety of figures in the spirit of a regression discontinuity design with a continuous running variable to show smoothness in both pre-determined covariates and discontinuities in the main outcome variables across the cutoff.

Figure B.1: Covariate Smoothness Across the Waitlist Cutoff



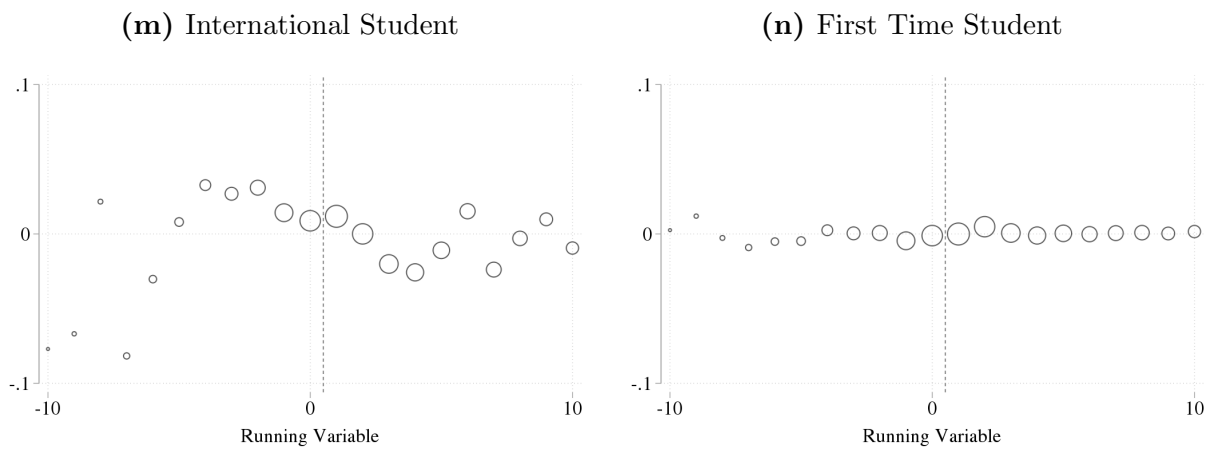
Notes. Each dot represents the mean of the residualized covariate, conditioned on the value of the running variable. The covariates are residualized by term by year fixed effects, registration priority fixed effects, and indicators for special student categories. The size of the dot reflects the number of observations in each bin.

Figure B.1: Covariate Smoothness Across the Waitlist Cutoff



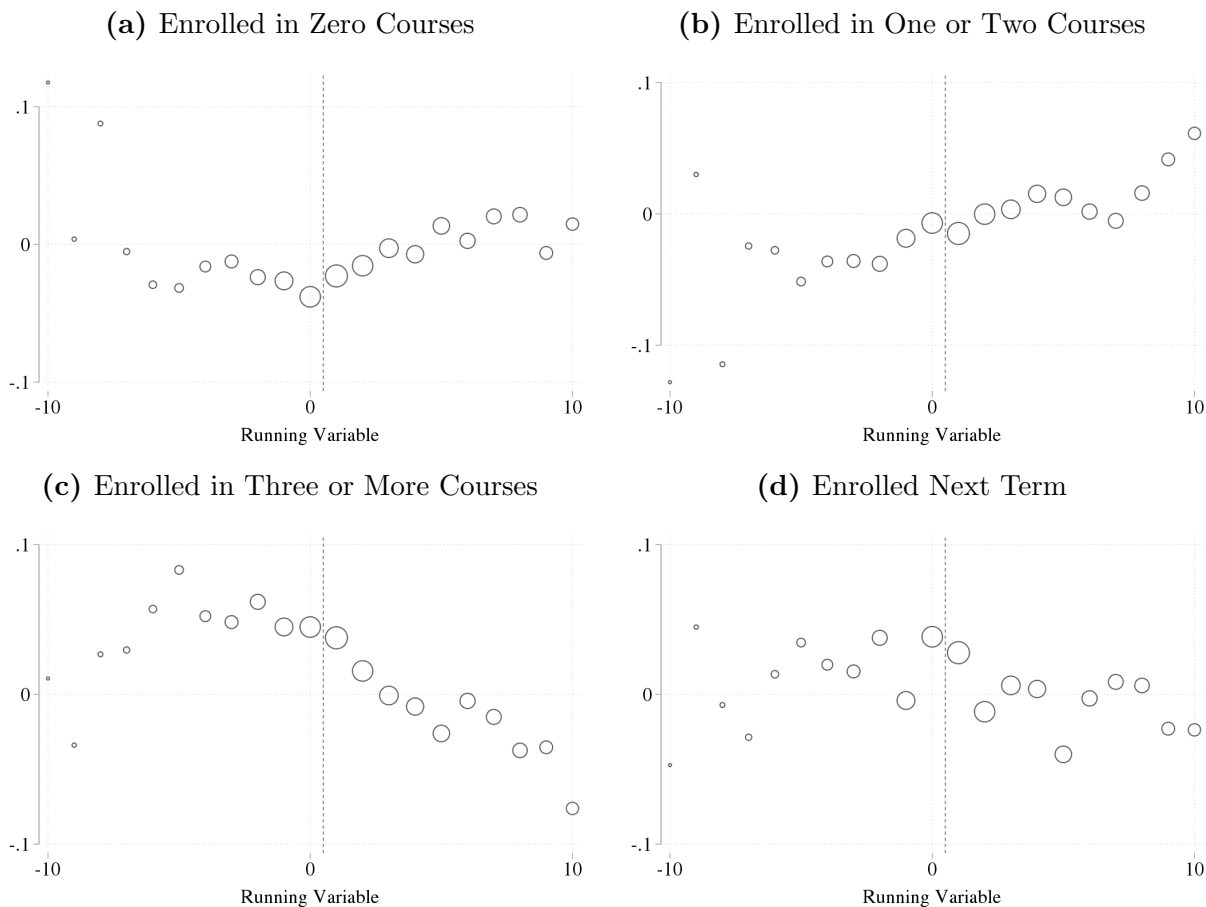
Notes. Each dot represents the mean of the residualized covariate, conditioned on the value of the running variable. The covariates are residualized by term by year fixed effects, registration priority fixed effects, and indicators for special student categories. The size of the dot reflects the number of observations in each bin.

