THE UNIVERSITY OF CHICAGO

ESSAYS ON EDUCATION AND HEALTHCARE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
ERIC ENGLISH

CHICAGO, ILLINOIS
JUNE 2020

I dedicate this dissertation to my friends and family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am deeply grateful for the feedback and guidance of the professors on my committee: Steven Levitt, Michael Dinerstein, and Seth Zimmerman. I am also grateful for comments from Derek Neal, Michele Carter, Ishan Nath, and Kimberly Conlon, and indebted to the New York City Department of Education which provided me with data.

ABSTRACT

This dissertation examines the impacts of public policy in the healthcare and education sectors. The first chapter examines a number of government policies that encourage successful charter networks to open new schools. However, there is little evidence on how quality changes within a network as it expands, and from a theoretical perspective, the direction of that change is ambiguous. I use student-level panel data from New York City to examine within-network changes in quality, and I find that charter networks are not able to maintain test-score value-added quality in expansion schools. Later schools in a network have lower quality, especially for English test score value-added. This decline increases in magnitude with a school's ordinal number in that network. In addition, I find that network value-added declines with a school's age, while quality for non-network charters improves with age. Last, I look into some mechanisms that could explain these trends, focusing on changes to the treatment group on the demand side and span of control issues on the supply side. Although I find heterogeneous returns to charter schools based on student demographics and prior test scores, I find no evidence that demographics or prior test scores changed in a way that could explain the patterns in quality that I observe. If anything, the prior test scores of incoming cohorts moved in a way that suggest I am underestimating the declines in quality with age. At the same time, I do not find any evidence that network declines in quality can be attributed to managers having increasing difficulty managing networks as they expand.

The second chapter investigates the Center for Medicare and Medicaid Services' attempts to reduce the high number of unplanned hospital readmissions. As part of the Affordable Care Act, CMS introduced the Hospital Readmissions Reduction Program (HRRP) which fined hospitals for excessive risk-adjusted readmissions. After readmissions fell, CMS declared the program a success. However, the HRRP was criticized by some because hospitals which served more minorities and low-income patients were more likely to be fined. This

led some to argue that demographics should be included in the readmission rate risk adjustment. However, it is unclear if patient demographics had a causal effect on readmission rates or if low-income and minority patients were just attending lower quality hospitals. In this chapter, I investigate both whether the HRRP has been effective and if the risk-adjustment calculations should be altered. To achieve the first goal, I use hospitals that were exempt from the program as a control group and uncover little evidence that the HRRP had much of an effect. In the years following the introduction of the program, the treatment and control hospitals saw nearly identical declines in readmission rates. To achieve my second goal, I identify ninety-nine hospital closings in the last eight years as an exogenous change to patient demographics at the closest neighboring hospital. I use difference-in-differences to first confirm that the closest hospital sees an increase in patients relative to the other nearby hospitals, and then determine if readmission rates changed based on the change in average patient demographics. I am unable to reject the null that the change in patient demographic mix had any effect on a hospital's readmission rates.

CHAPTER 1

CAN CHARTER SCHOOL NETWORKS SCALE UP EFFECTIVELY?

1.1 Introduction

In the last two decades, charter schools have been championed as a politically feasible way to improve the quality of education without increasing funding. Charter schools are publicly funded and privately run, and they are often a part of networks that are run by a charter management organization (CMO). In recent years, state and federal governments have adopted a number of policies that encourage successful networks to expand by replicating their schools. However, there is little evidence about how quality changes within a charter network as it expands, and from a theoretical perspective, the direction of this change is ambiguous. First, it is not clear if quality will be correlated across schools within a network since many of the determinants of school quality cannot be perfectly replicated, e.g. principals. Even if many features are replicable, there still may be declines in quality with expansion if certain inputs into the education production function are scarce. For example, good teachers might become more and more difficult to recruit as a network expands. Conversely, it is possible that quality could improve with expansion. There could be economies of scale if inputs such as the curriculum or certain course materials can be shared across schools in the same network. If the network then reinvested these cost savings, quality could improve with expansion.

There has not yet been any research on the evolution of quality within a network as it expands. Prior research on charter schools has mostly focused on either the school level or the market level. At the market level, these papers could be broadly classified as focusing on the question "are charter schools good?" Many find positive results. For example, Abdulkadiroğlu et al. (2011), Hoxby & Murarka (2009), Dobbie and Fryer (2011, 2015), Angrist et al. (2013), Tuttle et al. (2013) find that on average, charter schools improve test scores. Dobbie and Fryer (2015) and Angrist et al. (2016) both find that charter schools increase college enrollment. Dobbie & Fryer (2016) determine that on average, charter schools in Texas do not improve earnings later in life. Cohodes et al. (2019) determine that charters continue to improve scores by a similar amount in Boston when they account for a larger share of the market. However, other papers show a more mixed picture. CREDO (2013), Furgeson et al, (2012), Gleason et al. (2010), Zimmer et al. (2012) all find that when researchers examine areas broader than a single city, charter students do not do any better than traditional public students.

Because charter studies have found a mixed record of success, the existing research at the school level has focused on answering the question "what features make charter schools successful?" Angrist et al. (2013), Baude et al. (2014), and Dobbie & Fryer (2013, 2016) find that schools that utilize a "No Excuses" mantra improve both test scores and earnings more than those that do not. The "No Excuses" mantra includes a number of practices such as harsh discipline, longer school days, and data-driven instruction. Angrist et al. (2011) and Gleason et al. (2010) find that charter schools increase test scores much more in low income and urban areas. CREDO (2013, 2017) find that schools associated with non-profit management organizations tend to do better than those associated with for-profit management organizations.

Since prior research on charters has not been uniformly positive, it is understandable that a number of government policies favor the replication of successful charter schools as opposed to simply expanding the charter sector. However, given these policies, it is important to focus on networks as the unit of interest and determine how quality changes with expansion.

These policies take a few forms. First, there are a number of state and federal grants that are rewarded to networks that want to expand. Second, network affiliation has an impact on charter authorizers. In many states, it is easier for existing networks to open new charter schools than entrants. Third, governments often consider network affiliation when making oversight decisions. The SUNY Charter Institute states in its guidelines for the approval of new schools in an existing network that the breadth and depth of the review will be lower.

It is also important to focus on the network level because networks have reputations that likely affect decisions by donors and parents. Most of the large donations made to charter schools are made to networks and not specific schools; for example, in 2016, Julian Robertson gave $25 million to the Success Academy network – not to a specific school in that network[1]. Parents, too, are likely influenced by the network affiliation of a school when making enrollment decisions. Proving this empirically will be the subject of future work, however, it is clear that most networks emphasize the network affiliation of their schools. For example, the schools in the Success Academy network have names like Success Academy Harlem 1, Success Academy Harlem 2, or Success Academy Bronx 1. This suggests the network wants parents to focus on the link to the Success Academy network when choosing a school for their child.

In order to study charters at the network level, I use New York City as my setting. Charter school networks have expanded rapidly in New York City in the last twenty years, and it now has the second largest charter sector in the U.S. when ranked by the total number of charter students. By the 2018-19 school year[2], New York City had twenty-nine networks,

1. Success Academy (2016), *Success Academy Receives $25 Million Gift from the Robertson Foundation to Expand System of Top-Tier Public Charter Schools.*

2. From this point on, I will use the first year to identify school years – i.e. I will refer to the 2018-19 school year as 2018.

the largest of which had forty-five schools with more than 17,000 students. New York City is also a good setting because they have been collecting detailed student-level data for most of the period of interest. This data includes scores on state-required English and math tests in elementary, middle, and high school, as well as a rich set of demographics. Students are given a unique, numeric identifier, so I am able to track them throughout their entire time in New York City public schools. I take advantage of the panel nature of this data to calculate test score value-added measures in English and math for every public school in New York City in each year from 2007 to 2015.

From these estimates, I document four key facts that provide context to my analysis of charter network scale-up. First, at the beginning of my sample, network charters were better than traditional public schools on average, especially for math. In expectation, network charter schools improved math scores by 0.2 test score standard deviations as compared to the average public school[3]. They also boosted English scores by 0.05 standard deviations. Second, over my sample period, network test score value-added declined in English and even more so in math. Third, there is significant variation in quality across networks. The best networks consistently have some of the best schools in the city, while the worst are repeatedly below average. Last, network expansion was correlated with math test score value-added at the beginning of my sample, but conditional on a network's math quality, English quality has no relation to expansion. That being said, network English and math quality have a fairly high correlation.

With these facts in mind, I then investigate whether networks are able to maintain quality as they expand and find that they cannot. This decline takes two channels. First, schools that open later in a network are not as good as schools that open earlier in the same network,

3. This is not exactly the "average" public school, but it is very close. See Section 4 for a more complete discussion.

especially for English. This decline increases in magnitude with more and more expansion. On average, the second and third schools in a network have an English value-added that is 0.04 test score standard deviations lower than the first school in that network within a given year. That decline becomes 0.08 standard deviations for the sixth or later school to open within a given network. A similar pattern presents itself for math; however, the coefficients are smaller and not statistically significant. The second channel for decline is that network schools' value-added quality declines approximately 0.01 test score standard deviations in English with each year that that school has been open. Math again has a similar decline, but is not statistically significant.

Last, I investigate mechanisms that could explain these patterns. Possible mechanisms act on either the supply side or the demand side. On the demand side, heterogeneous treatment effects coupled with changes to the student body could result in the quality changes I find. Ladd et al. (2016) find that demographics changed at the market level in North Carolina over time, so if this has happened in New York City as well and new charter students belong to demographic groups that benefit less, network quality would decline with expansion as I observe. This explanation has been frequently studied in the context of scaling up successful randomized controlled trials, (Andrews and Oster, 2017; Campbell and Stanley, 1963; Cook and Campbell, 1979; Cronbach and Shapiro, 1982; Heckman and Vytlacil, 2005, 2007; Hedges and O'Muircheartaigh, 2011; Tipton, 2013; Stuart et al., 2011) though not in the context of charter network expansion. For this explanation to be valid, heterogeneous returns to charters are a necessary condition. Prior research on charter schools has found that different demographic groups tend to benefit more (Abdulkadiroğlu et al. (2011), Angrist et al. (2012, 2017)). I first verify that this happens in New York City. However, when I then look for evidence of systemic change within a network for either demographics or prior test scores, I do not find any. I also do not find any evidence that demographics change within a school as it ages, though I do find some evidence that prior test scores of incoming cohorts fall. Since these students typically benefit *more* from charter schools, this actually implies that I am underestimating the decline in quality with age.

The other possible mechanisms all act on the supply side. The first possibility is that there are scarce inputs that networks have difficulty obtaining as they expand. For example, quality teachers might become more and more difficult to recruit. The second possible mechanism is that networks initially spend more on quality in order to build a reputation, but after that reputation is established, they cut back. Last, it could be that as a network expands, it becomes more and more difficult to run. In regards to other industries, it has long been proposed that managers have a certain span of control (Lucas (1978), Rosen (1982)). It is fairly easy to imagine that a network with forty-five schools is much harder to manage than one with only three or four.

All three of these supply side mechanisms support the same policy implications. First, charter authorizers should reconsider policies that encourage network expansion such as increased funding and looser oversight. Second, policymakers should look at the marginal, not average, school in a network when approving new schools. In other words, if a network applies to open an eleventh school, authorizers should look at quality of the ninth and tenth school in that network when deciding on approval. Currently, authorizers look at the average quality of all previous schools in that network. This paper, does not imply, however, that there should be a cap on network size. There is a lot of heterogeneity in quality across networks, so even with declines in quality, later schools in the best networks are still quite good. In lower quality networks though, authorizers might want to consider not allowing them to get too large.

I arrange my paper as follows: in Section 2, I discuss institutional details of charter schools; in Section 3, I discuss my data; in Section 4, I discuss my empirical strategy; in Section 5, I discuss some contextual quality findings; in Section 6, I discuss my results; in Section 7, I investigate possible mechanisms for the patterns I observe; and Section 8 is my conclusion.

1.2 Institutional Details

1.2.1 Charter Laws

Oversight

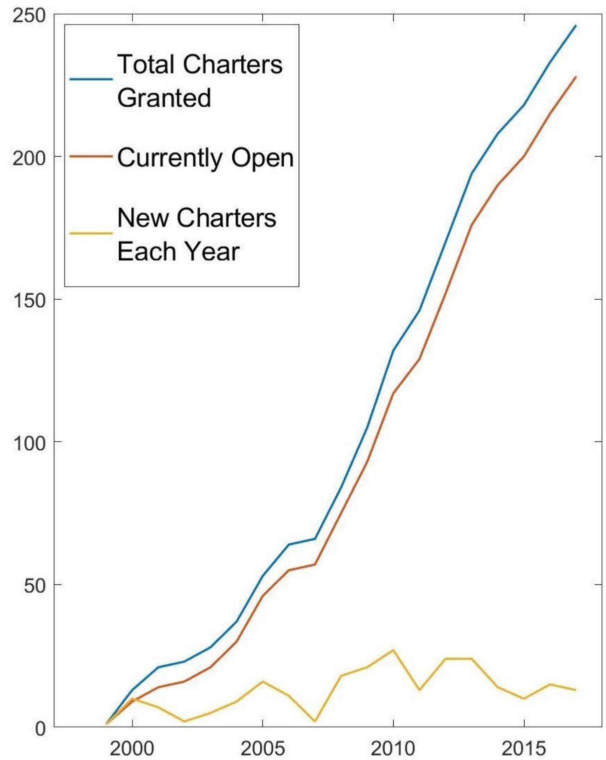
Charter schools emerged in the early 2000s as a politically feasible way to try to improve school quality. They are publicly funded and regulated, but privately run. This compromise allowed the introduction of the competitive market, without fully privatizing education. Since I am focusing on New York City, I will describe the charter laws in New York, though they are broadly similar in other states. The first charter schools began opening in New York after the passage of the Charter Schools Act of 1998.

If an organization wants to open a charter school, they must get approval from one of three authorizing bodies every five years: the New York State Board of Regents, the State University of New York (SUNY) Charter Institute, or the local board of education[4]. They must then be approved again by the Board of Regents if they applied to one of the other bodies first. These applications are very detailed and are often more than five hundred pages long. They include details such as the grade size, curriculum, verification that parents in the area want a charter school, and many others.

If approved, a charter school is subject to oversight throughout the term from the body

4. The New York City Department of Education approved charter schools when Michael Bloomberg was mayor, but stopped after Bill de Blasio was elected.

Figure 1.1: Stock and Flow of Approved Charters in NYC



which approved their initial application. This oversight is to ensure that the school is meeting various goals. The authorizing bodies can shut down charter schools if they are not maintaining high levels of achievement. Some charters get shut down in the middle of a term if they are doing very badly, while others are not given renewals. One network was also shut down because it's founder was stealing public money[5]. There have also been a few schools that voluntarily decided to leave because they could not attract enough students. Figure 1.1 shows the steady approval of charters from 2000-2015. As we can see from the figure, there are not many charters that were approved that are not still open. Throughout the whole sample, only twelve have been closed for the mix of reasons stated above.

5. https://www.wnyc.org/story/302621-believe-charter-founder-indicted/

Enrollment

Anyone who wants to attend a charter school can apply, even if they live in another part of the city. If a charter school is oversubscribed, i.e. it has more applicants than seats, it is required by law to run a randomized lottery. School and grade size are detailed in the initial application, and they cannot be changed without approval from the state, no matter how many students apply.

Standard Funding

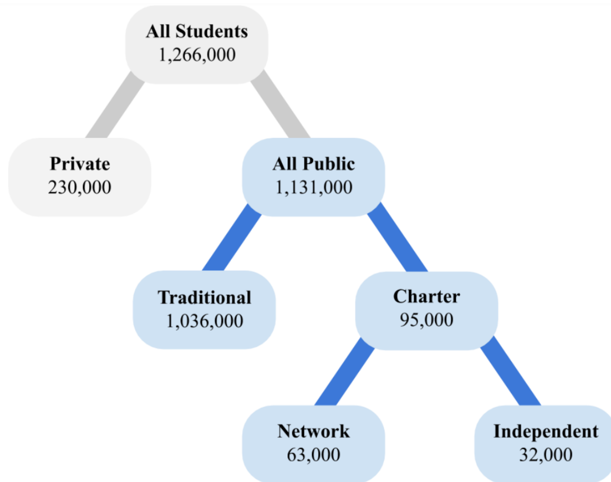
Charter schools are funded on a per-student basis. This funding is subject to certain increases for students who need an Individualized Education Program or are English language learners. In addition, in 2014, New York State passed a law that said that school districts must work with charter schools to help find space in a building already owned by the city or provide rental assistance. Before this law was passed, Mayor Michael Bloomberg he worked with charter schools to help them find space in city-owned buildings.

1.2.2 Public School Sector Organization

Distribution of Students

New York City is the largest school district in the country. Figure 1.2 shows the distribution of students in 2015. There were about 1.3 million K-12 students in New York City in 2015. Of those, 1.1 million attended public schools. I do not have any data on private school students, so I will not discuss them. Of the public school students, about ten percent, or one hundred thousand, attended charter schools. The term “all public schools” refers to both traditional public schools and charter schools. Within the charter sector, about two-thirds attended network charters and one-third attended independent charters.

