Essays in Applied Microeconomics

By

Claire Hug

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics at Brown University

Providence, Rhode Island

May 2023

This dissertation by Claire Hug is accepted in its present form by the Department of Economics as satisfying the requirement for the degree of Doctor of Philosophy

Date ________________ ________________________________
Emily Oster, Advisor

Recommended to the Graduate Council

Date ________________ ________________________________
Emily Oster, Reader

Date ________________ ________________________________
Anna Aizer, Reader

Date ________________ ________________________________
Bryce Millett Steinberg, Reader

Approved by the Graduate Council

Date ________________ ________________________________
Thomas A. Lewis, Dean of the Graduate School

Vita

Claire Hug received a B.S. in Economics and in Mathematics in 2017 from University of Kansas. She received her M.A. in Economics from Brown University in 2018 and her Ph.D. in 2023.

Preface

This dissertation explores topics in applied microeconomics with a particular focus on policy issues.

Chapter 1, *"Racial Bias in Child Protective Service Decisions and the Role of Emergency Care Providers"*, explores the referral patterns of emergency department providers to CPS when caring for patients with behavioral health conditions. Children of color are disproportionately over-represented in the child welfare system, and this chapter looks at one possible reason contributing to this over-representation. Using clustering techniques and regression analysis, I estimate the relationship between the probability of being placed with CPS on the same day as an ED visit and a child's race. Then, comparing children with similar medical histories and diagnoses, I find that race is not a significant predictor of referral, except in cases where children present with mild illnesses, offering ED providers greater discretion over the referral decision.

Chapter 2, *"Urban Renewal in Chicago"*, studies the impact of the urban renewal program, which demolished and rebuilt dilapidated neighborhoods in the 1950s and 1960s. This paper estimates the effect of these urban renewal projects on neighborhood demographic composition in Chicago. Using a difference-in-differences design to compare project census tracts to census tracts considered as project sites, I find an increase in the share of the population with a college degree in treated tracts. Looking for heterogeneous treatment effects, I find a significant negative impact on the Black population for early projects in predominantly Black neighborhoods.

Chapter 3, *"Post COVID-19 Test Score Recovery: Initial Evidence from State Testing Data*, is coauthored with Clare Halloran, Rebecca Jack, and Emily Oster. This chapter focuses on test scores in the wake of the COVID-19 pandemic. COVID-19 caused significant disruption in schooling in the U.S., and student test scores showed dramatic declines by the end of the 2020-21 school year. We use state test score data to analyze patterns of test score

recovery over the 2021-22 school year. On average, we find that 20% of test score losses are recovered in English language arts (ELA) by 2022, compared to 37% in math. These recovery rates do not significantly vary across demographic characteristics, baseline achievement rates, in-person schooling rates in the pandemic school year, or category-based measures of recovery funding allocations. We observe large state-level variation in recovery rates in ELA – from full recovery to further losses. This evidence suggests state-level factors play an important role in students' academic recovery, but we are unable to isolate particular state factors. Future work should focus on this variation to facilitate a broader recovery effort.

Acknowledgements

I am deeply grateful to my advisors Emily Oster, Bryce Steinberg, and Anna Aizer for their advice and guidance. I thank them for all of their insightful comments without which this research would not have been possible.

I would like to thank Emily Oster for her incredible mentorship and support throughout my time at Brown. I have learned so much from her and grown both professionally and personally. I am extraordinarily lucky to have known her and been able to call her my mentor.

I have also greatly benefited from many of the faculty members and students at Brown, in classes, in seminars and at breakfasts. I am grateful to all of my colleagues and classmates, and I am fortunate to have been in the presence of so many outstanding and thoughtful scholars. I want to thank Angelica Spertini and Ilise Hamilton for all of their help throughout the graduate school experience.

Lastly, I want to thank my friend Abby for always being there for me, throughout not only these past six years but for most of my life. And I would like to thank my dad. I know that I would not be here today without all of his love and encouragement. Thank you for always being my biggest supporter, love you dad.

Contents

List of Tables

List of Figures

CHAPTER 1

RACIAL BIAS IN CHILD PROTECTIVE SERVICE DECISIONS AND THE ROLE OF EMERGENCY CARE PROVIDERS

1.1 Introduction

It has been well-documented that children of color are over-represented in the child welfare system. Over 20% of children identified by Child Protective Services (CPS) as victims are Black, while Black children account for only 14% of the population of children in the United States (Children's Bureau Issue Brief, 2014). We see similar over-representation for American Indian and multiracial children, but less significant over-representation among Hispanic children. (Children's Bureau Issue Brief, 2014).

How much of this over-representation is a result of racial biases among CPS workers and providers responsible for CPS referrals and how much is a result of other systemic biases that cause non-white families to face worse home environments is an open question. This paper addresses the first possibility – that some portion of over-representation of non-white children in the social welfare system is due to disproportionate referral and placement rates by providers and caseworkers, rather than solely due to objectively worse circumstances at home. Specifically, I study racial bias in the referral decisions of emergency care providers.

The goal of this paper is to determine if race is a significant predictor of placement with CPS. I use Rhode Island emergency departments (EDs) as a case study. Emergency care providers are an important avenue for referral to CPS. Children visiting the ED with obvious injuries or significant behavioral health conditions may raise concern among providers and prompt them to refer the family to the Department of Child, Youth, and Family Services (DCYF) – the Rhode Island state agency in charge of investigating child maltreatment claims and placing children in the care of the state as necessary.

To study referral decision-making, I use Rhode Island state administrative records of children under 18 who visit the ED in 2017 for a primary behavioral health diagnosis. The main sources of data are Medicaid insurance claims from 2016-2018 and placement records from DCYF covering the same time period. Medicaid claims contain information about all healthcare children receive, including type of service, procedures performed, diagnoses received, cost, care provider, and date of service. I use the claims to identify ED visits and obtain a complete picture of children's medical histories.

DCYF placement records contain information about placement type, duration, and the date of placement. I use this information to match placements to ED visits based on date. My main outcome measure is the likelihood of placement with DCYF after an ED visit. I find a spike in placements on the day of an ED visit, indicating that children are more likely to be placed with DCYF after they visit the ED. A large proportion of these placements are for assessment and stabilization centers. These DCYF centers provide temporary, short-term placements during a crisis. The child's situation is evaluated in a neutral environment and then caseworkers determine if a longer-term placement, such as foster care, is warranted. This is important to note since children are not jumping directly from the ED to group homes or foster care. They are instead being temporarily removed and set up for an in-depth evaluation of their home-life conditions by DCYF.

Having established that placements do increase on the day of an ED visit, I use regression analysis to estimate the relationship between the probability of placement occurring on the same day as an ED visit and the race of the child. Simply regressing the likelihood of placement on a child's race will result in a biased coefficient. If characteristics that make children more likely to be placed - more severe behavioral illnesses or worse home circumstances - are correlated with race, the regression will capture this effect and overestimate the effect of race. To address this bias, I use cluster analysis to group children based on their medical histories in the 30 days prior to an ED visit. This allows me to compare children who have similar medical histories and present at the ED with similar symptoms. Children in each cluster should only differ in terms of their observable demographic characteristics.

Estimating the relationship between placement on the day of an ED visit and a child's race within each cluster, I find that race is not a significant predictor of placement. There are a two notable exceptions. I find significant effects of race on the likelihood of placement in clusters where children have less severe illnesses. In these clusters Hispanic children are 7.5% more likely to be placed than non-Hispanic children and non-white children are 1.4% more likely to be placed than white children. About 14% of the general population in Rhode Island is Hispanic, making Hispanics the largest minority population in the state. This could contribute to the larger impact of race for Hispanic children, since they are a frequently and easily observed minority.

Finding a significant impact of race on ED placements in two clusters of children with less severe illnesses suggests that we may observe racial bias when providers have more discretion over the referral decision. For example, children entering the ED with very severe behavioral health crises, such as a suicide attempt, or evidence of physical violence may clearly need intervention, regardless of their race. However, children who present with less severe conditions (e.g. a panic attack) may be "on the margin" of being placed, in which case ED providers default to race to help make the decision. If providers perceive non-white children has having more severe conditions or as having families that are less capable of caring for the child while giving white children and families the benefit of the doubt, even when white children exhibit similar symptoms, we will observe a larger share of non-white children being referred to DCYF.

The idea of marginal placements is particularly relevant for the sample studied in this paper. The children in my sample are negatively selected – they have all visited the ED for behavioral health conditions and nearly 28% have contact with DCYF during the study period, which is over five times the amount of contact in the general population (Administration for Children and Families Report, 2017). One possibility for the lack of evidence of racial bias in decision-making may be that the majority of cases in this sample are too severe for ED providers to exercise discretion based on race.

Given that significant results occur only in a few clusters, particularly among children with less severe illnesses, we may wonder if specific providers are driving these results and displaying a racial bias in referral rates. To explore this possibility, I break down placement rates by provider. Of the providers who have at least one placement occurring on the same day as an ED visit, there is variation ranging from 1% to 6% of ED visits ending in placement. This is not an extremely large spread of placement rates, and providers that refer children to DCYF see on average a worse mix of patients. They see a higher proportion of patients with significant behavioral conditions, suggesting that they are not exhibiting racial bias with most patients. The majority of cases in this sample are likely too severe to be able to pick up evidence of racial bias in decision-making, especially at the provider level.

The paper proceeds as follows. Section 2 discusses the current literature about the role of racial bias in clinical and CPS decision-making, and provides background about the DCYF referral process in Rhode Island. Section 3 describes the data. Section 4 explains the methods. Section 5 presents the results and section 6 concludes.

1.2 Background

1.2.1 Racial Bias in Healthcare and CPS Decisions

There is extensive work documenting implicit racial bias among healthcare providers (see Fitzgerald and Hurst 2017 and Maina et al. 2018 for a review). However, evidence on how this bias affects provider decision-making is mixed. Oliver et al. (2013) finds that general racial bias does not predict treatment decisions, but specific perceptions about cooperativeness of white and Black patients can influence which treatment options physicians pursue. Other studies find a significant positive correlation between implicit racial bias and lower quality care, including negative interpersonal interactions (Cooper et al. 2013) and worse long-term clinical relationships (Blair et al. 2013). On the other hand, a systematic review from Dehon et al. (2017) concludes that while physicians do exhibit racial biases, it generally does not influence clinical decision-making.

Turning to the effects of racial biases on children, Sabin et al. (2008) finds that pediatricians hold less implicit racial bias than other MDs, but do not reach a conclusion about the effect of bias on pediatricians' decision-making. Sabin and Greenwald (2012) again study pediatricians and find that racial bias does affect pediatricians' treatment decisions but only for pain management - they are less likely to prescribe pain medication to non-white patients. They find no significant effects on treatment plans for asthma and ADHD.

Considering if biases are exacerbated in emergency settings, Johnson et al. (2016) finds that the extreme stressors encountered when working in an ED increase racial biases, even though biases remain stable during times outside of an ED shift. They suggest that emergency care providers may be less well-equipped to confront their personal biases when providing care in stressful situations.

While studies of the effect of biases on clinical decision-making have yet to reach a consensus, evidence for the decisions of CPS workers is more clear. Font, Berger, and Slack

(2012) find that differences in outcomes for children referred to CPS are a result of case circumstances – Black families do not have higher rates of substantiation than white families. Putnam-Hornstein et al. (2012) show that after adjusting for health and socioeconomic factors other than race, non-white children are not more likely than white children to be removed and placed. However, there is evidence of welfare service providers exhibiting racial biases in the referral process. In California, Rodriguez and Shinn (2016) find that while Black families are more likely to be referred to CPS after entering a homeless shelter, children of these families are not more likely to enter the care of CPS than their white counterparts. Thus it appears that while racial biases may affect referral rates, they do not seem to affect placement decisions.

This paper sits at an intersection of the above literature. I study the effect of racial biases on referral decision-making in emergency settings. The effect is a priori ambiguous if physicians tend to exhibit less racial bias towards children, but racial biases tend to be heightened in high-stress situations. I compare children with similar medical histories who visit the ED for similar reasons, so even if CPS workers are generally unbiased in their placement decision-making, inequalities will persist if non-white children are disproportionately referred. Further, if non-white children are being unnecessarily referred, CPS resource allocation will be inefficient since caseworkers must investigate all referrals.

1.2.2 DCYF Referral Process in Rhode Island

This paper focuses on a specific decision point in the DCYF referral process, that of emergency care providers. There are two main ways an ED visit can result in a referral to DCYF. First, if the child or parent calls 911 in the event of an emergency, first responders can refer the child to DCYF. If first responders do not report the incident to DCYF or the family does not call 911, the ED physician can alert DCYF if they feel the child is in danger or the family is unable to mange the crisis appropriately.

Physician discretion is particularly relevant in the case of behavioral health conditions in teenagers. Conversations with agency workers about the placement process highlight the importance of the child's behavior in placements for teenaged children. Generally, when young children are removed from their home it is due to neglect or abuse from the parent. With teenagers, removal occurs when caseworkers feel the family is not equipped to handle the child's specific situation. The parents may be abusing or neglecting the teen, but this alone does not usually result in removal. Most teen removals are the result of behavioral issues. This is particularly true for my sample of children – not only are most of them teenagers, they have all been diagnosed with at least one behavioral health condition and visited the ED at some point during 2017 for a primary behavioral health diagnosis. I describe the sample in more detail in the next section.

Racial bias can have an impact in this setting if when viewed objectively white and non-white children have equally severe behavioral conditions, but ED providers view the non-white children as behaving worse than their white counterparts.

1.3 Data

The main sources of data for this paper come from administrative records of the state of Rhode Island. I use Medicaid claims data and data from DCYF placement records. The sample is composed of children under the age of 18, enrolled in Rhode Island Medicaid, who visited the emergency department (ED) in 2017 for a primary behavioral health diagnosis. Originally, this data was selected to study Medicaid children with significant behavioral health conditions. This should lend itself to studying placements after ED visits since placement with DCYF is a fairly common outcome among these children.

I use all Medicaid claims for these children from 2016-2018. In addition, I merge the corresponding placement records from DCYF to the Medicaid claims data set. This results in a sample of 2,150 children, 596 of whom have contact with DCYF during the study period.

Note that this is nearly 28% of the sample. Nationally, less than 5% percent of children were investigated by CPS in 2017 (Administration for Children and Families Report, 2017). The sample of children used in this paper are particularly negatively selected in terms of both behavioral health and home life conditions.

Finally, I use data at the zip code level from the 2017 wave of the American Community Survey and from the 2010 Census to measure the characteristics of the neighborhoods in which children live. This includes information about racial composition, median income level, unemployment rate, and education levels.

1.3.1 Medicaid Claims

Medicaid claims contain information about all healthcare billed to Medicaid. This includes diagnoses received, procedures performed, care provider, date of service, and cost of care. I use these claims to identify ED visits. In some cases, ED claims list the attending doctor but more often than not, the provider is the hospital. This is in contrast to other types of medical claims which do identify a specific attending provider.

In addition, the claims data set contains some basic demographic information of patients such as gender, race, age, and zip code of residence. Race is a combination of self-reports. Data scientists from the state combine information about race from all observations of a person to obtain the most consistent measure of race possible. From this I identify children as non-white, Black, and Hispanic. Non-white includes children that are Black, Hispanic, American Indian, Pacific Islander, and multiracial. It is not possible to identify which races multiracial children identify as, so non-white is meant as the broadest possible measure of race.

Claims billed to Medicaid not only include regular medical care (e.g. outpatient visits, ED visits, hospitalizations), but also include claims for care provided by schools and case managers. School Medicaid claims are for services provided at school for children with special needs, such as personal care aides, speech and hearing therapists, and transportation. These claims provide a reliable source for identifying children with illnesses that significantly impact their functioning – usually these are children with developmental disorders. Similarly, case management claims can help to identify children with significant behavioral conditions. Case managers are assigned to children who meet specific criteria for severe illnesses or instability and require help gaining access to necessary services (RI Executive Office of Health and Human Services). Thus, regular case management claims provide insight into the significance of a child's illness and the difficulties they face in obtaining and maintaining treatment. Further, these claim types offer a signal about the stability of a child's home life.

1.3.2 DCYF Placement Records

DCYF placement records contain information about all placements over the period 2016-2018. This includes the type and duration of placement. In addition, placement records contain the exact date of removal, which can then be matched to the dates of service for ED visits from the Medicaid claims. From the dates, I can observe if a placement occurred on the same day as an ED visit.

There are six categories of placement: foster care, group homes, residential treatment centers, independent living, adoption, and assessment and stabilization centers. Assessment and stabilization centers are available for temporary, short-term placements. Children will be removed and placed in an assessment and stabilization center during a crisis while their specific situation is evaluated. After evaluation, the agency decides whether to return the child to their home or move them to a longer term placement, such as foster care.

1.3.3 Sample Characteristics and Timing of Placements

Table 1 presents general summary statistics for my sample, both for children never placed with DCYF and children who have at least one placement during the study period. We can see that children who are placed with DCYF are significantly more likely to be non-white, have any behavioral health diagnosis, and visit the ED on average 1.3 times more per year. Children placed with DCYF do not seem to come from significantly worse zip codes, but this is not surprising given the overall negative selection in the sample. The zip codes in which children do reside tend to have a larger proportion of non-white children than the state total. Statewide, 80% of the population is white, 6% is Black, 8% is multiracial, and 14% is Hispanic.

Figure 1 shows the proportion of DCYF placements as they occur in the days surrounding an ED visit. We observe a spike in placements around the time of an ED visit, specifically on the day of the visit, with smaller increases both the day before and the day after an ED visit. As expected, DCYF placements and ED visits often occur simultaneously.

Figure 2 shows the proportion of DCYF placements as they occur around an ED visit broken down by race. Non-white children experience a larger increase than white children in placements on the same day as an ED visit, though not necessarily in the days immediately before and after an ED visit. Thus we may suspect that race plays a role in the decision to refer children to DCYF for ED providers.

We may also be interested in where children are being placed when they are referred to DCYF, specifically when they are referred after an ED visit. Table 2 shows the proportion of each placement type, broken down by day relative to the ED visit.

The most striking observation from this table is that the majority of placements occurring on the same day as an ED visit are in assessment and stabilization centers. These centers provide temporary, short-term placement during crises. Children's situations can be evaluated in a neutral environment and the results of this evaluation are used to determine if children should be returned to their homes or if they should be moved to longer term placements, such as foster care or a group home. This is important to note since children do not jump directly from the ED into foster care. Rather an ED visit resulting in referral triggers a temporary removal and an in-depth evaluation of the child's home life.

Table 2 illustrates that the general placement mix before ED visits is not very different from the mix of placements after ED visits. This suggests that the main role of ED providers in placement is the initial evaluation. It may be the case that evaluators at Rhode Island's assessment and stabilization centers demonstrate racial bias in further placement decisions, however this is not the focus of this paper and I do not observe evaluator decisions.

1.4 Methodology

The goal of this paper is to determine if race is a significant predictor of placement with DCYF after an ED visit. I use regression analysis to estimate the relationship between likelihood of placement and race. However, a simple regression of placements occurring at the same time as an ED visit on race will be biased. Patients visiting the ED with worse illnesses or worse family conditions are more likely to be placed than those who show up with only mild symptoms and supportive families. If race is correlated with more severe illnesses and more unstable home environments, any estimation will capture these effects and overestimate the effect of race. To address this bias, I group children based on the similarity of their medical histories and reason for visiting the ED. I then estimate the relationship between probability of placement occurring on the day of an ED visit and race for each group of children separately. This allows me to address the main source of bias and compare placement outcomes for children who should only differ in terms of their race. Thus, if race is a significant predictor of placement for each group of similar children, we can conclude that emergency providers are displaying racial biases in the decision to refer children to DCYF.

1.4.1 Cluster Analysis

I use cluster analysis to determine the similarity of children's medical histories. I use a k-medoid clustering algorithm from the R package "cluster." This algorithm is similar to k-means clustering, but can be used with non-numerical data, such as diagnoses. Both k-means and k-medoid clustering partition observations based on the distance to the nearest cluster mean value of the features chosen for clustering (see Peterson 2002 for further information about clustering).

1.4.2 Cluster Features

I cluster data based on the medical histories of children in the 30 days prior to an ED visit and their reason for visiting the ED. I use the following observables to account for medical history: up to 10 diagnoses since more than half my sample of children do not receive 11 or more diagnoses in the 30 days prior to an ED visit, visits to outpatient providers, other ED visits, inpatient hospitalizations, amount spent on Medicaid, school Medicaid claims, and case management Medicaid claims. I describe the relevance of each of these observable characteristics for clustering children below.

Outpatient visits, ED visits, inpatient hospitalizations, and amount spent in the 30 days prior to an index ED visit are meant to capture children's interaction with the medical system. Outpatient claims signal that children have routine primary care visits or receive care for their behavioral health conditions from mental health professionals on a regular basis. Children receiving routine care in the community may be better able to manage their conditions, and to have families that are more equipped to navigate the medical system.

Multiple ED visits in a short period of time on the other hand can signal poor management of illnesses or instability at home. These children might be more likely to be placed with DCYF, especially if ED providers notice the same child repeatedly showing up. Inpatient hospitalizations serve as a measure of the severity of children's illnesses. If children are hospitalized in the 30 days prior to an ED visit, their illnesses are likely quite severe or poorly managed.

Finally, the amount spent on Medicaid provides a measure of both number and quality of interactions with the healthcare system. Children who receive more care or more intensive care will have higher expenditures. Those children with large expenditures may have more severe chronic conditions that require frequent care, or they may have few but very expensive visits. The observables above should help sort these two types of children.

Medicaid claims billed to a child's school or to case management providers act as both a measure of significance of illness and of children's interactions with non-medical state providers. School Medicaid claims capture children with illnesses that are significant enough to impact their daily functioning, often these include developmental disorders. Similarly, case management claims can signal both the presence of a significant illness and unstable conditions at home. Case managers are assigned to help children receive appropriate services. These children may then be more likely to avoid DCYF placements via ED visits if the case manager is successful in helping children obtain necessary services. On the other hand, ED providers may feel pushed to contact DCYF if they notice a difficult situation in the home.

1.4.3 Cluster Results

The clustering algorithm assigns each child to a cluster made up of children that are most similar based on the observable features described above. Figure 3 shows the results of this clustering. Each color corresponds to a cluster and each dot is a child. The figure uses t-distributed stochastic neighbor embedding (t-SNE) to reduce the dimensions of the clustering features for visualization. This algorithm simply places similar points near each other and dissimilar points far away from each other with high probability. In an ideal figure, all dots of the same color would be next to each other, indicating that clusters are similar. This figure shows some well-defined clusters. However, there does exist overlap between clusters – the clustering algorithm is not perfect but it does match similar children to the same clusters. Figure 2 repeats figure 1 but adds ellipses over cluster boundaries to illustrate overlap more clearly.

I choose seven total clusters because this number maximizes the average silhouette distance between clusters. This distance metric tells us how dissimilar clusters are from each other – larger values of the silhouette distance indicate that a cluster is "further away" (or less similar) to its neighboring clusters.

Tables 3 and 4 show the characteristics of the children making up each cluster. Table 3 focuses on the demographic characteristics of children, while table 4 reports information about the observables on which children are clustered. From table 4 we can see evidence of clustering. Cluster 1 tends to have children with more significant illnesses - 40% are hospitalized and spend on average $2,861 in the month prior to an ED visit. Children in cluster 2 tend to be children with less severe illnesses, but who do not receive as much outpatient care as children in other clusters. This could be a result of mild illnesses that flare up only on occasion or a lack of access to primary care providers. Cluster 3 is the largest cluster, with the most variation, but it most closely resembles the general sample of children in terms of medical history. Similarly to cluster 2, clusters 4 and 7 tend to have children with less severe illnesses, differing only in their rates of outpatient care. Cluster 5 contains children with significant behavioral conditions, as evidenced by the school Medicaid claims and cluster 6 contains the highest percentage of children with case management claims, indicating significant illnesses and less stable home environments.

Table 3 reflects the results of clustering in demographic characteristics. Clusters 1, 5, and 6 have the highest percentages of non-white children, reflecting the fact that having more significant illnesses seems to be correlated with race in this sample. Percentage of children ever placed with DCYF also varies widely across clusters, again illustrating the correlation between severe behavioral conditions, unstable home environments, and placement. Finally, diagnoses received tend to match the overall severity level of a cluster. Children with mood disorders and child onset disorders (such as conduct disorders) tend to gravitate towards clusters made up of other children with significant behavioral disorders. This variation in characteristics and level of severity across clusters will be useful for interpreting the regression results below.

1.4.4 Regression Analysis

I use regression analyses to determine if race is a significant predictor of a DCYF placement occurring after an ED visit. I estimate the following equation within each cluster

$$Y_{vi} = Race_i + Age_{vi} + Female_i + \epsilon_{vi} \qquad (1)$$

where Y_{vi} is an indicator for a placement with DCYF occurring on the same day as an ED visit for visit v and child i, $Race_i$ is an indicator for race of the child, Age_{vi} is the age of child i during visit v, and $Female_i$ is an indicator for child i being female. I estimate this equation for each cluster individually. This controls for children's medical histories and diagnoses.

I use three separate racial indicators to determine if there are heterogeneous effects by race: non-white, Black, and Hispanic. As described in the data section, non-white is meant to broadly capture any race other than white, while Black and Hispanic are more specific indicators of race.

1.5 Results

Table 5 reports the results of estimating equation (1) for each cluster. Generally, race is not a significant predictor of placement occurring on the same day as an ED visit. There are a few notable exceptions. Hispanic children are significantly more likely to be placed than non-Hispanic in clusters 1 and 5, a 7.5% increase and a 0.5% increase in the likelihood of placement respectively. Non-white children are 1.4% more likely to be placed than white children in cluster 4. Finally, Hispanic children are 1.2% *less* likely to be placed than non-Hispanic children in cluster 3. I explore potential reasons for these results below.