Figure B.1: Covariate Smoothness Across the Waitlist Cutoff



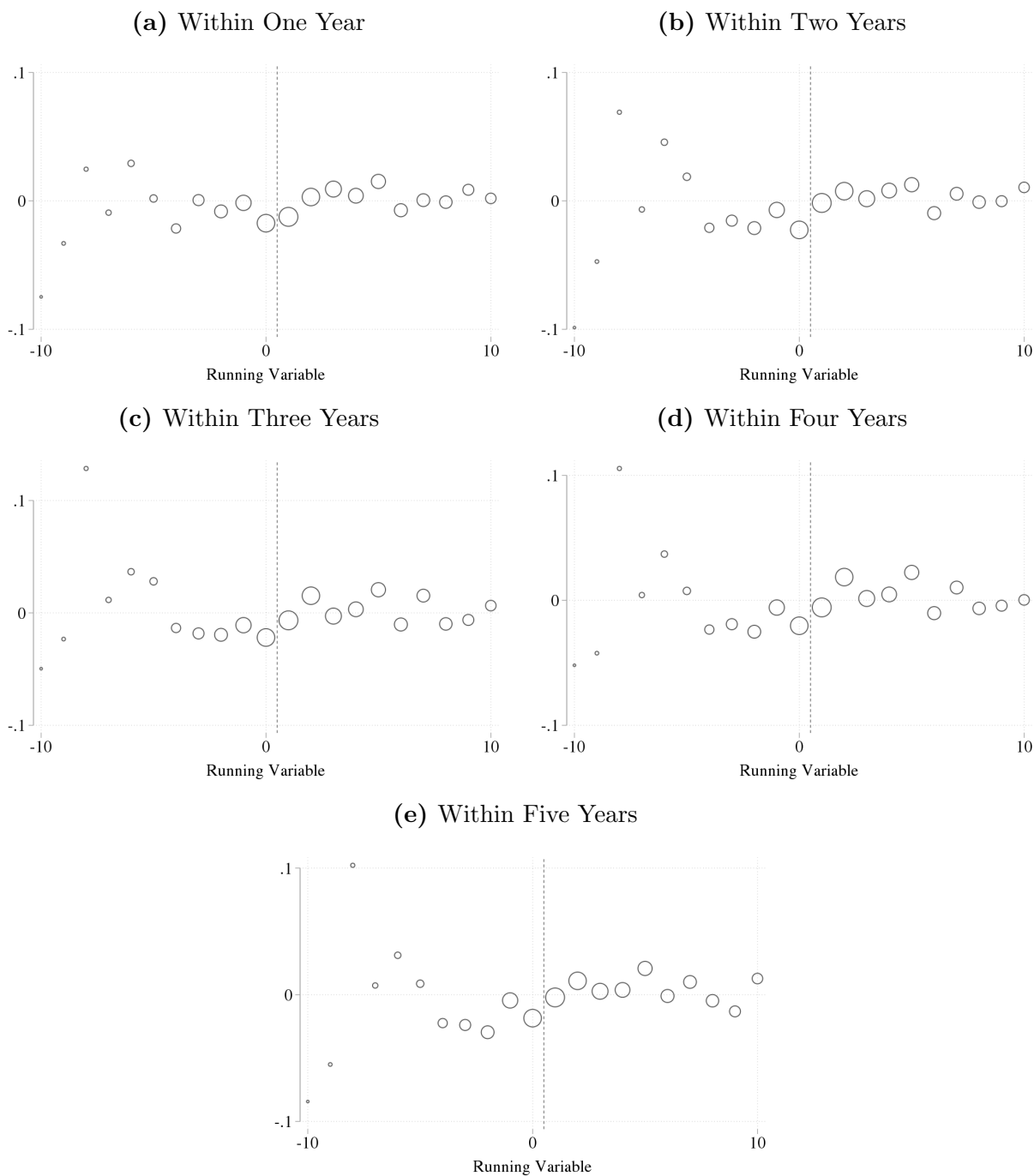
Notes. Each dot represents the mean of the residualized covariate, conditioned on the value of the running variable. The covariates are residualized by term by year fixed effects, registration priority fixed effects, and indicators for special student categories. The size of the dot reflects the number of observations in each bin.

Figure B.2: Reduced Form Effect of Missing the Waitlist Cutoff on Course Load and Persistence



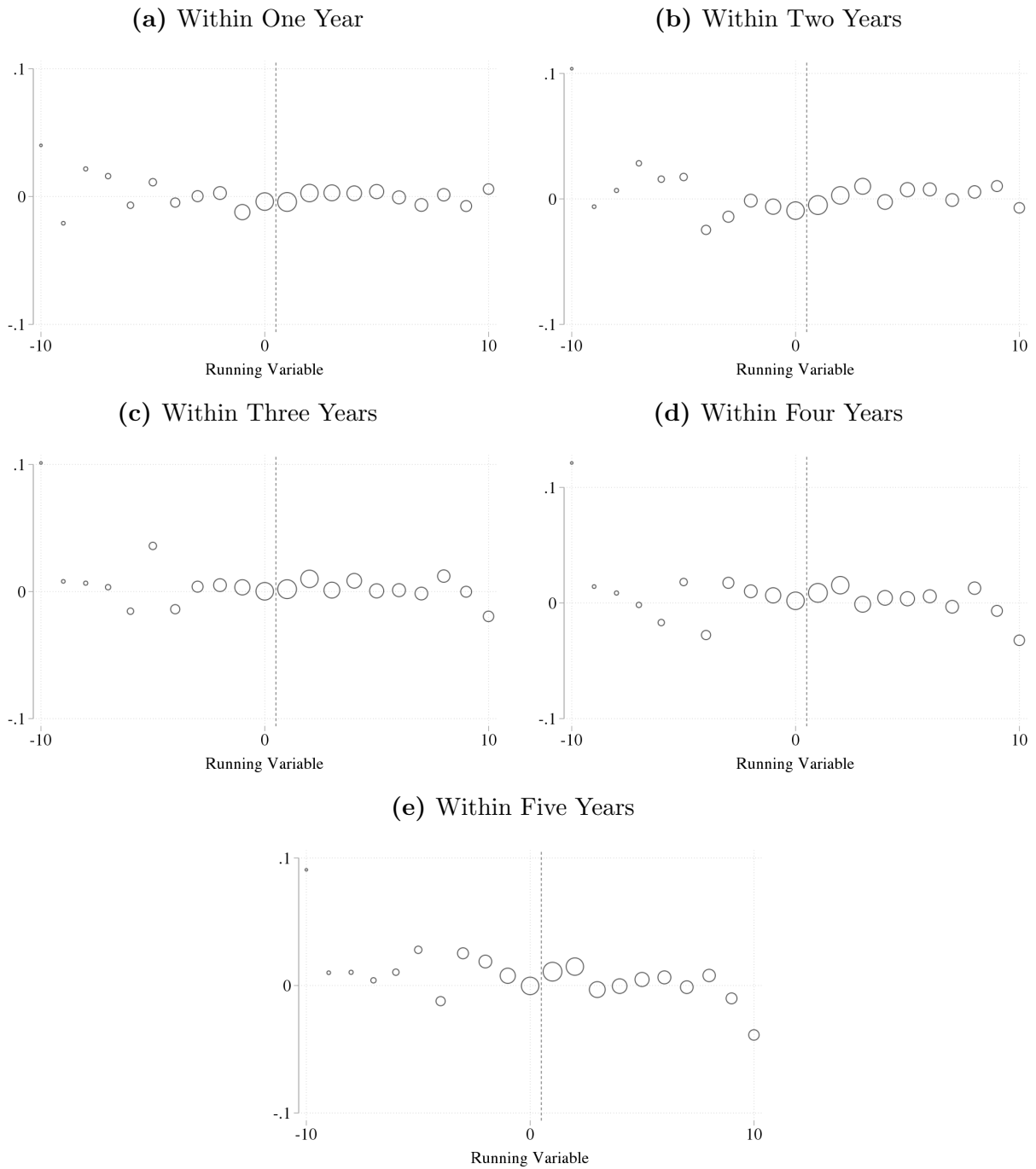
Notes. Each dot represents the mean of the residualized outcome, conditioned on the value of the running variable. The covariates are residualized by the control variables described in Table 2.4. Enrollment during the waitlisted term is defined as being enrolled after the add/drop period. The size of the dot reflects the number of observations in each bin.

Figure B.3: Reduced Form Effect of Missing the Waitlist Cutoff on Transfers to Another Two-Year School



Notes. Each dot represents the mean of the residualized outcome, conditioned on the value of the running variable. The covariates are residualized by the control variables described in Table 2.4. The size of the dot reflects the number of observations in each bin.

Figure B.4: Reduced Form Effect of Missing the Waitlist Cutoff on Transfers to a Four-Year School



Notes. Each dot represents the mean of the residualized outcome, conditioned on the value of the running variable. The covariates are residualized by the control variables described in Table 2.4. The size of the dot reflects the number of observations in each bin.

B.3 Using Time as the Running Variable

The analysis uses a highly discrete running variable, which necessitates local randomization assumptions. Alternatively, the running variable can be framed as a continuous measure if it is redefined in terms of registration time. The discrete running variable used in the main analysis is the “position RV” while this new continuous version is the “time RV.”