Figure 1.2: Distribution of New York City K-12 Students

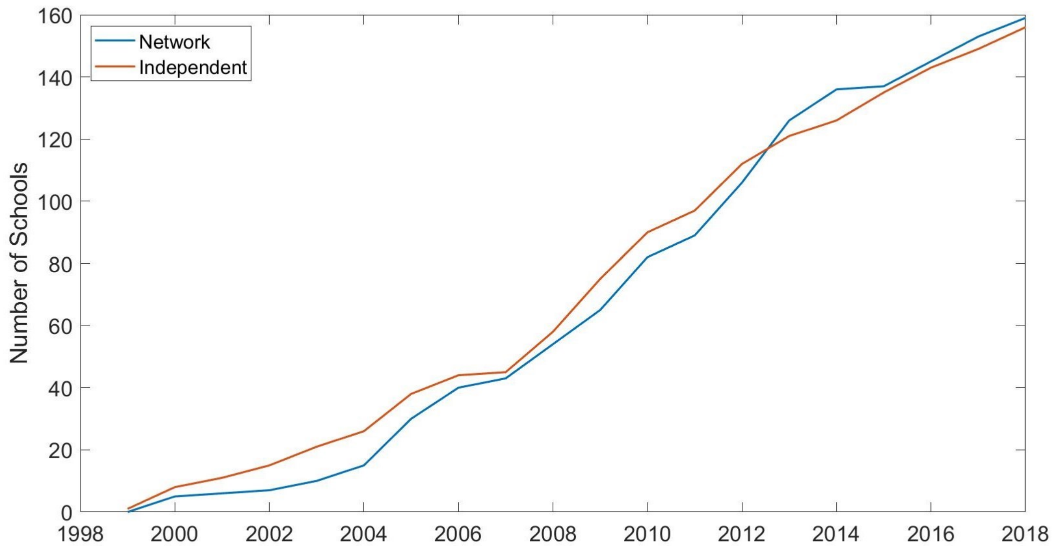


Types of Charter Schools

Within the charter sector in New York, there are two main types: schools that are run by a non-profit charter management organization (CMO) and independent charters[6]. As shown in Figure 1.3, there has been fairly similar growth in the number of independent and network charters in New York City. Independent charters are not a part of a network, though they can join one or expand to create their own network. This does not mean that the distinction between independent charters and networks is completely superficial. Independent charters are a bit more varied than CMO schools. They might be organized by a neighborhood group or by a successful businessman coming back to the neighborhood they grew up in. There are a number of differences in school characteristics and student demographics across the two types of schools that are not just a function of what year they were founded. I will discuss these differences in detail in Section 3. In addition, there are many independent charters that have been successful for many years that have not expanded. In my analysis,

6. There are also Education Management Organizations (EMOs) which are for-profit organizations. In 2010, the New York State Legislature revised the Charter Schools Act so that no new charters could be granted to EMOs. At the time, there were 12 schools in NYC managed by three EMOs. Two of these EMOs quickly left, meaning eight of these schools were no longer managed by a network. One EMO still operates four schools in NYC. In my analysis, I treat the one EMO that remains the same as the other CMOs. I drop the EMOs that left NYC for any network level longitudinal analysis

Figure 1.3: Number of Charter Schools by Type in New York City

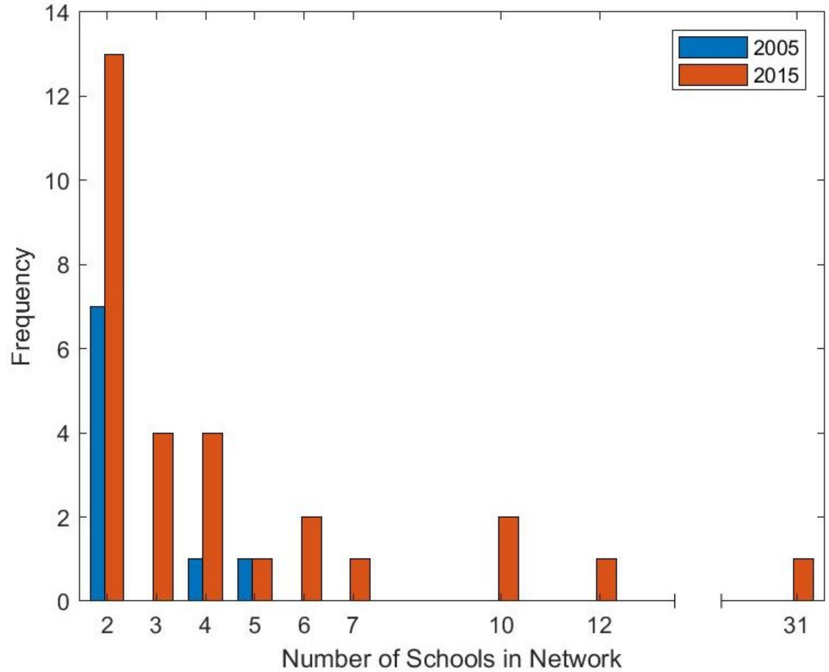


when I divide charters into network and independent, it is really "ever network" and "always independent." While I will present certain results on independent charters, this is just to provide context. My main analysis is focused on the evolution of networks.

1.2.3 Charter Growth

During the expansion of the charter sector in New York City, both the number and size of networks increased dramatically, as shown in Figure 1.4. In 2005, there were only two networks that had more than two schools in them, and the largest only had five. In 2015, there were sixteen networks with more than two schools, four that had ten or more, and one that had thirty-one. Figure 1.5 shows the steady growth in charter enrollment in New York City. Charter schools in New York were more likely to enroll elementary and middle school students than high school. By the end of my sample (the 2015 school year), there were nearly one hundred thousand charter students, eighty percent of which were elementary or middle school (EMS) students. Of New York City public elementary and middle school students in 2015, 10.4% attended a charter school. Of New York City high school students, 4.8% attended a charter school. These enrollment totals mean that New York City has the

Figure 1.4: Charter Network Sizes, 2005 vs. 2015

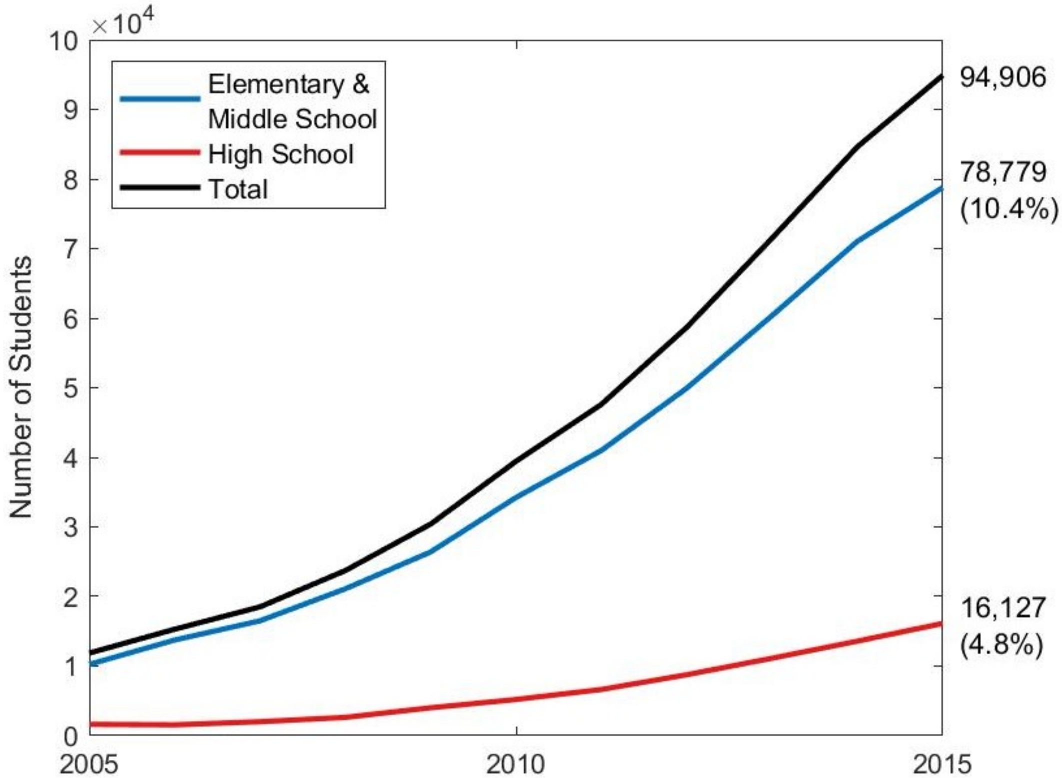


second highest number of charter students in any school district in the U.S.

1.2.4 Government Policies that Encourage Network Expansion

As the charter sector expanded, both state and federal governments began implementing policies that encouraged the replication of successful charter schools. At the federal level, the government awards Charter Schools Program Grants for Replication and Expansion of High-Quality Charter Schools. These grants are awarded to charter networks that have shown prior success and want to open new schools. Funding varies from year to year, but in the latest fiscal year, more than $100 million dollars were given out. The federal government gives out even more money in grants to the states, which can then allocate those funds as they see fit. New York State allotted some of this grant money to networks to replicate and expand. (The rest went towards schools serving various disadvantaged groups). The federal

Figure 1.5: Charter Enrollment Growth in New York City



government gave New York an award of $113 million in 2011 and a similarly sized award in 2018[7].

In addition, there are charter school oversight policies that make it easier for charters to replicate their schools. New York makes the application, renewal, and oversight process easier for established schools and networks. The SUNY Charter Institute's guidelines state that for the approval of new schools in a network or for their renewal "the scope and timing of the Institute's oversight process changes. In general, and consonant with the record of success...the frequency of the Institute's visits, as well as their breadth and depth decreases[8]." In addition, the performance of earlier schools in a network sometimes results in a more generous renewal process. For example, SUNY initially approved the renewal of three Success Academy charter schools that had only completed one year of their current five year charter term, essentially giving these schools ten year terms[9].

Although loosening oversight with prior success seems logical, it might not be good practice. With traditional public schools, it would be unusual for state regulators to assume that if a few schools in a district are good, then they all are. This might be unreasonable for large charter networks as well since the largest networks are the size of small urban school districts. In 2018, Success Academy operated forty-five schools in New York City with over 17,000 students. To put this in perspective, this is slightly larger than the Flint, Michigan school district. Of the more than 13,000 school districts in the U.S., Success Academy would

7. I was unable to find an exact breakdown of the awards in 2018 because many sections of the Department of Education's website do not work. However, in years that the data is available, the grants are fairly proportional to the state's population. Of the states that won a total of $315 million, New York accounted for about one-third of the total population

8. The Board of Trustees of the State University of New York. *Policies for the Renewal of Not-for-Profit Charter School Education Corporations and Charter Schools Authorized by the Board of Trustees of the State University of New York*. September 4, 2013.

9. The Board of Regents ultimately overturned this renewal, but many states do not have two levels of approval.

rank around 450 if it were its own school district[10]. Given the distinct possibility that scaling up a network to the size of a small urban school district is difficult, loosening oversight might not make sense.

1.3 Data

In my paper, I am using data from the New York City Department of Education (NYC DOE). They have provided me with four different datasets which include data on demographics, residence information, elementary and middle school test scores, and high school Regents Exam scores. Students are given a nine-digit unique identification number that remains the same across these datasets and over time. I also use school address data that is publicly available on NYC Open Data, combined with latitude and longitude data pulled from Google Maps.

1.3.1 Test Score Data

Every year, students in New York State in grades three through eight are required to take tests in English and math. I have these test scores back to 2003, which I normalize within each year, subject, and grade. Meaning, all fourth grade math tests in 2012 were normalized together. For high school students, I have scores from the New York State Regents Exams from 2000-2015. Students must pass five exams with a sixty-five or higher in order to get a diploma: English, US History, Global History, one science exam, and one math exam[11]. Math and science both have multiple options that students can choose from. For math, 98% of students take the Algebra exam first, though some students take more than one subject.

10. U.S. Department of Education, National Center for Education Statistics, Common Core of Data, "Public Elementary/Secondary School Universe Survey," 2008-2009

11. Dee et al. (2019) find evidence of bunching right above the cutoff for passing and attribute it to teachers manipulating test scores. I am able to replicate this finding, but find no difference between the rate of manipulation at charter schools and traditional public schools.

For science, 83% take biology. Unfortunately, the variation in test subject for science is mainly across schools, not within, so I would have to drop nearly one-fifth of the schools which are not randomly selected. For this reason, and because I only have English and math for elementary and middle school students, I focus on math and English. Because students can take the Regents Exams as many times as they want, I only use a student's first attempt for each subject (though I include all tests taken at that date when normalizing the scores). In addition, students are allowed to take these exams in any grade. While I control for this is my school quality regressions, these three aspects of the Regents exams might give a noisier measure of quality. Therefore, I will also estimate my main regressions using only elementary and middle schools as a robustness check.

1.3.2 Demographic Data

Table 1.1: Demographics — Elementary and Middle School Students, 2015

	Traditional Public Schools	Charters	Charters: Network	Charters: Independent
Average Grade Size	116	76	77	73
% Black	0.29	0.54	0.58	0.46
% Hispanic	0.42	0.37	0.33	0.43
% Asian	0.13	0.03	0.02	0.03
% White	0.14	0.05	0.04	0.05
% English Language Learner (ELL)	0.15	0.07	0.06	0.10
% Public Assistance	0.74	0.76	0.76	0.76
% Immigrant	0.13	0.06	0.05	0.07
N	674,120	78,257	53,851	24,406

The demographics data start in 2005. This dataset includes a wide array of demographic variables such as race, gender, and language spoken at home. The data also includes an

Table 1.2: Demographics — High School Students, 2015

	Traditional Public Schools	Charters	Charters: Network	Charters: Independent
Average Grade Size	161	80	69	98
% Black	0.36	0.55	0.65	0.37
% Hispanic	0.44	0.38	0.29	0.56
% Asian	0.09	0.01	0.01	0.02
% White	0.08	0.03	0.03	0.02
% English Language Learner (ELL)	0.13	0.05	0.03	0.06
% Public Assistance	0.74	0.77	0.74	0.79
% Immigrant	0.25	0.13	0.10	0.13
N	318,479	16,127	8921	7206

indicator for whether or not a student has been identified by the New York Human Resources Administration as qualifying for certain types of public assistance, such as for free or reduced-price lunch. Tables 1.1 and 1.2 display summary statistics for elementary and middle school students and high school students respectively, broken out by type of school. The main takeaway from the left side of these tables is that charter schools enroll more black students than traditional public schools (TPS) and fewer white and Asian students. They also enroll lower proportions of immigrants and English language learners (ELL) and similar numbers of students on public assistance.

There are also some key demographic differences between network charter students and independent charter students as shown on the right side of Tables 1.1 and 1.2. At the elementary and middle school level, network charters enroll more black students but fewer Hispanic students than independent charters. Independent charters also enroll more black students than traditional public schools, though the gap is smaller. On the other hand, network charters enroll fewer Hispanic students than traditional public schools or independent charters. Neither type enrolls many white or Asian students. These differences between network and independent charters are generally the same for high schools. However, the divergence in black and Hispanic enrollment between the two types of charter is even larger. One last thing to note is that while networks enroll two-thirds of elementary and middle school students, it is much closer to fifty-fifty for high schools.

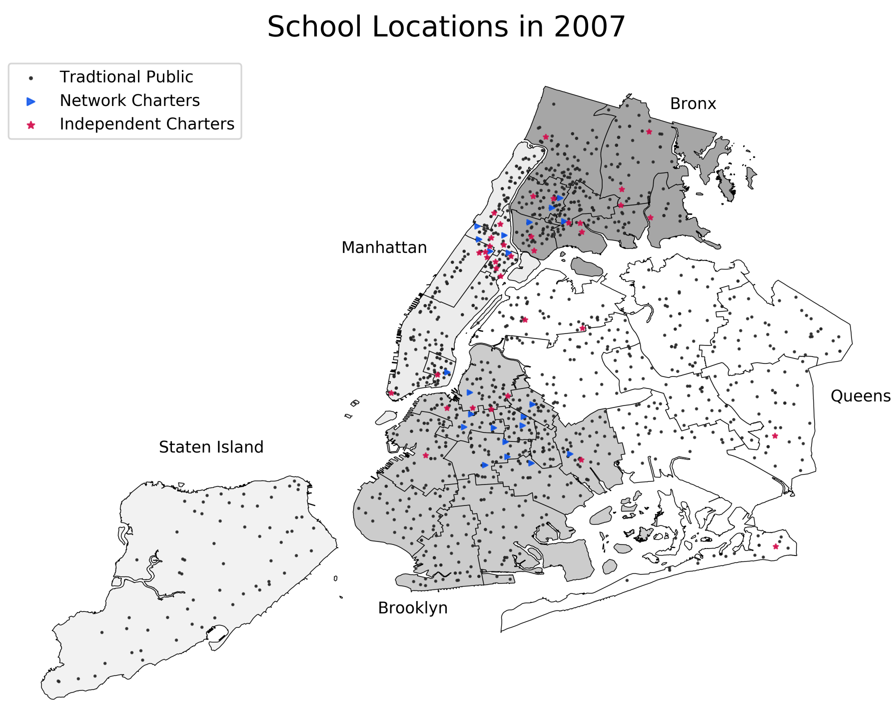
1.3.3 Student Residence Data

Starting in 2007, I have detailed student location data. This data includes which school a student is zoned for, their zip code, and census tract and block information. In New York City, census blocks are often literally one square block (see A.1), so the location data is fairly granular. I can link this data to publicly released census tract data on income and the educational attainment of adults to get more demographic information about the students. This is useful since I do not have data on parental education, and the income measure that the NYCDOE provided is very coarse. Based on the public assistance measure, traditional public and charter students seem fairly similar. However, Figure A.3 shows the distributions of the median household income of the census tracts that students live in. This figure shows that charter students tend to come from poorer neighborhoods.