In cluster 1, being Hispanic is related to an increase of 7.5% in the probability of being placed with DCYF on the same day as an ED visit. This is significant at the 1% level, suggesting that this is an important factor in the placement decision for the children in this cluster. Cluster 1 tends to have children with more severe behavioral disorders. Over half of the children in this cluster have a mood disorder or child onset disorder. These disorders often result in more serious ED visits, for example, suicide attempts from a depressive disorder or extreme aggression from conduct disorders. In these cases, the referral decision is likely obvious – ED providers would decide to refer the child regardless of race. However, in the case of non-mood, non-psychotic disorders, an example ED visit may be for a panic attack, where an ED provider may be able to choose to send a child home if they think the home environment is stable, or decide to involve DCYF if they think otherwise. For this cluster, we may see a significant result for Hispanic children if providers perceive Hispanic families as less capable of caring for a child after an ED visit. They may also know that the child has been hospitalized recently (40% of children in this cluster are inpatients at some point in the 30 days before an ED visit) and consider the step back up to the ED a signal of inability to provide a stable home environment among Hispanic families, while giving white families the benefit of the doubt. 7.5% is a relatively large increase in the likelihood of placement as well. This may be due in part to the overall visibility of Hispanic children in Rhode Island, since Hispanic people are the largest minority in the state.

The indicator for Black is significant only at the 10% level and the coefficient is very small in cluster 3, so this is not likely to be an important predictor of placement. But, here we see that being Hispanic actually significantly decreases the probability of placement on the same day as an ED visit. In this case, it is likely a result of the over-representation of Hispanic children in cluster 3. Nearly 20% of the children are Hispanic in this cluster, while in the sample population only about 12% of children are Hispanic. Since there is a large number of Hispanic children in this cluster, this regression is likely picking up the fact that there are relatively fewer Hispanic children ever placed in DCYF (only 9% of the children ever placed with DCYF are Hispanic, while 14% of the never placed children are Hispanic) rather than a racial bias favoring Hispanic children.

Cluster 4 shows that being non-white is associated with a 1.4% increase in the probability of being placed on the same day as an ED visit. This effect is small, but significant at the 5% level. When we examine the demographic makeup of cluster 4, we can see that a lower percentage of children have each subset of behavioral health diagnoses as compared to the other clusters. In addition, very few children in cluster 4 are hospitalized before their ED visit and a only a relatively small proportion have school and case management claims. Finally, 97% of these children have a claim for an outpatient visit in the 30 days preceding an ED visit. Taken together, this indicates the children in cluster 4 have relatively mild or moderate illnesses that are being well-managed with routine, community based care. In this case, we likely see that race is a significant predictor in the placement decision because providers have more discretion in the decision to alert DCYF about the child. They likely perceive non-white children to have more severe illnesses or less capable families, even as compared to white children with similar medical histories and diagnoses.

The indicator for Hispanic has a very small, but significant effect in cluster 6. It is associated with a 0.5% increase in the probability of placement on the day of an ED visit. A large percentage of children in this cluster have been diagnosed with a child onset disorder (e.g. conduct disorders, autism spectrum disorders). A relatively large percentage also have school claims, case management claims, and were hospitalized in the 30 days prior to an ED visit. In this cluster most children are coming into the ED with relatively more significant disorders that are also highly visible to providers – it will be obvious to a provider if a child is struggling with conduct or autism spectrum disorders. However, there is still a smaller percentage of children who have less significant disorders that are still highly visible to providers, in which case providers again judge illnesses of Hispanic children to be more severe or their families to be less capable of care than their white counterparts.

1.5.1 Provider Heterogeneity

While so far I have only shown only weak evidence of racial bias in ED provider referral decisions, we may still be interested to know if certain providers are driving the significant results and displaying a stronger racial bias. To explore this possibility, I break down referral rates by provider. Providers may exhibit differing placement rates for two reasons. They may display a racial bias and in the extreme refer all non-white children to DCYF, increasing their overall placement rate. Or they may see a worse mix of patients and encounter more children who objectively require an intervention.

Figure 5 shows the variation in referral rates among the 13 providers who have at least one placement occurring on the same day as an ED visit. This varies from less than 1% to almost 6% of ED visits resulting in referral. This is not an extremely large difference and most providers fall between a 1% and 3% referral rate.

Table 6 compares the composition of ED visits to providers with at least one simultaneous ED visit and placement to that of providers with no DCYF referrals. On average these two types of providers see a similar racial mix of patients. However, providers with at least one referral see on average worse patients. They have more behavioral health related ED visits and see a larger share of patients with mood and child onset disorders. This suggests that these providers are not frequently exhibiting racial biases, but rather referring patients that objectively require a placement.

In addition, providers with at least one placement see far more patients total. They make up only 11% of ED providers in the state, but account for 90% of the total ED visits. In fact, the providers with higher placement rates are the emergency departments at hospitals. Other providers are small emergency centers or individual doctors. Patients with more severe conditions tend to visit a hospital when they need emergency care, and usually hospital emergency departments list themselves as the provider on Medicaid claims, rather than an individual doctor.

Looking specifically within the clusters that do find significant effects of race on placement, I am unable to distinguish if a specific provider is driving the results. The hospital provider level is too aggregate. Since the majority of this sample is made up of children who visit the ED with fairly severe illnesses, and children with more severe emergencies tend to have hospitals listed as the provider, I can only observe that hospitals with higher placement rates see on average worse patients. I cannot distinguish the degree of racial bias among hospitals.

Given these observations in addition to the above results, it is likely that racial bias is not driving the majority of the variation among provider referral rates. We may see bias arise in cases where patients are on the margin of being placed, as the above results suggest, but it is not possible in this data set fully observe individual level decision-making. Future work with data that does allow for observation of individual providers may be able to pick up better evidence of racial bias affecting referral decision-making at the provider level.

1.6 Conclusion

Identifying the effects of racial bias is important for addressing systemic inequalities affecting non-white children and their families. Children of color continue to be over-represented in the child welfare system, which has long-term impacts on their outcomes. While I do not find strong evidence of racial bias affecting the referral decision-making of ED providers in my sample, future work should continue to explore this issue and identify points where racial bias may have an impact on decisions that affect the well-being of children. Given the negative selection affecting my sample, it will be valuable for future work to consider a more representative sample with overall less severe illnesses. The severity of children's conditions and instability at home may be masking racial bias that presents itself in less severe cases where providers have more discretion over the referral decision.

Tables

Table 1.1: *Summary Statistics*

	Non-DCYF	DCYF	Difference
Number of Children	1,554	596	
Proportion Non-white	0.64	0.80	0.16***
Proportion Black	0.09	0.10	0.01
Proportion Hipanic	0.14	0.09	0.05
Proportion Female	0.55	0.51	0.04
Mean Age	13.36	13.26	0.10
Proportion with Mood Disorder	0.69	0.78	0.09***
Proportion with Non-mood, Non-psychotic Disorder	0.82	0.89	0.07***
Proportion with Psychotic Disorder	0.09	0.17	0.08***
Proportion with Child Onset Disorder	0.60	0.85	0.25***
Proportion with Substance Use Disorder	0.10	0.25	0.15***
Proportion with Complex Chronic Condition	0.11	0.14	0.03
Average Number of ED Visits per Year	2.1	3.4	1.3***
Mean Percentage White in Zip Code	0.64	0.65	0.01
Mean Percentage Black in Zip Code	0.05	0.06	0.01***
Mean Percentage Hispanic in Zip Code	0.18	0.16	0.02***
Mean Median HH Income Over Zip	$57,588	$57,171	$417
Mean Unemployment Rate by Zip	7.9%	8.1%	0.02%***
Mean High School Degree by Zip	0.30	0.27	0.03***
Mean College or More Degree by Zip	0.36	0.41	0.05***
Mean Some College by Zip	0.18	0.17	0.01***

Notes: This table shows demographic summary statistics of the population of children in the analysis. Column (1) shows averages for children who did not had contact with DCYF during the 2016-2018 period, column (2) shows averages for children who did have contact with DCYF during the study period, and column (3) shows the difference. Diagnoses occurred any time throughout 2016-2018, but are only counted once per child, e.g. a child diagnosed with depression in 2016 by a primary care physician and 2017 by an ED provider will be marked as "diagnosed with a mood disorder," regardless of timing or place of diagnosis. This is because behavioral health conditions are often chronic. Mood disorders include depressive and bipolar disorders, non-mood, non-psychotic disorders include anxiety and eating disorders, psychotic disorders include disorders such as schizophrenia, and child onset disorders include conduct disorders and social disorders (e.g. autism). Complex chronic conditions are severe, life-affecting, non-behavioral medical conditions that require regular care to manage. Zip code statistics are weighted averages over all zip codes children live in, weighted by the number of children living in each zip code. *** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Table 1.2: *Types of Placements*

	On ED Visit	One day before ED visit	One day after ED visit	30 Days Before ED visit	30 Days After ED visit
Residential Treatment	0.11	0.21	0.14	0.20	0.28
Assessment and Stabilization Center	0.53	0.39	0.37	0.21	0.18
Foster Care	0.14	0.16	0.22	0.23	0.22
Group Home	0.17	0.19	0.23	0.25	0.23
Independent Living	0.04	0.05	0.04	0.07	0.08
Adoption	0	0	0	0.03	0.02
Juvenile Justice	0.11	0.06	0.10	0.13	0.13
Median Length of Placement	29 days	5 days	39 days	27 days	35 days
Mean Length of Placement	63 days	33 days	63 days	76 days	92 days
Number of Placements	141	62	92	595	608

Notes: This table shows the proportion of each type of DCYF placement corresponding to it's timing around an ED visit. Column (1) shows results for placements occurring on the same day as an ED visit; column (2) shows results for placements occuring one day before an ED visit; column (3) shows one day after an ED visit; column (4) shows results for all placements occurring 30 days before an ED visit, excluding the day before; and column (5) shows the results for all placements occuring 30 days after an ED visit, excluding the first day after a visit. Assessment and stabilization centers are utilized during crises to evaluate the severity and provide temporary placements if necessary. Juvenile justice refers to any juvenile justice incident occurring during a placement that began on the given date.

Table 1.3: *Cluster Comparison: Demographics*

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Number of Individuals	198	234	935	440	134	115	94
Percentage Non-white	67%	47%	65%	55%	75%	62%	71%
Percentage Black	4%	7%	7%	6%	10%	7%	12%
Percentage Hispanic	7%	12%	49%	9%	12%	12%	11%
Median Age	15	15	15	14	13	13	14
Percentage Female	45%	56%	60%	51%	28%	43%	63%
Percentage Ever Placed with DCYF	68%	16%	30%	13%	24%	34%	18%
Percentage with Mood Disorder Diagnosis	53%	26%	54%	27%	30%	40%	28%
Percentage with Non-mood, Non-psychotic Disorder Diagnosis	59%	34%	59%	39%	30%	53%	27%
Percentage with Psychotic Disorder Diagnosis	4%	1%	7%	0.5%	4%	9%	2%
Percentage with Child Onset Disorder Diagnosis	54%	18%	47%	32%	46%	74%	45%
Percentage with Substance Use Disorder Diagnosis	14%	2%	7%	2%	5%	9%	7%
Percentage with Complex Chronic Condition	6%	3%	5%	2%	4%	6%	1%

Notes: This table shows characteristics of each cluster, each column corresponds to a cluster. Diagnoses are from any medical claims in the 30 days prior to an ED visit. Mood disorders include depressive and bipolar disorders, non-mood, non-psychotic disorders include anxiety and eating disorders, psychotic disorders include disorders such as schizophrenia, and child onset disorders include conduct disorders and social disorders (e.g. autism). Complex chronic conditions are severe, life-affecting, non-behavioral medical conditions that require regular care to manage.

Table 1.4: *Cluster Comparison: Features and Medical Behavior*

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Number of Individuals	198	234	935	440	134	115	94
Average Silhouette Distance	0.35	0.27	-0.21	-0.12	0.22	0.03	0.15
Number of ED Visits in Cluster	1145	704	6094	1119	551	531	348
Percentage of ED Visits with Placement Occurring	3.4%	0.8%	1.2%	0.6%	2.8%	1.9%	0.9%
Percentage with ED Primary Diagnosis Non-Behavioral Health	56%	61%	65%	59%	55%	52%	63%
Percentage with Hospitalization 30 Days before ED visit	40%	6%	21%	3%	12%	18%	5%
Percentage with School Medicaid Claim in 30 Days before ED visit	26%	6%	24%	15%	100%	58%	7%
Percentage with Outpatient Visit Claim in 30 Days before ED visit	87%	59%	99%	97%	60%	50%	100%
Percentage with Case Management Claim in 30 Days before ED visit	51%	17%	50%	29%	34%	73%	22%
Mean Amount Spent on Medicaid in 30 Days before ED visit	$2,861	$1,526	$3,317	$1,409	$2,243	$2,247	$1,861
Median Amount Spent on Medicaid in 30 Days before ED visit	$1,852	$1,008	$2,413	$1,022	$1,460	$1,093	$1,134

Notes: This table shows characteristics of each cluster, each column corresponds to a cluster.. Outpatient claims are any medical claim filed by an outpatient provider. School Medicaid claims indicate care was provided at school, generally by personal aides or speech/hearing therapists. These claims are a reliable way to identify children with developmental disorders. Case management claims cover charges for case managers who assist children with severe behavioral health conditions and those with significant developmental delays. These claims present another measure of severity of illness. Children with case managers often face more unstable home life as well.

Table 1.5: *Cluster Regression Results*

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Non-White	-0.001	0.004	0.004	0.014**	-0.003	0.012	0.024
	(0.012)	(0.009)	(0.003)	(0.006)	(0.013)	(0.012)	(0.022)
Black	-0.019	-0.012	0.001*	-0.009	-0.011	0.025	-0.017
	(0.017)	(0.016)	(0.001)	(0.013)	(0.017)	(0.024)	(0.025)
Hispanic	0.075***	-0.011	-0.012**	0.007	0.001	0.005***	-0.007
	(0.022)	(0.011)	(0.005)	(0.009)	(0.001)	(0.002)	(0.020)
N	1145	704	6094	1119	551	531	348

Notes: This table shows the results of the regression equation of placement during an ED visit on each racial indicator. Each column corresponds to a cluster, and only one racial indicator is used at a time in a regression. All regressions include controls for age and gender. *** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

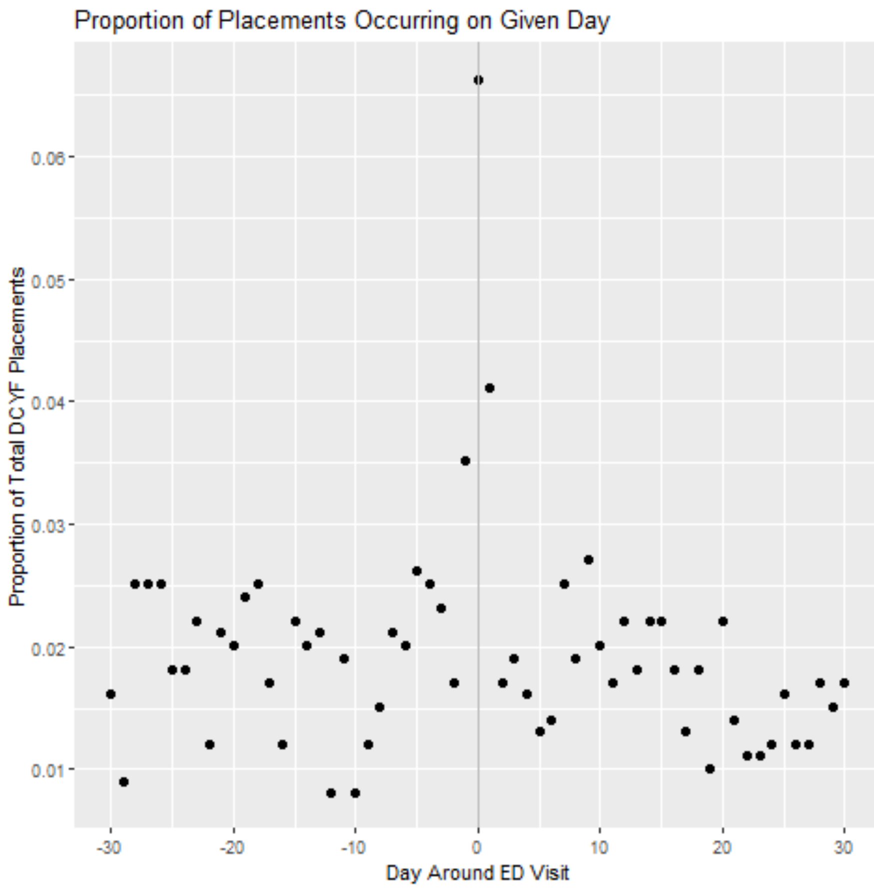
Table 1.6: *Provider Patient Mix*

	Providers with a Placement on an ED Visit	Providers with No Placement on an ED Visit
Number of ED Providers	13	101
Number of ED visits	9434	1044
Proportion Non-white ED Visits	0.52	0.50
Proportion Black ED Visits	0.06	0.08
Proportion Hipanic ED Visits	0.08	0.08
Proportion Female ED Visits	0.57	0.70
Mean Age of ED Visits	13.60	14.93
Proportion of ED Visits for Mood Disorder	0.14	0.08
Proportion of ED Visits for Non-mood, Non-psychotic Disorder	0.14	0.05
Proportion of ED Visits for Psychotic Disorder	0.01	0.003
Proportion of ED Visits for Child Onset Disorder	0.06	0.04
Proportion of ED Visits for Substance Use Disorder	0.02	0.03
Proportion of ED Visits for Complex Chronic Condition	0.001	0.004
Proportion of ED visits for Non-Behavioral Health Diagnoses	0.60	0.76

Notes: This table shows the patient mix of ED providers serving the study population. Characteristics of providers with a DCYF placement corresponding to an ED visit are shown in column (1). Column (2) shows providers that do not have a DCYF placement corresponding to an ED visit. Diagnoses are the primary diagnosis on the ED claim. Mood disorders include depressive and bipolar disorders, non-mood, non-psychotic disorders include anxiety and eating disorders, psychotic disorders include disorders such as schizophrenia, and child onset disorders include conduct disorders and social disorders (e.g. autism). Complex chronic conditions are severe, life-affecting, non-behavioral medical conditions that require regular care to manage.

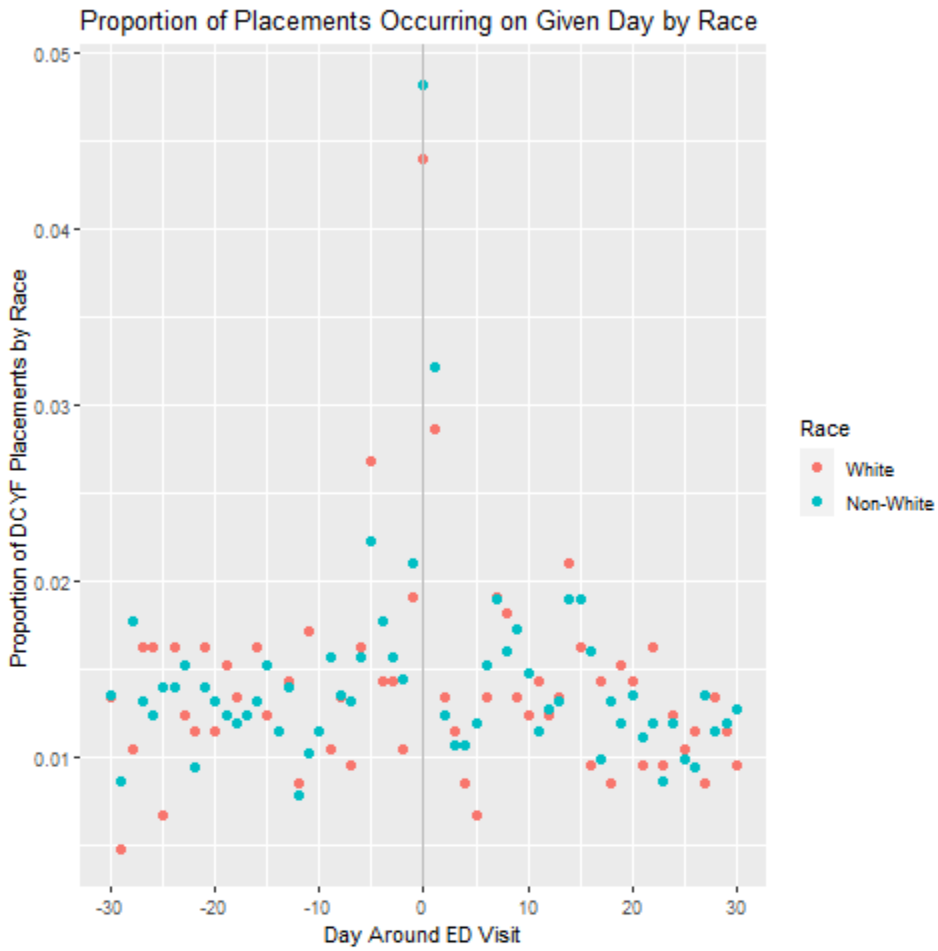
Figures

Figure 1.1: *Proportion of Placements*



Notes: This figure shows the proportion of total DCYF placements as the occur in reference to an ED visit. Time 0 corresponds to the date of the ED visit.

Figure 1.2: *Proportion of Placements by Race*

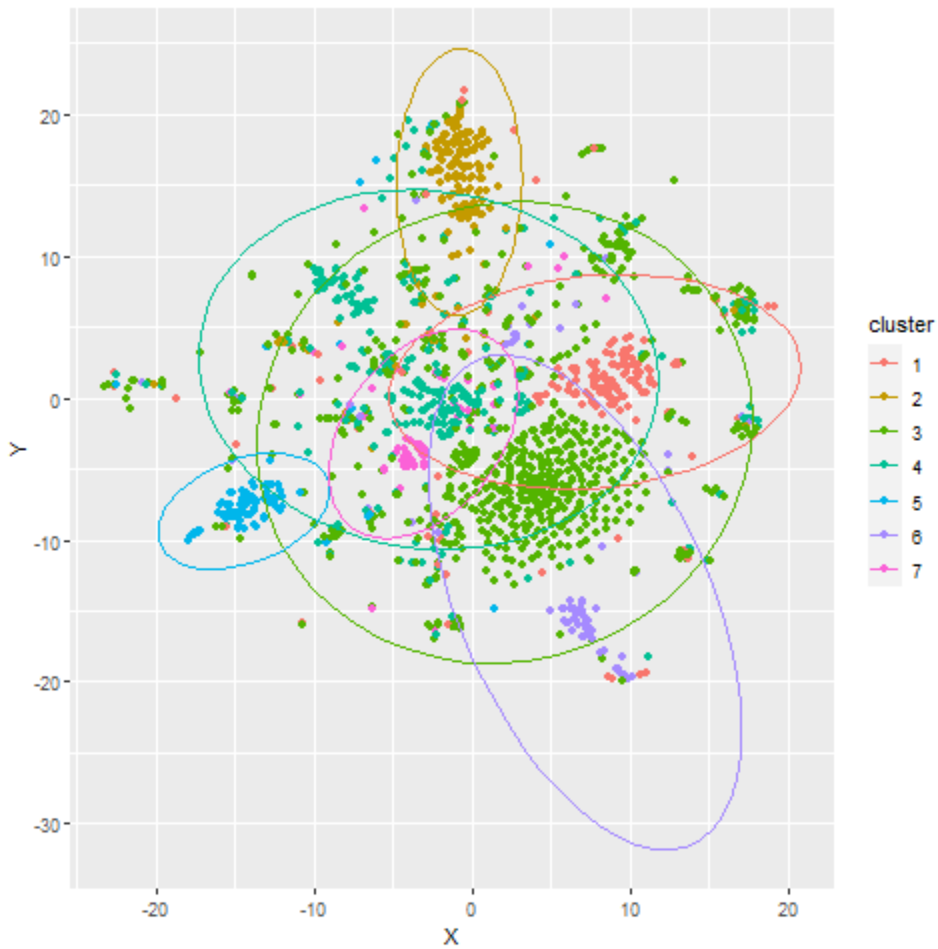


This figure shows the proportion of total DCYF placements as the occur in reference to an ED visit, broken down by race. Time 0 corresponds to the date of the ED visit.