Consider the time of day that each waitlisted student made her registration attempt. The time when the student with a position RV equal to zero signed up for the waitlist creates a cutoff in registration time. Students who signed up to the waitlist before this time could enroll in the section during the registration period (ie. had a negative position RV) while those who signed up after could not (ie. had a positive position RV). Therefore, the time RV is the amount of time, in hours, between when a student signed up for the waitlist and when the student with a position RV of zero registered. In this sense, the analysis compares students who missed the waitlist cutoff to those who just made it, within a window of hours around the cutoff time.¹

Figure B.5 shows the density of the time RV, using the analysis sample without the restriction of students being within one position of the cutoff. Note that there is a large spike at zero. This is a mechanical result due to the definition of the time RV. There is not a natural way to set the cutoff, therefore a position of zero is defined using the position RV from the main analysis. This forces many students to be at or near the cutoff artificially. For this reason, the density fails the manipulation test proposed in McCrary (2008) as well as the more recently proposed test in Cattaneo et al. (2017a). However, there is little chance that the density is a result of systematic manipulation rather than an artifact of the variable definition. The main argument for identification is that since the time RV, like the position RV, depends on the number of other students who drop, students cannot easily control it.

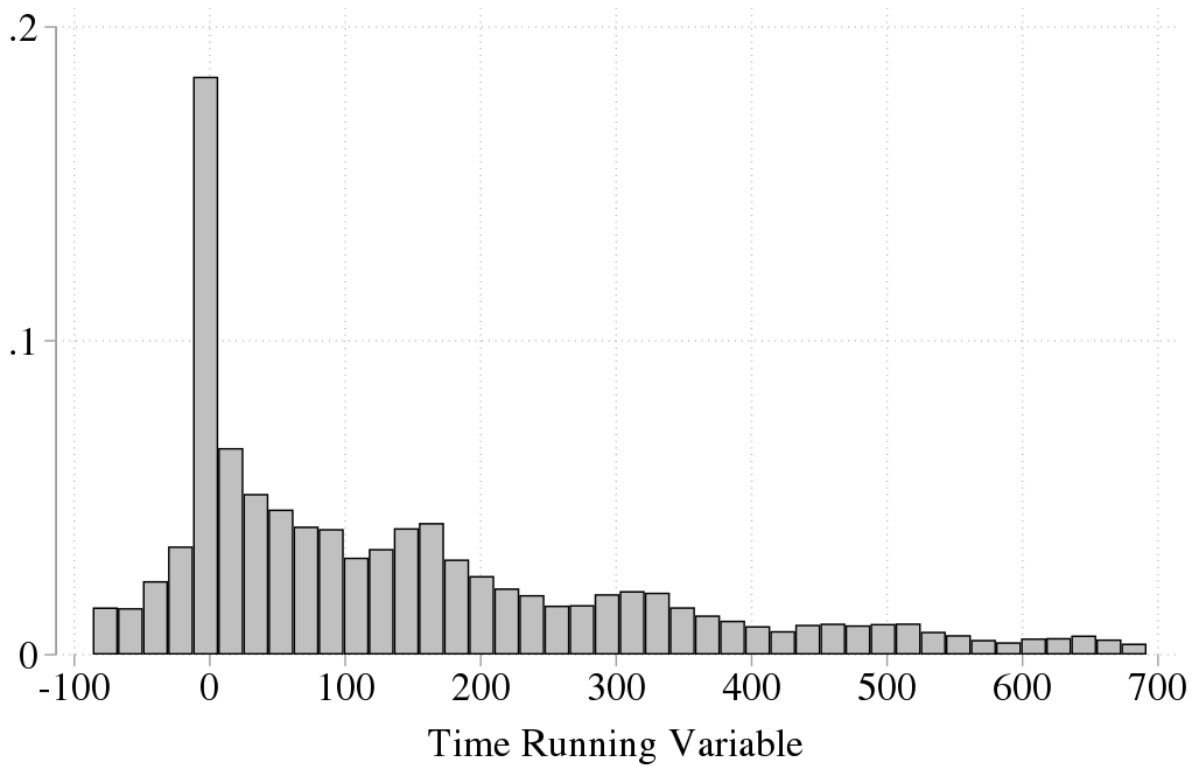
Figure B.6 plots section enrollment rates at the end of the advanced registration period binned by values of the time running variable. There is a clearly visible jump in enrollment to the left of the cutoff. Table B.12 shows formal estimates of the first stage and confirms that there is a discontinuity in the probability of section enrollment. Students who miss the waitlist cutoff are 82 percentage points more likely to be shut out of their desired section during the advanced registration period, and similarly unlikely to enroll in their desired course during advanced registration. These discontinuities are larger than those in the main

¹There are 2 edge cases in which it is not possible to compute a time RV for waitlisted students in a section. First, if enough previously enrolled students drop during the registration period such that everyone who signed up for the waitlist is able to get a seat, then there is no student with a position RV equal to zero. Second, if no previously enrolled students drop such that nobody who signed up to the waitlist is able to get a seat during the registration period, then there is also no student with a position RV equal to zero. The analysis drops these attempts, which amount to just over 4% of the registration attempts in the sample.

analysis, which were 64 and 65 percentage points, respectively.

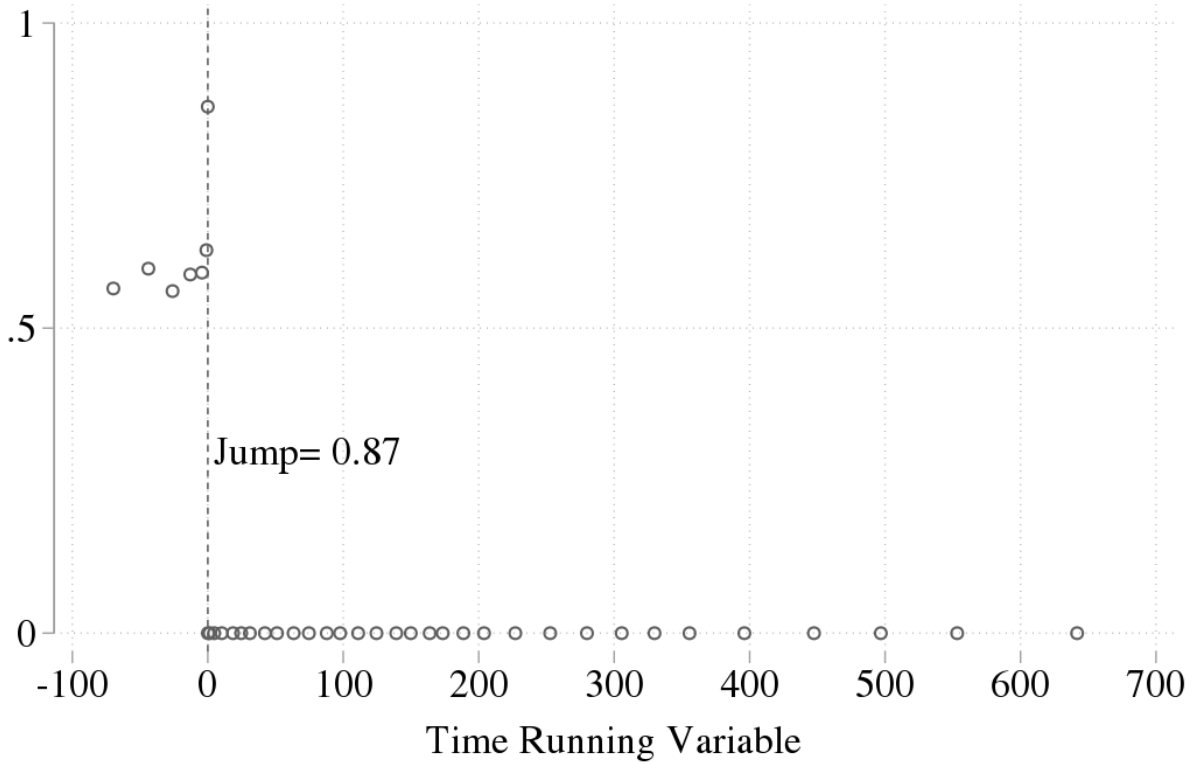
Table B.13 shows the estimates of the LATE on enrollment patterns in the concurrent term. The results are nearly identical to the main analysis. There is a 2.8 percentage point increase in the likelihood of taking no courses in the waitlisted term. The analysis cannot detect a change in the share of students who enroll part-time, or full-time, though the magnitudes of these are smaller than the drop-out estimate. These results almost perfectly line up with the main specification; not being able to enroll in a desired section leads to same-term drop-out.