1.3.4 School Location Data

Figures 1.6 and 1.7 show the expansion of charter schools throughout New York. The city is split into thirty-two different school districts, which appear on the figure as the black outlines within the boroughs. Charters did not locate randomly throughout the city, and are concentrated in certain neighborhoods. In 2015, charter schools were unevenly distributed across Brooklyn (85), the Bronx (72), Manhattan (52), Queens (22), and Staten Island (4). Figure 1.8 is a map of median household income at the zip code level. The colors represent the percentiles within the city, with dark red being the lowest income zip codes and dark green being the highest. As we can see from these maps, charters tend to locate in poorer neighborhoods.

Figure 1.6: Charter and Traditional Public School Locations, 2007



Some of the demographic differences detailed in Tables 1.1 and 1.2 are a result of the neighborhoods that charter schools are located in. This is because neighborhoods are fairly segregated, and distance has a large impact on which school a student attends, (Dinerstein & Smith (2015)). Network charters entered school districts that were 47% black and 39% Hispanic on average, while independent charters entered districts that were 35% black and 48% Hispanic. Like their student bodies, these ratios are flipped, and both are higher than the citywide average.

However, neighborhood location does not fully explain demographic differences. Figures 1.9 and 1.10 show how race at independent and network charters compare to the closest three schools with the same grade levels. The dashed lines represent the average. The x-axis

Figure 1.7: Charter and Traditional Public School Locations, 2015

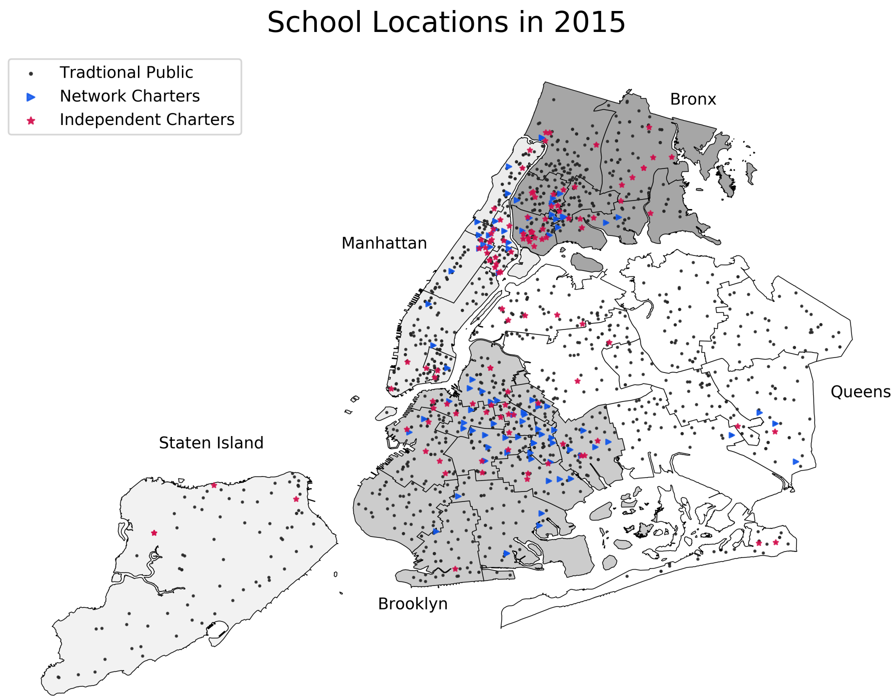
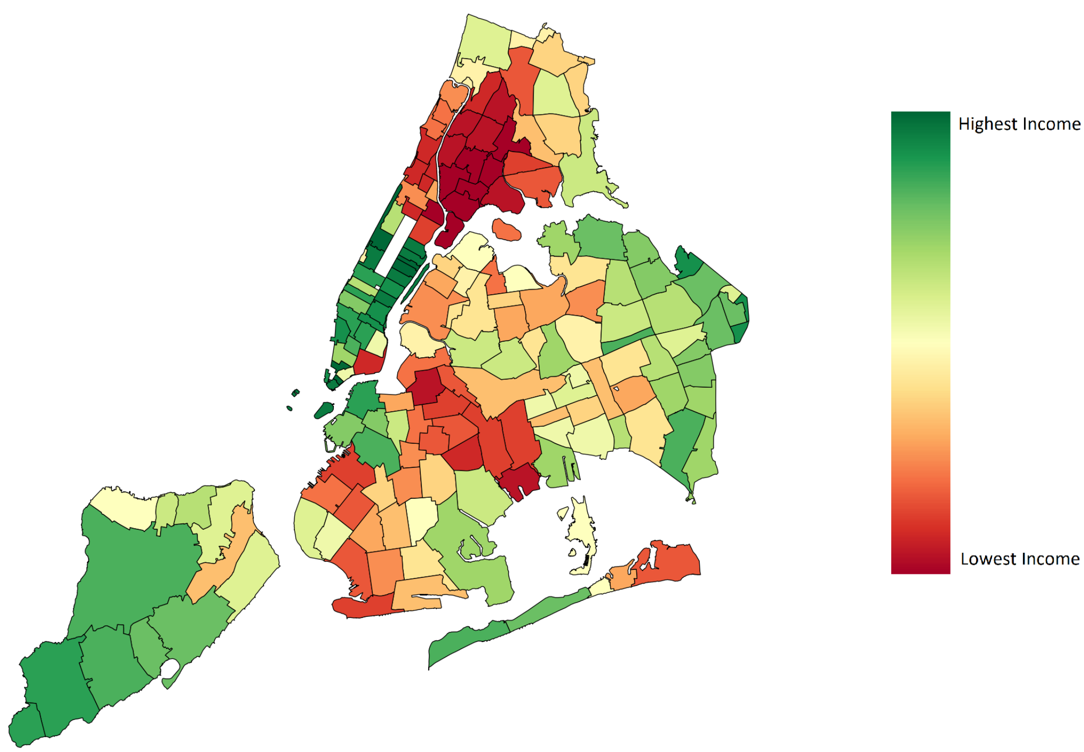


Figure 1.8: Zip Code Level Median Household Income Percentiles, 2010



in Figure 1.9 is the percent of students who are black at a charter school minus the percent of students that are black at the closest three schools. Figure 1.9 shows that most charters enroll more black students than nearby schools, especially for network charters. The average CMO charter enrolls 11 percentage points more black students than nearby schools. At the same time, charters actually enroll fewer Hispanic students than the nearest schools, especially for CMO charters.

Figure 1.9: Distribution of the Difference in Percent Black vs. Nearby Schools, 2015

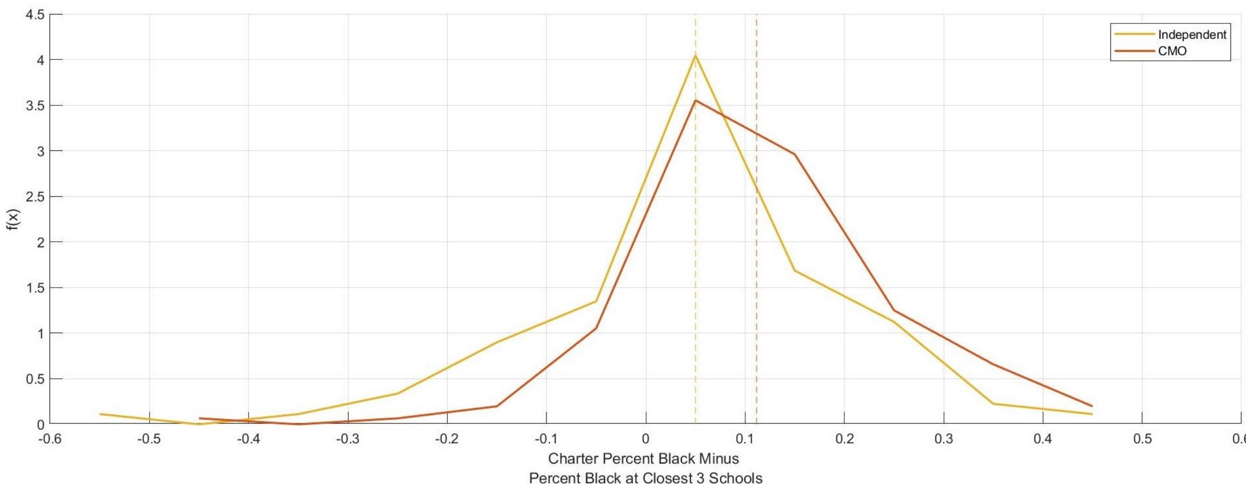
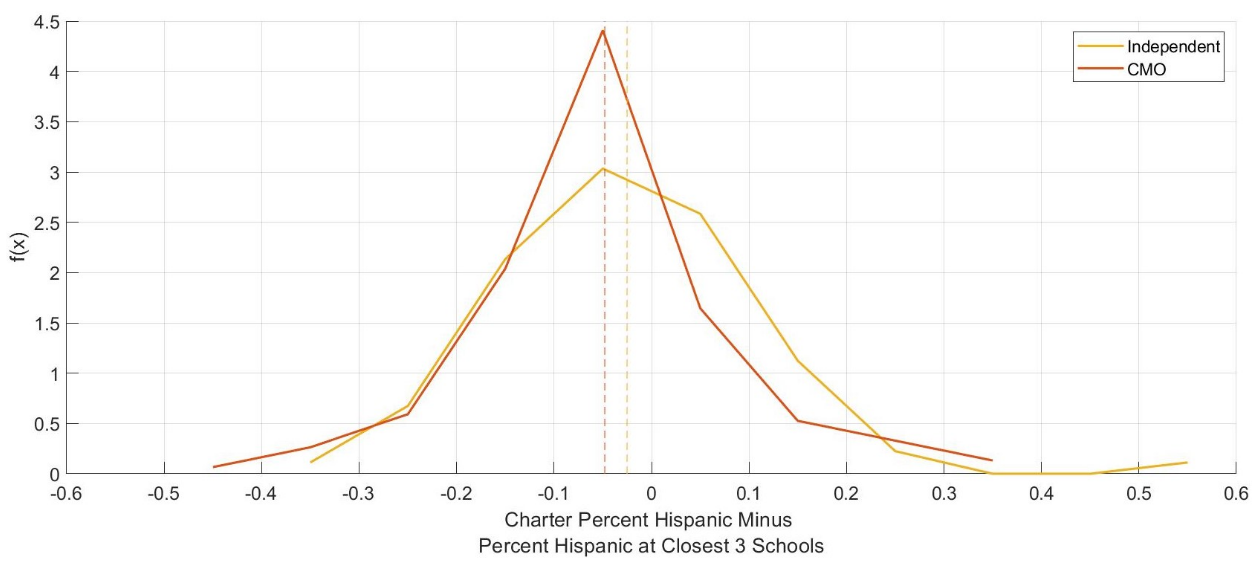


Figure 1.10: Distribution of the Difference in Percent Hispanic vs. Nearby Schools, 2015



1.4 Empirical Strategy

1.4.1 Methods of Measuring School Quality

The fist step I need to take to examine whether charter networks have been able to scale up their model is to measure school quality. My measure of school quality is test score value-added. Ideally, what I want to know is, what improvement in test scores will a school s contribute, relative to a hypothetical "average" public school[12]. We can represent this as:

$$Y_{is} - Y_{i0} = q_s I\{i \to s\}$$

where $I\{i \to s\}$ is an indicator for attending school s, Y_{is} is student i's test score s while Y_{i0} is how they would have done at a hypothetical average school. Here, q_s is the test score value-added of attending school s. Since we do not observe Y_{i0}, we can break it into observed $X_i'\beta$ and unobserved components ϵ_i:

$$\begin{aligned} Y_{is} &= Y_{i0} + q_s I\{i \to s\} && (1.1) \\ &= X'\beta + q_s I\{i \to s\} + \epsilon_i && (1.2) \end{aligned}$$

Estimating this equation will be problematic if the unobserved ability/characteristics of each student are correlated with the decisions to attend each school.

In the charter school literature, there are two main methods for dealing with this problem and estimating school quality: admissions lotteries and lagged test scores. Both of these methods have pros and cons. The main advantage of using admissions lotteries is that the students are chosen completely randomly. Charter schools that are oversubscribed, i.e. have more applicants than available seats, are required by law in most states to run random lotter-

12. To be precise, a hypothetical enrollment-weighted "average" public school.

ies. This means that within the pool of applicants, we do not have to worry about selection issues. However, it means that we are estimating a local average treatment effect, which might not be the same as the average treatment effect. In addition, the set of schools that we are able to calculate quality for is positively selected. Not all charter schools are oversubscribed, making it impossible to calculate quality for those schools. Intuitively, it would make sense if better schools were more likely to be oversubscribed. Indeed, Abdulkadiroğlu et al. (2011) found that schools that do not run lotteries are typically worse than those that do. Therefore, I would have to drop a negatively selected sample of schools which could lead to incorrect conclusions.

The next method is to use lagged test scores to control for unobservables. This method will give a causal estimate of school quality as long as selection into schools is only based on observables (which includes past test scores). The main benefit of this method is that I can calculate school quality measures for every school in every year of my sample. The main drawback is that selection into schools might be based on some unobservable characteristics that also affect test scores at time t. While there are certainly many unobserved characteristics that affect test scores at time t, these factors likely affect test scores in prior periods as well, meaning that lagged test scores should control for these effects. For example, while I cannot observe parental time investment, which likely affects a student's test scores at t, it also likely affects their scores in prior periods as well.

Whether lagged test scores adequately control for selection is a question that has received a lot of attention in the education literature. A number of papers have attempted to validate value-added measures as a part of their study. These papers take the subset of schools that run lotteries and compare those quality estimates to the estimates calculated using lagged test scores. Abdulkadiroglu et al. (2011) do this in Boston and describe the results from lotteries and lagged test scores as "remarkably similar." Dobbie and Fryer (2013) study New York City and find generally similar results as well. Angrist et al. (2017) focuses entirely on the question of bias in school value-added measures. While they do find some bias in lagged test score measures, they also find that these estimates are "highly correlated with school effectiveness" as measured using lotteries.

Two other methods that are used less frequently are matching or student level fixed effects. Matching relies on finding students with similar demographics that went to the same school at a lower level and comparing their outcomes. However, in New York City, more than sixty percent of charter schools start in kindergarten, so I would not be able to estimate quality for a large number of schools in my sample. In addition, some unobservable caused these students to attend different schools, and it might also affect test scores at time t, so it is not clear this would reduce selection bias. Student level fixed effects are rarely used because they are identified only by students that switch schools. Switching schools is not random and might be caused by a big family change, like a parent switching jobs, that would also affect test scores. In addition, some of the best schools have very few students switch out of them and into another New York City public school, making it impossible to estimate quality for those schools.

1.4.2 Specification

In this paper, my measures of school quality are year by year test score value-added for math and English. To calculate school value-added, I estimate the following equation separately by subject and school level:

$$y_{isbt} = \beta X_{it} + \alpha \vec{y}_{i,t-r} + q_{sbt} + u_{isbt} \tag{1.3}$$

where y_{isbt} is the test score for student i in subject b in school s at time t, X_{it} is a rich set of demographic variables, $\vec{y}_{it-r}$ is a vector of prior test scores in both English and math, and q_{sbt} is the school-by-year fixed effect for school s. This fixed effect estimate q_{sbt} is what I will be referring to when I talk about a school's value-added quality. Both English and math are included in $\vec{y}_{it-r}$ regardless of the dependent variable. At the elementary and middle school level, $\vec{y}_{i,t-r}$ includes the previous two years. For high school students, because there is not a specific grade that students have to take each Regents exam, $\vec{y}_{i,t-r}$ includes the last two years of test scores a student took before entering high school[13].

1.4.3 Which Networks Have Expanded?

Next, I want to determine whether higher quality networks have expanded more. I look at the correlation between expansion and quality by estimating the following:

$$\Delta E_n = \alpha_0 + \alpha_1 q_{\overline{nm,2007}} + \alpha_2 q_{\overline{nE,2007}} + \epsilon_n \tag{1.4}$$

where ΔE_n is the change in network n's enrollment from 2007 to 2015, $q_{\overline{nm,2007}}$ and $q_{\overline{nE,2007}}$ are the enrollment-weighted average value-added of schools in network n for math and English respectively in 2007, defined as:

$$\overline{q_{nbt}} = \frac{1}{E_n} \sum_{s \in n} E_s \hat{q}_{sbt}$$

and ϵ_n is the error term. The change in enrollment ΔE_n is defined as:

$$\Delta E_n = \frac{(E_{n,2015} - E_{n,2007})}{0.5(E_{n,2015} + E_{n,2007})}$$

13. This would be their seventh and eighth grade test scores if students never repeated grades. In addition, sometimes students get promoted to high school without passing eighth grade if they have had to previously repeat multiple grades.

where E_{nt} is enrollment at t. The benefit of using this measure for changes in network size as opposed to a simple percent change is that the values are bounded by -2 and 2. Using a simple percent change would result in a long right tail of values. One issue with using this transformation is that the values are somewhat harder to interpret. As a reference, $\Delta E_n > 0$ means the network grew, $\Delta E_n = 1$ means the network tripled in size, and $\Delta E_n = 2$ means the network had infinite growth, i.e. entered during the time period.