Figure 1.3: *Clusters*

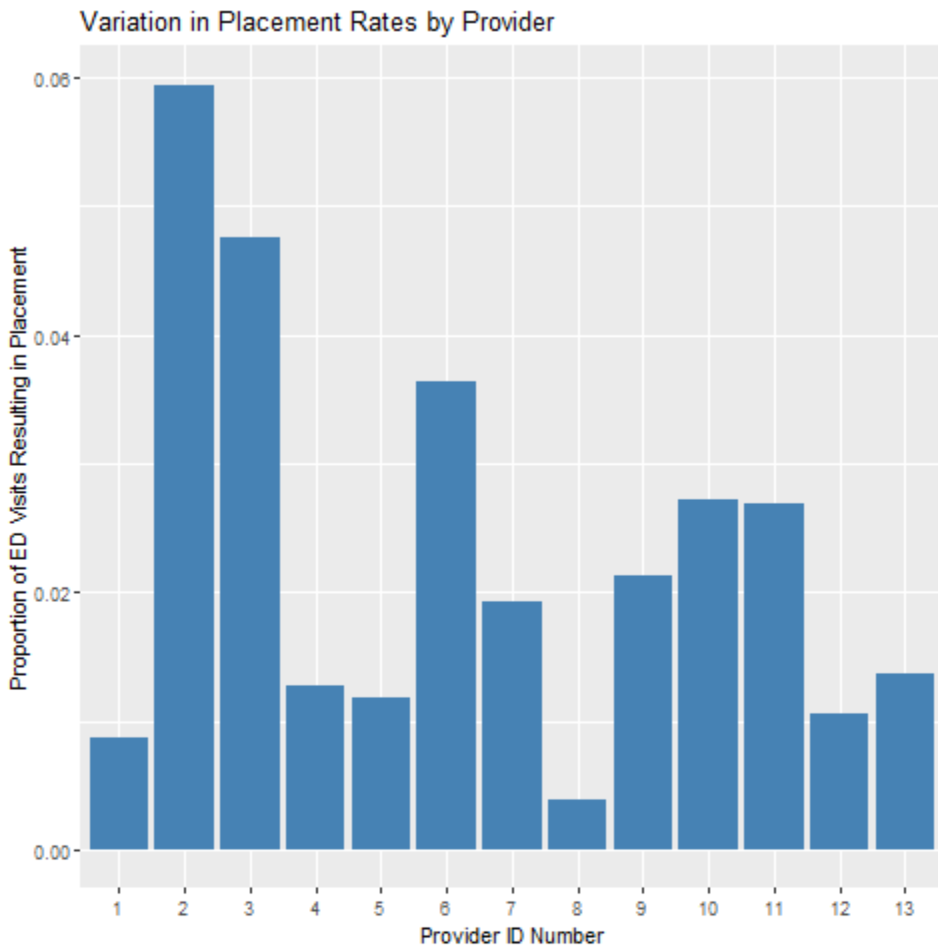
Notes: This figure uses t-SNE dimension reduction techniques to show the relative distance between observations. Colors correspond to the cluster number. Observations closer together are more similar than those farther apart.

Figure 1.4: *Clusters with Ellipses*



Notes: This figure uses t-SNE dimension reduction techniques to show the relative distance between observations. Colors correspond to the cluster number. Ellipses are drawn around cluster observations to highlight overlapping clusters. Observations closer together are more similar than those farther apart.

Figure 1.5: *Provider Placements*



Notes: This figure shows the proportion of total ED visits that correspond to a placement with DCYF for each of the 13 ED providers that have at least one DCYF placement corresponding to an ED visit.

Bibliography

Blair, Irene, John Steiner, Diane Fairclough, Rebecca Hanratty, David Price, Leslie Wright, Michael Bronsert, Elhum Karimkhani, David Magid, Holen Hirsh, Edward Havranek. 2013. "Clinicians' Implicit Ethnic/Racial Bias and Perceptions of Care Among Black and Latino Patients." *The Annals of Family Medicine,* 11(1): 43-52.

Child Maltreatment. 2017. Administration for Children and Families Report.

Chuang, Emmeline, and Rebecca Wells. 2010. "The role of interagency collaboration in facilitating receipt of behavioral health services for youth involved with child welfare and juvenile justice." *Child and Youth Service Review,* 32(12): 1814-1822.

Cooper, Lisa, Debra Roter, Kathryn Carson, Mary Catherine Beach, Janice Sabin, Anthony Greenwald, and Thomas Inui. 2012. "The Associations of Clinicians' Implicit Attitudes About Race With Medical Visit Communication and Patient Ratings of Interpersonal Care." *American Journal of Public Health,* 5: 979-986.

Dehon, Erin, Nicole Weiss, Jonathan Jones, Whitney Faulconer, Elizabeth Hinton and Sarah Sterling. 2017. "A Systematic Review of the Impact of Physician Implicit Racial Bias on Clinical Decision Making." *Acadmemic Emergency Medicine,* 895-904.

Feudtner, Charles, James Feinstein, Wenjun Zhong, Matt Hall, and Dingwei Dai. 2014. "Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation." *BMC Pediatrics,* 14: 199.

Fiztgerald, Chole, and Samia Hurst. 2017. "Implicit bias in healthcare professionals: a systematic review." *BMC Medical Ethics,* 18.

Font, Sarah, Lawrence Berger, and Kristen Slack. 2012. "Examining Racial Disproportionality in Child Protective Services Case Decisions." *Child Youth Service Review,* 34(11): 2188–2200.

Johnson, Tiffani, Robert Hickey, Galen Switzer, Elizabeth Miller, Daniel Winger, Margaret Nguyen, Richard Saladino, Leslie Hausmann. 2016. "The Impact of Cognitive Stressors in the Emergency Department on Physician Implicit Racial Bias." *Academic Energy Medicine,* 23(3): 297-305.

Johnson, Tiffani, Daniel Winger, Robert Hickey, Galen Switzer, Elizabeth Miller, Margaret Nguyen, Richard Saladino, and Leslie Hausmann. 2017. "Comparison of Physician Implicit Racial Bias Toward Adults Versus Children." *American Pediatrics,* 17(2): 120-126.

Knott, Theresa, and Sulaimon Giwa. 2012. "African American Disproportionality Within CPS and Disparate Access to Support Services: Review and Critical Analysis of the Literature." *Residential Treatment for Children and Youth,* 29(3): 219-230.

Lyons, John, and Laura Rogers. 2004. "The U.S. Child Welfare System: A De Facto Public Behavioral Health Care System." *Journal of American Child and Adolescent Psychology,* 43(8): 971-973.

Maina, Ivy, Tanisha Belton, Sara Ginzberg, Ajit Singh, and Tiffani Johnson. 2018. "A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test." *Social Science and Medicine,* 199: 219-229.

Oliver, Norman, Kristen Wells, Jennifer Joy-Gaba, Carlee Beth Hawkins, and Brian Nosek. 2014. "Do Physicians' Implicit Views of African Americans Affect Clinical Decision Making?" *Journal of American Family Medicine,* 27(2): 177-188.

Peterson, Jason. *Clustering Overview.* Accessed author's website: http:// citeseerx.ist.psu.edu/ viewdoc/download? doi=10.1.1.136.4170rep=rep1type=pdf.

Putnam-Hornstien, Emily, , Barbara Needell, Bryn King, and Michelle Johnson-Motoyama. 2012. "Racial and ethnic disparities: A population-based examination of risk factors for involvement with child protective services." *Child Abuse and Neglect.*

Racial Disproportionality and Disparity in Child Welfare. 2016. Children's Bureau Issue Brief.

Rodriguez, Jason, and Marybeth Shinn. 2016. "Intersections of Family Homelessness, CPS Involvement, and Race in Alameda County, California." *Child Abuse and Neglect,* 57: 41-52.

Sabin, Jaine, Frederick Rivara, and Anthony Greenwald. 2008. "Physician Implicit Attitudes and Stereotypes About Race and Quality of Medical Care." *Medical Care,* 46(7): 678-685.

Sabin, Janice, and Anthony Greenwald. 2012. "The Influence of Implicit Bias on Treatment Recommendations for 4 Common Pediatric Conditions: Pain, Urinary Tract Infection, Attention Deficit Hyperactivity Disorder, and Asthma." *American Journal of Public Health,* 102(5): 988-995.

CHAPTER 2

URBAN RENEWAL IN CHICAGO

2.1 Introduction

In recent years, place-based policies have become more popular (Neumark and Simpson 2014). Rather than targeting individuals, these types of policies focus on specific disadvantaged geographic areas to encourage economic growth, often through tax incentives or redevelopment programs. However, place-based policy design is not a new innovation. In one of the most widespread examples of a place-based policy, the United States embarked on a campaign to clear out and redevelop deteriorating neighborhoods during the 1950s and 1960s. Known as urban renewal, this effort was designed to revitalize cities suffering from overcrowding and dilapidation by building new housing, commercial centers, and public spaces. Cities across the country applied for federal funds to participate in the program.

This paper estimates the effects of the urban renewal program on neighborhood demographic composition using Chicago as a case study. Chicago provides an ideal case study: it is often thought of as the poster child for urban renewal in the United States. There was a relatively large number of projects throughout the city, and the program has generated controversy. By many historical accounts, urban renewal was positive for Chicago and resulted in better housing, better overall living conditions for families, and economic benefits for the city (Cress 1958; Stevens 1960; Getz 1966). However, other accounts have highlighted the negative impacts of urban renewal, especially for the city's Black population, as many were removed for demolition (Anderson 1964; Getz 1966). Qualitative reports looking back at urban renewal follow this discussion and view urban renewal largely as a failure for Chicago's poor and minority residents (Hirsh 1983; Pritchett 2003; Winling 2018).

Given the conflicting accounts and evidence, the effect of the program on neighborhood demographic composition is a priori ambiguous. On one hand, if Black communities were demolished and residents were forced to move, we may expect to see a decrease in the share of Black residents in project neighborhoods. We may also expect to see an increase in other types of residents, for example those with a college degree, if project neighborhoods are nicer and more expensive to live in after project completion. However, if residents returned to renewed neighborhoods, projects may not alter neighborhood demographics. Residents may have returned to project neighborhoods if they were unable to find housing in other neighborhoods due to discrimination (Fishback et al. 2021), or if the new housing and additions to neighborhoods did not attract new residents. I also estimate the effect of projects on homeownership rates and rental prices. Since the goal of urban renewal projects was largely to improve housing, we may expect that large-scale demolition projects led to an increase in rental prices. In this case, we may also expect that increased prices would attract wealthier, and by extension more educated, residents to project neighborhoods. Likewise, if new homes were built, we may see an increase in homeownership rates. If projects did not increase rental prices, this may be the result of people not valuing new housing or amenities in project neighborhoods, or projects may have made only small improvements to neighborhoods, as in the case of some later urban renewal projects.

To estimate the impact of projects, I use data from the Digital Scholarship Lab at Richmond University, which contains information about urban renewal projects, combined with data from the U.S. Census. I measure census tract level demographic composition with the share Black of the population and the share of the population with a college degree.

I use the share of owner occupied dwelling units to measure homeownership rates and the share of dwelling units below the national median rent as a of measure rental prices.

Simply regressing the above outcomes on whether or not a tract had an urban renewal project would result in a biased estimate since project sites were generally chosen based on the level of disrepair. So, to estimate the effects of urban renewal projects, I use a difference-in-differences design to compare census tracts that were the site of the urban renewal projects to those that were considered as project sites, but were ultimately not chosen for urban renewal.

These potential sites were surveyed by city planning agencies to collect information about levels of dilapidation. Planning agencies then worked with the city council, contractors, and residents to move forward with projects. Potential sites may have been dropped from the decision-making process for any of a various number of reasons, including insufficient dilapidation for demolition or city council and resident preferences.

Though I do not observe the reason one tract was chosen over another, potential site tracts provide a reasonable comparison group. They are on average similar to project tracts, and they may not have been chosen for relatively random reasons in many cases. However, one concern with using potential site tracts for comparisons is that of spillovers. Since people had to move for demolition, they may have moved to nearby potential site tracts. To address this, I also compare potential sites to census tracts that were never considered for urban renewal projects and were unlikely places for people to move after demolition of their original neighborhood. This allows me to see any demographic changes in potential site tracts as compared to unaffected tracts.

The difference-in-differences study design above depends on parallel trends between project tracts and potential site tracts. To test the validity of this assumption, I estimate an event study style specification of the main two-way fixed effects model to detect individual time period effects. I find that most pre-period effects are small and not significant, suggesting the assumption of parallel trends between the two groups is reasonable. In addition, I conduct a sensitivity analysis following the work of Rambachan and Roth (2022) to assess how much the estimates depend on the parallel trends assumption holding. Estimating the two-way fixed effects model I find a 5.3% increase in the share of the population with a college degree in census tracts with urban renewal projects.

Finally, I investigate the possibility of heterogeneous treatment effects. The laws governing urban renewal projects became more lenient over time. Initially, the federal government emphasized housing, but successive laws moved away from housing to include more mixed use projects, such as commercial centers and public parks. In addition, as time went on, the city moved away from large-scale demolition projects to smaller repairs and additions (Getz 1967). These different project types may have had different impacts on neighborhoods, and we may expect heterogeneous treatment effects to appear over time. I note that in the presence of heterogeneous treatment effects like this, the standard two-way fixed effects difference-in-differences model with staggered treatment timing may be biased (Roth et al 2021). To address this, I follow the work of Callaway and Sant'Anna (2021). Estimating their model for group-time effects, which splits estimation based on when units are treated, I find a significant decrease of 18.3% in the share Black of the population for the earliest treated tracts. I also find a significant increase of 8% in the college educated share of the population for both the earliest treated and middle treated tracts. Finally, I find a significant decrease in the number of dwelling units below the national median rent of 14.4% and 4.9% for the earliest and middle treated tracts respectively.

These effects suggest that early projects, which focused on demolition and rebuilding housing in predominantly Black neighborhoods, had significant impacts on both neighborhood demographics and rental prices – the original Black residents moved out and more educated residents moved into higher priced housing after project completion. The middle treated tracts target a more racially diverse group of neighborhoods, and likely experience a more muted effect on rental prices because this group is estimated with a wider range of project types – there were fewer demolition projects building new housing. However, more educated residents still moved to project neighborhoods. Finally, latest treated tracts do not experience any significant effects, suggesting that the original residents, or residents similar to the original residents, continued to live in renewed neighborhoods and benefited from smaller improvements that were not enough to attract new residents. I also note that the latest treated group had the smallest share of units below the national median rent prior to projects' start, indicating that these neighborhoods were richer to begin with. The overall lack of impact on homeownership likely stems from the fact that the housing that was built was predominantly rental properties (Hirsh 1983). Future work in this area could further investigate these heterogeneous treatment effects, particularly the lack of impact on the Black population in middle and latest treated tracts, and the lack of impact on rental prices in the latest treated tracts.

Broadly, this paper contributes to the literature exploring the effects of place-based policies (Busso et al. 2013; Freedman et al. 2021; Neumark and Simpson 2014; Kline and Moretti 2014) and to the literature studying policies that affect residential segregation in the United States (Shertzer et al. 2014; Fishback et al. 2020; Aaronson et al. 2021; Logan and Parman 2017). Work at the intersection of these two strands of literature often focuses on the impact that place-based policies have on minority communities. For example, Aaronson et al. (2021) study the long-term impact of redlining on neighborhoods and find that redlined neighborhoods have lower homeownership rates, lower home values, and increased residential segregation. Likewise, Carter (2019) shows that census tracts in Detroit where interstate highways were constructed experienced a decline in property values as well as a decline in the percentage of Black residents.

Considering urban renewal specifically, Collins and Shester (2013) estimate the effect of urban renewal projects on city-level income, property values, employment, and population for multiple cities in the United States. They find significant positive effects on income, property values, and population, which are not driven by any city-level changes in demographic composition. I extend the work of Collins and Shester by estimating the impact of projects at the census tract level to look for within city effects. It is not necessarily the case that city-level positive effects are evenly distributed across neighborhoods in a given city, and there may be changes in the demographic composition of neighborhoods that are not detectable at the city level.

This paper proceeds as follows: section 2 provides background and historical context, section 3 describes the data, section 4 describes the empirical method, section 5 presents the results, and section 6 concludes.

2.2 Background

2.2.1 Historical Context

Beginning with the Housing Act of 1949, the federal government offered funding to cities looking to demolish and rebuild dilapidated neighborhoods. This act emphasized public housing goals and required that any urban renewal projects be primarily residential. Title III of the act specifically set aside funds for up to 810,000 public housing units to be built over a six year period. In 1954, Congress passed a second Housing Act broadening the scope of the original. This act moved the emphasis from housing to more general urban redevelopment, including commercial centers and public facilities. A series of amendments to the 1954 act allowed for specific additional uses of federal funds, for example, funds awarded to universities and hospitals. This act also required cities to create planning departments and submit comprehensive redevelopment plans in project applications. The Housing Act of 1965 extended the urban renewal program and authorized grants for up to two-thirds of the cost of building neighborhood facilities. Finally, in 1974, the urban renewal program was replaced by the Community Block Development program. This program consolidated existing grant programs so cities could apply for funding after identifying community-specific needs.

In this paper, I focus on the city of Chicago. Chicago is often thought of as the poster child of urban renewal and numerous qualitative reports and newspaper accounts detail projects in different areas of the city. Contemporary reporting often covered urban renewal projects positively, highlighting the economic benefits to the city (Cress 1958). An article covering a study from the University of Chicago reported that urban renewal greatly reduced substandard housing and vacant land, allowing families to live in better conditions. It reports that 94% of those displaced moved from substandard to standard housing (Getz 1966). Newspaper coverage followed specific projects as well. For example, several articles in the Chicago Tribune covering a project in the Hyde Park-Kenwood area praise the project throughout the demolition and building process for eliminating blight and replacing overcrowded, unsafe buildings with new housing (Stevens 1960; Buck 1964; Busse 1967). Similarly, coverage of the project in Lake Meadows discusses how new apartments, a new shopping center, and a proposed park have revitalized the area (Unger 1956). The Hyde Park project was even heralded as leader of integration as new residents moved to the area (Yackley 1967).

While there may have been many benefits to urban renewal, not all coverage was positive. Some articles acknowledged that many saw urban renewal as a way to remove Black residents from neighborhoods and break up Black communities (Anderson 1964; Getz 1966). Others highlighted the struggles of business owners and homeowners after demolition (Hawkins, 1964; Bach, 1959). Coverage of the Near West Side project in particular described the difficulty the Italian Americans living in the Near West Side faced of maintaining their homes, community, and identity (Caputo 1968).

Many qualitative reports taking a retrospective view of urban renewal follow in this vein. They discuss how urban renewal was used to push out poor, minority residents and attract wealthier, white families back to the city from the suburbs (Hirsh 1983, "Renewing Inequality" 2018). For example, modern day descriptions of the Lake Meadows project note that it cleared out middle class property (Pritchett 2003) and displaced over displaced over 3,000 families of color ("Renewing Inequality" 2018), with only 900 of the original residents relocated to public housing or returned to the completed project (Reinberger 2021). Other work follows these displaced families to the nearby Hyde Park-Kenwood area, where they were eventually once again removed when the University of Chicago applied for federal funds to redevelop the area (Winling 2018).

2.2.2 Tract Selection

Originally, three major city agencies oversaw general urban renewal efforts. The Chicago Housing Authority built and operated public housing projects, the Community Conservation Board oversaw programs for the modernization of existing buildings, and the Chicago Land Clearance Commission (CLCC) cleared blighted tracts for redevelopment (The Chicago Tribune, 1962). In 1962 the CLCC was joined with the Community Conservation Board to create the Department of Urban Renewal (DUR) to streamline management of urban renewal projects (The Chicago Tribune, 1962).

The CLCC, and subsequently the DUR, selected areas of the city for renewal projects by collecting data on levels of dilapidation. The agencies commissioned studies to create reports on indicators of disrepair, such as overcrowding and lack of sanitary facilities (The Chicago Tribune 1965). Members of the agencies' boards then selected sites for demolition based on these reports. Site selections were sent to the city council, which held public hearings and voted on whether the area could become a project (The Chicago Tribune 1965). Residents and community groups had the opportunity to voice support and concerns at public meetings (The Chicago Tribune 1959; Getz 1967).

Once approved, land was acquired for projects either by purchasing it from tenants or through the use of eminent domain. Urban renewal agencies contracted with appraisers to determine the land's value, then sent offers to the land owners. The federal government also had to approve the land valuation (Priddy 1955). While eminent domain was an option, the agencies preferred to use it as a last resort and often tried negotiate with owners instead (Priddy 1955). The city offered relocation services to people in demolition areas, however few used these services beyond looking at real estate listings (Priddy 1955).

Urban renewal agencies then contracted with private developers to build projects. Agencies specified the use of the land, then developers submitted bids for the land along with plans for construction (Ziemba 1969). Agencies reviewed plans, and in a process similar to that for land clearance, sent their preferred plan to the city council and federal government for approval (Ziemba 1969). Often projects took many years to complete after the decision process (Buck, 1967). In certain cases, interest from an outside organization prompted the urban renewal process. For example, Michael Reese Hospital was located in what was considered a dilapidated neighborhood and the hospital board was interested in expanding. So, the hospital lobbied the CLCC and gathered business groups also interested in the land (Unger 1956). Similarly, a project in the Near West Side was spurred by the University of Illinois expressing an interest in buying any newly constructed housing units for the campus (The Chicago Tribune 1965).

The Chicago Department of Urban Renewal Records at the Chicago Public Library collects and maintains records from the CLCC and DUR. Some of these records as well as photographs from surveys of dilapidated neighborhoods have been digitized and are publicly available. I use these records to divide census tracts into three categories. Tracts chosen for urban renewal are those which survived the entire decision process and were used for urban renewal projects. Potential site tracts are those which the urban renewal agencies collected information about, but ultimately did not choose for urban renewal projects. Never considered tracts are those for which the agencies did not collect data. I do not observe the reason one tract was chosen over another, and potential site tracts may have been removed from consideration for urban renewal at any point in the decision-making process. The data collection may have revealed the tract was not in a bad enough state of disrepair, the city council may not have approved the tract for urban renewal, residents of the area may have resisted urban renewal successfully, or there may have been another reason a potential site was not chosen as a project site.

2.3 Data

2.3.1 Urban Renewal Projects

Data on urban renewal projects comes from the Digital Scholarship Lab at Richmond University. They have collected data from multiple sources, including the federal government's quarterly urban renewal project reports and the Department of Housing and Urban Development's annual reports, to compile a comprehensive data set of urban renewal projects. The data contains project names, project start and end dates, federal grant amounts, and project boundaries. The Digital Scholarship Lab has also collected information about number of families displaced, number of substandard dwelling units, and proposed land use, however, this data is more limited and I do not make use of it in this paper.

I use data only for Chicago in this paper. Figure 1a displays a map of the city with project areas highlighted. Figure 1b shows the same map with the addition of the potential site tracts, following the selection process described above. Potential project sites tend to be located near actual sites of urban renewal projects.

Table 1 shows the total average federal grant amounts approved and disbursed for projects. The grant amount approved is often much larger than the amount received for a project, and there is a large amount of variation in the grant amounts, indicating that some projects were far more expensive than others.

2.3.2 Census Data

Demographics and dwelling unit characteristics come from the Census. I use census tract level data from 1930-2010. I choose census tracts as the level of geographic aggregation because they are small enough to estimate within-city effects, and census tracts in Chicago tend to be relatively constant over time. This data contains total population, population by race, population by education level, total number of dwelling units, total dwelling units below the national median rent, and total number of owner occupied homes.

I combine the census data with the urban renewal projects data using a spatial merge. This ensures that even if tract boundaries do change over time, the physical areas of urban renewal projects will be constant in my data set. Since the census data is only available every 10 years, I match project start years to the closest available census wave: projects that began in 1949 and 1950 are matched to start with the 1950 census, projects that began after 1950 and before 1961 are matched to the 1960 census, and projects that began after 1960 are matched to the 1970 census.

Table 2 shows the pre-period means for census tracts by their project status. Project tracts tend to have a larger share Black of the population and a slightly larger college educated population. Project tracts also have a larger share of units below the national median rent and a lower share of owner occupied units than tracts that did not have urban renewal projects. Potential site tracts tend to be on average more similar to project tracts than are tracts that were never considered as project sites, however there are still significant differences between the two. Table 3 shows the means in the post-period.

2.4 Empirical Method

2.4.1 Difference-in-Differences Estimation

To estimate the effects of urban renewal projects, I use a difference-in-differences design, comparing urban renewal project tracts to potential site tracts. I use potential site tracts as the comparison group because simply regressing the outcomes on whether or not a tract had an urban renewal project will result in a biased estimate since poorer, run-down neighborhoods were chosen for an urban renewal projects. Potential sites provide a valid comparison group if potential site tracts were not chosen for as good as random reasons, and if there are no spillovers. While I do not observe the reason one tract was chosen over another, the tract selection process could have eliminated a tract at any point for a number of reasons unrelated to the level of dilapidation in the neighborhood. I address the possibility of spillovers below.

Tracts were treated between 1950 and 1970, and most projects took a long time to complete. Because of this I estimate both a standard two-way fixed effects model for a summary measure of the effects, as well as an event study version of the difference-in-differences model to see the effects in each time period. The estimates will give the average treatment effect on the treated (ATT).

The standard two-way fixed effects differences-in-differences model is

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 * treated_i + \beta_3 X_{it} \gamma_t + \delta_i + \epsilon_{it}$$

where D_{it} is a dummy variable that equals 1 for treated units in the post-treatment period, γ_t are year fixed effects, and δ_i are tract fixed effects. X_{it} are time-varying controls for total population and total dwelling units. Y_{it} are the outcome variables: share Black of the population, the share of the population with a college degree, the share of owner occupied dwelling units, and the share of dwelling units below national median rent. Robust standard errors are clustered at the tract level.

I also estimate the event study version of the difference-in-differences model,

$$Y_{it} = \alpha + \sum_{k=-4}^{k=-1} \beta_k treat_{ik} + \sum_{k=0}^{k=5} \beta_k treat_{ik} + \theta X_{it} + \gamma_t + \delta_i + \epsilon_{it}$$

where $treat_{ik}$ is an indicator for treatment in time period k, with time period -1 as the reference period. In this data, time periods correspond to census waves so the reference period is 10 years before the project start census wave, time period 1 is 10 years after the project start wave, and so on. γ_t are year fixed effects, and δ_i are tract fixed effects. X_{it} are time-varying controls for total population and total dwelling units. Y_{it} are the outcome variables listed above. Robust standard errors are clustered at the tract level. In this equation the β_k are the coefficients of interest, measuring the effect of urban renewal projects in each time period.

Since people had to move for demolition, and they often moved to nearby, affordable housing, they may have moved to the potential sites, which would cause the estimation to pick up spillover effects. To investigate this possibility, I estimate the two-way fixed effects model above comparing potential sites to never considered sites. It is unlikely that many people from demolished neighborhoods moved to never considered tracts because these tracts were more expensive to live in. They also tended to be located farther away from urban renewal projects than potential site tracts. Finding no effect would increase confidence that the estimation is not simply picking up spillovers.