Figure B.5: Density of the Time Running Variable



Notes. The time running variable is the amount of time, in hours, between when a student signed up for the waitlist and when the student with a position RV of zero registered. The figure censors time running variables smaller than the 10th and greater than the 90th percentile in order to make it more easily interpretable.

Figure B.6: First Stage Effect of Missing the Waitlist Cutoff on Enrollment in the Waitlisted Section



Notes. Each dot represents section enrollment binned in forty quantiles by the value of the time running variable, where enrollment is equal to one if the student was enrolled in the waitlisted section at the end of the advanced registration period. Section enrollment is equal to zero for students with a time running variable greater than zero by construction. The time running variable is the amount of time, in hours, between when a student signed up for the waitlist and when the student with a position RV of zero registered. The figure censors time running variables smaller than the 10th and greater than the 90th percentile in order to make it more easily interpretable.

Table B.12: First Stage Effect of Missing the Waitlist Cutoff on Enrollment in Waitlisted Section, Using Time Running Variable

	(1)	(2)
	Enrolled in Section	Enrolled in Section
Missed WL Cutoff	-0.818*** (0.009)	-0.817*** (0.009)
Observations (N_l/N_r)	2404/1285	2403/1282
CCT BW	10.797	10.748
Controls	N	Y

Notes: Results are from a local linear regression using time as the continuous running variable. The dependent variable is enrollment in the waitlisted section, where enrollment is equal to one if the student was enrolled at the end of the advanced registration period. The bandwidth is calculated according to the CCT optimal bandwidth selection procedure. The first column does not include controls while the second controls for the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table B.13: Effect of Missing the Waitlist Cutoff on Course Load, Using Time Running Variable

	# Courses Enrolled in Concurrent Term		
	Zero (1)	One or Two (2)	Three or More (3)
2SLS	0.028* (0.016)	-0.010 (0.023)	-0.017 (0.023)
Reduced Form	0.022* (0.013)	-0.008 (0.018)	-0.013 (0.018)
Observations (N_l/N_r)	3,057/2,582	3,054/2,574	3,175/2,831
CCT BW	29.877	29.684	35.336

Notes: Results are from a local linear regression using time as the continuous running variable. The outcome is an indicator for whether the student took no courses in the concurrent term in Column (1), took one or two courses in Column (2), or took three or more courses in Column (3). A course is counted if the student is enrolled after the add/drop date. The bandwidth is calculated according to the CCT optimal bandwidth selection procedure. The first column does not include controls while the second controls for the covariates listed in Table 2.4. Standard errors are robust to heteroskedasticity. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

APPENDIX C

Appendix to The Effect of Summer Employment on the Educational Attainment of Under-Resourced Youth

Table C.1: Balance Tests of Differences Between Junior Police and Fire Cadet Participants and Age-Eligible Non-Participants

	(1)	(2)
	Comparison Mean	Participated in JPC
Special Education	0.107	0.012 (0.027)
Low Income	0.863	0.025 (0.030)
% Below Poverty Line in Neighborhood	34.678	3.414** (1.513)
Attendance Rate	0.919	0.005 (0.007)
Chronically Absent	0.283	-0.015 (0.045)
Took 8th Grade Math Test	0.911	-0.006 (0.021)
Proficient on 8th Grade Math Test	0.115	0.029 (0.024)
Standardized 8th Grade Math Score	-0.580	0.051 (0.067)
Took 8th Grade Reading Test	0.913	-0.001 (0.022)
Proficient on 8th Grade Reading Test	0.450	-0.018 (0.036)
Standardized 8th Grade Reading Score	-0.454	-0.089 (0.070)
N	5,321	

Notes. This table reports the results from a weighted regression of a baseline characteristic on an indicator for participated in Junior Police and Fire Cadets (JPC) and match group fixed effects, where participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. The age-eligible non-participants are applicants who were born between 1/1/1999 and 7/1/2001 and did not participate in JPC. Column 1 shows the weighted mean of the baseline characteristic for non-participants and column 2 shows the coefficient on the indicator for participation in JPC . All baseline characteristics are measured in the match year, except for 8th grade test scores. Youth who did not reach 8th grade before the program are not included in the analysis of 8th grade test scores. Standard errors are clustered by match school. Stars indicate: *p<0.1, ** p<0.05, *** p<0.01.

Table C.2: Description of the Sample for Each Outcome and the Formula for Outcomes in Follow-up Years One and Two

	Sample Description		Formula for FY1 and FY2
	Follow-up Year One	Follow-up Year Two	
Enrolled in K12	Grade 11 or below in 2014-2015 SY	Grade 10 or below in 2014-2015 SY	Avg of FY1 and FY2
Attendance Rate	Grade 11 or below in 2014-2015 SY and enrolled during 2015-2016 SY	Grade 10 or below in 2014-2015 SY and enrolled during 2016-2017 SY	Avg of FY1 and FY2
Chronically Absent	Grade 11 or below in 2014-2015 SY and enrolled during 2015-2016 SY	Grade 10 or below in 2014-2015 SY and enrolled during 2016-2017 SY	Avg of FY1 and FY2
Expelled	Grade 11 or below in 2014-2015 SY and enrolled during 2015-2016 SY	Grade 10 or below in 2014-2015 SY and enrolled during 2016-2017 SY	Avg of FY1 and FY2
Took SAT	Grade 10 in 2014-2015 SY	Grade 9 in 2014-2015 SY or Grade 10 in 2014-2015 but did not take SAT in 2015-2016	Max of FY1 and FY2
SAT Composite Score	Took SAT in 2015-2016 SY and Grade 10 in 2014-2015 SY	Took SAT in 2016-2017 and Grade 9 in 2014-2015 SY or Grade 10 in 2014-2015 and did not take SAT in 2015-2016	Max of FY1 and FY2
Graduated HS	Grade 11 in 2014-2015 SY and did not Graduate HS in 2014-2015 SY	Grade 10 during 2014-2015 SY or Grade 11 in 2014-2015 SY but did not Graduate HS in 2015-2016	Max of FY1 and FY2

Notes. This table describes the samples for each outcome variable in follow-up year one and follow-up year two, as well as the formula used to calculate the outcome in follow-up years one and two. FY1 indicates follow-up year one and FY2 indicates follow-up year two.

Table C.2: Description of the Sample for Each Outcome and the Formula for Outcomes in Follow-up Years One and Two
(Continued)

	Sample Description		Formula for FY1 and FY2
	Follow-up Year One	Follow-up Year Two	
College from HS	Grade 12 in 2014-2015 SY	Grades 11 or 12 in 2014-2015 SY	Avg of FY1 and FY2
2 Year College from HS	Grade 12 in 2014-2015 SY	Grades 11 or 12 in 2014-2015 SY	Avg of FY1 and FY2
4 Year College from HS	Grade 12 in 2014-2015 SY	Grades 11 or 12 in 2014-2015 SY	Avg of FY1 and FY2
College as a HS Grad	Graduated HS before 2014-2015 SY	Graduated HS before 2014-2015 SY	Avg of FY1 and FY2
2 Year College as a HS Grad	Graduated HS before 2014-2015 SY	Graduated HS before 2014-2015 SY	Avg of FY1 and FY2
4 Year College as a HS Grad	Graduated HS before 2014-2015 SY	Graduated HS before 2014-2015 SY	Avg of FY1 and FY2

Notes. This table describes the samples for each outcome variable in follow-up year one and follow-up year two, as well as the formula used to calculate the outcome in follow-up years one and two. FY1 indicates follow-up year one and FY2 indicates follow-up year two.