1.4.4 Within Network Dynamics

Last, I examine how within network quality has evolved as networks have expanded. To do this, I regress estimated school value-added on age, a school's ordinal number in its network, and network by subject by year fixed effects:

$$\hat{q}_{sbt} = \psi_1 A_{st} + \psi_2 R_{sn} + \gamma_{nbt} + I\{HS\} + \mu_{sbt} \tag{1.5}$$

A_{st} is the age of school s at time t, R_{sn} is a bin capturing a school's ordinal number in the network, γ_{nbt} is a network by subject by year fixed effect, I{HS} is an indicator for whether or not s is a high school, and μ_{sbt} is the error term. To clarify R_{sn}, if a school is the fifth school opened by a network, the indicator for the four and five bin ($I\{\#4-5\}$) would equal one. If two schools in the same network are founded in the same year, then I used the same value for R_{sn}. Because there is no reason to think that the effect is linear and because networks often open more than one school in a year, my main specification uses bins for R_{sn} as opposed to making it a linear term. When running this regression, I only include network charters.

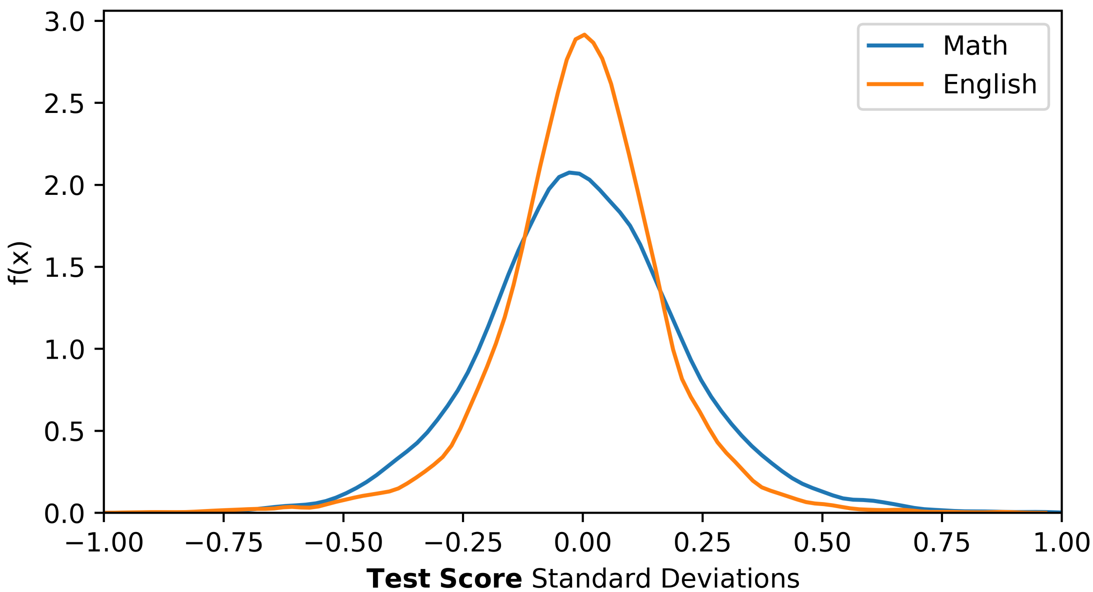
1.5 Contextual Quality Findings

Figure 1.11 shows the distribution of the estimated $\hat{q}_{sbt}$'s. The x-axis units are in test-score standard deviations. As we can see from the graph, there is slightly more variance in estimated math quality than English. One standard deviation in school math quality is roughly equal to 0.2 test score standard deviations: $\sigma_{school} \approx 0.2\sigma_{test}$. This relationship is a little closer to 0.175 for English. The "average" public school would be equal to zero if all schools had the same enrollment. Obviously, they do not, but the quality of the average public school still ends up being very close to zero. The correct way to think about these quality values is: imagine two students, i and j, that are identical at $t-1$. At time t, i attends school s where $\hat{q}_{sEt} = 0.1\sigma$ and student j attends school s' where $\hat{q}_{s'Et} = 0.2\sigma$. In expectation, school s' should raise student j's English test scores 0.1σ more than school s raises student i's scores at time t.

To put the size of these estimated $\hat{q}_{sbt}$'s in context, if an education intervention, such as lowering student-teacher ratios, raised test scores by 0.1 standard deviations, it would be considered a success. At the same time, relative to other measures, such as the black-white test score gap, 0.1 test score standard deviations would be very small. Throughout my sample period, the gap between the average white student's test scores and the average black student's test scores was about 0.75 standard deviations for both English and math respectively. Appendix Figure A.2 presents a plot of the black white test gap throughout my sample.

Table 1.4 presents a summary of the $\hat{q}_{sbt}$'s by school type. The interpretation of the (Charter - TPS) value on the left side of Table 1.4 is that in expectation, charter schools raised a student's math test score by 0.158 test score standard deviations in 2007 relative to a school with

Figure 1.11: Distribution of $\hat{q}_{sbt}$ for *All* Public Schools



$\hat{q}_{sbt} = 0$. By 2015, this had fallen to 0.121 σ_{test}. Looking at the right side of the table, it is clear that most of the decline in math quality from 2007 to 2015 came from network charters.

For English, the gains from charter schools are smaller, but have improved slightly over the sample period. In 2007, charter schools raised English test scores by 0.034 σ_{test} in expectation, which rose to $0.055\sigma_{test}$ by 2015. Looking at the right side of the table, we can see that although network charters were initially more than $0.1\sigma_{test}$ better than independent charters, they are now slightly worse. Figures 1.12 and 1.13 show a time series of the enrollment weighted average $\hat{q}_{sbt}$'s for English and math by sector[14]. As we can see from these graphs, independent charters have narrowed the gap to network charters in both English and math. Since test scores are normalized at the city level, the slight decline in network math value-added should be interpreted as a narrowing of the gap between network charters and traditional public schools. This could mean that network schools got worse, traditional

14. The graph doesn't quite match the table because it is an enrollment weighted average, while the table is a simple average across schools.

Table 1.3: Percentiles for School-Level Test Score Value-Added, 2015

Percentile	English	Math
10th	−0.16	−0.24
25th	−0.07	−0.12
50th	0.03	0.00
75th	0.13	0.13
90th	0.25	0.26

public schools got better, or some combination of the two.

Table 1.4: Summary of School Value-Added by School Type

	TPS	All Charters	(Charter–TPS)	Indep.	Netw.	(Netw.–Indep.)
2007						
English	0.016	0.050	0.034	-0.013	0.093	0.106
	(0.164)	(0.200)		(0.163)	(0.214)	
Math	-0.009	0.149	0.158	0.029	0.241	0.212
	(0.199)	(0.274)		(0.222)	(0.278)	
N	1336	44		18	26	
2015						
English	0.032	0.087	0.055	0.095	0.081	-0.015
	(0.173)	(0.178)		(0.205)	(0.155)	
Math	-0.006	0.115	0.121	0.024	0.181	0.157
	(0.198)	(0.253)		(0.254)	(0.232)	
N	1550	194		82	112	

Standard deviation of value-added estimates in parentheses

1.6 Results

1.6.1 Heterogeneity in Network Quality

Using these school level quality estimates, I can now examine charters at the network level. Figures 1.14 and 1.15 show the enrollment weighted average quality of each network in 2015.

Figure 1.12: Average Math School Value-Added by Charter Classification

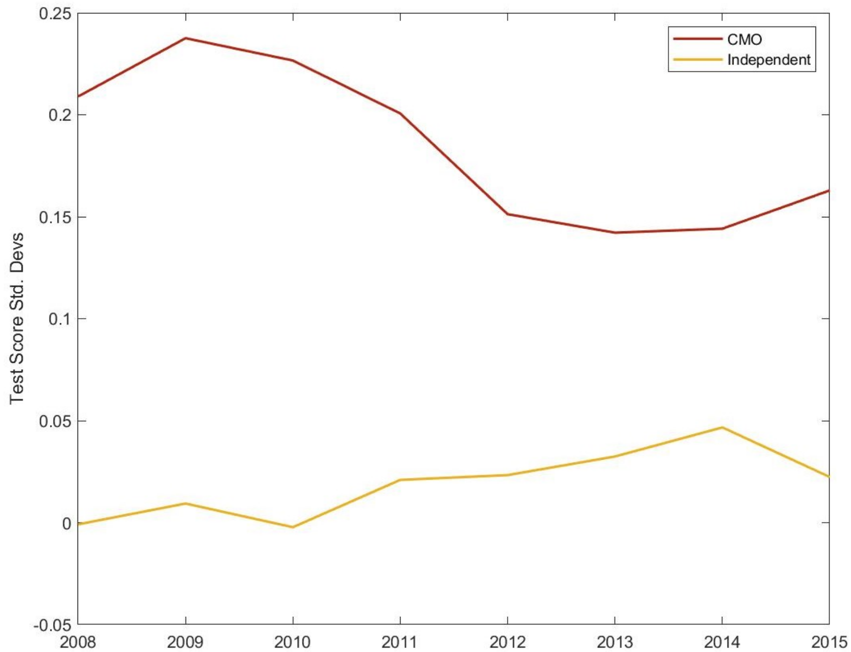
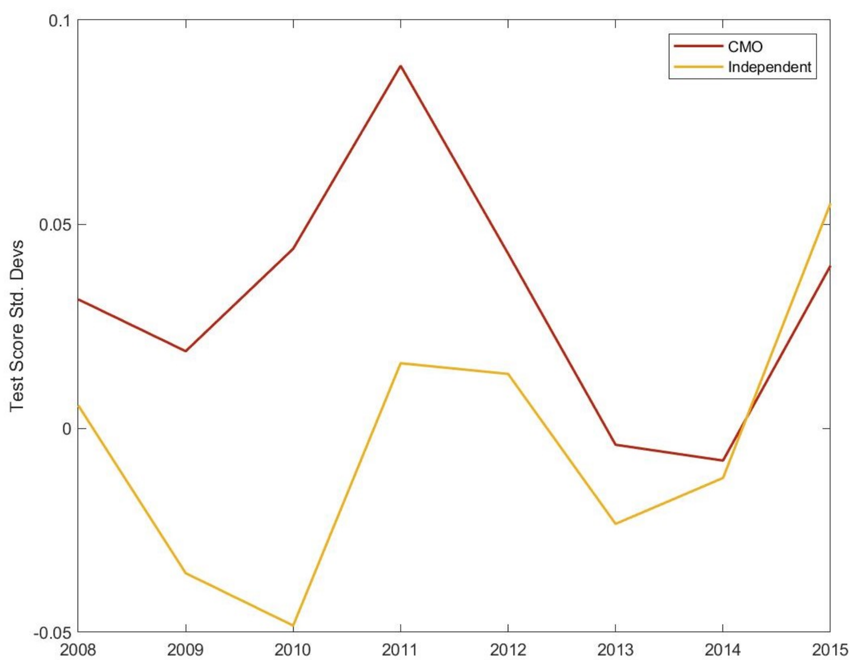


Figure 1.13: Average English School Value-Added by Charter Classification



They reveal significant heterogeneity in both English and math. For English, a hypothetical average school in the best networks would be in the top decile of all public schools, while an average school in the worst networks would be in the bottom decile. Overall, most networks are slightly above the average New York City public school. For math, the distribution of networks is similar, except the top of the distribution goes much higher. Six networks average value-added is higher than the 90^{th} percentile public school, and the best two networks are above the 98^{th} percentile.

1.6.2 Network Expansion

With such a wide range in quality across networks and government policies that encourage high quality networks to expand, it is important that the networks that are expanding are actually the high quality ones. Table 1.5 shows the results of my estimation of Equation (1.4) regarding this expansion. I estimate this equation with English and math value-added as separate regressors in column (1), and with only the mean value-added in column (2). There are a couple of key findings. First, the coefficient for English value-added quality is close to zero and not statistically significant, while the math coefficient is large and statistically significant. The interpretation of this coefficient is that a network that is one school level standard deviation ($\sigma_{school} \approx 0.2\sigma_{test}$) better than the average network in math in 2007 will grow an additional thirty percentage points. Although the coefficient for English value-added is close to zero, there is a positive correlation between network math and English value-added ($Corr(q_{\overline{nm,2007}}, q_{\overline{nE,2007}}) = 0.74$), so there is still an unconditional correlation between expansion and English quality. However, conditional on a given level of math quality, English does not have a significant relation to expansion.

Figure 1.16 shows a scatter plot of network average quality in English versus math, with both the size and color of the points representing the amount of growth. A large green dot