2.4.2 Potential for Heterogeneous Treatment Effects

Since there is substantial variation in urban renewal projects, there is potential for heterogeneous treatment effects. For example, building a housing development vs. an industrial project could affect people's ability and desire to move to the renewed neighborhood. There may also be heterogeneous treatment effects across time, for example if earlier treated neighborhoods have larger demographic changes because people have a longer period of time to move. In the case of urban renewal projects, differences across time and project type will be closely related because the laws governing which types of projects cities were allowed to build became more lenient over time.

Because tracts were treated in different years, the above models estimate a case of staggered treatment timing. In this case, when there is potential for heterogeneous treatment effects across groups, the estimates may be biased (Roth et al 2021). In particular, the standard two-way fixed effects model may use units that are already treated as “control” units in comparisons (Callaway and Sant’Anna 2021). Since the standard two-way fixed effects model produces a weighted average of some underlying treatment effect parameters, these comparisons may result in negative weights, potentially even changing the sign of the estimate (Goodman-Bacon 2021). Even if these weights are positive, they are sensitive to the size of each group, the timing of treatment, and the total number of time periods (Goodman-Bacon 2021).

To address this, I estimate an additional model following Callaway and Sant’Anna (2021) to obtain an appropriately weighted ATT. Callaway and Sant’Anna’s method estimates group-time effects: the effect for each set of units treated in a particular time period. In my case, there are three groups: tracts treated in 1950, tracts treated in 1960, and tracts treated in 1970, following the project start year assignment described above. I follow their method using a never-treated group as the comparison, in this case the potential project sites. See Callaway and Sant’Anna (2021) for more detail of the process.

2.5 Results

2.5.1 Means Over Time

Figures 2, 3, and 4 display mean characteristics over time for each tract type (urban renewal tracts, potential site tracts, and never considered tracts). The project time period takes place between 1950 and 1970. Figure 2 shows the mean total population and the mean number of total dwelling units. Both begin to decrease in the project period, as a natural consequence of demolition, and continue to stay at lower levels until 2010.

Figures 3 and 4 show the mean share of various outcomes. The share Black of the population starts out highest in urban renewal tracts and increases during the project period, potentially because earlier projects focused on housing. It appears to stay relatively constant in the post-project period before decreasing in 2010. The share of the population with a college degree starts out at a similar level for all tract types, then urban renewal tracts experience a larger increase over time relative to the other census tracts, suggesting that more educated residents moved to urban renewal tracts.

The share of owner occupied dwelling units increases over time for all tract types, with urban renewal tracts experiencing a relatively steeper increase in the post-project period. The share of units below national median rent decreases over time for all tract types, and it does not appear that urban renewal tracts are substantially different than the other groups of tracts for most of the post period.

Figures 5, 6, and 7 show the means over time for urban renewal tracts centered around event time zero. Again, there is a decrease in the total population and total dwelling units after projects begin. The share Black of the population increases until event time zero, after which point it seems to level off and eventually decrease slightly, suggesting Black people were moving to project neighborhoods up until the time of the project. The share of the population with a college degree increases slowly before event time zero, then increases more steeply in the post-period. The share of owner occupied units is increasing throughout all time periods, but increases more steeply after the project begins. Finally, the share of units below median rent decreases leading up to event time zero, experiences a small increase, then decreases again.

2.5.2 Difference-in-Differences Estimation

The difference-in-differences models rely on the parallel trends assumption, that in the absence of treatment, the difference between urban renewal tracts and potential site tracts would be constant over time. The validity of this assumption can be seen in the event study style graphs displayed in figure 8. Of the pre-period coefficients for the share Black of the population, two are not significant and one is close to being insignificant. None of the coefficients are large. Figure 3 shows that project tracts experience a slightly faster increase than potential site tracts in the final pre-period time period, however these groups appear to follow a similar trend for most of the pre-period. The share of the population with a college degree does not appear to exhibit any significant pre-trend. The pre-period coefficients are not significant and are very near zero. In addition, figure 3 shows potential site tracts and urban renewal tracts following a similar trend in the pre-period.

The pre-period coefficients for the share of owner occupied homes are very close to zero, and only one is slightly significant. Figure 4 suggests there may be a slight difference in trends in the pre-period, with potential site tracts increasing at a faster rate than urban renewal tracts. Finally, the pre-period coefficients for the share of units below the national median rent are both not significant, and one is zero. Figure 4 also suggests that urban renewal tracts and potential site tracts were following a a similar pre-treatment trend.

The estimates of the effects in the post-period are fairly imprecise, however, there is an initial increase in the Black population, followed by decreasing coefficients, which level off at 30 years post-project. The coefficients for the population with a college degree are positive and increasing after the initial effect. The initial effect on the share of owner occupied units is positive. The coefficients then decrease before beginning to increase after 30 years post project. Finally, the coefficients for the share of units below the national median rent are all not significant. There are positive coefficients until 30 years post project, then the coefficients begin to decrease.

Column (1) of Table 4 contains the results of the two-way fixed effects estimation specification described above. Only one of the outcome variables has a significant effect: the share of the population with a college degree. There is an increase of 5.3% in the share of the college educated population, which is significant at the 10% level.

2.5.3 Heterogeneous Treatment Effects

The results from the Callaway Sant'Anna (2021) estimation are shown in Table 5. The overall ATT is an average of the group-time coefficients, weighted by group size. The overall ATT for the share of the population with a college degree is now a 7.9% increase, significant at the 1% level. This is being driven mostly by the middle treated tracts, which experience an 8.7% increase. The earliest treated tracts also have a an 8% increase the share of the college educated population, which is significant at the 5% level. The coefficient for the latest treated tracts is much smaller and not significant.

The overall ATT for the share of dwelling units below median rent is also now significant, although only at the 10% level. Treated tracts on average experience a 4.4% decrease in the share of units below national median rent. Again this is driven by the earliest and middle treated tracts. The earliest treated tracts experience a 14.4% decrease, and the middle treated tracts experience a 4.9% decrease. There is no significant effect for the latest treated tracts.

Most coefficients for the share Black of the population are not significant, though they are all negative. However, the earliest treated group experiences an 18.3% decrease, significant at the 10% level. Finally, there are no significant effects on the share of owner occupied units.

2.5.4 Discussion of Group Effects

The differences in the effects across treatment groups might be driven by several factors, for example by project type and timing or by pre-existing neighborhood composition. Table 6 contains the pre-treatment means of the outcome variables for each group. The most notable difference is in the share Black. The mean share Black for group 1 is 0.78, while groups 2 and 3 have a mean of 0.19 and 0.15 respectively. This suggests the negative effect on the Black share for group 1 could be because this group contained the highest proportion of Black residents prior to demolition. For example, group 1 contains projects like Lake Meadows, which is well known for demolishing a Black neighborhood, including some middle class property (Pritchett 2003, Reinberger 2021). In addition, early projects focused heavily on land clearance and building new housing. The large, negative effect on the share of dwelling units below median rent is likely a result of these project goals – early projects built new housing. Taken together with the increase in the college educated population, the group 1 estimates suggest that the early projects likely followed the classic story of urban renewal: Black residents were removed for more educated, white residents to move back to the city.

Group 2 is the largest group and contains a wider mix of projects. It seems that urban renewal projects in this group targeted a more racially diverse mix of neighborhoods, and there is no significant effect on the Black population. However, there is a strong positive effect on the college educated population, suggesting still that more educated people moved to renewed neighborhoods. Many projects in group 2 still focused on housing. For example, it contains the Hyde Park-Kenwood and Near West Side projects, both of which built housing purchased by universities. However, laws governing urban renewal had become more lenient by this time, so this group also contains projects that focused more on commercial construction and neighborhood amenities (Cress, 1958). The project diversity likely accounts for the smaller magnitude of the effect on the number of units below median rent as compared to group 1. Though new housing would have higher rent prices, it could be that adding other types of amenities to neighborhoods did not have as drastic of an impact on rental prices.

Finally, as the latest treated group, group 3 contains fewer demolition projects. As time went on, the city began to move away from large land clearance projects to smaller demolition zones and addditions (Getz, 1967). The lack of significant effects on the Black population and college educated population are likely driven by the decrease in demolition – fewer people were removed from their neighborhoods. However, this also suggests that the new construction that was completed during this time did not attract new residents or result in significant changes in rental prices. I also note that group 3 had a relatively low share of dwelling units below national median rent prior to projects, only 0.24 on average. This suggests that group 3 neighborhoods were already home to wealthier residents, thus projects had a more limited effect.

Given the lack of impact for the latest treated group, and the smaller impact on rental prices for the middle treated group, I also estimate the effect of projects on tract level income measures for groups 2 and 3. The census does not have a consistent measure of income for the entire time period, so I estimate the effect for group 2 using the share of people below national median personal income from 1950-1980, and for group 3 using the share of families below national median family income from 1960-2000. The results are in table 7. There are no significant impacts for either group and the coefficients are small, suggesting projects did not necessarily attract wealthier people in the later time periods.

2.5.5 Additional Analysis and Sensitivity Analysis

To investigate the presence of spillover effects, I conduct an analysis comparing potential site tracts to never considered tracts. I estimate the two-way fixed effects model above. Table 8 contains the results. None of the effects are significant, and they are all relatively small.

The estimates for share Black of the population and the share with a college degree are 2.6% and 0.4%. This suggests that the above results are not being driven solely by demographic changes in the potential site tracts.

Finally, to assess the sensitivity of the above estimates to violations of the parallel trends assumption, I conduct an analysis following Rambachan and Roth (2022). The parallel trends assumption may be violated in two main ways. First, it can be violated if treated and untreated groups are affected differentially by shocks. For example, negative economic shocks may be exacerbated in poorer neighborhoods. In this case, project tracts may be more susceptible to negative shocks than potential site tracts, since, on average, project tracts tend to be worse off. Second, the parallel trends assumption may be violated by smoothly evolving trends that affect the treated and untreated groups differently. For example, while the population attaining a college degree increases for all groups over time, it may increase faster for richer people, who also live in better neighborhoods.

Rambachan and Roth (2022) propose two analyses which place restrictions on post-treatment differences between treated and comparison groups, given the pre-treatment differences in trends. The relative magnitude analysis addresses the first type of violation in the parallel trends assumption: that of differential shocks. This analysis assumes that any post-period shocks are not too different from shocks in the pre-period, leaving the researcher to choose the bounds on how much shocks are allowed to differ. The smoothness analysis addresses the second type of violation: that of smoothly evolving trends. This analysis restricts the extent to which the slope of a differential trend may change in consecutive periods. See Rambachan and Roth (2022) for further details.

I conduct the analysis for all three outcome variables for which I find at least one significant effect: the share Black of the population, the share of the population with a college degree, and the share of units below median rent. I opt to use a weighted average of the event study coefficients from the Callaway-Sant'Anna analysis, both for ease of viewing and given that individual period effects are often insignificant. I set the significance level to be 10% to match the least significant of the above results.

Figure 9 displays the results for the share Black of the population. Panel A shows the results from the relative magnitude analysis. The Mbar parameter dictates the size of post-treatment shocks. The estimate breaks down with shocks only one-fourth as large as those in the pre-period, indicating that this estimate is sensitive to this type of violation in parallel trends. Panel B shows the results of the smoothness analysis. Here, the M parameter dictates how non-linear the differences in trends can be, with M = 0 indicating a linear trend. The estimate breaks down with M larger than 0.01.

Figure 10 displays the results for the share of the population with a college degree. Again, Panel A displays the relative magnitude analysis, which indicates that the estimate breaks down with violations one-half as large as any pre-treatment shocks, so this estimate is slightly less sensitive to this particular violation. Panel B displays the smoothness analysis results, showing that the estimate breaks down with the allowance of a trend beyond linearity.

Finally, figure 11 shows the results for share of units below national median rent. Panel A shows that this estimate breaks down with violations one-fourth as large as pre-treatment shocks, similar to the estimate for the share Black of the population. Panel B shows that the estimate breaks down with the allowance of a linear trend. This indicates this estimate is particularly sensitive to smoothly evolving differential trends.

2.6 Conclusion

This paper estimates the impact of urban renewal projects in Chicago, finding that in some cases projects had a large, negative impact on the Black population and in many cases had a positive effect on the college-educated population. The analysis suggests that large scale demolition and reconstruction changed neighborhood demographic composition, but that other projects not focused on land clearance and housing did not have the same impact. This is also reflected in the decrease in the share of units below national median rent – project tracts became more expensive to live in when housing demolition and reconstruction was the focus, but rental prices do not appear to change much with smaller additions of amenities.

There are some limitations in the interpretation of these results. I find they are sensitive to violations of the parallel trends assumption. I am also limited by data availability of project type. While I attempt to look for heterogeneous effects, I am unable to estimate the regressions by project type – the effects of housing, industrial, and commercial projects are all estimated together. Separating projects by treatment year offers some insight, but does not fully correct this issue.

Finally, I only study Chicago in this paper. Urban renewal projects were widespread throughout the United States, including many projects in smaller cities. Future work then could consider the effects of urban renewal in other cities to see if the results found here hold. Future research efforts could also investigate the lack of impact later projects had on neighborhoods, particularly the fact that there does not appear to be changes in housing prices. Understanding the impact of demolition and reconstruction projects is important for policymakers facing issues of dilapidation and residential segregation. Studying historical policies can provide insight into current city structures, as well as give urban planners guidance in designing future policies.

Tables

Table 2.1: *Urban Renewal Projects: Mean Grant Amounts*

	Mean and SD
Total federal grants approved	$38,200,000
	($39,800,000)
Total federal grants disbursed	$15,800,000
	($23,800,000)

Notes: This table shows the mean total federal grant amounts approved and disbursed for urban renewal projects in Chicago. It also includes the standard deviation in parentheses under the mean.

Table 2.2: *Pre-Period Mean Characteristics By Tract Type*

	Urban Renewal	Potential Site Tracts	Never Considered
Total population	3733.68	3541.87	3988.84
Share population Black*	0.22	0.12	0.03
Share population with college degree*	0.06	0.04	0.04
Total dwelling units	1210.86	1067.94	1203.68
Share owner occupied units*	0.12	0.29	0.43
Share units below national median rent*	0.50	0.38	0.23
N	507	964	1,092

Notes: This table shows the mean characteristics for each census tract type in the pre-project period. Column (1) shows urban renewal tracts, which includes all tracts that were part of urban renewal projects. Column (2) shows potential tracts, which were those tracts considered as project sites, but were ultimately not chosen. Column (3) shows the means for tracts that were never considered for urban renewal. Statistically significant differences in the means are indicated with *.

Table 2.3: *Post-Period Mean Characteristics By Tract Type*

	Urban Renewal	Potential Site Tracts	Never Considered
Total population*	2604.73	3422.34	4198.43
Share population Black*	0.51	0.43	0.26
Share population with college degree*	0.30	0.18	0.18
Total dwelling units	1221.71	1372.51	1560.35
Share owner occupied units*	0.20	0.36	0.50
Share units below national median rent*	0.33	0.26	0.16
N	635	1,342	1,896

Notes: This table shows the mean characteristics for each census tract type in the post-project period. Column (1) shows urban renewal tracts, which includes all tracts that were part of urban renewal projects. Column (2) shows potential tracts, which were those tracts considered as project sites, but were ultimately not chosen. Column (3) shows the means for tracts that were never considered for urban renewal. Statistically significant differences in the means are indicated with *.

Table 2.4: *Main Specification Results*

	DID	Pre-period Mean Project Tracts
Black Population	-0.069	0.22
	(0.053)	
Population with College Degree	0.053*	0.06
	(0.030)	
Owner Occupied Dwelling Units	0.038	0.12
	(0.023)	
Dwelling Units Below National Median Rent	0.040	0.48
	(0.043)	
N	3,351	

Notes: This table shows the results for the main two-way fixed effects difference-in-differences specification with outcomes reported as shares. Column (1) shows the results of the two-way fixed effects difference-in-differences model. Column (2) contains the pre-period mean for the outcome variables in treated tracts. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

Table 2.5: *Callway and Sant'Anna (2021) Specification Results*

	Black Population	College Degree	Owner Occupied	Below Median Rent	N
Overall ATT	-0.070	0.079***	0.016	-0.044*	3,351
	(0.043)	(0.018)	(0.016)	(0.024)	
Group 1	-0.183*	0.080**	-0.038	-0.144**	2,171
	(0.105)	(0.041)	(0.031)	(0.053)	
Group 2	-0.051	0.087***	0.026	-0.049*	3,033
	(0.047)	(0.019)	(0.018)	(0.028)	
Group 3	-0.088	0.034	0.002	0.067	2,277
	(0.121)	(0.052)	(0.036)	(0.045)	

Notes: This table shows estimates of the average treatment effect (ATT) obtained using the Callaway and Sant'Anna (2021) model for staggered treatment timing with heterogeneous effects. Outcomes are reported as shares. Column (1) shows the results for the Black share of the population. Column (2) shows the results for the share of population with a college degree. Column (3) shows the results for the share of owner occupied dwelling units. Column (4) shows the results for the share of dwelling units below the national median rent. Column (5) shows the number of observations in the regression. Group 1 is the earliest treated census tracts. Group 2 is the middle treated census tracts. Group 3 is the latest treated tracts. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

Table 2.6: *Callaway and Sant'Anna (2021) Specification Pre-treatment Means*

	Group 1	Group 2	Group 3
Proportion Black	0.78	0.19	0.15
Proportion with College Degree	0.03	0.06	0.10
Proportion Owner Occupied	0.05	0.13	0.10
Proportion Rent Below National Median	0.66	0.54	0.24

Notes: This table shows the pre-treatment mean characteristics for each group of census tracts in the Callaway-Sant'Anna (2021) specification. Column (1) shows means for Group 1, the earliest treated tracts. Column (2) shows means for Group 2, the middle treated tracts. Column (3) shows means for Group 3, the latest treated tracts.

Table 2.7: *Results for Income*

	(1)	(2)
Group 2	0.043	1,763
	(0.065)	
Group 3	-0.029	2,189
	(0.021)	

Notes: This table shows estimates for the effect of urban renewal projects on the tract-level share of people below median income. Column (1) shows the estimate and column (2) shows the number of observations. Group 2 is the middle treated census tracts. Group 3 is the latest treated tracts. The outcome for group 2 is the share of people below national median personal income and covers the time period 1950-1980. The outcome for group 3 is the share of people below national median family income and covers the time period 1960-2000. The two outcomes are because the census lacks a consistent income measure for the entire study period. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

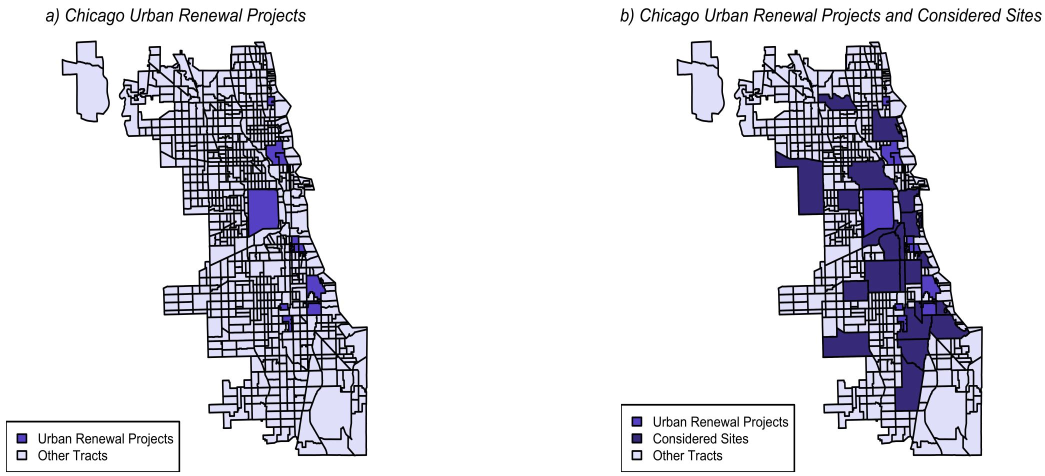
Table 2.8: *Placebo Test Results*

	Placebo DID
Black Population	0.026
	(0.020)
Population with College Degree	0.004
	(0.007)
Dwelling Units Below National Median Rent	-0.022
	(0.015)
N	5,213

Notes: This table shows the results for the placebo test comparing potential site tracts to never considered tracts using the two-way fixed effects model. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

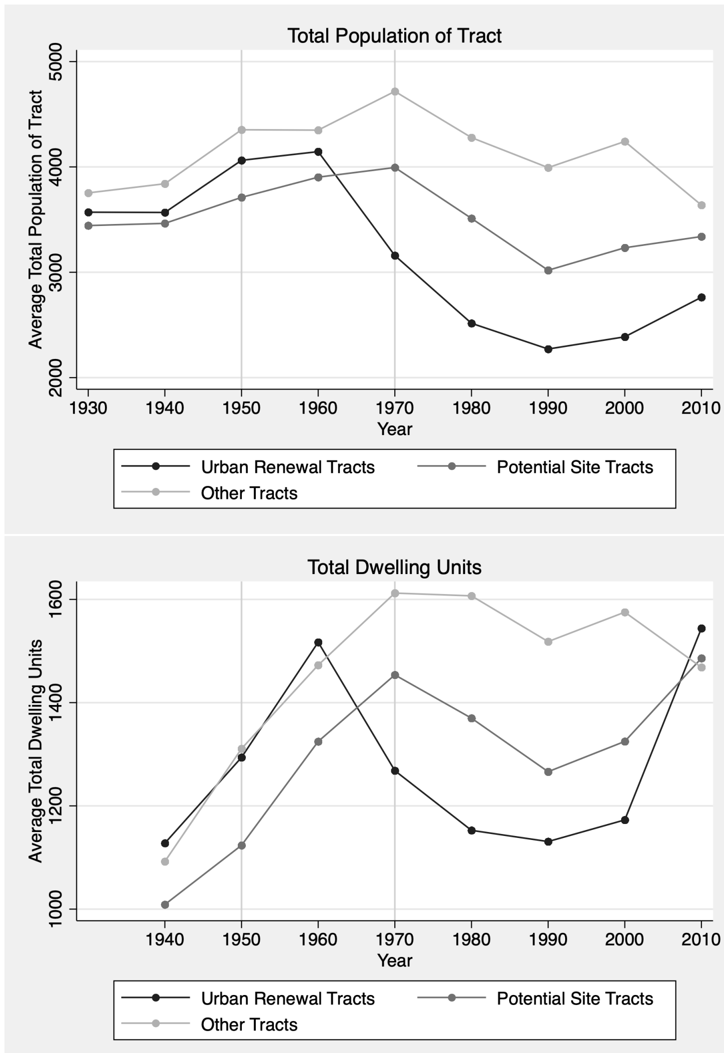
Figures

Figure 2.1: *Maps of Chicago*



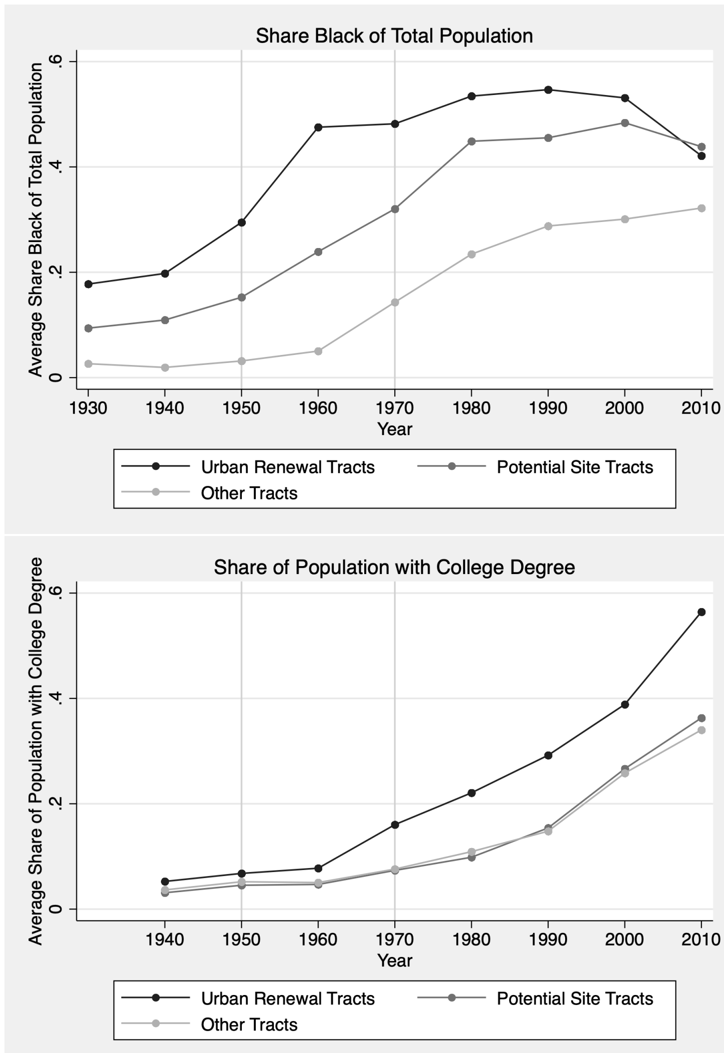
Notes: This figure shows a map of census tracts in Chicago. Panel A shows Chicago with urban renewal areas highlighted. Panel B shows Chicago with both urban renewal areas and potential site areas highlighted. Potential sites are those that were considered but ultimately not chosen for urban renewal. Never considered sites are all tracts that are not urban renewal tracts and not potential sites.