Table C.3: The Effect of Participation in GDYT on Educational Outcomes in Follow-up Year One and Follow-up Year Two

	Follow-up Year One			Follow-up Year Two		
	(1) N	(2) Comparison Mean	(3) Participated	(4) N	(5) Comparison Mean	(6) Participated
Enrolled in K12	6,340	0.965	0.013*** (0.005)	4,929	0.914	0.014 (0.010)
Attendance Rate	5,034	0.889	0.002 (0.004)	3,684	0.878	0.004 (0.006)
Chronically Absent	5,034	0.375	0.007 (0.015)	3,684	0.398	-0.013 (0.022)
Expelled	6,138	0.002	-0.000 (0.001)	4,547	0.004	-0.002 (0.002)
Took SAT	1,956	0.729	0.023 (0.026)	2,593	0.568	0.038* (0.021)
SAT Composite Score	1,487	862.427	0.423 (6.603)	1,525	858.083	-0.530 (5.196)
Graduated HS	1,375	0.861	0.006 (0.019)	1,871	0.797	0.038* (0.022)
College from HS	1,073	0.481	0.022 (0.052)	2,484	0.404	-0.001 (0.028)
2 Year College from HS	1,073	0.247	0.035 (0.047)	2,484	0.189	0.004 (0.024)
4 Year College from HS	1,073	0.234	-0.013 (0.030)	2,484	0.215	-0.005 (0.020)
College as a HS Grad	1,701	0.516	0.042 (0.041)	1,701	0.381	0.035 (0.045)
2 Year College as a HS Grad	1,701	0.246	0.014 (0.037)	1,701	0.164	0.015 (0.033)
4 Year College as a HS Grad	1,701	0.270	0.028 (0.027)	1,701	0.217	0.020 (0.033)

Notes. This table reports the results from a weighted regression of an educational outcome on an indicator for participated, a vector of control variables as listed in the text and match group fixed effects, separately for follow-up year one and follow-up year two. Participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. For follow-up year one, column 1 reports the number of observations in the sample, column 2 reports the weighted mean of the baseline characteristic for non-participants and column 3 shows the coefficient on the indicator for participation. Columns 4 through 6 show similar results for follow-up year two. Standard errors are clustered by match school. Stars indicate: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Balance Tests of Differences Between Participants and Non-Participants, by Subgroup

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	9th Grade and Below	10th and 11 Grade	Male	Female	Above Median Math Score	Below Median Math Score	Not Chronically Absent	Chronically Absent
Special Education	0.010 (0.021)	0.003 (0.016)	0.010 (0.018)	0.016 (0.011)	0.009 (0.009)	-0.000 (0.021)	0.012 (0.013)	0.016 (0.032)
Low Income	-0.022 (0.014)	-0.000 (0.017)	0.006 (0.016)	-0.001 (0.012)	-0.021 (0.019)	0.012 (0.012)	0.007 (0.017)	-0.011 (0.026)
% Below Poverty Line in Neighborhood	-0.495 (0.887)	0.187 (0.883)	0.449 (0.879)	-0.395 (0.665)	0.719 (0.860)	-1.835* (1.013)	0.059 (0.948)	0.334 (1.584)
Attendance Rate	0.005 (0.006)	0.022*** (0.005)	0.013*** (0.004)	0.016*** (0.005)	0.006** (0.003)	0.021*** (0.007)	0.002* (0.001)	0.027*** (0.009)
Chronically Absent	0.021 (0.029)	-0.077*** (0.024)	-0.067** (0.026)	-0.027 (0.020)	-0.023 (0.021)	-0.091*** (0.034)		
Took G8 Math Test	0.001 (0.016)	-0.007 (0.016)	-0.004 (0.016)	-0.009 (0.012)			-0.006 (0.011)	-0.000 (0.024)
Proficient on G8 Math Test	0.035** (0.015)	-0.003 (0.019)	0.035** (0.015)	-0.002 (0.012)	0.027 (0.021)	0.000 (0.000)	0.001 (0.015)	0.021 (0.020)
Standardized G8 Math Score	0.043 (0.047)	-0.024 (0.041)	0.035 (0.036)	0.008 (0.032)	0.030 (0.032)	-0.006 (0.016)	-0.017 (0.029)	0.035 (0.059)

Notes. This table reports the results from a weighted regression of a baseline characteristic on an indicator for participated and match group fixed effects, separately for each subgroup. Participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. All baseline characteristics are measured in the match year, except for 8th grade test scores. Youth who did not reach 8th grade before the program are not included in the analysis of 8th grade test scores. Columns 5 and 6 are restricted to youth who reached 8th grade before the program since we define prior achievement by 8th grade test scores. Standard errors are clustered by match school. Stars indicate: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Balance Tests of Differences Between Participants and Non-Participants, by Subgroup (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	9th Grade and Below	10th and 11 Grade	Male	Female	Above Median Math Score	Below Median Math Score	Not Chronically Absent	Chronically Absent
Took G8 Reading Test	0.001 (0.017)	-0.006 (0.017)	-0.006 (0.015)	-0.007 (0.011)	-0.001 (0.003)	0.001 (0.003)	-0.002 (0.011)	-0.011 (0.022)
Proficient on G8 Reading Test	0.014 (0.022)	0.019 (0.029)	0.015 (0.028)	0.017 (0.018)	0.050* (0.027)	0.006 (0.022)	0.023 (0.026)	0.025 (0.023)
Standardized G8 Reading Score	-0.003 (0.045)	-0.013 (0.052)	0.003 (0.049)	-0.020 (0.036)	0.029 (0.053)	0.007 (0.049)	-0.023 (0.044)	-0.024 (0.050)
N	3900	4036	4789	6748	4949	4901	5842	3204

Notes. This table reports the results from a weighted regression of a baseline characteristic on an indicator for participated and match group fixed effects, separately for each subgroup. Participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. All baseline characteristics are measured in the match year, except for 8th grade test scores. Youth who did not reach 8th grade before the program are not included in the analysis of 8th grade test scores. Columns 5 and 6 are restricted to youth who reached 8th grade before the program since we define prior achievement by 8th grade test scores. Standard errors are clustered by match school. Stars indicate: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: Falsification Tests of the Effect of Participation in GDYT on 6th Grade Educational Outcomes, by Subgroup

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	9th Grade and Below	10th and 11 Grade	Male	Female	Above Median Math Score	Below Median Math Score	Not Chronically Absent	Chronically Absent
Attendance Rate	-0.002 (0.005)	0.015*** (0.004)	0.008 (0.005)	0.001 (0.004)	0.001 (0.004)	0.009 (0.006)	0.004 (0.003)	-0.004 (0.010)
Chronically Absent	0.011 (0.023)	-0.051** (0.023)	-0.023 (0.024)	-0.003 (0.014)	0.005 (0.018)	-0.053** (0.026)	-0.019 (0.018)	-0.026 (0.031)
Took Math Exam	-0.006 (0.014)	0.007 (0.006)	0.003 (0.009)	0.005 (0.008)	-0.005 (0.008)	0.002 (0.013)	-0.005 (0.006)	-0.009 (0.014)
Proficient on Math	-0.023 (0.020)	-0.001 (0.021)	-0.025 (0.018)	-0.000 (0.013)	-0.018 (0.019)	-0.007 (0.016)	-0.008 (0.018)	-0.007 (0.027)
Std Math Score	-0.073 (0.049)	-0.036 (0.030)	-0.038 (0.040)	-0.042 (0.026)	-0.032 (0.029)	-0.067 (0.050)	-0.035 (0.037)	-0.028 (0.044)
Took Reading Exam	-0.012 (0.014)	0.005 (0.006)	-0.004 (0.008)	0.001 (0.009)	-0.010 (0.008)	-0.002 (0.011)	-0.003 (0.007)	-0.019 (0.014)
Proficient on Reading	-0.027 (0.024)	-0.004 (0.020)	-0.007 (0.021)	-0.001 (0.020)	-0.002 (0.021)	-0.006 (0.023)	0.007 (0.023)	0.026 (0.038)
Std Reading Score	-0.083** (0.038)	0.012 (0.043)	0.001 (0.045)	-0.034 (0.031)	-0.030 (0.030)	-0.037 (0.046)	-0.005 (0.039)	-0.001 (0.069)
N	3900	4036	4789	6748	4949	4901	5842	3204

Notes. This table reports the results from a weighted regression of a 6th grade educational outcome on an indicator for participated, a vector of control variables as listed in the text and match group fixed effects, separately for each subgroup. Participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. Columns 5 and 6 are restricted to youth who reached 8th grade before the program since we define prior achievement by 8th grade test scores. Standard errors are clustered by match school. Stars indicate: *p<0.1, ** p<0.05, *** p<0.01.