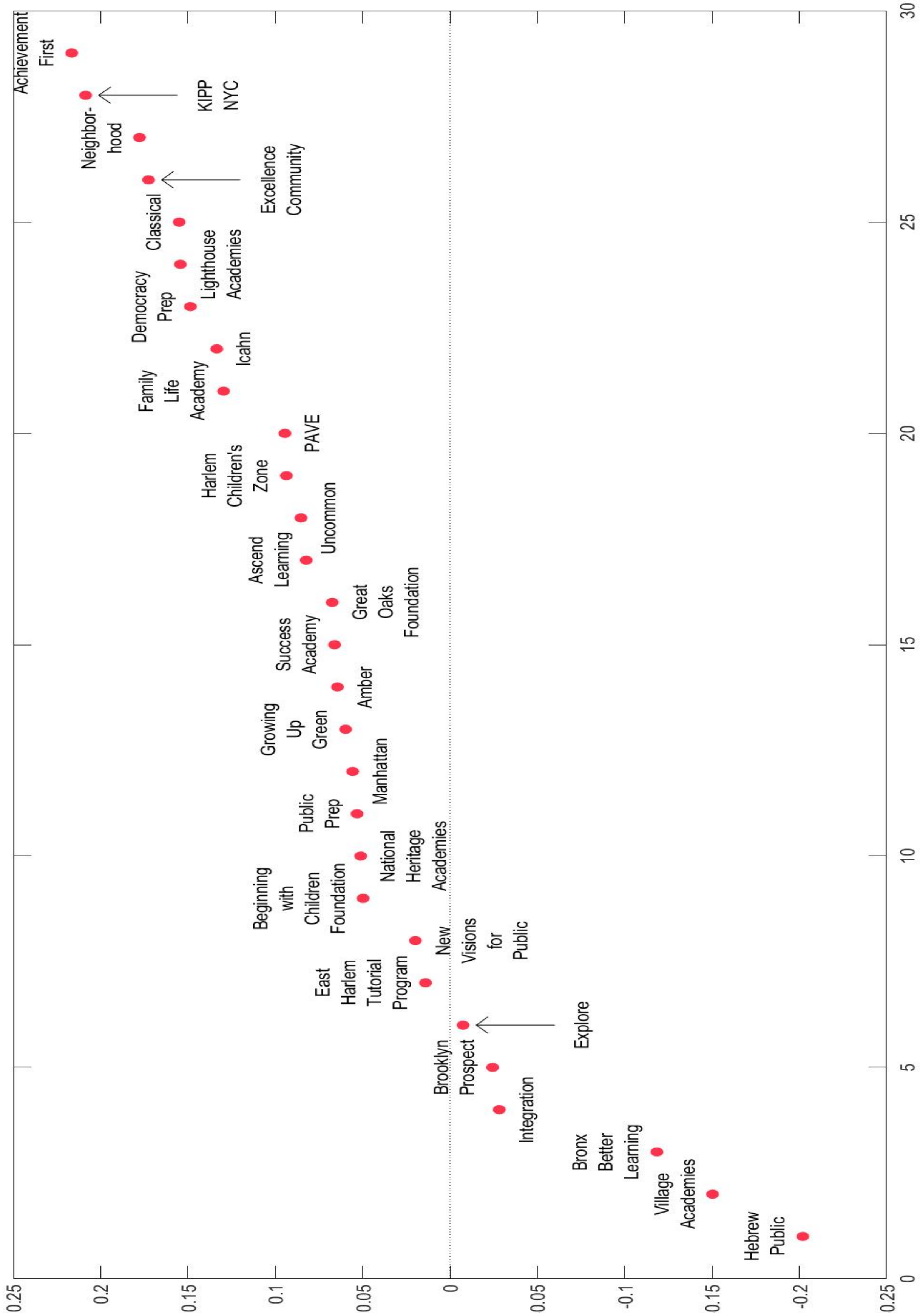


Figure 1.14: English – Network Average Value-Added, 2015

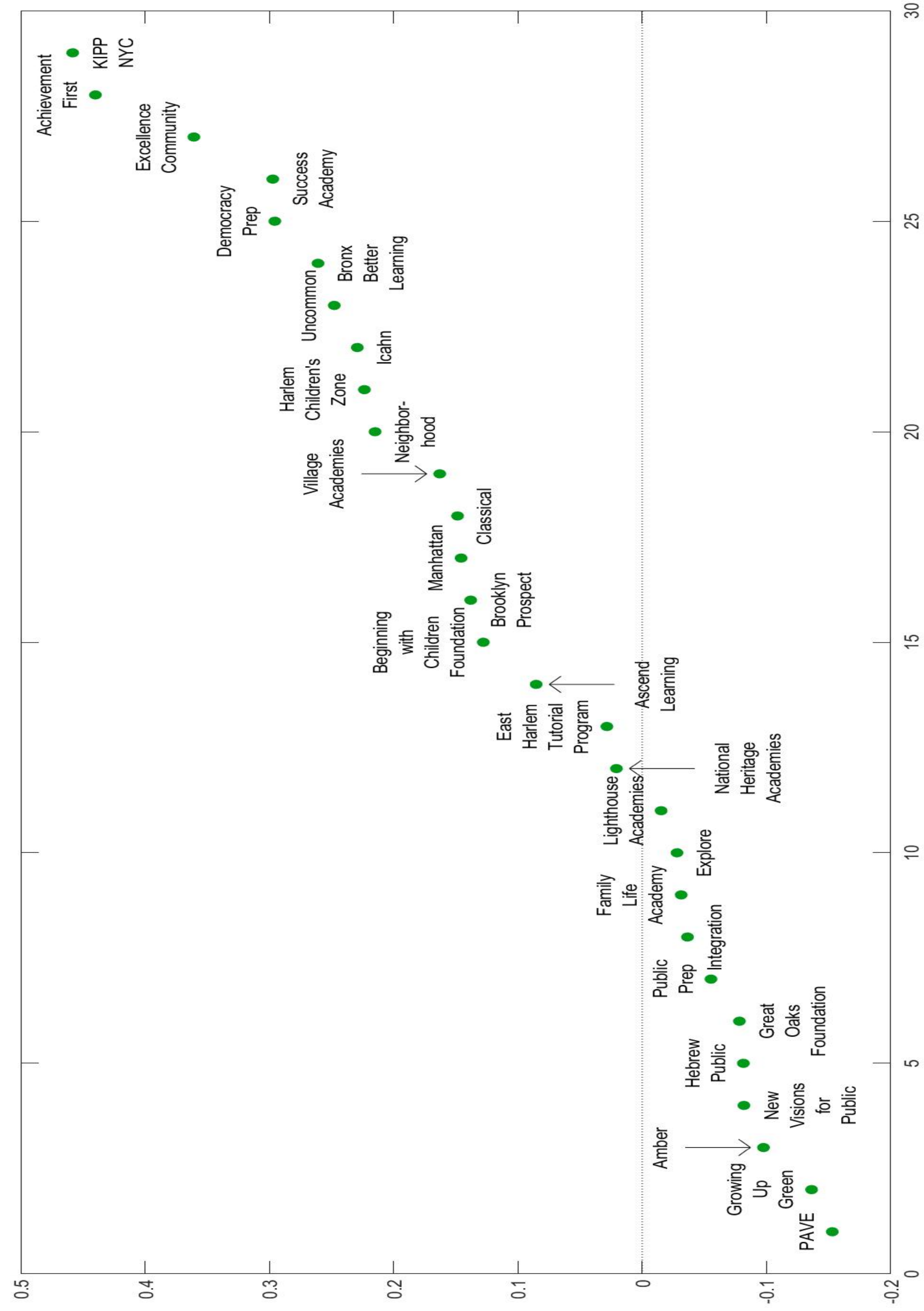


Figure 1.15: Math – Network Average Value-Added, 2015

represents a network that grew a lot, while a small red dot represents a network that grew very little. One thing to note is that the network that grew the least still grew over this time period.

Table 1.5: Expansion and Network Quality in 2007

Dependent Variable: ΔE_n	(1)	(2)
Constant	0.649***	0.709***
	(0.169)	*(0.150)*
$q_{\overline{nE},2007}$ Network English Value-Added	-0.125	
	(1.315)	
$q_{\overline{nE},2007}$ Network Math Value-Added	1.562*	
	(0.915)	
Network Value-Added Mean of English & Math, 2007		1.782**
		(0.765)
N	15	15

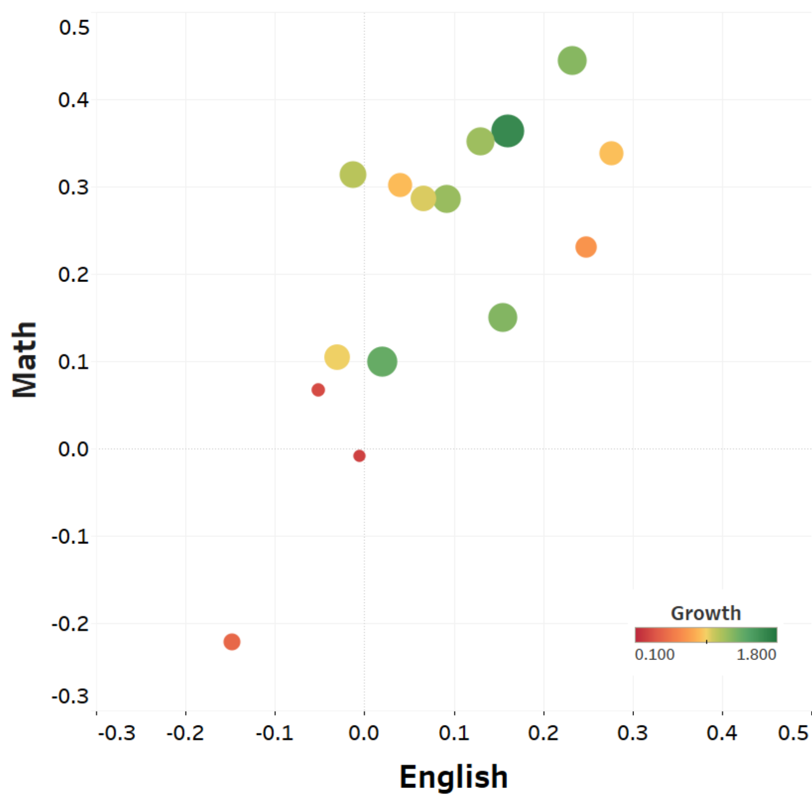
Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

1.6.3 Within Network Changes in Quality

I next look at how quality has changed within a given network to determine if charter networks have maintained consistent levels of quality with expansion. Table 1.6 presents the results of the regression presented in Equation (1.5). These regressions only include network charters since independent charters do not have multiple schools. In columns (2) and (4), I used an integer for "number in network," meaning if a school is the third in a network, this variable would equal three. In regressions (1) and (3) I used an indicator to place schools into various bins since there is no reason to assume that the effect of expansion would be linear.

Figure 1.16: 2007 Network Average Math Quality vs. English Quality, with Growth 2007-2015



As we can see from Table 1.6, as networks expanded, English value-added quality declined. Since I am including network-by-year fixed effects, the coefficient on I{#2-3} should be interpreted as: the value-added of the second and third school will be $-0.0422\sigma_{\text{test}}$ lower than the first school in that network in that year in expectation. Quality decreases monotonically with the I{#4-5} and I{#>=6} bins. One standard deviation in the distribution of school quality is equal to 0.175 test score standard deviations for English ($\sigma_{school} \approx 0.175\sigma_{test}$). This implies that the sixth school in a network is about one-third of a standard deviation lower in the distribution of schools than the first school in that network. Another thing to note is that the coefficient on age is negative and statistically significant. Over time, this effect will be fairly large. This coefficient implies that a school that was open at the start of my sample will fall $0.5\sigma_{\text{school}}$ in expectation by the end. For math, the results are qualitatively similar, although much smaller, and I cannot reject the null hypothesis that later schools in a network have the same quality as earlier schools. Neither can I reject the null that school value-added does not change with age. Since schools that opened earlier in a network are necessarily older, it could be the case that in a given year, later schools are not much worse. However, later schools in the network will be starting from a much lower level, and then declining as earlier schools in that network do.

Robustness Checks

For robustness, because the measurement of high school test score value-added introduces some complications (see Section 3), I also run this regression restricting my sample to elementary and middle schools. The results are very similar, and in fact, the declines in math quality more closely match the English declines (Appendix Table A.1).

In addition, I also estimated this equation including age nonparametrically. Appendix Figure A.4 and A.5 show a plot of the coefficients for each age bin when estimated nonparametrically. These coefficients show a generally similar pattern to the linear specification.

Because a negative coefficient for the age of a school is somewhat surprising, I also examined the relation between age and quality for independent charters to see if I found similar results. Table 1.7 shows that for independent charters, English value-added *increases* with age, although this coefficient is not statistically significant when year fixed effects are also included. The coefficient for age is also positive but insignificant for math.

Table 1.6: Within Network Changes in Value-Added

	English		Math	
	(1)	(2)	(3)	(4)
Age	-0.0117**	-0.00712*	-0.00648	-0.00363
	(0.00456)	(0.00413)	(0.00593)	(0.00541)
EMS	0.0136	0.0133	0.0771**	0.0746**
	(0.0218)	(0.0218)	(0.0334)	(0.0333)
I{#2-3}	-0.0422*		0.00976	
	(0.0225)		(0.0295)	
I{#4-5}	-0.0691*		-0.0233	
	(0.0386)		(0.0509)	
I{#>=6}	-0.0756		-0.0290	
	(0.0472)		(0.0621)	
Number in Network		-0.00451		0.00219
		(0.00630)		(0.00844)
Network-Year FEs	Yes	Yes	Yes	Yes
Observations	598	598	590	590

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 1.7: Independent Charters and Age

	English		Math	
	(1)	(2)	(3)	(4)
Age	0.0108**	0.00878	0.0088	0.0086
	(0.0046)	(0.0066)	(0.0055)	(0.0081)
School FEs	Yes	Yes	Yes	Yes
Year FEs	No	Yes	No	Yes
N	380	380	380	380

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

1.7 Mechanisms

As discussed in Davis et al. (2017) and Al-Ubaydli et al. (2017), when social programs are scaled up from the pilot phase, declines in quality can occur as a result of either demand side or supply side mechanisms.

1.7.1 Demand Side

One possible mechanism that could explain the declines in quality that I observe is that there are large changes to the student body. It is easy to imagine that the students that attend charter schools first are the ones who benefit the most. As a network ages and/or expands, it is possible that later students benefit less and less. Since charters only made up ten percent of the market by the end of my sample, and charter and traditional public school demographics are very different, there is a lot of room for the student body to change. A necessary condition for this explanation is that there are heterogeneous returns to charter schools. Other papers, such as Abdulkadiroglu et al. (2011), have shown that poor and minority students benefit more from charter schools, as well as students who have previously done poorly in school. Angrist et al. (2012) find heterogeneous returns and determine that differences in student demographics account for more than half of the difference between the value-added of urban charter schools and non-urban charters in Boston.

My first step is to verify that there are heterogeneous returns to charter schools in New York City, as in other settings. To do this, I will follow the method used in Angrist et al. (2012) and Angrist et al. (2017), which is to re-run my quality estimations using only students from the subgroup that I am interested in. Table 1.8 presents the average difference between school value-added estimates from estimating Equation (1.3) using only the subsample of interest versus using the entire sample. Each column presents the difference from a separate subgroup regression. My results suggest that black and Hispanic students benefit more from charters, as well as those with below average test scores at $t-1$[15]. These effects are generally larger in math than English. I also find that English language learners benefit more from charters in English, with a smaller impact for math. My results generally line up with the rest of the literature. The notable exception is that I find that students who qualify for free lunch or other public assistance do not seem to benefit more from charters, while other papers find that they benefit more. Since measures of poverty vary across cities, this doesn't necessarily indicate a conflict.

Table 1.8: Heterogeneous Returns to Charter Schools

	Black	Hispanic	Poverty	ELL	Below Average English $t-1$	Below Average Math $t-1$
Math						
Network	0.040	0.024	-0.006	0.013	0.054	0.059
Independent	0.032	0.013	-0.005	-0.006	0.030	0.043
English						
Network	0.018	0.015	-0.003	0.078	0.037	0.027
Independent	0.020	0.004	0.003	0.013	0.033	0.033

15. For high school students $t-1$ is actually the last test they took in middle school.

My next step is to see if demographics are changing in a way that matches the declines in quality that I observe. To do this, I estimate the same relationship as in Equation (1.5), except I use the demographic of interest as the dependent variable:

$$D_{st} = \theta_1 A_{st} + \theta_2 R_{sn} + \kappa_{nt} + I\{HS\} + \xi_{st} \tag{1.6}$$

D_{st} is the demographic of interest for school s at time t, A_{st} is the age, R_{sn} is the bin for the ordinal number in the network, κ_{nt} is a network-by-year fixed effect, and I{HS} is an indicator for whether s is a high school. The results are presented in Table 1.9.

Table 1.9: Within Network Changes to Demographics

	Black	Hispanic	White	Asian	Pub. Assist.
Age	0.0041	-0.0037	-0.0002	-0.0006	-0.0012
	(0.0046)	(0.0043)	(0.0012)	(0.0004)	(0.0023)
I{High School}	0.0143	-0.0025	-0.0048	-0.0018	-0.0646***
	(0.0212)	(0.0199)	(0.0054)	(0.0020)	(0.0105)
I{#2-3}	-0.0756***	0.0675***	0.0040	0.0020	0.0212*
	(0.0236)	(0.0222)	(0.0060)	(0.0022)	(0.0117)
I{#4-5}	0.0341	-0.0402	0.0017	0.0006	0.0209
	(0.0397)	(0.0374)	(0.0101)	(0.0037)	(0.0197)
I{#>=6}	-0.0225	-0.0290	0.0313**	0.0101**	-0.0040
	(0.0480)	(0.0452)	(0.0123)	(0.0045)	(0.0238)
Network-Year FEs	Yes	Yes	Yes	Yes	Yes
N	836	836	836	836	836

Standard errors in parentheses, * $p < .1$, ** $p < .05$, *** $p < .01$

Although there are some changes to demographics, there are no patterns to these changes, and certainly none that match the declines that we observe in quality. For example, the second and third schools in a network tend to have fewer black students and more Hispanic students than the first school in that network in a given year, however, this pattern does not continue with later schools. The correct way to interpret the coefficient on $I\{\#2-3\}$ is that the second and third school in a network have 7.5 percentage points fewer black students than the first school in that network in a given year.

Students with low prior test scores or who are English language learners also benefit more from charter schools. I have to analyze changes to these characteristics with expansion in a slightly different matter. Students have to *have* prior test scores in order for me to look at prior test scores. As a result, I have to drop many of the schools in my sample. In addition, I only use the incoming cohort of students to a given charter school. The English language learner classification is something that can change with time, so I will also only use incoming grades for those regressions.

Because I have to drop a number of schools, I am no longer able to estimate equation (1.6). Instead, I estimate a similar equation, except I use network and year fixed effects separately as follows:

$$\overline{y_{sb,t-1}} = \theta_1 A_{st} + \theta_2 R_{sn} + \omega_{nb} + \Omega_t + \delta_{st} \tag{1.7}$$

$\overline{y_{sb,t-1}}$ represents the average test scores at $t-1$ of the incoming cohort, ω_{nb} is the network fixed effect, and Ω_t is the year fixed effect. The results are presented in table 1.10. There do not seem to be any changes to the prior test scores of incoming cohorts with later schools in the same network. It does seem that the prior test scores of incoming cohorts are falling with age and the fraction of English language learners is increasing. However, since these students typically benefit more from charter schools, this suggests my negative age coefficient might be biased upwards.

Table 1.10: Within-Network Changes, Incoming Cohort Only

	Math Scores	English Scores	English Language Learner
Age	-0.0191***	-0.0162***	0.0100***
	(0.0057)	(0.0050)	(0.0015)
I{#2-3}	0.0764	0.0236	0.0285**
	(0.0493)	(0.0431)	(0.0122)
I{#4-5}	0.0227	0.0337	0.0104
	(0.0605)	(0.0529)	(0.0150)
I{#>=6}	-0.0788	-0.0830	-0.0004
	(0.0755)	(0.0660)	(0.0188)
Network FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
N	165	165	163

Standard errors in parentheses, * $p < .1$, ** $p < .05$, *** $p < .01$

1.7.2 Supply Side

There are a few supply side mechanisms that could account for the patterns that I find, which all have similar policy implications.

Span of Control

Lucas (1978) and Rosen (1982), as well as many others since, postulated that different firm managers have different abilities to manage large numbers of subordinates. A part of this theory was that firms become more and more difficult to run as they grow larger. This literature mostly attempts to relate manager wages to the number of subordinates they oversee and their average productivity, but I can extend this idea to charter school networks as well. As a charter network expands, it may become more and more difficult to manage. It is easy to imagine that a network with ten or twenty schools is harder to manage than one with only three or four. Using a similar setup it is easy to show that if the span of control of managers is an issue for network expansion, then quality at earlier schools in the network should decline with expansion.

Span of Control Model Setup:

- Charter network n has a manager m_n with ability a_m drawn from distribution $M(\cdot)$.
  - Network managers oversee the set of schools $s \in n$ where the number of schools in the network is denoted S_n.
  - Manager skill is not divisible. Two managers with skill $a_m/2$ will not result in the same output as a manager with skill a_m.
- Each school has a principal p_s with ability a_p drawn from a distribution $P(\cdot)$.
- Managers have a fixed amount of time T which they can split up however they want across schools in their network: $\sum_{s \in n} t_s = T$.