Figure 2.2: *Graphs Over Time: Population and Dwelling Units*



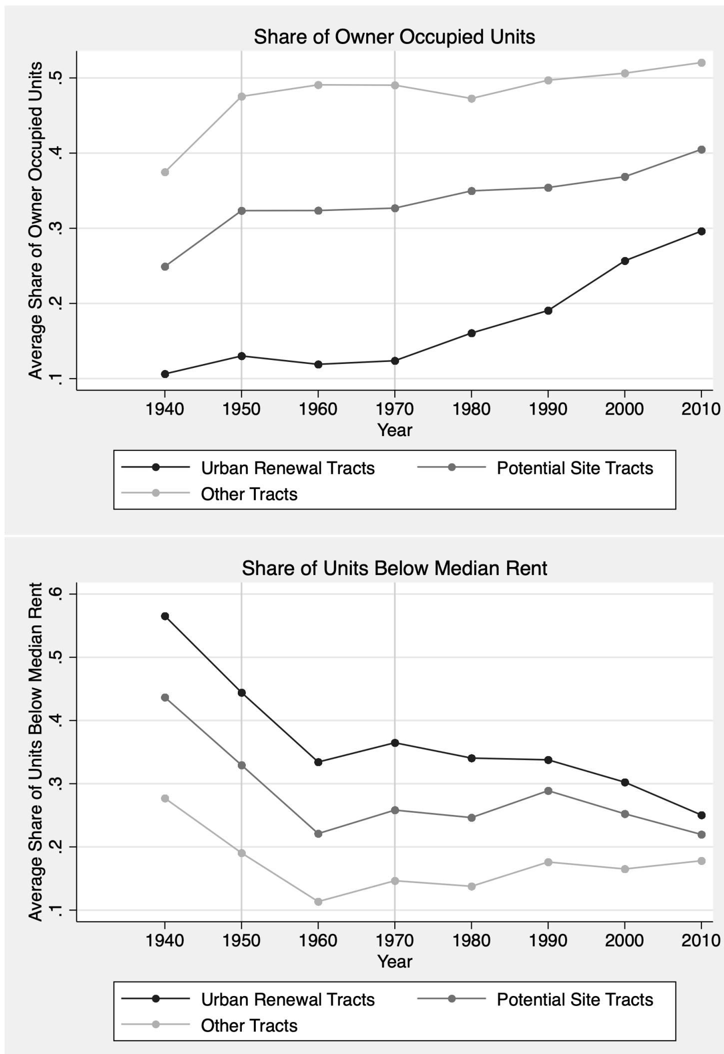
Notes: This figure shows mean characteristics over time for census tract characteristics by tract type. The first graph displays the mean total population of tracts. The second shows mean total dwelling units in tracts.

Figure 2.3: *Graphs Over Time: Demographics*



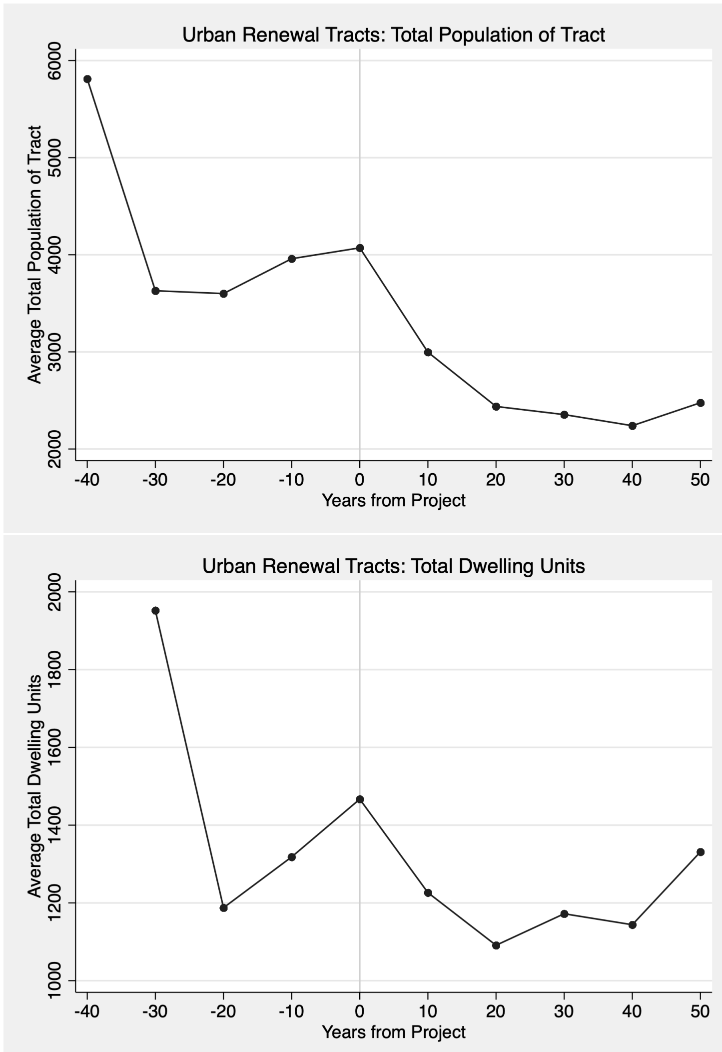
Notes: This figure shows the mean characteristics over time for census tract characteristics by tract type. The first graph displays the mean share Black of the total population. The second shows the mean share of the population with a college degree.

Figure 2.4: *Graphs Over Time: Housing*



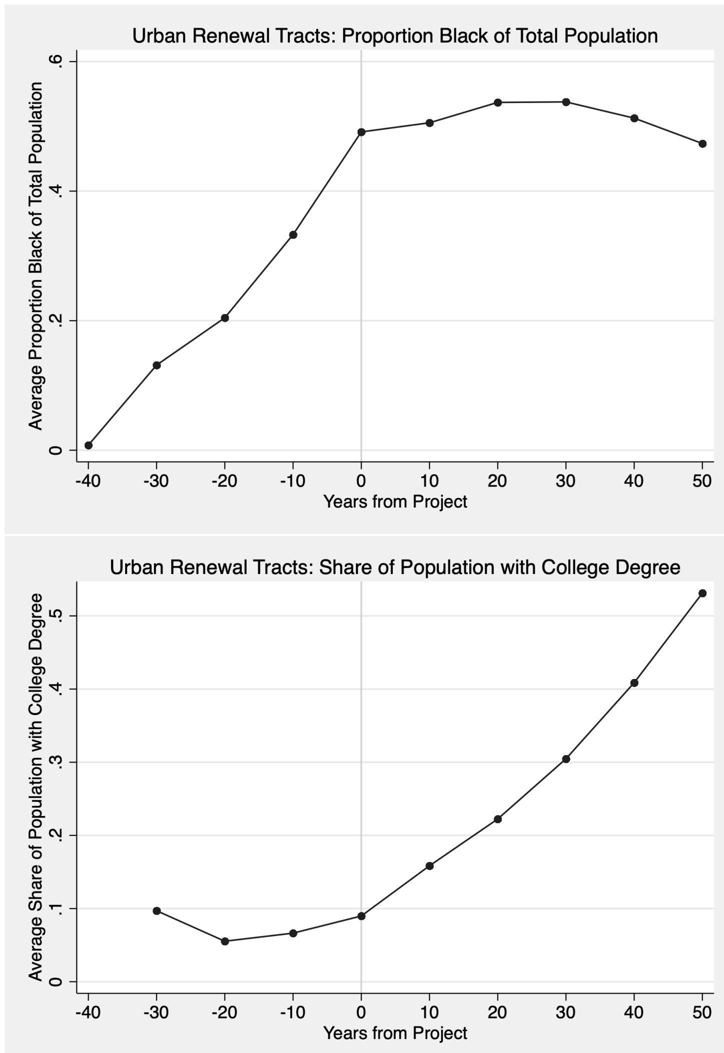
Notes: This figure shows the mean characteristics over time for census tract characteristics by tract type. The first shows the mean share of owner occupied dwelling units, and the second graph shows the mean share of units below the national median rent,.

Figure 2.5: *Graphs Over Time for Project Tracts, Centered at Zero: Population and Dwelling Units*



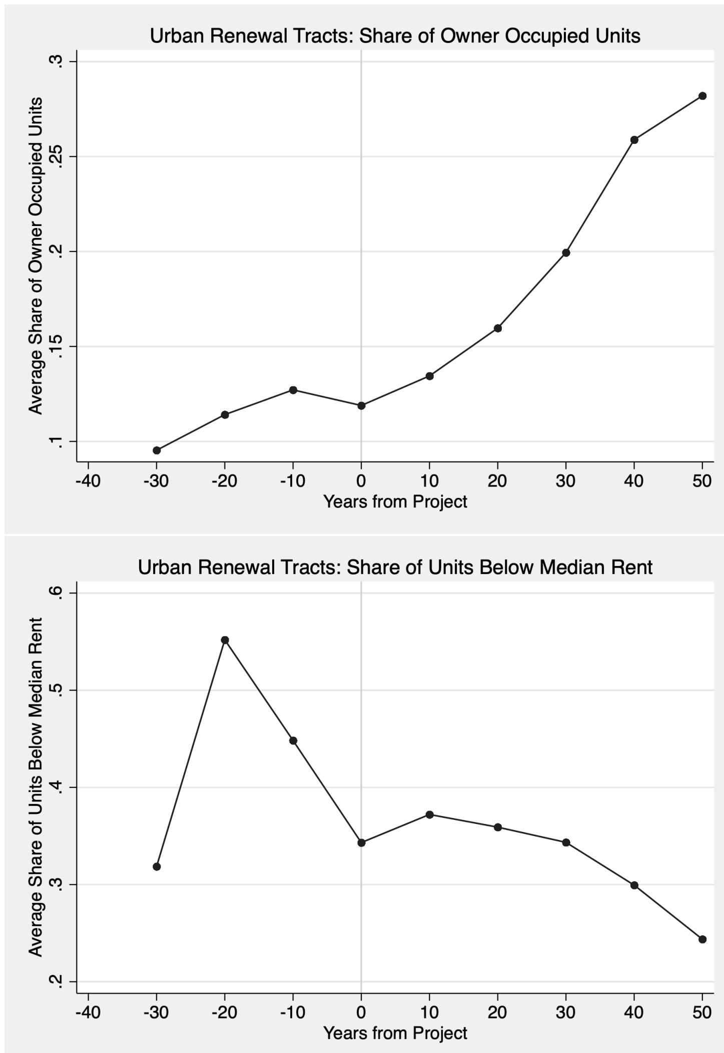
Notes: This figure shows mean characteristics over time for census tract characteristics for urban renewal tracts centered at event time zero. The first graph displays the mean total population of tracts. The second shows mean total dwelling units in tracts.

Figure 2.6: *Graphs Over Time for Project Tracts, Centered at Zero: Demographics*



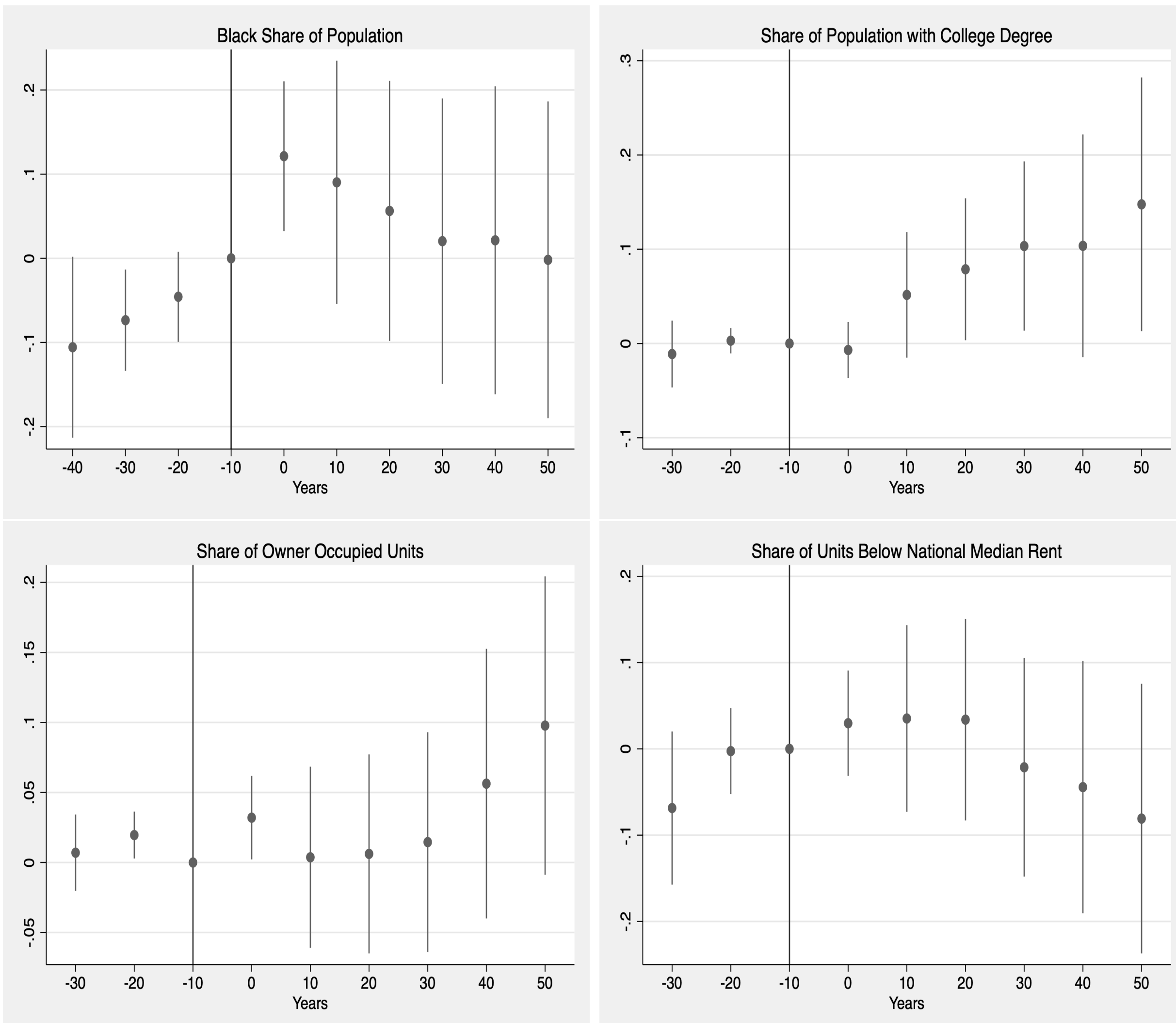
Notes: This figure shows the mean characteristics over time for census tract characteristics for urban renewal tracts, centered at event time zero. The first graph displays the mean share Black of the total population. The second shows the mean share of the population with a college degree.

Figure 2.7: *Graphs Over Time for Project Tracts, Centered at Zero: Housing*



Notes: This figure shows the mean characteristics over time for census tract characteristics for urban renewal tracts, centered at event time zero. The first shows the mean share of owner occupied dwelling units, and the second graph shows the mean share of units below the national median rent.

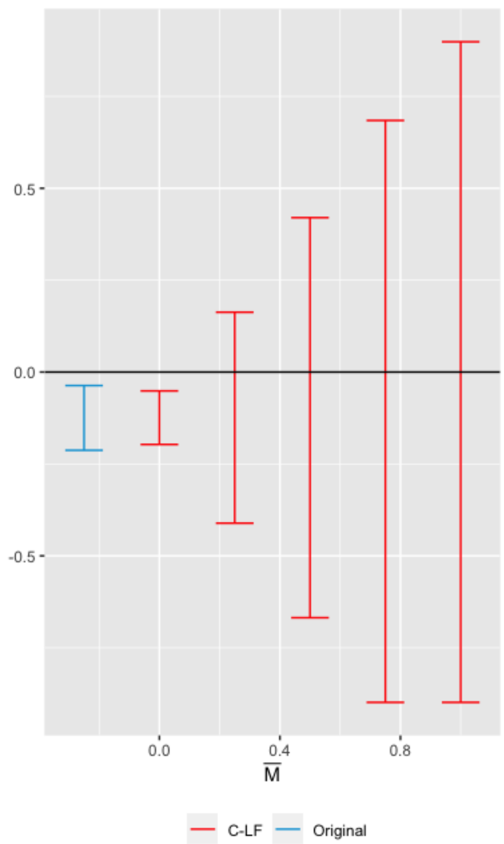
Figure 2.8: *Event Study Results*



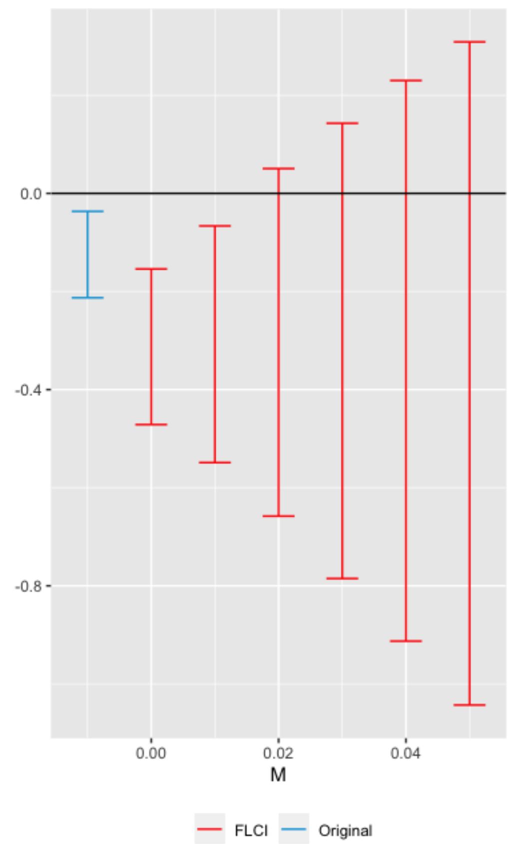
Notes: This figure shows the results of the event study style difference-in-differences specification. The first graph shows the effect on the share Black of population. The second shows the effect on the share of the population with a college degree. The third shows the effect on the share of owner occupied homes. The last shows the effect on the share of units below the national median rent.

Figure 2.9: *Sensitivity Analysis: Share Black*

(a) *Relative Magnitude Analysis*

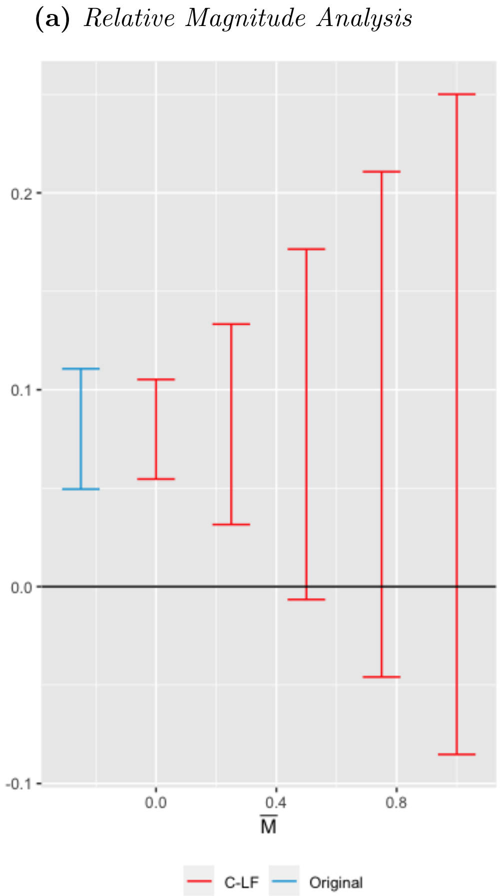


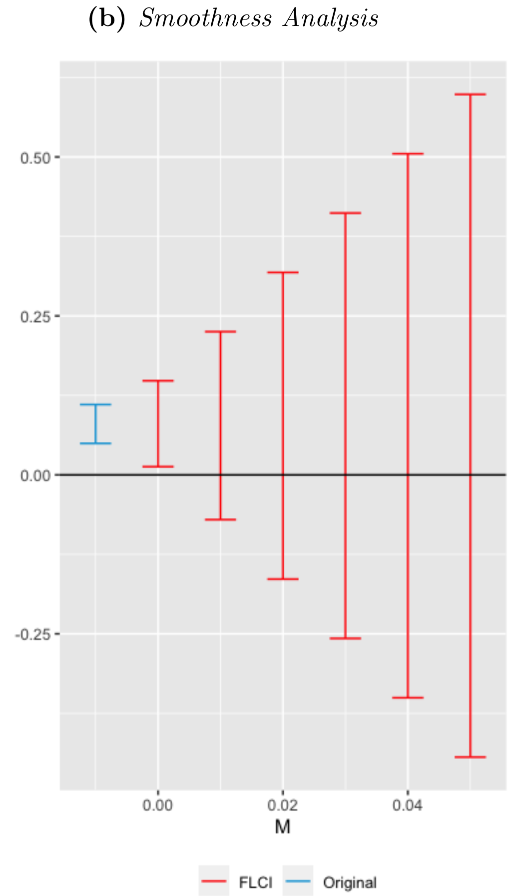
(b) *Smoothness Analysis*



Notes: This figure displays the results of the sensitivity analysis following Rambachan and Roth (2022) for the share Black of the population. Panel A shows the results of the relative magnitude sensitivity analysis, where post-treatment shocks are restricted to be no larger than pre-treatment differences in trends. The Mbar parameter, which dictates the size of post-treatment differences in trends relative to the pre-period, increases in steps of 0.25. Panel B shows the results of the smoothness restriction sensitivity analysis, which imposes that differential trends evolve smoothly over time. The Mbar parameter, which bounds the extent to which trends' slopes may change across consecutive periods, increases in steps of 0.01. I use a weighted average of the event study coefficients and the significance level is set to 0.10.

Figure 2.10: *Sensitivity Analysis: Share College Degree*

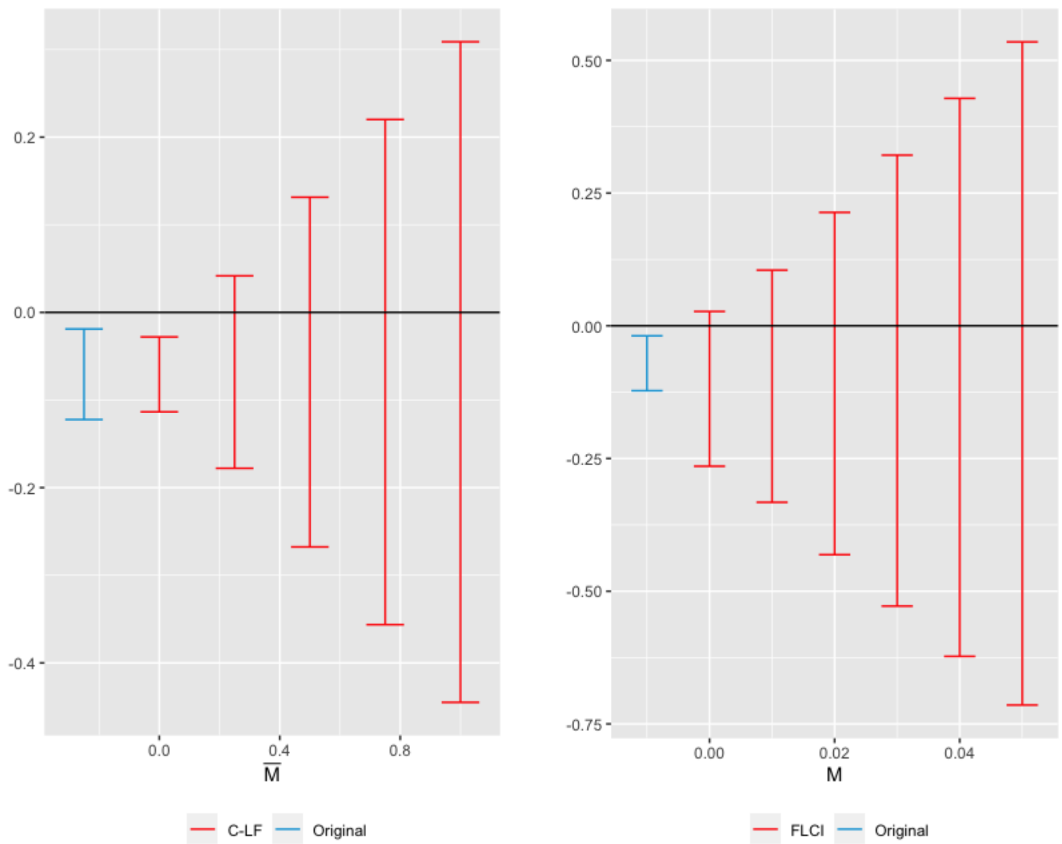




Notes: This figure displays the results of the sensitivity analysis following Rambachan and Roth (2022) for the share of the population with a college degree. Panel A shows the results of the relative magnitude sensitivity analysis, where post-treatment shocks are restricted to be no larger than pre-treatment differences in trends. The Mbar parameter, which dictates the size of post-treatment differences in trends relative to the pre-period, increases in steps of 0.25. Panel B shows the results of the smoothness restriction sensitivity analysis, which imposes that differential trends evolve smoothly over time. The Mbar parameter, which bounds the extent to which trends' slopes may change across consecutive periods, increases in steps of 0.01. I use a weighted average of the event study coefficients and the significance level is set to 0.10.