Table C.6: The Effect of Participation in GDYT on Additional Educational Outcomes, by Subgroup

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	9th Grade and Below	10th and 11 Grade	Male	Female	Above Median Math Score	Below Median Math Score	Not Chronically Absent	Chronically Absent
Attendance Rate	0.002 (0.005) [0.885]	0.001 (0.005) [0.867] {0.771}	0.002 (0.006) [0.879]	0.002 (0.005) [0.873] {0.981}	0.001 (0.005) [0.900]	0.003 (0.009) [0.865] {0.894}	-0.002 (0.004) [0.925]	-0.002 (0.016) [0.781] {0.965}
Expelled	-0.002 (0.002) [0.004]	-0.000 (0.001) [0.001] {0.545}	-0.001 (0.002) [0.002]	-0.001 (0.001) [0.003] {0.782}	-0.002 (0.002) [0.002]	0.001 (0.002) [0.003] {0.149}	-0.000 (0.001) [0.001]	0.001 (0.002) [0.002] {0.565}
Took SAT	0.038 (0.025) [0.713]	0.029 (0.024) [0.745] {0.781}	0.033 (0.024) [0.710]	0.033 (0.023) [0.743] {0.995}	-0.003 (0.025) [0.831]	0.049* (0.029) [0.707] {0.183}	0.011 (0.022) [0.870]	0.011 (0.051) [0.577] {0.993}
SAT Composite Score	-1.045 (5.054) [859.284]	0.921 (6.404) [861.239] {0.817}	6.529 (5.683) [853.922]	-5.610 (5.234) [865.057] {0.123}	7.539 (6.954) [937.142]	2.957 (6.007) [793.735] {0.603}	3.097 (6.043) [884.968]	-6.487 (15.635) [821.039] {0.476}
2 Year College from HS		0.016 (0.029) [0.197]	0.010 (0.037) [0.199]	0.017 (0.031) [0.219] {0.883}	0.035 (0.044) [0.213]	-0.005 (0.039) [0.219] {0.420}	-0.031 (0.038) [0.235]	0.053 (0.044) [0.196] {0.068}
4 Year College from HS		-0.012 (0.024) [0.256]	0.036 (0.034) [0.176]	-0.041 (0.025) [0.275] {0.095}	-0.024 (0.025) [0.344]	0.014 (0.037) [0.128] {0.338}	-0.004 (0.029) [0.319]	-0.045 (0.034) [0.139] {0.307}

Notes. This table reports the results from a weighted regression of an educational outcome on an indicator for participated, a vector of control variables as listed in the text and match group fixed effects, separately for each subgroup. Participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. Standard errors, which are clustered by match school, are reported in parentheses and the weighted mean of the baseline characteristic for non-participants is shown in brackets below. In the even columns, we report the p-value from a test of whether the point estimates are equal across subgroups in curly braces. Columns 5 and 6 are restricted to youth who reached 8th grade before the program since we define prior achievement by 8th grade test scores. Stars indicate: *p<0.1, ** p<0.05, *** p<0.01.

Table C.6: The Effect of Participation in GDYT on Additional Educational Outcomes, by Subgroup (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	9th Grade and Below	10th and 11 Grade	Male	Female	Above Median Math Score	Below Median Math Score	Not Chronically Absent	Chronically Absent
College as a HS Grad			-0.013 (0.056) [0.389]	0.052 (0.051) [0.487] {0.323}	0.030 (0.046) [0.550]	0.042 (0.089) [0.356] {0.879}	0.007 (0.049) [0.532]	0.083 (0.059) [0.352] {0.162}
2 Year College as a HS Grad			-0.057 (0.056) [0.232]	0.038 (0.032) [0.206] {0.091}	0.022 (0.045) [0.231]	-0.003 (0.069) [0.215] {0.664}	-0.040 (0.040) [0.215]	0.077 (0.053) [0.223] {0.024}
4 Year College as a HS Grad			0.044 (0.050) [0.163]	0.014 (0.036) [0.295] {0.614}	0.001 (0.039) [0.336]	0.051 (0.057) [0.144] {0.375}	0.043 (0.047) [0.329]	0.006 (0.039) [0.138] {0.423}

Notes. This table reports the results from a weighted regression of an educational outcome on an indicator for participated, a vector of control variables as listed in the text and match group fixed effects, separately for each subgroup. Participants have a weight of one and non-participants have a weight equal to the ratio of participants to non-participants in their match group. Standard errors, which are clustered by match school, are reported in parentheses and the weighted mean of the baseline characteristic for non-participants is shown in brackets below. In the even columns, we report the p-value from a test of whether the point estimates are equal across subgroups in curly braces. Columns 5 and 6 are restricted to youth who reached 8th grade before the program since we define prior achievement by 8th grade test scores. Stars indicate: *p<0.1, ** p<0.05, *** p<0.01.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Abdulkadiroglu, A., Pathak, P. A., and Walters, C. R. (2018). Free to choose: Can school choice reduce student achievement? *American Economic Journal: Applied Economics*, 10:175–206.
- AECF (2017). Kids count data center. Technical report, The Annie E. Casey Foundation. <https://datacenter.kidcount.org>.
- Aizer, A. and Doyle, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, 130(2):759–803.
- Alfeld, C. (2016). Career technical education is growing; research must follow. Technical report, Inside IES Research.
- Altonji, J., Elder, T., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184.
- Anderson, M. L., Gallagher, J., and Ritchie, E. R. (2018). School meal quality and academic performance. *Journal of Public Economics*, 168:81–93.
- Angrist, J., Cohodes, S., Dynarski, S., Pathak, P., and Walters, C. R. (2016). Stand and deliver: Effects of boston’s charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*, 34(2).
- Arteaga, C. (2019). The cost of bad parents: Evidence from the effects of parental incarceration on children’s education. Working paper.
- Autor, D., Figlio, D., Karbownik, K., Roth, J., and Wasserman, M. (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economics Journal: Applied Economics*, 11(3):338–381.
- Bald, A., Chyn, E., Hastings, J. S., and Machelett, M. (2019). The causal impact of removing children from abusive and neglectful homes. National Bureau of Economic Research Working Paper 25419.
- Barnow, B. S. and Smith, J. (2015). Employment and training programs. In *Economics of Means-Tested Transfer Programs in the United States, Volume 2*.

- Barrat, V. X. and Berliner, B. (2013). The invisible achievement gap, part 1: Education outcomes of students in foster care in california's public schools, part one. Technical report, WestEd.
- Barrow, L., Richburg-Hayes, L., Rouse, C. E., and Brock, T. (2014). Paying for performance: the education impacts of a community college scholarship program for low-income adults. *Journal of Labor Economics*, 32(3):563–599.
- Belotti, J., Rosenberg, L., Sattar, S., Esposito, A. M., and Ziegler, J. (2010). Reinvesting in america's youth: Lessons from the 2009 recovery act summer youth employment initiative. Technical report, Mathematica Policy Research.
- Berrick, J. D. (2018). *The Impossible Imperative: Navigating the competing principles of child protection*. Oxford University Press.
- Bertrand, M. and Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economics Journal: Applied Economics*, 5(1):32–64.
- Berzin, S. C. (2010). Understanding foster youth outcomes: Is propensity scoring better than traditional methods? *Research on Social Work Practice*, 20(1):100–111.
- Bettinger, E. P. and Baker, R. B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, 36(1):3–19.
- Bhuller, M., Dahl, G. B., Loken, K. V., and Mogstad, M. (2018). Intergenerational effects of incarceration. In *AEA Papers and Proceedings*, volume 108, pages 234–40.
- Billings, S. B. (2019). Parental arrest and incarceration: How does it impact the children? Working paper.
- Bohn, S., Reyes, B., and Johnson, H. (2013). The impact of budget cuts on california's community colleges. Technical report, Public Policy Institute of California.
- Bosk, E. A. (2015). *All Unhappy Families: Standardization and Child Welfare Decision-Making*. PhD thesis, University of Michigan.
- Bound, J., Lovenheim, M., and Turner, S. (2010). Why have college completion rates declined? an analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, 2:129–157.
- Bound, J., Lovenheim, M., and Turner, S. (2012). Increasing time to baccalaureate degree in the united states. *Education Finance and Policy*, 7:375–424.
- Bound, J. and Turner, S. (2007). Cohort crowding: How resources affect collegiate attainment. *Journal of Public Economics*, 91:877–899.
- Carrell, S. E. and Hoekstra, M. (2014). Are school counselors an effective education input? *Economics Letters*, 125(1):66–69.