Let $k(t_s, a_m, p_s)$ be an aggregate educational production function for each school. To make this model as flexible as possible, I will not specify a functional, but will only make two assumptions.

Assumptions:

1. $k_1(\cdot, \cdot, \cdot) > 0$
2. $\lim_{t_s \to 0^+} k_1(t_s, a_m, p_s) = +\infty$.

The first property implies that outcomes are increasing in manager ability and manager time spent on that school. The second rules out that a manager will spend no time with any of the schools in the network. One could imagine that if a manager spent absolutely no time on a school, it would not even be able to complete basic functions like securing its funds. The optimization problem for the network manager becomes:

$$\max_{(t_1,t_2,....,t_S)} \left(\sum_{s\in n} k(t_s, a_m, p_s) + \lambda(T - \sum_{s\in n} t_s) \right)$$

where λ is a Lagrange multiplier.

The first order conditions imply that:

$$T = \sum_{s\in n} t_s$$

and:

$$k_1(t_s, a_m, p_s) = \lambda, \forall s \in n \tag{1.8}$$

We can set condition (1.8) for all schools equal to each other:

$$k_1(t_1, a_m, p_1) = k_1(t_2, a_m, p_2) = ... = k_1(t_S, a_m, p_{S_n})$$

This model gives me a test for span of control issues. If the span of control of the manager is a problem in the context of charter school networks, then as a network expands, all schools that previously existed should decline in quality. The proof is as follows:

Suppose network n expands and adds one school so that it now has $S_n + 1$ schools. I will denote the manager's time choices for each school now by t'_s. To differentiate the network in the two periods, I will use n_t and n_{t+1}. The manager's maximzation problem becomes:

$$\max_{(t'_1,t'_2,....,t'_{S_n+1})} \left(\sum_{s\in n_{t+1}} k(t'_s, a_m, p_s) + \lambda'(T - \sum_{s\in n_{t+1}} t'_s) \right)$$

The first order conditions imply:

$$k_1(t'_1, a_m, p_1) = k_1(t'_2, a_m, p_2) = ... = k_1(t'_{S_n+1}, a_m, p_{S_{n+1}})$$

and:

$$T = \sum_{s \in n_{t+1}} t'_s$$

.

Given the properties of $k(\cdot, \cdot, \cdot)$, there are two conclusions that we can make about this new set of time choices.

1. $t'_s > 0, \forall s \in n_{t+1}$

2. $t'_s < t_s, \forall s \in n_{t+1}$

The first conclusion is straightforward. By the Assumption 2, the derivative goes to infinity as the time spent at that school goes to zero. Thus, it must be the case that $t'_s > 0$.

The second conclusion is also easy to prove. Since T is the same in both scenarios, we know that $\sum_{s \in n_t} t_s = \sum_{s \in n_{t+1}} t'_s$. If we couple that with the fact that $t'_{S_n+1} > 0$, we know that:

$$\sum_{s=1}^{S_n} t'_s < \sum_{s=1}^{S_n} t_s$$

Thus, we know that at least one $t'_s < t_s$. Let's refer to this hypothetical school as $s = \ell$. Then we also know that:

$$k_1(t_\ell, a_m, p_\ell) < k_1(t'_\ell, a_m, p'_\ell)$$

because of our first assumption – i.e. $k_1(\cdot,\cdot,\cdot) > 0$.

Coupled with the FOCs of the manager's maximization, if $k_1(t'_s, a_m, p'_s)$ is smaller for one school, it must be smaller for all previously existing schools in the network. Since a_m and p_s are fixed, this must mean that $t'_s < t_s$ for all schools that were in the network in both periods. Because the aggregate educational output increases with manager time, a decrease in time for those schools should cause a decline as well.

Thus, if span of control is a significant issue, I should find that the quality of all the pre-existing schools will decline with expansion. In order to test this, I will look at how school quality changes with the addition of new schools in the network. I will estimate the following:

$$\hat{q}_{sbt} = \omega_s + A_{st} + \delta E_{nt} + \mu_{sbt} \tag{1.9}$$

where ω_s is a school fixed effect, A_{st} is the school's age at time t, E_{nt} is the total enrollment at network n at time t, and μ_{sbt} is the error term. I use three different measures of enrollment, which each have their pros and cons: total number of schools, total enrollment and total grades served. One problem with total number of schools is that charter schools age in–i.e. in their first year they usually only have one grade, then the next year, those students move up and they admit another class. Thus, there are sometimes fairly substantial changes to the number of students a network serves, even in years that they didn't really expand. Total enrollment is also problematic though, since it might be related to demand for that network. Total number of grades is a compromise between the two. The results are presented in the following table:

As we can see from the table, regardless of how we specify network size, I cannot reject the null that schools do not change in quality with the size of the network as a whole. Thus, I do not find any evidence that span of control issues are causing problems as networks

Table 1.11: Changes to Quality with Expansion for Previously Existing Schools

	English			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Age	-0.0067	-0.0092*	-0.0082	-0.0111*	-0.0157**	-0.0129*
	(0.0043)	(0.0054)	(0.0054)	(0.0057)	(0.0073)	(0.0073)
# Schools in Network	0.0025			-0.0015		
	(0.0063)			(0.0085)		
Total Network Enroll. (in 1000s)		0.0096			0.0116	
		(0.0115)			(0.0158)	
# Grades Served			0.0005			0.0003
			(0.0009)			(0.0012)
School FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	372	372	372	359	359	359

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

expand. Because I am now including school fixed effects, the fact that I have an unbalanced panel would cause some problems, so I restrict the sample to only schools that existed in my entire sample. In Appendix Tables (A.2, A.3, A.4) I try another few specifications, such as only using elementary and middle schools, or year fixed effects, but the results are very similar.

Networks Focus on Building a Reputation

It is difficult for economists to determine school quality, let alone for parents. Abdulkadiroğlu et al. (2019) find that parents choose schools based on peer achievement levels, not on value-added. It is possible that parents use a network's reputation as another proxy for quality. If this is the case, then networks might focus on building a reputation early on and then cut back spending on quality once that reputation is established. Testing this theory will be the focus of future work.

Scarce Inputs

Another possibility is that there are certain inputs that are scarce that a network has more and more difficulty obtaining as they expand. Teachers and principals typically have a large impact on quality and could certainly be scarce. However, all of the charter networks in New York City expanded on very different time frames. When Success Academy first entered New York, some networks already had more than eight schools. The declines that I see are with age or number in network, not year. So, it would have to be a very specific scarcity, i.e. it would have to be a scarcity of teachers or principals that match only with that network.

1.8 Conclusion

In my investigation of charter network scale up in New York, I find evidence that policy makers should reconsider or re-work policies that encourage the replication of successful charter schools. Although I find that networks that improve math test scores more are the ones expanding the most, I also find that as these networks expand, English test score value-added declines both for later schools in a network and with age. I investigate one of the most likely interpretations which is that there is a change in the types of students that charter schools are serving. With heterogeneous returns to charter schools, this would cause my measured value-added to decline. Although I find heterogeneous returns to charters, I find no evidence of changes to observable characteristics that match the declines in quality I observe.

This evidence also suggests that charter authorizers should spend more time looking at the quality of the marginal school in a network, not the average quality of schools in a network. Currently, when an authorizer is reviewing applications for new charter schools, they focus on the average. However, if a network were trying to open an eighth school, authorizers should look at quality of the sixth and seventh school in that network. This does not imply that there should be a cap on network size. There is a lot of heterogeneity in quality across networks, so if quality declines somewhat for the best networks, those schools will still be very good. In lower quality networks though, authorizers might want to considering limiting network size.

In addition, other states might want to keep a closer watch on networks that expand. New York State has a fairly well-regulated charter sector, so some of these patterns of decline might be worse in other states. For example, in Michigan, any university or community college can be a charter authorizer, and they get to collect a percent of the state funding that will go to the schools they approve. This clearly decreases the incentive for authorizers to reject new charter schools or shutdown existing ones. In addition, there are some authorizer specific policies which encourage expansion that might exacerbate the declines in quality I find. For example, the application to open a replication school through Grand Valley State University, has an explicit *maximum* of only ten pages, and there is only one question that deals with academic results at the original schools in that network. In future work, I plan to determine if the patterns I witness in New York hold in other states.

CHAPTER 2

EVALUATING THE HOSPITAL READMISSIONS REDUCTION PROGRAM

2.1 Introduction

Hospital readmissions have been a focus of research since Anderson and Steinberg (1984) found that between 1974 and 1977 readmissions accounted for twenty-four percent of Medicare spending. This amounted to $2.4 billion dollars during that period. This figure has only grown with time, and Jencks, et. al (2009) found that in 2004, unplanned readmissions cost Medicare $17 billion. Recent estimates by Boozary (2015) have risen as high as $26 billion. Early work on readmissions attempted to determine whether or not readmissions were preventable, resulting in a wide array of results. On the low end, Frankl, Breeling, and Goldman (1991) found that nine percent of all readmissions at a university hospital were preventable. On the high end, Chaput-Toupin et al. (1996) found that fifty percent could have been avoided. Kelly, et al. (1992), Gautam et al. (1996), Graham and Livesley (1983), and Oddone et al. (1996) all found that anywhere between fifteen and thirty-four percent of readmissions were preventable.

Given the high cost and the fact that at least some, if not many, readmissions could be prevented, policymakers have looked for ways to fix the problem. In 2008, the CMS suggested implementing payment incentives to reduce the rate of hospital readmissions. Afterwards, the Hospital Readmissions Reduction Program (HRRP) made its way into the Affordable Care Act, which was signed into law on March 23, 2010. Starting in October 2012, the CMS began adjusting all payments made to hospitals in the Inpatient Prospective Payment System (IPPS) if they had excess readmissions rates for specific conditions. In the beginning, the only conditions included were pneumonia, acute myocardial infarction (AMI)[1], and heart failure. In 2015, the standards were expanded to include chronic obstructive pulmonary disease and total hip or knee anthroplasty, and for 2017, coronary artery bypass graft surgery will be added. Readmission rates are risk-adjusted using comorbidities, age, and sex, and then hospitals are given a benchmark based on their case mix. If they exceed this benchmark, then all payments by Medicare to the hospital are adjusted by a maximum of three percent (this maximum was one percent in the first year and two percent in the second year). In the past three years, almost two-thirds of hospitals have been fined a total of nearly $1 billion, with an estimated total of $428 million in fiscal year 2015.

As soon as the HRRP details were announced, many criticized the fact that readmission rates were not risk adjusted based on the race or socioeconomic status of the patients. Indeed, even the expert panel that the CMS convened when crafting the HRRP suggested that they be taken into account. Since then, MedPAC advised Congress in 2013 that they should be considered, and numerous journal articles, such as Joynt and Jha (2012, 2013a, 2013b), Joynt et al. (2011), and Shih et al. (2015), have documented the fact that hospitals that serve more low income and minority patients tend to have higher readmission rates. As a result, bills have been introduced in both the House and the Senate to include socioeconomic factors, and the CMS is currently conducting a two-year trial to test the use of socioeconomic factors for risk adjustment.

Although there has been a push to include race or socioeconomic status in risk-adjustment, the CMS's justification for not including socioeconomic status—i.e. that poor patients likely attend worse hospitals on average—seems highly plausible. In Figure 1, I plot the correlation between the percent of patients at a hospital that are eligible for Supplemental Security In-

1. More commonly known as a heart attack.

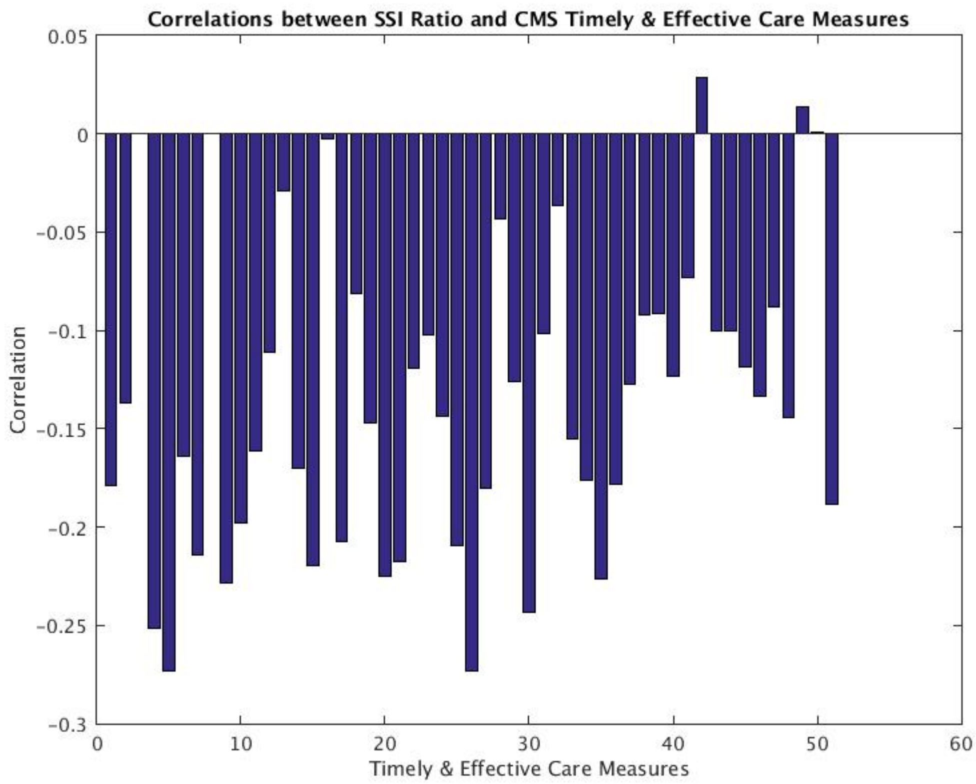


Figure 2.1: Correlation between SSI Ratio and Hospital Quality

come (SSI) and the scores on fifty-one different CMS process of care quality metrics. These metrics should not be affected by the socioeconomic status of the patients (as mortality rates or readmission rates could be) because they measure adherence to basic procedure. For example, one metric measures whether or not patients who arrive with a possible heart attack are given aspirin, and another measures what percent of the hospital staff has been given a flu shot. Figure 1 shows that the SSI ratio has a negative correlation with almost all of the process of care measures[2]. Since a *higher* SSI ratio means there are more poor patients, this figure suggests that indeed, poor patients are more likely to attend worse hospitals. Other papers have found similar results, such as Hasnain-Wynia, et al. (2007) which found that the majority of differences in treatments across races was due to site of care.

2. For some of the measures, a low score is better, so I took the negative of the measure

When studying this relationship, the majority of other papers, such as the ones cited above, have just looked at correlations between patient income and readmissions. Some papers have tried to control for the fact that there is also likely a negative correlation between hospital quality and patient income, but there are not that many that I could find. Joynt, et al. (2011) compare the outcomes of blacks at predominately non-minority serving hospitals and whites at predominantly minority-serving hospitals. To make this comparison, they just split hospitals into these two broad categories, and thus do little to control for site of care. I was also able to find two papers that focus on individual hospitals to control for this issue. First, Hu et al. (2014) look at a major hospital in Detroit and find somewhat mixed results. Because they do not have individual socioeconomic variables, they must rely on a patient's address to estimate them. They find that patients from areas with a large portion of families below the poverty line tend to have higher readmissions, but that there is no effect for patients who are from areas with lower median incomes. Second, Amarasingham et al. (2010) look at a hospital in Texas and find a weakly significant effect of living in the lowest quintile census tract. (With both of these papers, it is unclear how much socioeconomic variation there is for just one hospital, and neither explicitly addresses this.)

There also have been a limited number of studies that have analyzed the effects of the HRRP on readmissions. The only one that I could find was Zuckerman et al. (2016), which found that from 2007 to 2015, readmissions rates for targeted conditions (AMI, heart failure, and pneumonia) declined from 21.5% to 17.8%, and that for non-targeted conditions, they fell from 15.3% to 13.1%. However, they do not claim that HRRP caused this drop, and because non-targeted conditions fell as well, it is possible that other changes from the ACA caused the decrease in readmissions. I will build on this study by identifying a group of hospitals that were not subject to the HRRP, and use them as a control group. Next, I investigate whether or not a hospital's proportion of poor patients has an effect on a hospital's readmission rates. To do so, I identified hospitals that have closed in the last eight years. A number of papers have found that distance is one of the most important determinants of hospital choice, such as Gaynor and Vogt (2000, 2003) and Town and Vistnes (2001), Chandra et al. (2016), and Romley and Goldman (2011). Thus, after a hospital that treats a large number of poor patients closes, nearby hospitals should have an increase in the fraction of poor patients that they are treating. I can use this to look at how the changes in patient population affect the readmission rates at the closest hospitals.

2.2 Data

2.2.1 Sources

Beginning in 2005, CMS started its Hospital Compare datasets. These allowed both the public and researchers to compare the quality of different hospitals. The site began with quality metrics, but in 2009, CMS started posting 30-day risk adjusted readmission rates and mortality rates for heart failure, pneumonia, and AMI. The CMS defines a readmission as an admission to a hospital within thirty days of being discharged from another hospital. CMS risk-adjusts these measures for the presence of comorbidities, age, and sex. The readmission rates are updated annually in July on a one year lag, and are on a rolling three year average. So, in July 2009, CMS posted the data covering July 2005- June 2008. The CMS does not calculate readmission rates for any condition for which the hospital saw fewer than 25 patients. In 2015, the CMS widened the definition of what constituted a pneumonia admission, so I dropped that year for pneumonia.