Figure 2.11: *Sensitivity Analysis: Share Units Below Median Rent*

(a) *Relative Magnitude Analysis* **(b)** *Smoothness Analysis*



Notes: This figure displays the results of the sensitivity analysis following Rambachan and Roth (2022) for the share of dwelling units below the national median rent. Panel A shows the results of the relative magnitude sensitivity analysis, where post-treatment shocks are restricted to be no larger than pre-treatment differences in trends. The Mbar parameter, which dictates the size of post-treatment differences in trends relative to the pre-period, increases in steps of 0.25. Panel B shows the results of the smoothness restriction sensitivity analysis, which imposes that differential trends evolve smoothly over time. The Mbar parameter, which bounds the extent to which trends' slopes may change across consecutive periods, increases in steps of 0.01. I use a weighted average of the event study coefficients and the significance level is set to 0.10.

Bibliography

Aaronson, Daniel, Daniel Hartley, annd Bhashkar Mazumder. 2021. "The Effects of the 1930s HOLC "Redlining" Maps." *American Economic Journal: Economic Policy,* 13(4): 355-92.

Amazaki, Barbara. 1967. "Study 36-Acre Clybourn-Ogden Plan." *The Chicago Tribune,* 129.

Anderson, Martin. 1964. "Urban Renewal Under Attack." *The Chicago Tribune,* 186.

Bach, Erwin. 1957. "Renewal Plan Studied for 10 More Areas." *The Chicago Tribune,* 97.

Bach, Erwin. 1959. "Angry Owner Attacks Hyde Park Project." *The Chicago Tribune,* 223.

Balboni, Clare, Gharad Bryan, Melanie Morten, and Bilal Siddiqi. 2021. "Could Gentrification Stop the Poor from Benefiting from Urban Improvements?" *AEA Papers and Proceedings,* 111: 532-37.

Belden, David. 2017. "Urban Renewal and the Role of the University of Chicago in the Neighborhoods of Hyde Park and Kenwood." *College of Education Theses and dissertations.* 93.

Biles, Roger. 2011. The Fate of Cities: Urban America and the Federal Government: 1954-2000. University Press of Kansas.

Brinkman, Jeffrey and Jeffrey Lin. 2022. "Freeway Revolts! The Quality of Life Effects of Highways." No. WP 22-24. Federal Reserve Bank of Philadelphia.

Buck, Thomas. 1960. "Lake Meadows Final Bulding Ready July 1." *The Chicago Tribune,* 99.

Buck, Thomas. 1964. "Nation Eyes Hyde Park-Kenwood Project." *The Chicago Tribune,* 10.

Busse, Richard. 1967. "Gives Renewal Success Formula." *The Chicago Tribune,* 325.

Busso, Matias, Jesse Gregory, and Patrick Kline. 2010. "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." No. 16096. National Bureau of Economic Research.

Callaway, Brantly and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics,* 225(2): 200-30.

Caputo, Philip. 1968. "West Side Italian-Americans Struggle to Maintain Identity." *The Chicago Tribune,* 152.

Carriere, Michael. 2012. "Chicago, the South Side Planning Board, and the Search for (Further) Order: Toward an Intellectual Lineage of Urban Renewal in Postward America." *Journal of Urban History,* 39(3): 411-432.

Carter, Chelsea. 2019. "The Road to Urban Interstates: A Case Study from Detroit."

Chicago Department of Urban Renewal Records. Digital Collections. Chicago Public Library.

"City Council O.K. Expected on Sale of Renenwal Site." 1965. *The Chicago Tribune,* 137.

"City Leaders Hoping for an End to Slums." 1962. *The Chicago Tribune,* 4.

Collins, William and Katharine Shester. 2013. "Slum Clearance and Urban Renewal in the United States." *American Economic Journal: Applied Economics,* 5(1):239-73.

Cress, Harold. 1958. "See Renewal as 'Recession Stopper'. Seek More Funds for Englewood" *Southtown Economist,* 21.

Dragan, Kacie, Ingrid Ellen, and Sherry Glied. 2019. "Does Gentrification Displace Poor Children? New Evidence from New York City Medicaid Data." No. 25809. Nationa Bureau of Economic Research.

Fernandez, Lilia. 2012. Race, Class, Housing, and Urban Renewal: Dismantling the Near West Side. *Brown in the Winndy City: Mexicans and Puerto Ricans in Postwar Chicago.* University of Chicago Press.

Fishback, Price, Jessica LaVoice, Allison Shertzer, and Randall Walsh. 2021. "The HOLC Maps: How Race and Poverty Influenced Real Estate Professionals' Evaluation of Lending Risk in the 1930s." No. 28146. National Bureau of Economic Research.

Freedman, Matthew, Shantanu Khanna, and David Neumark. 2021. "JUE Insight: The Impacts of Opportunity Zones on Zone Residents." *Journal of Urban Economics,"* Article 103389.

Getz, Anne. 1966. "Study Shows Renewal Gains." *The Chicago Tribune,* 105.

Getz, Anne. 1967. "Plan Less Razing in City Renewal." *The Chicago Tribune,* 144.

Getz, Anne. 1967. "DUR Offers to Sell Plots for Light Industry." *The Chicago Tribune,* 92.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics,* 225(2): 254-77.

Hawkins, Robert. 1964. "Businesses Hit By Renewal Can Get Aid." *The Chicago Tribune,* 69.

Hertz, Daniel Kay. 2018. *The Battle of Lincoln Park: Urban Renewal and Gentrification.* Belt Publishing.

Hirsh, Arnold. 1983. Making the Second Ghetto: Race and Housing in Chicago, 1940-1960. University of Chicago Press.

Jacobs, Jane. 1961. The Death and Life of Great American Cities.

Jacobs, Jane. 1962. "Chicago's Woodlawn - Renewal by Whom?" *Forum,* CHM.

"Joliet Tells Plans for 554 Unit Project." 1959. *The Chicago Tribunen,* 9.

Kline, Patrick, and Enrico Moretti. 2014. "Local Economic Development, Agglomeration Economics, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics,* 129(1): 275-331.

Logan, Trevon and John Parman. 2017. "Segregation and Homeownership in the Early Twentieth Century." *The American Economic Review,* 107(5): 410-14.

Neumark, David and Helen Simpson. 2014. "Place-Based Policies." No. 20049. National Bureau of Economic Research.

"Plan Would Modernize Southtown." 1955. *Suburbanite Economist,* 25.

Priddy, Gladys. 1955. "Hyde Parkers Ponder Moves; Homes to Go." *The Chicago Tribune,* 9, cont. 179.

Pritchett, Wendell. 2003. "The "Pubilc Menace" of Blight: Urban Renewal and the Private Uses of Eminent Domain." *Yale Law & Policy Review.*

Rambachan, Ashesh and Jonathan Roth. 2022. "A More Credible Approach to Parallel Trends." Working Paper.

"Report Homes to be Saved Renewal Head Assures Woodlawn Group." 1963. *The Chicago Tribune,* 43.

Reinberger, Mark. 2021. "Lake Meadows, Chicago, an experiment in early redevelopment." *Planning Perspectives,* 36(3): 583-608.

"Renewal and Revival." *The University and the City: A Centennial View of the University of Chicago.* The University of Chicago Library.

"Renewing Inequality." 2018. Digital Scholarship Lab. Richmond University.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe. 2022. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature."

"Seek to Vacate Street for Big Homes Project." 1950. *The Chicago Tribune,* 2.

Shertzer, Allison, Tate Twinam, Randall Walsh. 2014. "Race, Ethnicity, and Discriminatory Zoning." No. 20108. National Bureau of Economic Research.

Stevens, Robert. 1960. "Community Leads Way to Rehabilitation of Cities." *The Chicago Tribune,* 72.

Unger, Rudolph. 1956. "Michael Reese Shows Faith in Near S Side." *The Chicago Tribune,* 7.

Unger, Rudolph. 1956. "New Replaces Old in South Side." *The Chicago Tribune,* 9.

U.S. Department of Housing and Urban Development. Major Legislation on Housing and Urban Development Enacted since 1932.

"West Side Renewal Due". 1965. *The Chicago Tribune,* 137.

Winling, LaDale. 2018. Building the Ivory Tower: Universities and Metropolitan Development in the Twentieth Century. University of Pennsylvania Press.

Yackley, Sel. 1967. "Hyde Park is Leader in Integration." *The Chicago Tribune,* 154.

Yackley, Sel. 1967. "Integration is Underway in High Rise Housing Complexes." *The Chicago Tribune,* 38.

Ziemba, Stan. 1969. "DUR Gets 2 Proposals for Housing Construction." *The Chicago Tribune,* 145.

CHAPTER 3

POST COVID-19 TEST SCORE RECOVERY: INITIAL EVIDENCE FROM STATE TESTING DATA

with Clare Halloran, Rebecca Jack, and Emily Oster

3.1 Introduction

The COVID-19 pandemic caused a significant disruption in student learning, both within the U.S. and globally. Virtually all U.S. schools closed in March 2020, and school re-opening approaches varied widely throughout the 2020-21 school year. Even students whose schools were open during the 2020-21 school year experienced pandemic-related disruptions.

Student test scores suffered during this period. In the spring of 2021, both state-administered tests and MAP assessments from NWEA showed large declines in both math and reading compared to prior years (Halloran et al., 2021; Lewis et al., 2021). In the spring of 2022, National Assessment of Educational Progress (NAEP) scores showed historic drops relative to 2019. In these tests, as in the state assessments, we observed the largest declines in math (notably an 8 point decline in Grade 8 math, comparable to scores from 2003) and smaller declines in reading (3 point declines in both Grade 4 and Grade 8, comparable to scores dating to 2005 and 1998, respectively) (NCES 2022a, NCES 2022b).

These large losses have led to crucial questions about recovery. So far, the U.S. Department of Education has allocated a historic $122 billion dollars in relief funds to state education agencies (SEAs) to encourage academic recovery as part of the federal American Rescue Plan Elementary and Secondary School Emergency Relief Fund, or "ARP ESSER." SEAs received ARP ESSER funds in proportion to each state's funding allocation as part of Title I, Part A of the ESEA (U.S. Department of Education, 2021). States, in turn, were required to administer 90% of their allotment to school districts, also in proportion to each district's Title I funding. Notably, ARP ESSER requires that school districts reserve at least 20 percent of their total funding allocation to address learning loss through interventions that respond to students' academic, social, emotional, and mental health needs (U.S. Department of Education, 2021). Understanding what approaches can lead to improved learning outcomes will both be central to recovery from the pandemic learning losses *and* may inform academic interventions more broadly.

In this paper, we provide initial estimates of the extent of test score recovery and its correlates. We use data from states that administered assessments in the Spring of 2021 and the Spring of 2022. First, we evaluate declines in the percentage of students achieving proficiency (as measured by each state's assessment) between Spring 2019 and Spring 2021 to better understand the extent of pandemic learning loss among the states in our sample. Consistent with existing literature (Halloran et al., 2021; Lewis et al., 2021; Kuhfeld et al., 2022), we show large declines in test scores between 2019 and 2021. On average, ELA and math proficiency rates among the states in our sample declined by 6 and 11 percentage points, respectively, over this period. These declines were larger in lower income districts and in districts with less in-person learning during 2020-21. Test score declines were more severe in math than in ELA among the states in our sample, on average.

Second, we evaluate initial recovery rates, which we define as the share of the 2019-

2021 test score decline that is recovered by Spring 2022. On average, approximately 20% of the losses in ELA and 37% of the losses in math were recovered among the states in our sample. Initial test score losses are strongly correlated with recovery: that is, the districts in which student proficiency rates declined the most see the largest absolute recovery. We focus, however, on the *percent* of the losses recovered. Using this approach to recovery, we observe wide variation in recovery across the sample, suggesting the potential for using these data to better understand what factors may play a role in students' academic recovery.

We first estimate the relationship between recovery rates and district demographics, remote learning during the 2020-21 school year, and pre-pandemic test scores. We find limited evidence of correlation with recovery rates. For example, in the tercile of districts with the highest share of Black students, math scores recovered 35% between 2021 and 2022, while the recovery was 39% in districts with the lowest share.

We then estimate the relationship between recovery rates and planned district-level ARP ESSER spending priorities. This is an important test because the goal of this spending was, at least in part, to improve test score performance. We use data that indicates if districts designated using ARP ESSER funding across the following categories: academic interventions, equity and at-risk learners, mental and physical health, professional development, technology, facilities and operations, general staffing needs, transportation, and other (Burbio, 2022). The regressions are weighted by district enrollment, and we again estimate separate regressions for ELA and math scores. We find no evidence that district-level spending priorities correlate with recovery as of Spring 2022. However, we note that current investments may produce longer-term academic outcomes, which may explain, at least in part, a lack of initial findings here.

In contrast to these null results, we find significant variation in academic recovery rates at the state level. Several states *continued to decline* in ELA scores between 2021 and 2022, while others show a full recovery to their 2019 ELA levels. We see much less variation in math across states, and every state in our sample shows at least some recovery in math scores. The state-level variation occurs even across states with similar initial declines in ELA scores. For example, Arkansas experienced an average of a 9 percentage point decline in ELA scores, with a 41% recovery, while Connecticut had an 8 percentage point decline with a 19% recovery. Notably, we observe this state-level variation on either side of state borders, suggesting demographic differences do not drive these findings. Despite our efforts, we were unable to identify particular state-level policies which correlate directly with recovery, though we have an initial indicator that reading legislation may be related to future ELA outcomes. The large state-level variation suggests that there likely *are* better and worse policies for recovery, and we hope future work will be better able to elucidate them.

Broadly, this paper contributes to recent literature on learning loss in the wake of the COVID-19 pandemic (Lewis et al., 2021; West et al., 2021). This includes papers on inequality in access to in-person instruction (Goldhaber et al., 2022; Oster et al., 2021; Parolin & Lee, 2021) and evidence on the relationship between in-person learning and test score losses (Halloran et al., 2021; West et al., 2021). More specifically, this paper adds to literature studying academic recovery after the pandemic. For example, Kuhfeld and Lewis (2022) explored learning loss and recovery at a national level from the 2021-22 school year based on NWEA MAP Growth data and found that, overall, student achievement continued to lag relative to a typical year and that declines were greater in math compared to reading. In comparing Spring 2019 to Spring 2022 outcomes across districts on state assessments, Fahle and colleagues (2022) found that test score declines were greater among districts with more remote learning during 2020-21, but that this was not the main factor and that substantial variation was observed among districts. We expand on these analyses by using data on changes in academic proficiency relative to the recovery baseline year of 2020-21, and by looking at factors affecting recovery outside of schooling mode in 2020-21.

Finally, we add to the small but growing literature studying variation in recovery levers and student outcomes. In particular, we consider the impact of ARP ESSER funding. As districts are still using these funds, research on this program is limited. However, one recent study compared academic outcomes with recovery interventions that four school districts implemented with ARP ESSER funding during the 2021-22 school year (Carbonari et al., 2022). The specific interventions examined include tutoring, small group interventions, out-of-school-time programs, virtual learning programs, and extended school year approaches. The researchers found that interventions did not meet desired outcomes in terms of scale or impact due to a wide variety of implementation challenges, such as issues engaging the targeted students, and issues related to staffing and scheduling. We expand this work by conducting analyses of district funding decisions using a larger sample and more funding categories to examine the relationship between funding priorities and academic outcomes.

In the sections that follow, we describe the data used for our analyses, and present our results related to the correlates of academic recovery. We follow this with a discussion of our finding relating to state-level variation in recovery, and conclude with some promising next steps for future research.

3.2 Data

We use the following sources of data: 1) district-level state standardized assessment data from Spring 2017–2022; 2) district-level ARP ESSER planned expenditure data; 3) district-level schooling mode data from the 2020-21 school year; and 4) additional data including district-level demographic data from NCES, county-level COVID-19 transmission level data for 2021-2022 from the CDC, and data on states' reading curricula policies from multiple sources.

3.2.1 Assessment Data

We base our measure of students' academic proficiency over time on state standardized assessment data during Spring 2017–2022; data for Spring 2020 are not available due to cancelled assessments resulting from the COVID-19 pandemic. States are required to administer assessments in reading/English language arts and mathematics to students in Grades 3–8 and once in high school (as well as science in select grades) as part of the Every Student Succeeds Act (ESSA). These tests therefore provide a comprehensive and critical look at within-state changes over time that NAEP data cannot provide. However, these assessment results cannot be compared across states, as many states administer their own unique tests. We use these assessments to look at changes over time for each state, and discuss these changes relative to each state's own assessment and proficiency criteria. See Appendix A for greater detail of each state's assessment.

We focus our analyses on changes in students' levels of proficiency as measured by each state's assessment in English language arts (ELA) and math within Grades 3–8. We include assessment data in our analyses if a) the state has been using the current assessment since at least 2018 (to be able to look at pre-pandemic trends); b) the state has not changed cut score criteria that would affect the number of students considered proficient; and c) state-level participation rates were at least 70 percent in 2021. Our final sample includes assessment data from the following 21 states: Arkansas, Colorado, Connecticut, Georgia, Idaho, Kansas, Louisiana, Massachusetts, Minnesota, Mississippi, Missouri, New Hampshire, Ohio, Pennsylvania, Rhode Island, South Carolina, South Dakota, Virginia, West Virginia, Wisconsin, and Wyoming.

3.2.2 ARP ESSER Data

We use information on district-level ARP ESSER planned expenditures as a measure of district-level priorities for addressing student learning loss. Our research team accessed data from Burbio's School Budget Tracker Database (2022), which documents approximately 5,000 school district plans for ARP ESSER spending. This database categorizes planned district expenditures into over 100 categories within the topic areas such as academic intervention and learning loss, physical and mental health, facilities and operations, technology, and staffing and retention. Reported expenditure allocations were drawn from district ARP ESSER plans when available from district websites, or state e-grant portals if not. One data limitation is that states had different planned expenditure reporting requirements for districts. For example, some states only asked districts to designate which categories would receive ARP ESSER funding, while other states required districts to provide specific funding allocation amounts. Moreover, not all states listed the same funding categories. Finally, the database lacks information on smaller and more rural districts with lower student populations. Nonetheless, the database represents the most detailed district-level information available on planned expenditures by category (rather than overall allocations). We do not use measures of current actual expenditures given that districts are currently in different phases of using planned ARP ESSER funds.

3.2.3 Additional Data Sources

Schooling Mode Data: We draw district-level schooling mode data from the COVID-19 School Data Hub (CSDH, 2023), which is the only public database that centralizes schooling mode data provided directly from state agencies (typically State Education Agencies, or SEAs) for the 2020-21 school year (see Appendix B for more detailed information about data levels and time periods). States are included in our analyses if schooling mode data were available at least monthly during the 2020-21 school year. For each time period in each state's schooling mode data, districts are categorized as either 1) "in-person" (all or most students had access to traditional daily in-person instruction); 2) "virtual" (all or most students received daily instruction online/off campus); and 3) "hybrid" (schooling modes that did not fall into one of these approaches). Similar to our approach in other analyses (Halloran et al., 2021), we determined the share of the school year that students had access to each schooling mode by using each time period's schooling mode classification, the length of the time period, and the school or district's K–12 enrollment. We did not include the week of Thanksgiving 2020 or the last two weeks of December 2020 for any district in this calculation, even when districts reported a schooling mode for those weeks. We acknowledge that even though a school or district may have offered a particular schooling mode during a given month, there were likely variations across grade levels and classrooms due to individual student and family choices, local case rates, district quarantine procedures, and other factors. However, we believe this measure presents the best estimate available regarding district schooling mode data.

Student Demographic Data: To capture student demographic information from the 2020-21 school year, we draw from the National Center for Education Statistics (Urban Institute, 2022). Specifically, we used district-level information on total enrollment and the share of enrolled students by race and ethnicity. Additionally, we use school-level information from the 2019-2020 school year on eligibility for free and reduced price lunch (FRPL) aggregated to the district level (this was the most recent year available).

COVID-19 Community Levels by County: We use data from the Center for Disease Control on county-level COVID-19 transmission for the period from September 2021 to June 2022, or the baseline recovery year (CDC, 2022). We use the four CDC classifications of average community transmission level during the 2021-2022 school year: low, moderate, substantial, and high.

Reading Curricula: We collect data from two sources on states' reading curricula and methods of teaching. Schwartz (2022) provides information about when states adopted laws or policies related to the science of reading; these policies encompass teacher preparation and certification, professional development, assessment, materials, and instruction or intervention. American Public Media provides information on the level of SEA involvement in determining what curricula schools can use to teach reading (Peak, 2022). We use this data to look for patterns in states' reading policies given the variation we find in ELA test score recovery.

3.3 Results

We begin this section by describing the patterns of test score decline and initial recovery overall. We then turn to the correlates of recovery.

3.3.1 Descriptive Data

We present basic summary statistics for each of the 21 states included in our sample in Table 1. We include the number of districts within each state included in the sample, the percentage of the school year districts offered each schooling mode (in-person, hybrid, and virtual) for the 2020-21 school year, and district demographic characteristics, including share of enrolled students who are Black, Hispanic, and eligible for free and reduced price lunch. Overall, our sample includes nearly 5,000 school districts. On average, districts offered full in-person instruction for 43% of the 2020-21 school year, compared to 38% hybrid instruction and 19% virtual instruction. Districts in the sample were characterized by student populations that were 18% Black and 15% Hispanic, and 45% of students were eligible for FRPL. However, states in our sample vary across predominant schooling mode, as well as across demographic characteristics.

We next present test score summary statistics by state in Table 2. We observe that all states in our sample experienced declines in ELA and math proficiency levels between 2019 and 2021, with an overall average decline in test score pass rates of 6.4 percentage points in ELA and 12.5 percentage points in math. ELA declines in 2021 ranged from 9 percentage points in Arkansas, Pennsylvania, and Virginia to 2 percentage points or less in Idaho, Kansas, and Wyoming. Math declines in 2021 ranged from 32 percentage points in Massachusetts to 4 percentage points in Idaho and Wyoming.

With regard to academic recovery (i.e., the percent of test score declines in 2021 that were recovered by by 2022), we observe evidence of recovery in both ELA and math, with greater recovery in math (see Table 2). On average, weighted by district enrollment, 20% of test score declines between Spring 2019 and 2021 were recovered in ELA in Spring 2022, compared to 37% in math. Recovery varied widely by state, particularly in ELA. Specifically, in ELA, we observe that six states *continued to decline* in 2022 (most notably Kansas and Massachusetts, among others), while two states fully recovered their pandemic losses (Mississippi and South Carolina). In math, we observe that all states had at least some recovery in the number of students reaching proficiency, but no states reached 2019 levels as of 2022. Mississippi and Rhode Island observed the greatest test score recovery (over 70% of scores recovered), while Arkansas and Minnesota experienced the least recovery (less than 20% of scores recovered).

In Figure 1, we illustrate the distribution of changes in test score proficiency rates by school district compared to the prior assessment year, between Spring 2017 and Spring 2022. Pre-pandemic, looking at changes from Spring 2017 to 2018 or from Spring 2018 to 2019, we observe these changes fairly tightly centered near zero. Thus, while test scores appear to increase in some locations and decrease in others, consistent with noise and possibly other factors, we do not observe systematic increases or decreases in scores. However, between Spring 2019 and 2021, we observe large declines in test score pass rates, as has been documented elsewhere (Halloran et al., 2021; Lewis et al., 2021). This is shown in the leftward shift of the distribution. Although some school districts in our sample showed test score increases, this share was small: 13% of districts demonstrated gains in ELA, compared to 9% in math.

3.3.2 Correlates of Recovery

It is clear from Figure 1 that there is variation in recovery across school districts. A key policy question is what determines that recovery rate. In Figure 2, we present a binscatter of the relationship between district-level test score percentage point changes between 2019–2021 and between 2021–2022 for both ELA and math. There is a negative relationship in both cases, indicating that districts with the largest test score declines in 2021 had the largest score increases by 2022. We again see a larger decrease in math scores during the pandemic, and a subsequent larger increase post-pandemic as compared to ELA scores.

In Figure 3, we show variation in math and ELA recovery overall (among all states in our sample) and by state, share of in-person instruction during the 2020-21 school year, district demographics, test score declines in 2019–2021, and baseline achievement. As a measure of district baseline achievement, we calculate each districts' average proficiency rate across Spring 2017, 2018, and 2019 separately for ELA and math and then we regress these averages on the following district demographic characteristics: the share of Black students enrolled, the share Hispanic students enrolled, and the share of students enrolled who are eligible for free and reduced-price lunch (FRPL). The regression is weighted by district enrollment and we estimate separate regressions for ELA and math scores. We use the residuals from this regression as the measure of baseline achievement.

Figure 3 helps to demonstrate two findings. First, there we observe significantly more variation in recovery in ELA than in math. Second, for the most part, there is relatively little systematic variation in recovery by schooling mode in 2020-21, demographic groups, or test score groups. This is especially true in math, but even in ELA the variation is small across in-person shares, race/ethnicity shares, baseline pre-pandemic test scores, and test score declines during the pandemic. Where we do see systematic variation is across states, as observed in Table 2, especially in ELA. Here, we can see the full recovery in ELA scores in Mississippi and South Carolina, compared to continued declines in 2022 in states such as Massachusetts and Kansas. We return to this state-level variation below.

Next, we estimate a regression for the outcome district proficiency rate in 2022 to assess the correlates of recovery:

$$Y_i = \alpha + \beta_1 * prof2021_i + \beta_2 * prof2019_i + \beta_3 * achieve_i + \beta_4 * X_i + \beta_5 * covid_i + \gamma_s + \epsilon_i$$

where $prof2021$ and $prof2019$ are districts' proficiency rates in the respective years, $achieve_i$ is the district baseline achievement described above, X_i are the demographic characteristics (the share of students who are Black, Hispanic, and eligible for FRPL), $covid_i$ is the average county-level COVID-19 transmission level as defined by the CDC, and γ_s is a state fixed effect. We also estimate this regression with the addition of indicators for each of the main ARP ESSER funding categories (academic intervention, equity and at-risk learners, mental and physical health, professional development, technology, facilities and operations, and general staffing needs), with the transportation and other categories combined as the comparison group. The regressions are weighted by district enrollment, and we again estimate separate regressions for ELA and math scores.