- Cattaneo, M., Idrobo, N., and Titiunik, R. (2018). A practical introduction to regression discontinuity designs: Volume i and ii. In Alvarez, R. M. and Beck, N., editors, *Cambridge Elements: Quantitative and Computational Methods for Social Science*. Cambridge University Press, Cambridge.
- Cattaneo, M., Jansson, M., and Ma, X. (2017a). Simple local regression distribution estimators. *Working Paper*.
- Cattaneo, M., Titiunik, R., and Vasquez-Bare, G. (2017b). Comparing inference approaches for rd designs: A reexamination of the effect of head start on child mortality. *Journal of Policy Analysis and Management*, 36:643–681.
- CCCCO (2012). Focus on results– accountability reporting for the california community colleges. A Report to the Legislature, Pursuant to AB 1417 (Pacheco, Stat. 2004, Ch. 581).
- Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers 1: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.
- Chetty, R., Hendren, N., and Katz, L. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.
- ChildrensRights (n.d.). Class actions. <https://www.childrensrights.org/our-campaigns/class-actions/>.
- ChildTrends (2017). Michigan foster care, federal fiscal year 2015. Technical report, Child Trends. https://www.childtrends.org/wp-content/uploads/2017/01/Michigan-Foster-Care-Factsheet_2015.pdf.
- ChildTrends (2018). Foster care. Technical report, Child Trends Databank.
- Chyn, E. (2018). Moved to opportunity: The long-run effect of public housing demolition on labor market outcomes of children. *American Economic Review*, 108(10):3028–3056.
- Clifford, S. and Silver-Greenberg, J. (2017). Foster care as punishment: The new reality of ‘jane crow’. *The New York Times*. <https://www.nytimes.com/2017/07/21/nyregion/foster-care-nyc-jane-crow.html>.
- Cohodes, S. and Goodman, J. (2014). Merit aid, college quality and college completion: Massachusetts? adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4).
- Collinson, R. and Reed, D. (2019). The effects of evictions on low-income households. Working paper.
- CRHE (2017). Homeschooling by the numbers. Technical report, Coalition for Responsible Home Education. <https://www.responsiblehomeschooling.org/homeschooling-101/homeschooling-numbers/>.

- Dahl, G. B., Kostøl, A. R., and Mogstad, M. (2014). Family welfare cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752.
- Davis, J. and Heller, S. B. (2017). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *The National Bureau of Economic Research Working Paper No. 23443*.
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *The Review of Economics and Statistics*, 86(1):195–210.
- Deming, D., Cohodes, S., Jennings, J., and Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics*, 98(5):848–862.
- Deming, D., Goldin, C., and Katz, L. (2012). The for-profit postsecondary school sector: Nimble critters or agile predators? *Journal of Economic Perspectives*, 26:139–164.
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640.
- Deming, D. J. and Walters, C. R. (2017). The impact of state budget cuts on u.s. postsecondary attainment. Working Paper 23736, National Bureau of Economic Research.
- Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–40.
- Dollarwise (2016). Summer youth jobs survey summary. Technical report, U.S. Conference of Mayors.
- Doyle, J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review*, 97(5):1583–1610.
- Doyle, J. J. (2008). Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care. *Journal of political Economy*, 116(4):746–770.
- Dunbar, A., Hossler, D., and Shapiro, D. (2011). National postsecondary enrollment trends: Before, during, and after the great recession. Technical report, National Student Clearinghouse Research Center.
- Eren, O. and Mocan, N. (2017). Juvenile punishment, high school graduation and adult crime: Evidence from idiosyncratic judge harshness. National Bureau of Economic Research Working Paper 23573.
- Fairlie, R. W., Hoffmann, F., and Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, 104:2567–2591.
- Figlio, D. and Winicki, J. (2005). Food for thought: The effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89:381–394.

- Fortin, N. (2006). Higher education policies and the college premium: Cross-state evidence from the 1990s. *American Economic Review*, 96:959–987.
- Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity when the running variable is discrete. In *Regression Discontinuity Designs*, volume 38 of *Advances in Econometrics*, pages 281–315.
- Frandsen, B. R., Lefgren, L. J., and Leslie, E. C. (2019). Judging judge fixed effects. National Bureau of Economic Research Working Paper 25528.
- Frisvold, D. E. (2015). Nutrition and cognitive achievement: An evaluation of the school breakfast program. *Journal of Public Economics*, 124:91–104.
- Gelber, A., Isen, A., and Kessler, J. B. (2016). The effects of youth employment: Evidence from new york city lotteries. *The Quarterly Journal of Economics*, 131(1):423–460.
- Goldin, C., Katz, L. F., and Kuziemko, I. (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives*, 20(4):133–156.
- Goodman, J. (2014). Flaking out: Student absences and snow days as disruptions of instructional time. National Bureau of Economic Research Working Paper 20221.
- Gurantz, O. (2015). Who loses out? registration order, course availability, and student behaviors in community college. *Journal in Higher Education*.
- Hayduk, I. (2017). The effect of kinship placement laws on foster children’s well-being. *The B.E. Journal of Economic Analysis & Policy*, 17(1):1–23.
- Heller, S. B. (2014). Summer jobs reduce violence among disadvantaged youth. *Science*, 346(6214):1219–1223.
- Herbst, D. (2018). Liquidity and insurance in student loan contracts: Estimating the effects of income-driven repayment on default and consumption. Working paper.
- HRC (2015). Lgbtq youth in the foster care system. Technical report, Human Rights Campaign.
- Humphries, J. E., Mader, N. S., Tannenbaum, D. I., and van Dijk, W. L. (2019). Does eviction cause poverty? quasi-experimental evidence from cook county, il. National Bureau of Economic Research Working Paper 26139.
- IES (2017). What works clearinghouse version 4.0 standards handbook. Technical report, Institute of Education Sciences.
- IHE (2009). California calamity.
- Imberman, S. A. and Kugler, A. D. (2014). The effect of providing breakfast on student performance. *Journal of Policy Analysis and Management*, 33(669-699).

- Infoplease (2017). Population of the 20 largest u.s. cities, 1900-2012.
- Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review*, 21(6):589–598.
- Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *American Economic Review*, 94(1):233–258.
- Jacob, B. A. (2017). What we know about career and technical education in high school. Technical report, Brookings Institution: Evidence Speaks.
- Kawano, L., Sacerdote, B., Skimmyhorn, W., and Stevens, M. (2017). On the determinants of young adult outcomes: An examination of random shocks to children in military families. Working paper.
- Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review*, 96(3):863–876.
- Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.
- Kling, J. R., Ludwig, J., and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Quarterly Journal of Economics*, 120(1):87–130.
- Kolesar, M. and Rothe, C. (2016). Inference in regression discontinuity designs with a discrete running variable. *Working Paper*.
- Kurlaender, M., Jackson, J., Howell, J. S., and Grodsky, E. (2014). College course scarcity and time to degree. *Economics of Education Review*, 41:24–39.
- Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142:655–674.
- Leos-Urbel, J. (2014). What is a summer job worth? the impact of summer youth employment on academic outcomes. *Journal of Policy Analysis and Management*, 33(4):891–911.
- Leos-Urbel, J., Schwartz, A. E., Weinstein, M., and Corcoran, S. (2013). Not just for poor kids: The impact of universal free school breakfast on meal participation and student outcomes. *Economics of Education Review*, 36:88–107.
- Lovett, N. and Xue, Y. (2018). Family first or the kindness of strangers? foster care placements and adult outcomes. Working paper.
- Mack, J. (2017). Where michigan children attended school in 2016-2017 – public and private. Technical report, MLive. https://www.mlive.com/news/2017/09/where_michigan_children_attend.html.

- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142.
- Mitchell, M., Leachman, M., and Masterson, K. (2016). Funding down, tuition up: State cuts to higher education threaten quality and affordability at public colleges. Technical report, Center on Budget and Policy Priorities.
- Modestino, A.-S. (2017). Reducing inequality summer by summer: An analysis of the short-term and long-term effects of boston's summer youth employment program. Technical report, Dukakis Center for Urban and Regional Policy at Northeastern University.
- Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. Working paper.
- Mulhern, C. (2019). Beyond teachers: Estimating individual guidance counselors' effects on educational attainment. Working paper.
- Neering, K. F. (2018). Course closed: The short- and long-run impacts of course shutouts on university students. Submitted.
- Neilson, S. (2019). More kids are getting placed in foster care because of parents' drug use. *NPR*. <https://www.npr.org/sections/health-shots/2019/07/15/741790195/more-kids-are-getting-placed-in-foster-care-because-of-parents-drug-use>.
- Norris, S. (2019). Examiner inconsistency: Evidence from refugee appeals. Working paper.
- Norris, S., Pecenco, M., and Weaver, J. (2019). The effects of parental and sibling incarceration: Evidence from ohio. Working paper.
- OCA (2018). Examining connecticut's safety net for children withdrawn from school for the purpose of homeschooling. Technical report, Office of the Child Advocate, State of Connecticut.
- O'Hare, W. P. (2007). *Census Bureau Plans to Eliminate 'Foster Child' Category*. Population Reference Bureau. <https://www.prb.org/censusbureauaufosterchildcategory/>.
- Oster, E. (2016). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economics Statistics*, pages 1–18.
- Pears, K. and Fisher, P. A. (2005). Developmental, cognitive, and neuropsychological functioning in preschool-aged foster children: Associations with prior maltreatment and placement history. *Journal of Developmental & Behavioral Pediatrics*, 26(2):112–122.
- Pecora, P. J., Kessler, R. C., O'Brien, K., White, C. R., Williams, J., Hiripi, E., English, D., White, J., and Herrick, M. A. (2006). Educational and employment outcomes of adults formerly placed in foster care: Results from the northwest foster care alumni study. *Children and youth services review*, 28(12):1459–1481.

- Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., and DiSalvo, R. (2017). Stanford education data archive (version 2.0).
- Ringler, D. A. (2018). Office of the auditor general, performance audit report, children's protective services investigations. Technical report, Michigan Department of Health and Human Services.
- Roberts, K. V. (2019). Foster care and child welfare. Working paper.
- Ryan, J. P., Jacob, B. A., Gross, M., Perron, B. E., Moore, A., and Ferguson, S. (2018). Early exposure to child maltreatment and academic outcomes. *Child maltreatment*, 23(4):365–375.
- Ryan, J. P. and Testa, M. F. (2005). Child maltreatment and juvenile delinquency: Investigating the role of placement and placement instability. *Children and youth services review*, 27(3):227–249.
- SAMHSA (2009). Treatment episode dataset (teds) highlights 2007, national admissions to substance abuse treatment services. Technical report, Substance Abuse and Mental Health Services Administration, Office of Applied Studies.
- Sanbonmatsu, L., Kling, J. R., Duncan, G. J., and Brooks-Gunn, J. (2006). Neighborhoods and academic achievement: Results from the moving to opportunity experiment. *Journal of Human Resources*, 41(4):649–691.
- Scherr, T. G. (2007). Educational experiences of children in foster care: Meta-analyses of special education, retention and discipline rates. *School Psychology International*, 28(4):419–436.
- Schwartz, A. E., Leos-Urbel, J., and Wiswall, M. (2015). Making summer matter: The impact of youth employment on academic performance. *The National Bureau of Economic Research Working Paper No. 21470*.
- Schwartz, A. E. and Rothbart, M. W. (2017). Let them eat lunch: The impact of universal free meals on student performance. Working paper.
- Scrivener, S., Weiss, M. J., Ratledge, A., Rudd, T., Sommo, C., and Fresques, H. (2015). Double graduation rates: Three-year effects of cuny's accelerated study in associate programs (asap) for developmental education students. Technical report, MDRC.
- Sengupta, R. and Jepsen, C. (2006). California's community college students. California Counts: Population Trends and Profiles 2, Public Policy Institute of California.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Snyder, T. D., de Brey, C., and Dillow, S. A. (2018). Digest of education statistics 2016.

- Sullivan, C., Sommer, S., and Moff, J. (2001). Youth in the margins: A report on the unmet needs of lesbian, gay, bisexual, and transgender adolescents in foster care. Technical report, Lambda Legal Defense and Education Fund.
- Talbot, M. (2017). The addicts next door. *The New Yorker*. <https://www.newyorker.com/magazine/2017/06/05/the-addicts-next-door>.
- Trout, A. L., Hagaman, J., Casey, K., Reid, R., and Epstein, M. H. (2008). The academic status of children and youth in out-of-home care: A review of the literature. *Children and Youth Services Review*, 30(9):979–994.
- U.S. Department of Education (2017). Digest of education statistics. Technical report, Institute of Education Sciences, National Center for Education Statistics, Washington, DC.
- USDHHS (2003). Child welfare outcomes 2001: Annual report. Technical report, United States Department of Health and Human Services.
- USDHHS (2016a). Child welfare outcomes 2016 report to congress. Technical report, Children’s Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- USDHHS (2016b). *Reunification: Bringing Your Children Home From Foster Care*. Children’s Bureau. <https://www.childwelfare.gov/pubPDFs/reunification.pdf#page=9&view=What%20can%20I%20expect%20after%20my%20children%20come%20home?>
- USDHHS (2017a). Child maltreatment 2015. Technical report, United States Department of Health and Human Services.
- USDHHS (2017b). Child welfare outcomes report data. Technical report, Children’s Bureau, Administration for Children and Families, U.S. Department of Health and Human Services. <https://cwoutcomes.acf.hhs.gov/cwodatasite/byState>.
- USDHHS (2018a). The afcars report: Preliminary fy 2017 estimates. Technical report, United States Department of Health and Human Services, Administration for Children and Families.
- USDHHS (2018b). Child maltreatment 2016. Technical report, Administration for Children and Families, Administration on Children, Youth and Families, Children’s Bureau.
- Valentine, E., Anderson, C., Hassain, F., and Unterman, R. (2017). An introduction to the world of work: A study of the implementation and impacts of new york city’s summer youth employment program. Technical report, MDRC.
- Walters, C. R. (Forthcoming). The demand for effective charter schools. *Journal of Political Economy*.
- White, R. and DeGrow, B. (2016). A survey of michigan’s private education sector. Technical report, Mackinac Center for Public Policy.

- Wildeman, C. and Emanuel, N. (2014). Cumulative risks of foster care placement by age 18 for us children, 2000–2011. *PloS one*, 9(3):e92785.
- Wilson, S., Newell, M., and Fuller, R. (2010). Ready or not, here they come the complete series of undergraduate enrollment demand and capacity projections, 2009–2019. Technical Report 10-08, California Postsecondary Education Commission.
- Wiltz, T. (2018). This new federal law will change foster care as we know it. *PEW Charitable Trusts*. <https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2018/05/02/this-new-federal-law-will-change-foster-care-as-we-know-it>.
- Wulczyn, F., Smithgall, C., and Chen, L. (2009). Child well-being: The intersection of schools and child welfare. *Review of research in education*, 33(1):35–62.
- Zlotnick, C., Tam, T. W., and Soman, L. A. (2012). Life course outcomes on mental and physical health: the impact of foster care on adulthood. *American Journal of Public Health*, 102(3):534–540.