To measure socioeconomic status at each hospital, the CMS posts the ratio of patients that qualify for Supplemental Security Income for all hospitals in the IPPS. This data is not included in the Hospital Compare datasets. Because the IPPS does not include the critical access hospitals, I do not have data on the SSI ratios for those hospitals, and thus cannot include them when I look at the effect of an influx of poor patients at a hospital.

CMS also publishes a yearly IPPS Impact File which contains the fines that will be levied against the hospital in the upcoming year. This fine is a downwards adjustment that will be applied to all payments made by Medicare. The size of the fine is determined by how much the hospital exceeds a benchmark the CMS determines. This benchmark is calculated by first running a hierarchical logistic regression of readmissions on comorbidities, age, and sex. From this regression, they generate a predicted readmission rate for each patient. They then sum these probabilities and determine a predicted readmission rate. If the hospital exceeds this predicted level, then they are fined. The fine is the payment weighted average of the excess readmission rates across the conditions CMS is using at the time. Because CMS continually updates the benchmarking process, as hospitals across the country lower their readmission rates, the predicted values of readmission fall which makes it harder for the hospitals to avoid getting fined. Zhang et al. (2016) theorize that this setup might cause hospitals to end up in a Nash equilibrium such that no hospitals are investing in reducing readmissions rates.

I also used some county-level data in my regressions. First, I used county population estimates from the Census. The Census estimates county population by five-year age group every year, so I used the total age 65 and over population for each county. I also used county-level unemployment rates. The Bureau of Labor Statistics calculates employment data at the county-level at a monthly frequency.

2.2.2 IPPS vs. CAH Hospitals

The IPPS was setup in 1983 to reorganize the way Medicare made payments to hospitals. In 1997, after a string of rural hospital closings, Congress created a critical access hospital (CAH) designation for some rural hospitals. For a rural hospital to be deemed a critical access hospital, it needs to maintain certain services, such as a 24/7 emergency room, and meet certain criteria. For example, the hospitals need to be the only one within a thirty-five mile radius. If the hospital does qualify, it gets reimbursed at 101% of reasonable costs. These two types of hospitals cover the majority of hospitals in the country, though Maryland has its own all-payer system, and thus hospitals there do not participate in either program [3]. These critical access hospitals are spread throughout the U.S., though they are the most concentrated in the Midwest. Figure 2 shows the distribution of the hospitals. Most states have a number, though five states do not have any (Rhode Island, Connecticut, New Jersey, Maryland, and Delaware). This is not too surprising given that a critical access hospital is required to be 35 miles away from the closest hospital, and Rhode Island itself is only 38 miles by 47 miles in area. Table 1 shows some summary statistics comparing the two types of hospitals.

3. In 2014, Maryland started its own readmission reduction program, so I dropped them from my sample.

Figure 2.2: Map of Critical Access Hospitals

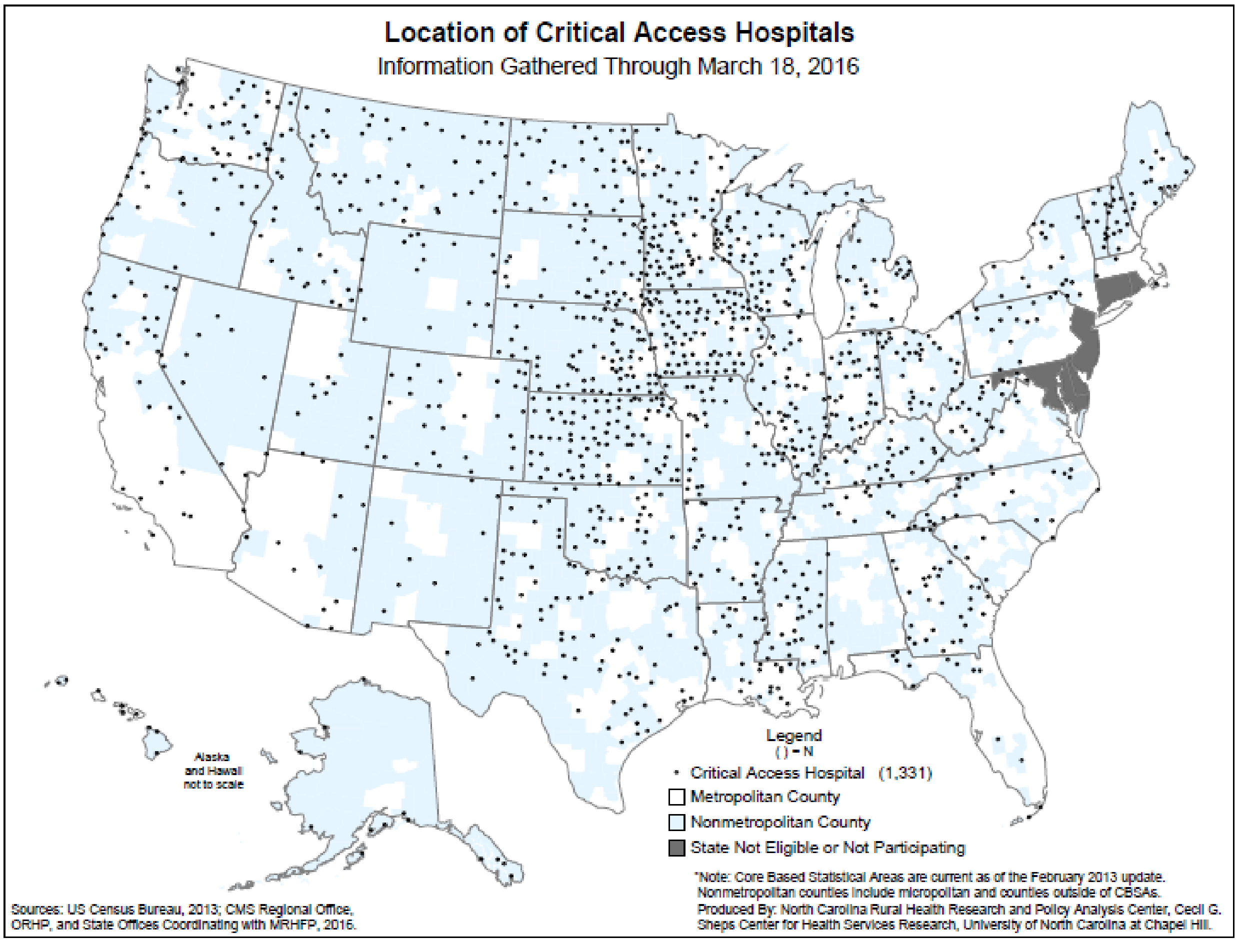


Table 2.1: IPPS vs. CAH Hospitals

	IPPS	CAH
Hospital Average 3-year Total Number of Patients (July 2007 - June 2010)		
Heart Failure	401.4	79.8
Pneumonia	335.9	106.2
AMI	223.65	62.4
Average Readmission Rates (July 2007 - June 2010)		
Heart Failure	24.9	24.7
Pneumonia	18.5	18.2
AMI	19.8	20.3
Std. Dev Readmission Rates (July 2007 - June 2010)		
Heart Failure	2.07	1.57
Pneumonia	1.61	1.34
AMI	1.40	1.22
Number of Hospitals in Sample (July 2007 - June 2010)		
Heart Failure	3096	965
Pneumonia	3109	1136
AMI	2294	125

Because the Hospital Compare data is a three-year rolling window, the first row in Table 1 is the average over all hospitals of each individual hospital's total from 2007 to 2010—the last data point before the ACA was signed. As we can see from the table, CAHs are substantially smaller than IPPS hospitals. However, there are still about one thousand hospitals with data for heart failure and pneumonia. In addition, though the sizes are very different, their readmission rates are nearly identical. When the ACA was passed, there were no provisions that linked performance to pay for these critical access hospitals[4]. Because critical access hospitals are not subject to HRRP, I will use them as a control group.

Though there are differences between IPPS and CAH hospitals, I think they can be used as a viable control group. Though they are smaller than the average IPPS hospital, their average readmission rates before the ACA was passed were nearly identical. Though somewhat more concentrated in the Midwest, they are spread throughout the map, and thus would be affected by regional trends similar to IPPS hospitals. Also, although all critical access hospitals are rural, not all rural hospitals are critical access hospitals. In fact, there are 1,058 rural IPPS hospitals. The big difference between the two sets of hospitals is that the critical access hospitals were not subject to the payment system changes that those in the IPPS were. Because a number of other IPPS pay-for-performance measures began at the same time, such as the Hospital Acquired Condition Reduction Program, any differences between the two groups will not be directly attributable to the HRRP, but the pay-for-performance measures more generally. At the same time, the HRRP was the only one of these which focused on readmissions.

4. There has been some discussion of moving critical access hospitals to a more similar payment system, but there have not yet been any moves made, only discussions about how it will be difficult to adapt the program to these hospitals.

2.2.3 Identifying Closed Hospitals

To identify hospitals that had closed within the last eight years, I looked at hospitals that dropped out of the Hospital Compare dataset, and found three hundred and one. However, a closure was not the only reason that a hospital might drop out of the sample. For example, a few hospitals dropped out of the sample because the facility was outdated, and a new one had been built across the street or close by (clearly, I did not include these). In order to verify which hospitals had truly closed, I used local news reports and hospital press releases. Many of these hospitals had closed in a manner that made them unusable, but I identified ninety-nine that had. Many closed because they were losing money, but there were others reasons too, such as fraud investigations. A few of hospitals had not fully closed, but had instead downgraded the amount of care that they offered. To make a decision on whether or not to include these hospitals, I looked at how much the hospital had cut back. If they still ran an emergency room and had services to treat major ailments, I did not include them. However, a few were turned into urgent care facilities, which do not treat any serious ailments, have limited hours, and cannot provide adequate care for pneumonia, AMI, or heart failure. Last, there were also some that closed too early or too late for them to be usable given the data I have.

Because the data I used was only updated annually, I had to decide which data point would be the first that the hospital closing affected. I decided that if a hospital closed before December 31, I considered it to have closed at the beginning of that year of data; otherwise, I used the next year. For example, if a hospital closed on November 15, 2011, I said that its closing would have affected nearby hospitals in the new data covering July 2011 - June 2012.

Of the ninety-nine hospitals I found, I then only included those that had at least twenty five patients for each specific condition within the last three years. This is the cutoff that

CMS uses as the number of patients necessary to calculate a readmission rate. In Table 1, I present summary statistics comparing the closed hospitals to the full sample. The average closed hospital was about half the size as the average hospital in the full sample, and the median was about sixty percent as large. The hospitals that closed also had about twice as many patients who qualified for SSI. Because SSI ratios at each hospital are not included in the Hospital Compare dataset, I also had to drop one more hospital from my regressions on hospital closings. This brings down the total number of closed hospitals to 75 and 50 for heart failure and pneumonia respectively. Because there were so few closed hospitals that had enough AMI patients, I focus on pneumonia and heart failure going forward, though I still present the results for AMI.

Table 2.2: Hospital Summary Statistics

	Closed Hospitals	Full Sample
Number of Hospitals:		
Heart Failure	76	4089
Pneumonia	51	4328
AMI	23	2469
Hospital-Average 3-year Total Number of Patients (July 2005- June 2008)		
Heart Failure	176.8	354.37
Pneumonia	181.6	306.7
AMI	109.0	232.9
Hospital Median 3-year Total Number of Patients per (July 2005- June 2008)		
Heart Failure	143	229
Pneumonia	149	234
AMI	81	137
Mean SSI Ratio	14.6%	7.1%

In Figure 3, I plotted the hospitals that closed that had enough patients to be a part

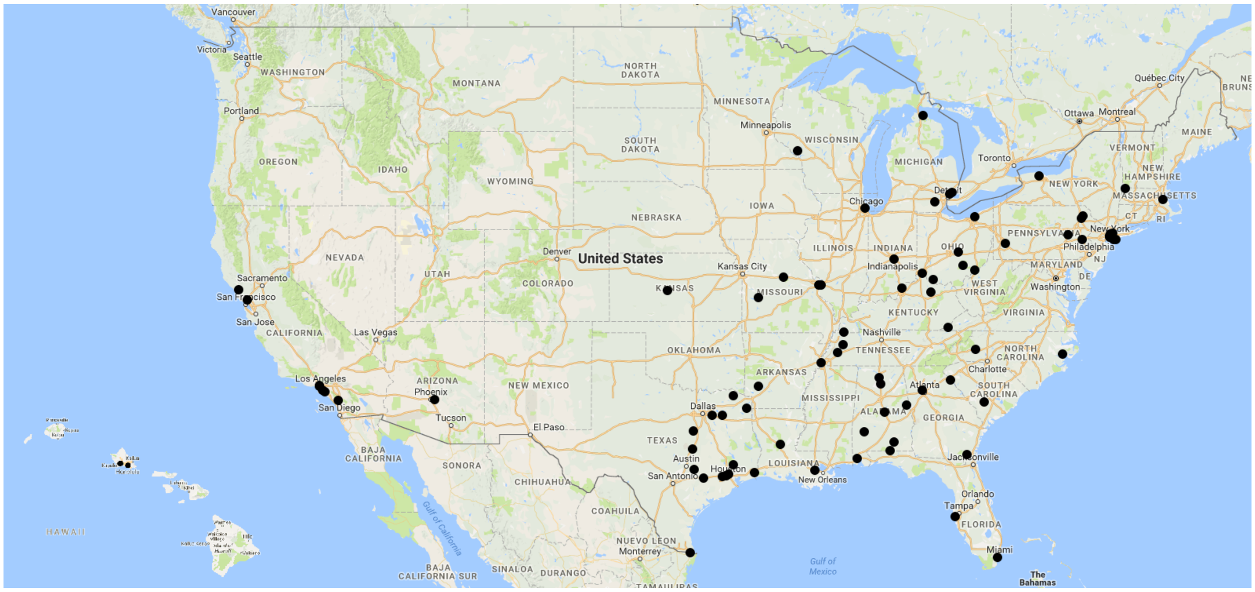


Figure 2.3: Map of Closed Hospitals in Sample

of my sample. Most of the hospitals are in the eastern half of the United States, though within the eastern half of the United States, they are fairly spread out. After creating a dataset of closed hospitals, I needed to determine the next closest hospital. First, I assigned each hospital to its Hospital Referral Region (HRR) as defined by the Dartmouth Atlas of Health Care. These regions were defined by "determining where patients were referred for major cardiovascular surgical procedures and for neurosurgery." The HRRs are frequently used to delineate health care markets. Figure 3 shows the map of the 306 HRRs in the country. After assigning each hospital to its HRR, I then used Python to query Google Maps as to the distance between the hospital that had closed and every other hospital in the same HRR. I took these results and then was able to identify the closest three hospitals to the hospital that had closed. Of the seventy-six closed hospitals that I used, they come from sixty different hospital referral regions. The majority of these HRRs only experienced one hospital closing in the sample period, though four HRRs had three or more. Three of these four regions included Los Angeles, Houston, and New York City. (Alabama was the main area consisting of the fourth, and there were fifty hospitals in total in that HRR.) On

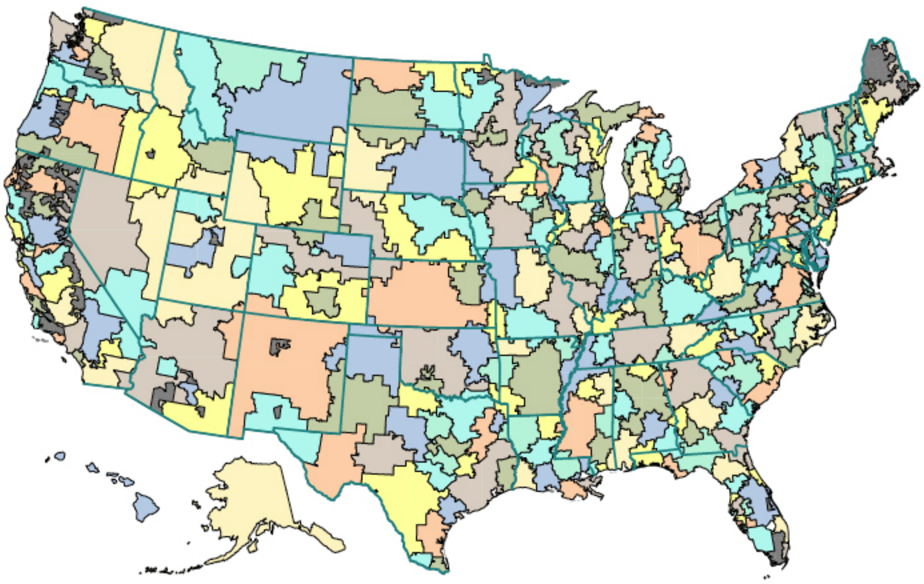


Figure 2.4: Map of Dartmouth Atlas Hospital Referral Regions

average, HRRs that had at least one hospital close had 29.2 total hospitals. Of the hospitals that closed, there was only one hospital that was closest to two separate hospitals that had closed. Of the top three closest hospitals for all seventy-six closings, there were only nine that were also on another hospital's list in the top three, and no closed hospital was the closest to another hospital that closed.