In Table 3, we show the results of the above regressions, with ELA results in Column (1) and math results in Column (3), both without the funding variables. For both subjects, the state fixed effects are jointly significant. This aligns with our earlier plot highlighting the variation in test score recovery by state (Figure 3). Districts' past performance also has a significant relationship with their 2022 proficiency rate: both proficiency coefficients are statistically significant at the 1% level and the coefficients are of similar magnitude in each year for both ELA and math. However, baseline achievement, which measures the amount that mean pre-COVID scores vary from the expected mean score level based on district demographics, matters little. The demographic characteristics also do not seem to have a significant relationship with 2022 proficiency rate for either subject. Likewise, the coefficient for community COVID-19 transmission is not significant in either regression. For math scores, the share of hybrid class time is significant at the 5% level, but the coefficient is very small, showing very little impact on proficiency rates.

We add results of the regression in Table 3 to include indicators for each ARP ESSER funding category in Column (2) for ELA and Column (4) for math. Again, the state fixed effects are highly jointly significant for both subjects, as are the previous years' proficiency rates. For math, none of the coefficients for the funding categories reach statistical significance, and they are all small in magnitude. Similarly for ELA, the coefficients are all negligible, and only two are statistically significant at the 5% level: funding for technology and for facilities and operations.

3.4 Unpacking State-Level Variation

Our analyses indicate that of the correlates that we were able to explore, the state where the district is located is the most important factor in a district's post-pandemic recovery. We can further illustrate this with a border design. We consider two sets of states in our sample – South Carolina/Georgia and Massachusetts/New Hampshire – that share borders. This provides an opportunity to look at whether the variation appears even in areas geographically close and similar.

In Figure 4, we illustrate an example of this cross-state variation in score recovery. Panel A shows district-level ELA recovery in Georgia and South Carolina. It is clear that districts in South Carolina have a larger percentage of ELA scores recovered than those in Georgia along the border. On average, border districts in Georgia actually experienced a loss of test scores in 2022 – they lost an additional 45% of their original decline in test scores. In South Carolina, on the other hand, many border districts fully recovered their ELA pandemic losses – the average percentage of the score decline recovered was 134%. Panel C similarly shows the border of Massachusetts and New Hampshire. Again, we see that New Hampshire's border districts experienced a better recovery, with an average of 10% of scores recovered. Massachusetts border districts' average percent recovery was -189%, indicating a loss in 2022 that was even greater than the 2019–2021 decline in scores.

We repeat these maps for math scores in Panels B and D of Figure 4, but we do not see stark differences in recovery along either border. Mean percent recovery in Georgia and South Carolina for border districts was 32% and 31%, respectively. Similarly, Massachusetts' average recovery in math for border districts was 43%, and in New Hampshire's border districts, average recovery was 44%.

Given the large amount of cross-state variation, it seems worthwhile to try to understand whether there are state-level characteristics or policies that correlate with this variation. To do so, we first collected information on states' educational priorities and interventions from their ARP ESSER plans. We coded plans according to what the states identified as their top strategies that have been effective in supporting the needs of students during the pandemic, and by how states reported they would use their funding to address the impact of lost instructional time during the pandemic. State plans highlighted a wide range of priority areas, including investments in remote learning, tutoring, health and wellness, capacity-building efforts, and data use and management, among many other topics. We ranked states by the percentage of score recovery in 2022 in both ELA and math and compared plans of the top one-third to those from the bottom two-thirds. No clear patterns of recovery interventions or supports emerged between the high recovery or low recovery states. A limitation here is that some state plans focused on a few key priority areas, while other state plans focused more on breadth. In this way, we could not always identify which areas were the greatest priority for a given state.

To further explore the variation in state-level ELA recovery, we collected information on how long states have had legislation related to the science of reading (SOR) (Schwartz, 2022). The science of reading refers to research findings that reading comprehension is a product of students' ability to sound out the letters of a word (decoding) and knowledge of what the words mean (language comprehension) (Hanford, 2019). Schwartz (2022) categorized state reading policies according to how many of the following six areas were addressed: teacher preparation, teacher certification or license renewal, professional development/coaching, assessment, materials, and instruction/intervention. Four states in our sample adopted reading policies relating to at least one of these categories prior to 2019 (Arkansas, Mississippi, Missouri, South Carolina), while nine states did so between 2019–2022 (Arkansas, Colorado, Connecticut, Kansas, Louisiana, Minnesota, Pennsylvania, Rhode Island, Virginia, West Virginia). The remaining eight states in our sample had not yet adopted such legislation as of 2022.[1]

Most notably, we find that the two states in our sample with the earliest adoption of SOR legislation (Mississippi in 2013 and South Carolina in 2014) are the only two states to fully recover their pandemic learning losses by 2022. Legislation from both states addressed, at a minimum, teacher preparation, professional development, and instruction. While we cannot identify science of reading legislation as the only lever here, or even the extent to which legislation has impacted scores, the correlation may indicate that such legislation could be an important component in academic recovery.

That said, we do not observe additional patterns in ELA recovery by whether or not states have enacted SOR-related legislation. For example, Arkansas is a state with relatively early legislation (2017; updated in 2021), and experienced a 9 percentage point decline in 2021 with a 41% recovery. Meanwhile, Ohio has no legislation and had a similar 8 percentage point decline and 44% recovery.

It is possible that a simple measure of adoption is not enough to uncover the effect of reading curricula on test score recovery. Many of these policies have only been enacted in the past few years, meaning that the full effects of the legislation may not yet have had sufficient time to fully or even partially reach districts and students. Some policies may be significantly less comprehensive than others, or may only affect teacher licensure programs,

1. Kansas was not included in documentation by Schwartz (2022) but passed the Every Child Can Read Act in May 2022, which will go into effect for the 2023-24 school year; we include Kansas here as having adopted SOR legislation

for example, meaning that the impacts on student learning would not apparent until these teachers enter the classroom. In addition, a lack of legislation does not mean that states or individual districts do not have programs related to the science of reading, only that they are not necessarily enacted in legislation. Nonetheless, this can provide an initial look at state approaches to addressing student progress related to English language arts.

We also consider a measure of state involvement in district reading curricula (Peak, 2022). There are four categories: minimal state involvement, some state guidance, state advisory list, and state mandated list of reading curricula. We again find no systematic patterns between level of state involvement and ELA recovery, though there are some hints toward trends. For example, as noted, Arkansas recovered 41% of their pandemic score loss in ELA (with a 9 pp decline) and has the highest level of state involvement in reading curricula. However, Wisconsin recovered a similar 45% of their loss (with a 7 pp decline), with minimal state involvement in curricula. Both high and low levels of ELA recovery were experienced by states of minimal levels of involvement *and* states with mandated lists of curricula. However, we note, again, that many of SOR-related policies and approaches to ELA curricula are relatively new and may not be fully implemented by districts at this time. We might expect to see more clear findings in the future.

3.5 Conclusion

Post-pandemic learning recovery remains an important concern for education leaders. This paper documents both the decline in test scores between Spring 2019 and 2021 as well as the subsequent recovery in 2022. We find that scores have recovered more in math than in ELA, but in both cases, many districts have not fully regained their pre-pandemic scores. We also show substantially more variation in ELA score recovery as compared to math.

In assessing the correlates of recovery, a district's state appears to be the most significant factor. District demographic characteristics, schooling mode from 2020-21, and community

COVID-19 transmission levels from 2021-22 do no not seem to impact recovery, nor do districts show meaningful variation in recovery across these categories.

While we highlight the importance of states in students' ELA academic recovery, we are not able to pinpoint which specific factors may be responsible for some states experiencing stronger recoveries. Investigating state funding decisions has not yet revealed clear patterns in effectiveness, though we did observe that the two states with the longest legislation in place related to the science of reading have also experienced there strongest recovery in ELA scores. Moreover, state and districts are still in the process of using ARP ESSER funding, and many states are in the midst of responding to state-level legislation related to evidence-based reading curricula, meaning that impacts on student achievement may not yet be evident. Thus, this paper serves as a starting point for future research into potential determinants of test score recovery, particularly at the state level. Knowing which interventions and policies are most effective in combating learning loss would provide critical information for policymakers and education leaders as they seek to ensure that students across the country recover after a significant educational disruption.

Tables

Table 3.1: *Summary Statistics by State*

	Districts	% In-Person	% Hybrid	% Virtual	% Black	% Hispanic	% FRPL
AR	247	87.0	12.4	0.6	19.5	13.7	64.7
CO	140	27.3	42.7	29.9	4.6	34.3	42.3
CT	171	50.1	39.4	10.5	12.7	27.7	42.1
GA	202	48.3	19.1	32.7	36.5	17.2	62.1
ID	139	62.8	32.8	4.4	1.1	19.1	40.2
KS	274	64.7	23.2	12.2	6.8	20.7	47.7
LA	69	72.0	19.1	8.9	39.3	8.7	53.9
MA	352	27.5	52.3	20.2	9.5	22.5	0.0
MN	439	14.6	65.5	19.9	11.4	10.1	36.8
MO	510	52.6	34.1	13.4	15.3	7.3	51.5
MS	127	67.8	18.7	13.5	47.3	4.4	78.3
NH	169	44.6	42.7	12.7	2.2	6.6	29.1
OH	606	50.0	32.3	17.7	14.6	6.5	26.4
PA	596	17.4	54.8	27.7	14.4	12.9	52.4
RI	49	48.2	43.0	8.8	8.4	26.5	46.5
SC	75	44.5	50.4	5.2	32.1	11.7	63.5
SD	144	98.8	1.1	0.1	3.5	7.3	34.7
VA	131	14.4	61.7	23.9	22.1	17.5	45.5
WI	413	54.8	23.7	21.5	9.0	12.9	41.6
WV	55	37.1	41.6	21.3	4.1	2.0	53.3
WY	48	92.7	6.4	0.9	0.9	14.4	36.7
Overall	4956	42.5	38.2	19.3	18.3	14.7	45.4

Notes: This table shows summary statistics for the 21 states included in the sample. "Districts" presents the number of districts included in the sample. Schooling mode variables ("% In-Person", "% Hybrid", "% Virtual") are drawn from the COVID-19 School Data Hub and represent the average percent of the school year that the state's school districts offered each schooling mode. Demographic variables are from the NCES data and include: the share of enrolled students who are Black, the share of enrolled students who are Hispanic, and the share of students who are eligible for free and reduced-price lunch (FRPL). Massachusetts does not report FRPL, so it is not included here.

Table 3.2: *Test Score Summary Statistics by State*

	ELA				Math			
	2019	2021	2022	2022	2019	2021	2022	2022
	% Pass	% Pass	% Pass	% Recovery	% Pass	% Pass	% Pass	% Recovery
AR	45.6	36.7	40.3	40.8	52.7	40.2	42.4	17.8
CO	46.1	43.0	43.7	24.1	35.0	27.4	32.1	62.0
CT	55.8	47.4	49.0	18.8	48.2	36.2	40.0	31.9
GA	43.8	38.0	38.3	4.5	43.9	34.2	37.5	34.3
ID	55.5	53.5	54.3	37.8	45.3	41.5	44.1	69.3
KS	38.1	36.6	33.4	-208.9	34.2	29.2	31.1	36.3
LA	45.0	40.6	42.0	31.3	35.0	27.9	30.7	39.7
MA	52.0	45.7	41.0	-76.8	48.6	33.1	38.7	36.6
MN	59.1	51.0	50.3	-8.6	56.6	43.5	45.7	16.8
MO	47.1	43.5	41.4	-61.0	41.1	34.9	38.4	57.0
MS	42.1	35.5	42.4	104.7	47.9	36.1	44.9	74.4
NH	55.1	48.3	48.7	5.8	48.8	35.5	40.6	38.2
OH	66.3	58.0	61.6	43.7	66.1	51.4	55.2	26.1
PA	62.0	53.4	55.0	18.9	43.2	31.0	34.7	30.0
RI	38.8	32.8	31.1	-28.3	32.2	21.4	29.0	70.5
SC	45.8	42.2	46.9	130.7	45.6	36.8	39.3	28.6
SD	53.7	49.7	50.1	11.7	47.4	42.0	44.1	38.3
VA	76.0	67.0	70.7	41.1	79.2	47.2	60.4	41.2
WI	41.1	33.7	37.0	44.5	43.7	33.6	39.3	56.0
WV	45.4	38.6	40.6	29.1	41.1	28.9	34.5	46.2
WY	56.7	54.5	53.9	-28.8	54.1	49.9	50.9	23.8
Overall	53.3	46.9	48.2	19.8	49.5	37.0	41.6	37.2

Notes: This table shows summary statistics of ELA and math proficiency rates between 2019 and 2022 for the 21 states included in the sample. "% Pass" represents the percent of students reaching proficiency on the state assessment in the given year. "% Recovery" represents the percent recovery defined as the percentage of the decline in test scores between Spring 2019 and Spring 2021 that was recovered by 2022. Columns 2-5 show ELA scores and recovery, and columns 6-9 show math scores and recovery.

Table 3.3: *Determinants of 2022 Proficiency Rates*

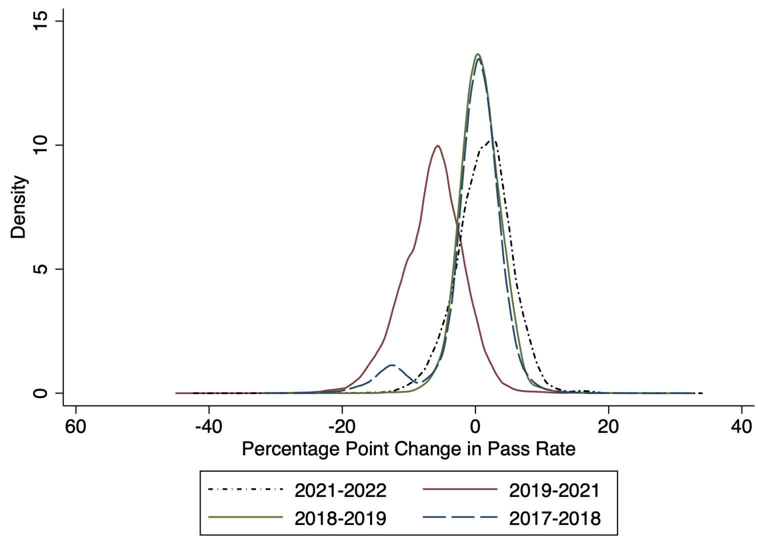
	ELA		Math	
	(1)	(2)	(3)	(4)
Proficiency rate 2021	0.684	0.674	0.661	0.648
	(0.020)	(0.027)	(0.018)	(0.024)
Proficiency rate 2019	0.289	0.311	0.370	0.452
	(0.033)	(0.051)	(0.035)	(0.054)
Baseline achievement	0.008	-0.001	-0.037	-0.104
	(0.036)	(0.057)	(0.034)	(0.056)
% in-person, 2020-2021	-0.003	0.002	-0.005	-0.000
	(0.005)	(0.005)	(0.006)	(0.006)
% hybrid, 2020-2021	0.003	0.006	0.014	0.019
	(0.005)	(0.005)	(0.005)	(0.006)
% Black	0.004	0.007	-0.010	0.008
	(0.008)	(0.011)	(0.009)	(0.014)
% Hispanic	0.006	0.010	-0.007	0.012
	(0.008)	(0.013)	(0.009)	(0.014)
% FRPL	-0.011	-0.006	0.007	0.034
	(0.011)	(0.018)	(0.011)	(0.017)
COVID levels, 2021-2022	0.006	0.005	0.004	0.002
	(0.005)	(0.005)	(0.005)	(0.007)
Fund academic intervention		0.001		-0.001
		(0.003)		(0.003)
Fund equity and at-risk learners		0.001		-0.000
		(0.002)		(0.002)
Fund mental and physical health		0.003		0.003
		(0.002)		(0.002)
Fund professional development		-0.001		0.000
		(0.002)		(0.002)
Fund technology		-0.004		0.000
		(0.002)		(0.002)
Fund facilities		0.004		-0.002
		(0.002)		(0.003)
Fund staffing		0.005		-0.001
		(0.003)		(0.003)
Observations	4913	1794	4910	1787
State FE joint F-test	0.000	0.000	0.000	0.000

Notes: This table shows the relationship between state test score proficiency rates in 2021 and 2019 on state test score proficiency rates in 2022 at the district level separately for ELA and math. We control for district baseline achievement (the residual from regressing the average pre-COVID pass rate from 2017-2019 on district demographics), the share of time spent in-person and in hybrid learning during the 2020-2021 school year, district race/ethnicity shares, district share of students eligible for free and reduced-price lunch (FRPL), and community COVID-19 transmission levels in 2021-2022. FRPL counts are not available in MA so we code districts missing FRPL as 0 and include a missing FRPL binary variable. Regressions in columns (2) and (4) only include districts in the Burbio ARP ESSER dataset (2022) and we include binary variables for the seven main categories of funding. All regressions include state fixed effects, which are jointly highly significant for all regressions.

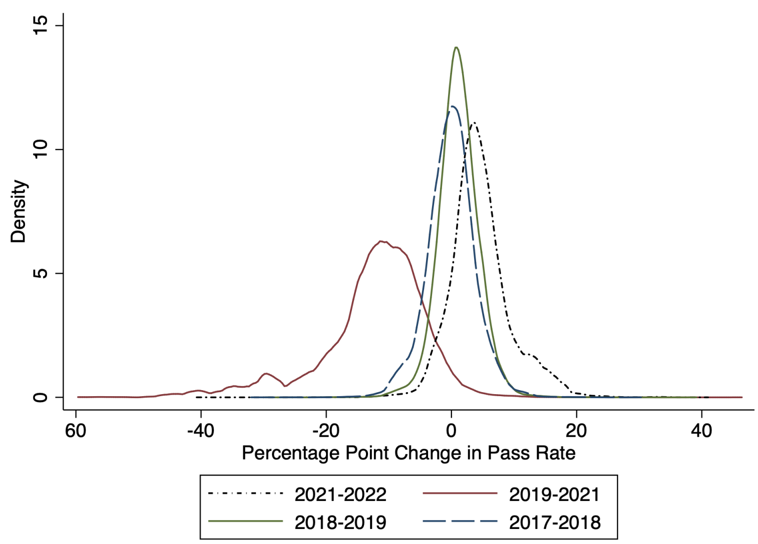
Figures

Figure 3.1: *Distribution of Changes in ELA and Math Test Score Pass Rates Compared to Prior Year, 2017-2022*

(a) *Distribution of Changes in ELA Test Score Pass Rates by Year*



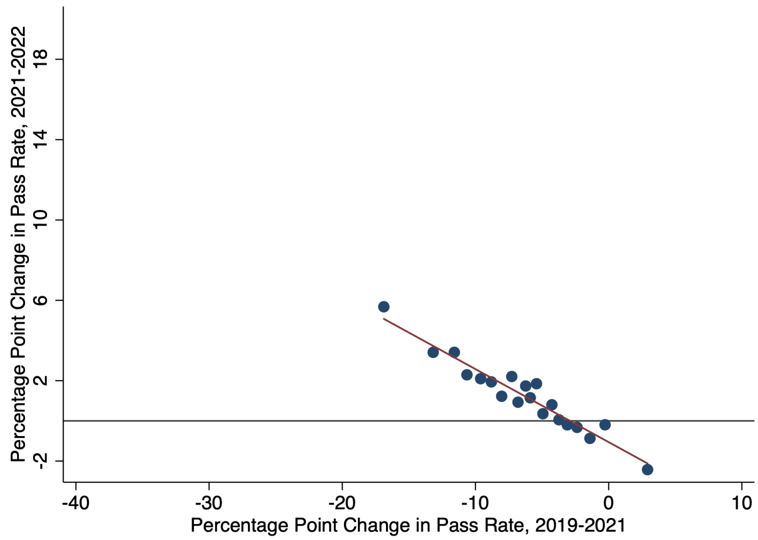
(b) *Distribution of Changes in Math Test Score Pass Rates by Year*



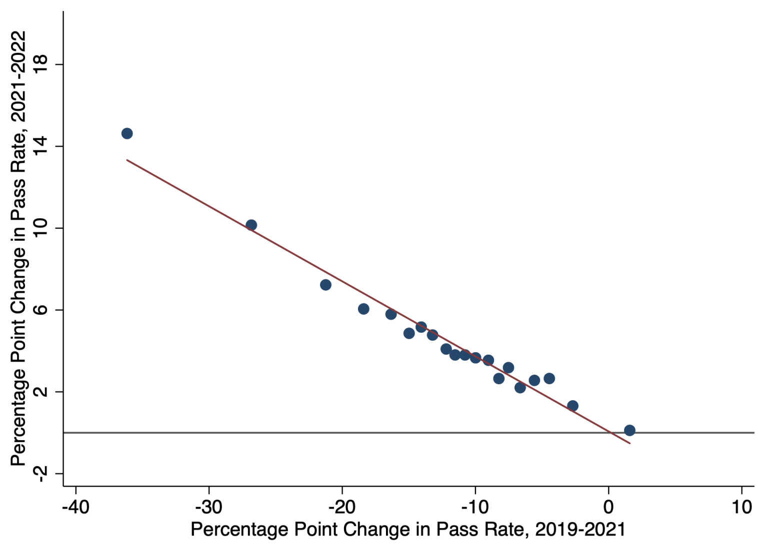
Notes: This figure shows the distribution of test score changes in pass rates across districts compared to the prior year for Spring of 2018, 2019, 2021, and 2022 in percentage point changes. Pass rates represent the percent of students reaching proficiency on state assessment in a given school year. Spring 2021 is compared to Spring 2019, as assessments were not administered in Spring 2020 due to the pandemic. Panel (a) (top) shows the distributions for ELA pass rate changes and Panel (b) (bottom) shows the distributions for math pass rate changes.

Figure 3.2: *Changes in ELA and Math Pass Rates from Spring 2021 to 2022, Relative to Changes from Spring 2019 to 2021*

(a) *ELA Test Score Changes*

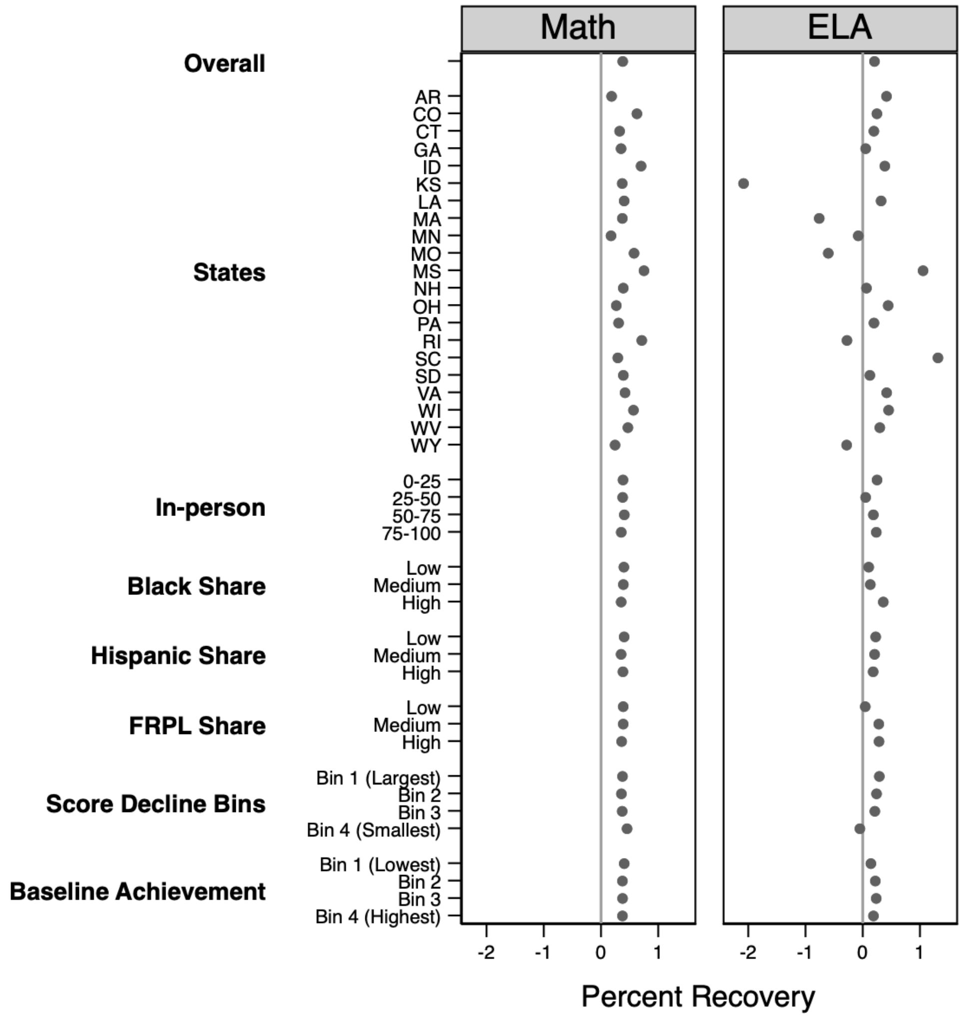


(b) *Math Test Score Changes*



Notes: This figure shows binscatters of changes in test score pass rates from Spring 2021 to Spring 2022, relative to test score pass rate declines in between Spring 2019 and 2021. Panel (a) (top) shows the binscatter for ELA score changes and Panel (b) (bottom) shows the binscatter for math score changes. Pass rates represent the percent of students reaching proficiency on state assessment in a given school year. Changes are shown in percentage points. Each bin represents districts within the relevant 20 quantiles of change between 2019 and 2021 (n=4913).

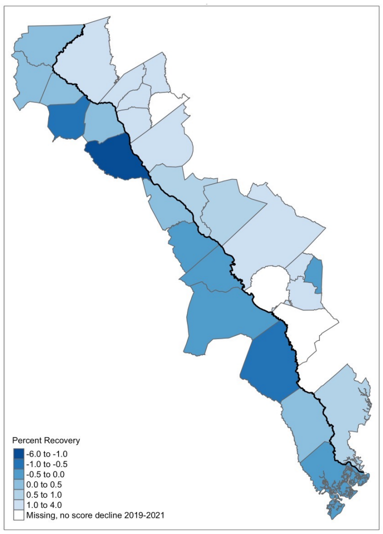
Figure 3.3: *District Characteristics and Percent Recovery in ELA and Math*



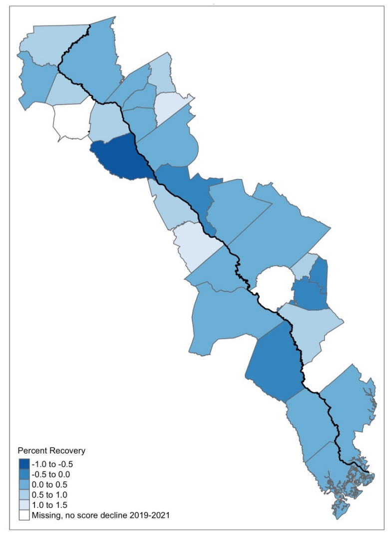
Notes: This figure shows the average percent recovery in math (left panel) and ELA (right panel) for students in Grades 3–8 on state standardized assessments, weighted by enrollment. Percent recovery is defined by the percentage of the test score decline in proficiency rates from 2019 to 2021 recovered by 2022. Comparisons are presented: a) for all students in the sample (overall); b) by state; c) by the percent of in-person instruction offered by districts over the 2020-21 school year; d) by the share of students who are Black or Hispanic (based on NCES 2020-21 data); e) by the share of students who are eligible for free and reduced price lunch (FRPL) (based on NCES 2019-20 data due to changes in reporting requirements in 2020-21; MA does not report FRPL and is excluded here); f) by the test score decline quartiles between Spring 2019 and 2021; and g) district baseline achievement as defined as the residuals from a regression of districts' pre-pandemic test scores on district demographics characteristics. Ranges for % In-person groups include the lower bound of each range.

Figure 3.4: *State Border Maps of Variation in Percent Recovery in ELA and Math*

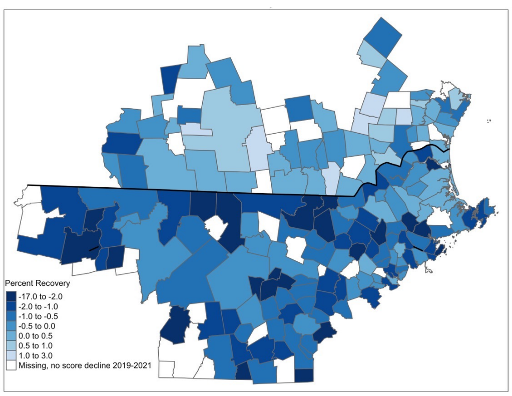
(a) *ELA Recovery in GA and SC*



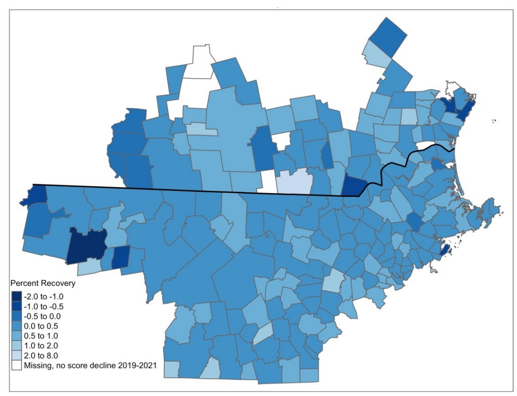
(b) *Math Recovery in GA and SC*



(c) *ELA Recovery in MA and NH*



(d) *Math Recovery in MA and NH*



Notes: This figure shows percent recovery in ELA and math pass rates for districts along the border of selected states. For the states of Georgia and South Carolina, we present the percent recovery in ELA in Panel (a) (top left) and in math in Panel (b) (top right). For the states of Massachusetts and New Hampshire, we present the percent recovery in ELA in Panel (c) (bottom left) and in math in Panel (d) (bottom right).

Bibliography

Arkansas Division of Elementary Secondary Education. 2022. “District and School Results: ACT Aspire.” Accessed August 12, 2022, https:// dese.ade.arkansas.gov/ Offices/learning-services/assessment-test-scores/2022.

Burbio. 2022. Burbio’s School Budget Tracker. Accessed September 14, 2022.

Carbonari, Maria V., Miles Davison, Michael DeArmond, Daniel Dewey, Elise Dizon-Ross, Dan Goldhaber, Ayesha K. Hashim, et al. 2022. “The Challenges of Implementing Aca- demic Interventions During COVID: Evidence from the Road to Recovery Project.” CALDER Working Paper No. 275-1122.

Center for Disease Control (CDC). 2022. United States COVID-19 Community Levels by County. Accessed November 29, 2022. Center for Disease Control. https://data.cdc. gov / Public - Health - Surveillance / United - States - COVID - 19 - Community - Levels - by - County/3nnm-4jni.

Colorado Department of Education (CDE). 2022. “CMAS - Mathematics, English Language Arts, Science and Social Studies Data and Results.” Accessed August 19, 2022, https: //www.cde. state.co.us/assessment/cmas-dataandresults.

Connecticut State Department of Education (CSDE). 2021. Student Assessment News: Au- gust 25, 2021. Hartford, Connecticut: CSDE. https://portal.ct.gov/-/media/SDE/ Student-Assessment/Student-Assessment-News/StudentAssessmentNews-082521.pdf.

2022. “Smarter Balanced by All Students.” Accessed September 1, 2022, http:// data.ctdata.org /visualization/smarter-balanced-by-all-students.

COVID-19 School Data Hub (CSDH). 2023. Percentage of School Year Spent In-Person, Hybrid, or Virtual, District Overall Shares. Accessed March 2, 2023. https:// www. covidschooldatahub.com /data-resources.

Fahle, Erin, Thomas J. Kane, Tyler Patterson, Sear F. Reardon, and Douglas O. Staiger. 2022. Local Achievement Impacts of the Pandemic. Education Recovery Scorecard.

Goldhaber, Dan, Thomas J. Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O. Staiger. 2022. "The Consequences of Remote and Hybrid Instruction During the Pandemic." Research Report. Cambridge, MA: Center for Education Policy Research, Harvard University.

Governor's Office of Student Achievement. 2022. "Georgia Milestones End-of-Grade (EOG) Assessments." Accessed September 1, 2022, https://gosa.georgia.gov/dashboards-data-report-card/downloadable-data.

Halloran, Clare, Rebecca Jack, James C. Okun, and Emily Oster. 2021. "Pandemic Schooling Mode and Student Test Scores: Evidence from US States." NBER Working Paper 29497, National Bureau of Economic Research, Cambridge, MA.

Hanford, Emily. 2019. Opinion: There Is a Right Way to Teach Reading, and Mississippi Knows It. The New York Times. https://www.nytimes.com/2019/12/05/opinion/mississippi-schools-naep.html.

Idaho State Department of Education. 2022. "ISAT Results." Accessed October 28, 2022, https:// www.sde.idaho.gov/assessment/accountability/.

Kansas State Department of Education. 2022. "Longitudinal Performance Level Reports-Full Results." Accessed October 24, 2022, https://ksreportcard.ksde.org/ assessment results.

Kuhfeld, Megan, and Karyn Lewis. 2022. Student Achievement in 2021–2022: Cause for Hope and Continued Urgency. NWEA Research Brief. Portland, Oregon: NWEA.

Kuhfeld, Megan, James Soland, and Karyn Lewis. 2022. Test Score Patterns Across Three COVID-19-Impacted School Years. EdWorkingPaper: 22-521, Annenberg Institute at Brown University, Providence, RI.

Lewis, Karyn, Megan Kuhfeld, Erik Ruzek, and Andrew McEachin. 2021. Learning During COVID-19: Reading and Math Achievement in the 2020-21 School Year. Portland, Oregon: NWEA. Louisiana Department of Education. 2022. “LEAP 2025 Results.” Accessed August 5, 2022, https : / / https : / / www . louisianabelieves . com / resources / %20library / elementary - and - middle-school-performance.

Massachusetts Department of Elementary and Secondary Education (MA DESE). 2021. 2021 MCAS Reporting Call for Superintendents, Principals and Test Coordinators. PowerPoint Presentation, September 2021. Malden, Massachusetts: DESE. https://www.doe.mass.edu /mcas/2021/results/reporting-call.pptx.

2022. “Massachusetts Comprehensive Assessment System Results.” Accessed Octo- ber 3, 2022, https://www.doe.mass.edu/mcas/results.html. Minnesota Department of Education (MDE). 2021. “Minnesota Participation 2021.” Re- ceived via data request November 9, 2021.

2022. “Data Reports Analytics: MCA Assessment Files.” Accessed August 25, 2022, https:// public.education. mn.gov/ MDEAnalytics/ DataTopic.

Mississippi Department of Education. 2022. “Assessment Files.” Accessed August 19, 2022, https://www.mdek12.org/OPR/Reporting/Assessment.

Missouri Department of Elementary Secondary Education. 2023. “District MAP Results.” Received via data request January 23, 2023.

National Center for Education Statistics, U.S. Department of Education. 2022a. 2022 NAEP Mathematics Assessment: Highlighted Results at Grades 4 and 8 for the Nation, States, and Districts. NCES 2022124. https://www.nationsreportcard.gov/highlights/mathematics/2022/.

2022b. 2022 NAEP Reading Assessment: Highlighted Results at Grades 4 and 8 for the Nation, States, and Districts. NCES 2022126. https:// www.nationsreportcard.gov /

highlights/reading/2022/.

New Hampshire Department of Education. 2023. "NH Statewide Assessment System Results, Grades 3–8." Received via data request January 22, 2023.

Ohio Department of Education (ODE). 2022. "Ohio School Report Cards: Download Data." Accessed September 15, 2022, https://reportcard.education.ohio.gov/download.

Oster, Emily, Rebecca Jack, Clare Halloran, John Schoof, Diana McLeod, Haisheng Yang, Julie Roche, and Dennis Roche. 2021. "Disparities in Learning Mode Access Among K–12 Students during the COVID-19 Pandemic, by Race/Ethnicity, Geography, and Grade Level—United States, September 2020–April 2021." Morbidity and Mortality Weekly Report 70 (26): 953.

Parolin, Zachary, and Emma K Lee. 2021. "Large Socio-economic, Geographic and Demographic Disparities Exist in Exposure to School Closures." Nature Human Behaviour 5 (4): 522–528. Peak, Christopher. 2022. How Legislation on Reading Instruction is Changing Across the Country. APM Reports. https:// www.apmreports.org/ story/2022/ 11/17/reading- instruction-legislation-state-map.

Pennsylvania Department of Education (PDE). 2022. "PSSA Results." Accessed November 28, 2022, https://www.education.pa.gov/DataAndReporting/Assessments/Pages/ PSSA-Results.aspx.

Rhode Island Department of Education (RIDE). 2022. "Assessment Data Portal: RICAS." Accessed November 4, 2022, https://www3.ride.ri.gov/ADP.

Schwartz, Sarah. 2022. Which States Have Passed 'Science of Reading' Laws? What's in Them? EducationWeek. https://www.edweek.org/teaching-learning/which-states- have-passed-science-of-reading-laws-whats-in-them/2022/07.

South Carolina Department of Education. 2022. "SC READY." Accessed September 9, 2022, https://ed.sc.gov/data/test-scores/state-assessments/sc-ready/.

South Dakota Department of Education. 2022. "South Dakota Aggregated Smarter Balanced Assessment Results, Grades 3–8." Received via data request November 7, 2022.

U.S. Department of Education (ED). 2021. Frequently Asked Questions: Elementary and Secondary School Emergency Relief Programs, Governor's Emergency Education Relief Programs. Washington, DC: U.S. Department of Education. https://oese.ed.gov/ of f ices / education - stabilization - f und / elementary - secondary - school - emergency - relief - f und/.

Urban Institute Education Data Portal (Version 0.16.0). 2022. Common Core of Data database. Accessed November 16, 2022. Urban Institute. https://educationdata.urban.org/documentation/. Virginia Department of Education (VDOE). 2021. "Virginia Participation." Received via data request November 18, 2021. . 2022. "SOL Test Results." Accessed August 19, 2022, https://www.doe.virginia. gov/statistics reports/sol-pass-rates/index.shtml.

West, Martin R., Robin Lake, Julian Betts, Sarah Cohodes, Brian Gill, Andrew Ho, Susanna Loeb, et al. 2021. How Much Have Students Missed Academically Because of the Pandemic? A Review of the Evidence to Date. Seattle, Washington: Center on Reinventing Public Education (CRPE).

West Virginia Department of Education (WVDE). 2022. "Assessment Achievement." Accessed August 12, 2022, https://zoomwv.k12.wv.us/Dashboard/dashboard/7301. Wisconsin Department of Public Instruction (DPI). 2022. "WISEdash Data Files by Topic: Forward." Accessed October 3, 2022, https://dpi.wi.gov/wisedash/download-files/ type?field wisedash upload type value=Forward.

Wyoming Department of Education (WDE). 2022. "WY-TOPP and WY-ALT Results grades 3 - 10 - District Level - Aggregated." Accessed October 26, 2022, https://portals. edu . wyoming . gov / Reports / (S(upztr0lvig411by0mehp0o4n)) / Public / wde - reports - 2012/public-reports/assessment/pawsresultsdistrict-levelaggregated.

3.A Appendix A: State Assessment Data

Table 3.A.1: *State Assessments: AR, CO, GA, ID*

AR (ARKANSAS)

Assessment Name:	ACT Aspire
Source:	Arkansas Division of Elementary & Secondary Education (2022)
Years Included in Analysis:	2018–2022 (2020 not administered)
Proficiency Levels:	• Proficient: Ready / Exceeding • Not proficient: In Need of Support / Close
2021 Participation Rate:	97%
Additional Information:	In 2021, tests were administered in person. The 2022 data used for this report reflect preliminary scores released in August 2022.

CO (COLORADO)

Assessment Name:	Colorado Measures of Academic Success (CMAS)
Source:	Colorado Department of Education (CDE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	71.6%
Proficiency Levels:	• Proficient: Met / Exceeded Expectations • Not proficient: Did Not Yet Meet / Partially Met Expectations
Additional Information:	In 2021, instead of all students testing in all subjects as in prior years, Grades 3, 5, and 7 were tested in ELA, and Grades 4, 6, and 8 were tested in math (parents could choose to have their children take both tests). In 2021, tests were administered in person. Data in this report reflect all grades tested in other school years.

CT (CONNECTICUT)

Assessment Name:	Smarter Balanced Assessment Consortium (SBAC)
Source:	Connecticut State Department of Education (CSDE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	88.3%
Proficiency Levels:	• Proficient: Meets / Exceeds the Achievement Standard (Standards 3 & 4) • Not proficient: Does Not Meet / Approaching the Achievement Standard (Standards 1 & 2)
Additional Information:	In 2021, in-person testing and remote testing options were available for students to take the state assessments and approximately 12% of students completed the assessment remotely (of these students, over 90% also used a fully or mostly remote schooling mode during the school year (CSDE, 2021). Data in this report reflect outcomes for both in-person and virtual test-taking approaches.

GA (GEORGIA)

Assessment Name:	Georgia Milestones End-of-Grade (EOG) Assessments
Source:	Governor's Office of Student Achievement (GOSA, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	71.4%
Proficiency Levels:	• Proficient: Proficient/Distinguished Learners • Not proficient: Beginning/Developing Learners
Additional Information:	In 2021, tests were administered in person.

ID (IDAHO)

Assessment Name:	Idaho Standards Achievement Test (ISAT)
Source:	Idaho State Department of Education (2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	98%
Proficiency Levels:	• Proficient: Proficient / Advanced • Not proficient: Below Basic / Basic
Additional Information:	In 2021, tests were administered in person.

Table 3.A.2: *State Assessments: KS, LA, MA, MN, MS*

KS (KANSAS)

Assessment Name:	Kansas Assessment Program (KAP) General Education Assessments
Source:	Kansas State Department of Education (2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	93.3%
Proficiency Levels:	• Proficient: Effective Ability / Excellent Ability • Not proficient: Limited Ability / Basic Ability
Additional Information:	In 2021, tests were administered in person.

LA (LOUISIANA)

Assessment Name:	LEAP 2025 (Louisiana Educational Assessment Program)
Source:	Louisiana Department of Education (2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	98.5%
Proficiency Levels:	• Proficient: Mastery / Advanced • Not proficient: Unsatisfactory / Approaching Basic / Basic
Additional Information:	In 2021, tests were administered in person.

MA (MASSACHUSETTS)

Assessment Name:	Massachusetts Comprehensive Assessment System (MCAS)
Source:	Massachusetts Department of Elementary and Secondary Education (MA DESE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	95%
Proficiency Levels:	• Proficient: Meeting / Exceeding Expectations • Not proficient: Not Meeting / Partially Meeting Expectations
Additional Information:	In 2021, students took one session of each subject area test instead of two (DESE, 2021a). Additionally, in-person testing and remote testing options were available for students to take the state assessments in 2021, and approximately 20% of students completed the assessment remotely (DESE, 2021).

MN (MINNESOTA)

Assessment Name:	Minnesota Comprehensive Assessments (MCA)
Source:	Minnesota Department of Education (MDE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	78.2% (MDE, 2021)
Proficiency Levels:	• Proficient: Meets / Exceeds Standards • Not proficient: Does Not Meet / Partially Meets Standards
Additional Information:	In 2021, tests were administered in person.

MS (MISSISSIPPI)

Assessment Name:	Mississippi Academic Assessment Program (MAAP)
Source:	Mississippi Department of Education (2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	96.9%
Proficiency Levels:	• Proficient: Proficient /Advanced (Levels 4-5) • Not proficient: Minimal / Basic / Passing (Levels 1-3)
Additional Information:	In 2021, tests were administered in person.

Table 3.A.3: *State Assessments: MO, NH, OH, PA, RI, SC*

MO (MISSOURI)

Assessment Name:	Missouri Assessment Program (MAP)
Source:	Missouri Department of Elementary & Secondary Education (2023)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	91%
Proficiency Levels:	• Proficient: Proficient / Advanced • Not proficient: Below Basic / Basic
Additional Information:	In 2021, tests were administered in person.

NH (NEW HAMPSHIRE)

Assessment Name:	NH Statewide Assessment System (NH SAS)
Source:	New Hampshire Department of Education (2023)
Years Included in Analysis:	2018–2022 (2020 not administered)
2021 Participation Rate:	80% in ELA; 81% in math
Proficiency Levels:	• Proficient: Levels 3-4 • Not proficient: Levels 1-2
Additional Information:	In 2021, tests were administered in person.

OH (OHIO)

Assessment Name:	Ohio Achievement Assessment (OAA)
Source:	Ohio Department of Education (ODE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	94%
Proficiency Levels:	• Proficient: Proficient / Accelerated / Advanced / Advanced Plus • Not proficient: Limited / Basic
Additional Information:	In 2021, tests were administered in person.

PA (PENNSYLVANIA)

Assessment Name:	Pennsylvania System of School Assessment (PSSA)
Source:	Pennsylvania Department of Education (PDE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	71%
Proficiency Levels:	• Proficient: Proficient / Advanced • Not proficient: Below Basic / Basic
Additional Information:	In 2021, tests were administered in person.

RI (RHODE ISLAND)

Assessment Name:	Rhode Island Comprehensive Assessment System (RICAS)
Source:	Rhode Island Department of Education (RIDE, 2022)
Years Included in Analysis:	2018–2022 (2020 not administered). The state first administered the RICAS assessment in 2018.
2021 Participation Rate:	88.9%
Proficiency Levels:	• Proficient: Meeting / Exceeding Expectations • Not proficient: Not Meeting / Partially Meeting Expectations
Additional Information:	In 2021, tests were administered in person.

SC (SOUTH CAROLINA)

Assessment Name:	South Carolina College-and Career-Ready Assessments (SC READY)
Source:	South Carolina Department of Education (2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	87.9%
Proficiency Levels:	• Proficient: Meets / Exceeds Expectations

Table 3.A.4: *State Assessments: SD, VA, WI, WV, WY*

	• Not proficient: Does Not Meet / Approaches Expectations
Additional Information:	In 2021, tests were administered in person only.

SD (SOUTH DAKOTA)

Assessment Name:	Smarter Balanced Assessment Consortium (SBAC)
Source:	South Dakota Department of Education (2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	95%
Proficiency Levels:	• Proficient: Levels 3 / 4 • Not proficient: Levels 1 / 2
Additional Information:	In 2021, tests were administered in person only.

VA (VIRGINIA)

Assessment Name:	Standards of Learning (SOL)
Source:	Virginia Department of Education (VDOE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	78.7% (VDOE, 2021)
Proficiency Levels:	• Proficient: Measured by VA's "Pass Rate"
Additional Information:	In 2019, the state updated its math cut scores to reflect the 2016 mathematics content standards. In 2021, the state updated its reading cut scores to reflect the 2017 English content standards. In 2021, tests were administered in person.

WI (WISCONSIN)

Assessment Name:	Forward Exam
Source:	Wisconsin Department of Public Instruction (WI DPI, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	87.0%
Proficiency Levels:	• Proficient: Proficient / Advanced • Not proficient: Below Basic / Basic
Additional Information:	In 2021, tests were administered in person.

WV (WEST VIRGINIA)

Assessment Name:	West Virginia General Summative Assessment (WVGSA)
Source:	West Virginia Department of Public Instruction (WVDE, 2022)
Years Included in Analysis:	2017–2022 (2020 not administered)
2021 Participation Rate:	83.9%
Proficiency Levels:	• Proficient: Meets / Exceeds Standard • Not proficient: Does Not Meet / Partially Meets Standard
Additional Information:	In 2021, tests were administered in person.

WY (WYOMING)

Assessment Name:	Wyoming Test of Proficiency and Progress (WY-TOPP)
Source:	Wyoming Department of Education (WDE, 2022)
Years Included in Analysis:	2018–2022 (2020 not administered). The state first administered the WY-TOPP in 2018 to replace the state's former PAWS assessment
2021 Participation Rate:	97.0%
Proficiency Levels:	• Proficient: Proficient / Advanced • Not proficient: Below Basic / Basic
Additional Information:	In 2021, tests were administered in person.

Note. Participation rates reflect the districts included in the states in our sample, but align with state-reported participation rates as well.

3.B Appendix B: Schooling Mode Data

Table 3.B.1: *COVID-19 School Data Hub Schooling Mode Data for States Included in Analyses*

#	State	Original Data Source	Data Level	Time Period Interval
1	AR	Arkansas Department of Education (ADE)	School	Monthly, 10/1/20- 5/31/21
2	CO	Colorado Department of Education (CDE)*	District	Monthly, 8/1/20- 6/31/21
3	CT	Connecticut State Department of Education (CSDE)	District	Weekly, 8/30/20- 6/5/21
4	GA	Georgia Policy Labs (GPL)*	School	Monthly, 8/1/20- 6/31/21
5	ID	Idaho State Department of Education (SDE)*	District	Weekly, 8/9/20- 6/19/21
6	KS	Kansas Department of Education (KSDE)	District	Weekly, 8/16/20 - 5/29/21
7	LA	Louisiana Department of Children and Family Services (DCFS)*	School	Monthly, 8/1/20- 6/31/21
8	MA	Massachusetts Department of Elementary and Secondary Education (MA DESE)	District	Bi-weekly, 10/1/20 - 5/26/21
9	MN	Minnesota Department of Education (MDE)	District	Weekly, 9/1/20 - 5/31/21
10	MS	Mississippi Department of Human Services (MDHS)*	School	Monthly, 8/1/20- 5/31/21
11	MO	Missouri Department of Elementary and Secondary Education (MO DESE)	District	Monthly, 9/1/20 - 5/31/21
12	NH	New Hampshire Department of Health and Human Services (NH DHHS)*	School	Monthly, 9/1/20 – 6/30/21
13	OH	Ohio Department of Education (ODE)	District	Weekly, 8/2/20- 5/22/21
14	PA	Pennsylvania Department of Human Services (DHS)*	District	Monthly, 9/1/20 - 5/31/21
15	RI	RI Department of Elementary & Secondary Education (RIDE)	School	Weekly, 9/13/20- 6/19/21
16	SC	South Carolina Department of Education (SCDE)	District	Monthly, 9/1/20 - 5/31/21
17	SD	South Dakota Department of Education (SDDOE)*	District	Monthly, 9/1/20 - 5/31/21
18	VA	Virginia Department of Education (VDOE)	District	Monthly/Weekly, 9/8/20-5/9/21
19	WI	Wisconsin Department of Public Instruction (DPI)*	School	Monthly, 8/1/20- 6/30/21
20	WV	West Virginia Department of Education (WVDE)	District	Weekly, 9/6/20-6/12/21
21	WY	Wyoming Department of Education (WDE)	District	Weekly, 8/16/20-6/12/21

Note. All data files were sourced from the COVID-19 School Data Hub, a central database for state schooling model data. Schooling mode information provided by state agencies designated with an (*) was collected as part of the state's plan to determine and disburse benefits as part of the Pandemic Electronic Benefit Transfer (P-EBT) program through the U.S. Department of Agriculture. This program was designed to offer temporary emergency nutrition benefits to eligible students who were not able to receive meals at school (that they would have otherwise received from the National School Lunch Program) due to school closures or schools operating with reduced hours.