2.3 Empirical Strategy

Because the HRRP does not cover all hospitals, but only those in the IPPS, the critical access hospitals are exempt. I therefore used the CAHs as a control group in the following procedure. In order to investigate the changes caused by the HRRP, I estimated the following:

$$\Delta R_{h,m,t} = y_t + y_t \cdot \mathbb{I}_h\{IPPS\} + HRR_k + g_h + \kappa \Delta u_{c,t} + \beta f_{h,t} + \epsilon_{h,t} \qquad (2.1)$$

I used the difference in readmission rates, $\Delta R_{h,m,t}$[5], in order to remove hospital level fixed effects. I also include stand alone year dummies, y_t, and year dummies interacted with whether or not the hospital is in the IPPS program. I included hospital referral region dummies, HRR_k, and the change in county level unemployment, $\Delta u_{c,t}$, in order to determine whether or not any changes were more correlated with regional developments than the HRRP. I also include any fine the hospital may have received, $f_{h,t}$. The fines are announced directly preceding a fiscal year, and are applied to all Medicare reimbursement in the upcoming year. Last, the variable g_h represents classification type. The CMS classifies its hospitals as either large urban, other urban, rural, or critical access.

After investigating the overall effect of the HRRP, I then try and determine the effect that having low socioeconomic patients has on a hospital's readmission rate. To do so, I identified seventy-six hospitals that have closed in the US since 2008. My first step after identifying these hospitals was to verify that nearby hospitals saw an increase in patients after a hospital closed down. To investigate this, I estimated the following for each condition separately:

$$\Delta N_{h,t} = HRR_k + \delta_{yg} + \alpha' Q_{h,t-1} + \gamma \Delta p_{c,t} + \sum_{d=1}^{3} \mathbb{I}_{d,j} \cdot N_{j,t'} + \mu_{h,t} \tag{2.2}$$

where $\Delta N_{h,t}$ is the change in the number of patients at hospital h and δ_{yg} is a year-classification dummy (the four classifications being rural, other urban, large urban, and critical access, which Medicare designates). The sum at the end before the error term includes indicators for whether or not the hospital was the d^{th} closest hospital to hospital j that shut down in period t'. Since the data are three-year totals, I include this indicator

5. Because the rates in the data are three year averages, when looking at two individual differences for two different hospitals, comparing the two could be misleading if the base number of patients changed in a very different manner for the two hospitals. If IPPS hospitals were consistently growing and CAH hospitals were consistently shrinking, using equation 2 to compare these differences would be much more difficult. However, the overall percent change in the number of patients at CAH and IPPS hospitals tracks fairly closely, and the average difference in the growth rates by condition are less than one percent, so this mitigates this issue.

for the three years after the hospital closed. I interacted this with the size of the closest hospital since a larger hospital should send more patients to nearby hospitals after it shuts down. When trying to determine the effect of hospital closings on the number of patients, there are a number of issues that could arise. First, the fact that hospital closings are not random[6] could cause issues. A hospital in a region that was shrinking might be more likely to close, which would also likely affect the nearest hospital. This is why I included the HRR dummies, as well as the change in the elderly population in the county, $\Delta p_{c,t}$. Another reason that hospital j might be more likely to close is if hospital h is a high quality hospital. Since Chandra et al. (2016) showed that hospital quality is associated with a growing market share, not controlling for quality would bias this regression. Therefore, I included $Q_{h,t-1}$ which is a vector of the thirty day readmission rates and the thirty day mortality rates for the three conditions (these are two of the quality metrics Chandra et al. (2016) were looking at).

After showing there seems to be an increase in patients at nearby hospitals after a shutdown, I then move on to the main regression—to see if this increase results in a change in the readmission rate. To investigate this, I estimate the following equation:

$$\Delta R_{h,t} = \delta_{yg} + HRR_k + \alpha f_{h,t} + \eta \Delta u_{c,t} + \sum_{d=1}^{3} \mathbb{I}_{d,j} \cdot w_j + \xi_{h,t} \tag{2.3}$$

The variable $\Delta u_{c,t}$ represents changes in county level unemployment, and w_j represents whether or not the closed hospital, hospital j, served a different share of poor patients than the hospital h that it was close to. I divide nearby hospitals into three categories: those which had richer patients than the nearby hospital, poorer, and about the same. When I classified the closest three hospitals as either richer, poorer, or about the same, the cutoff that I used to classify the three groups was half a standard deviation of the SSI distribution.

6. One hospital did close because it was destroyed by Hurricane Sandy.

Again, there are a number of endogeneity concerns at first glance, since hospital closings are not random. Area economic conditions could both drive a hospital out of business and cause an increase in readmissions, which is why I am controlling for changes in the county unemployment rates. I also include HRR fixed effects in case there are other regional trends correlated with hospital closings and readmission rates. Another concern is that many hospitals are owned by companies that manage more than one hospital in the region. These companies often consolidate regional hospitals when one is struggling financially. Since this consolidation might come with some kind of planned modernization or expansion of the nearby hospital which could affect readmission rates, I purposefully left off hospital closings that were announced as a part of a regional consolidation. Next, hospitals compete with each, so there might be a concern that if a nearby hospital is of much higher quality (as measured by readmissions rates), this hospital could drive the other out of business. While this surely sometimes happens, since I looked at differences in readmissions rates, I am differencing out a hospital quality fixed effect that might be causing this. For there to be a problem, there would need to be some hospital specific intervention that both decreased readmission rates and drove other nearby hospitals out of business in the same time horizon. Since most readmission rates interventions involve following up better with patients, it seems unlikely that such an intervention would both take a while to bring down rates, and cause an nearly immediate closing of a nearby hospital.

Another concern in a different vein is that since patients will have to travel farther, in the case of an emergency, like a heart attack, it will take the patient longer to get to the hospital, so their readmission rates will be higher because they will be in worse health when they arrive. However, Joynt et al., (2015) studied areas that had hospital closings, thus forcing patients to travel further, and found that they were not associated with a worse outcomes.

Last, one concern might be that an influx of patients would cause a rise in the readmission rate because the hospital was overwhelmed by the new patients. By including hospital closings where the nearby hospital treats a similar number of SSI patients or even fewer, if the influx were overwhelming the hospital, we would expect to see a positive coefficient on all three types of closings.

2.4 Results

2.4.1 Effect of the HRRP

The first part of my investigation involves using the critical access hospitals to look at the effect of the HRRP. Table 3 presents the results equation 1. When running the regression, I left off the IPPS designation of "rural"[7]. So, the other region dummies are measured in comparison to that baseline. Fines are measured in percentage points, and the average fine for each year is at the bottom of the table.

Table 3 mostly shows very little effect of the HRRP. Readmission rates seem to have fallen, especially between 2012 and 2014. Although readmission rates fell by a fair amount for all three conditions, the difference between the IPPS hospitals that were affected by the HRRP and the critical access hospitals which were not a part of the program is negligible. In most years, the difference is statistically insignificant. In 2012, there is some evidence that IPPS hospitals saw a larger decline in readmissions rate, though compared to the overall drop that year, it is not very large. Though the fines have a statistically significant coefficient, after the fines started, the difference between the IPPS hospitals and CAH hospitals becomes positive. This somewhat negates the impact of the fine, though the worst hospitals still improved slightly more than the CAH hospitals those years. The trend also seems to be fairly uniform across the country. Of the 305 coefficients on the HRR dummies (one was left

7. This is not the same as a critical access designation.

Table 2.3: Change in Readmission Rates and HRRP

	2009	2010	2011	2012	2013	2014	2015	Full Sample
Pneumonia								
Year Dummy	0.06	0.06	0.02	-0.86**	-0.32**	-0.39**		
Year-IPPS Interaction	0.02	-0.05	0.01	-0.10**	0.20**	0.12**		
Fines								-0.61**
Other Urban								0.06**
Large Urban								0.03
Heart Failure								
Year Dummy	0.24**	0.14	-0.02	-1.55**	-0.36 **	-0.63**	-0.05	
Year-IPPS Interaction	-0.04	-0.02	-0.06	-0.16**	0.13**	0.01	0.18**	
Fines								-0.43**
Other Urban								0.09*
Large Urban								0.05
AMI								
Year Dummy	0.06	-0.04	-0.04	-1.2**	-0.60**	-0.81**	-0.20	
Year-IPPS Interaction	-0.01	-0.06	-0.10	-0.15	0.19*	0.01	0.18	
Fines								-0.22**
Other Urban								0.04
Large Urban								0.03
Average Fine					0.27	0.25	0.49	

* $p < 0.05$, ** $p < 0.01$

out as the baseline), only 3 are statistically significant, which is within the range we would expect if there were no true differences between the HRRs. Hospitals that the CMS classifies as "other urban" seem to have had slightly less of a decline. The findings seem to show most of the decline coming in an abrupt fall in 2012 followed but smaller but still significant declines the next two years. This does seem a little strange, but it is in line with what Zuckerman et al. (2016) found using Medicare Part A and B claims forms. They theorize it was because hospitals were anticipating the upcoming fines, but because they do not include the critical access hospitals, they miss that these hospitals fell by a similar amount.

Figures 2.5, 2.6, and 2.7 show the cumulative effect of changes in readmissions at IPPS and critical access hospitals for pneumonia, heart failure, and heart attacks respectively. From these graphs, it is striking how similar the declines in readmissions are across these two types of hospitals. Readmissions for these three conditions fell between 1.5 percentage points and 3 percentage points, while the difference between the two types of hospitals was never greater than 0.2 percentage points.

It is possible that another program that only affected CAHs came into effect during a similar time period that caused a simultaneous decline, but I was unable to find any. While I was able to find a some other programs that targeted readmissions, I was not able to find any that only targeted critical access hospitals. It is possible that some of these other programs are more responsible for the decline in readmissions than the HRRP is. For example, in April 2011, the Partnership for Patients campaign was launched in an attempt to reduce readmissions and hospital acquired conditions. The program included both IPPS and critical access hospitals. This program too, though, was studied by Gerhardt (2013), and hospitals that did and didn't participate saw similar declines in readmissions. Another possibility is the Health Information Technology for Economic and Clinical Health (HITECH) Act of

Figure 2.5: Cumulative Change in Pneumonia Readmissions Rates

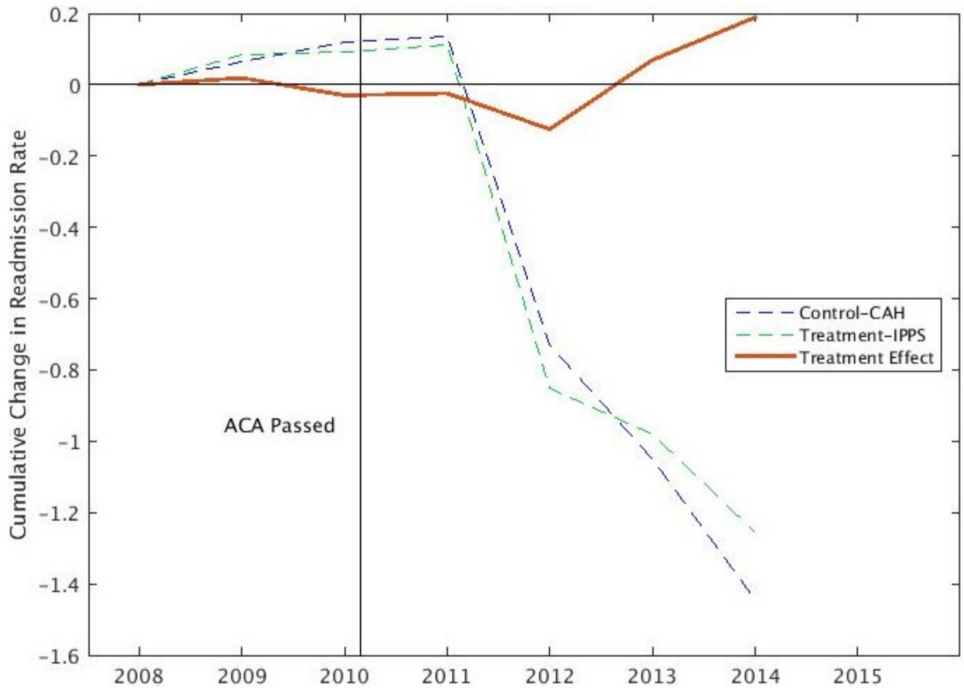


Figure 2.6: Cumulative Change in Heart Failure Readmissions Rates

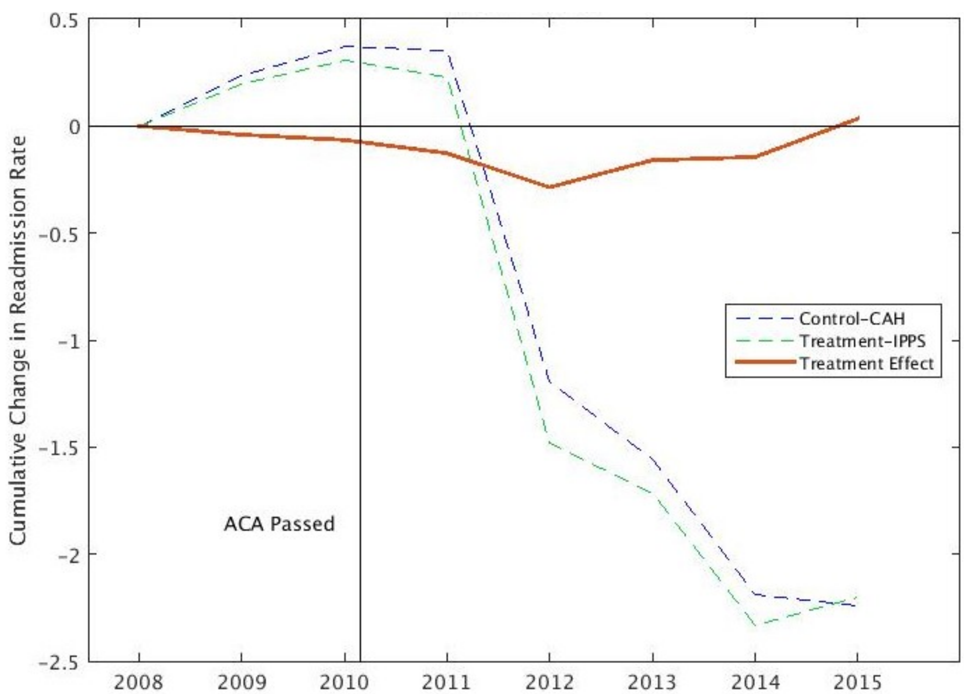
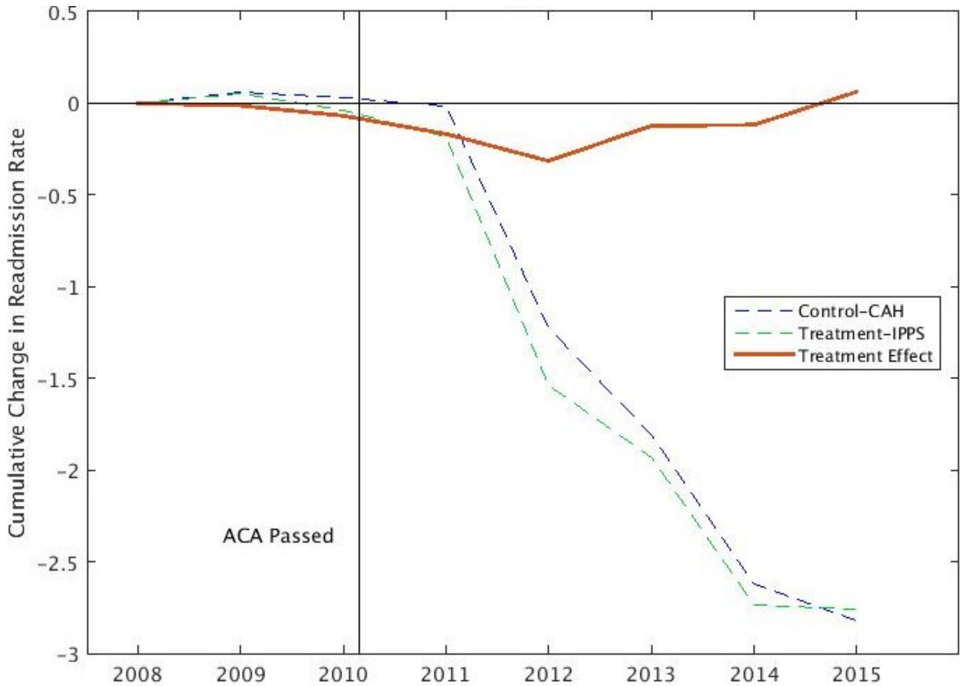


Figure 2.7: Cumulative Change in AMI Readmissions Rates



2009. This act incentivized the transition to electronic health records and took place in a similar time frame as the HRRP fines. The switch to electronic health records also often brings with it systems that facilitate communication between doctors and patients. Because many readmissions are caused by things like a patient not taking their medication correctly or missing a follow-up appointment, a clear print out of instructions could theoretically help reduce readmissions rates. It is clear that determining why readmissions declined requires further investigation.

2.4.2 The Effect of Hospital Closings

Next, I attempted to determine if after a hospital closes, whether or not the nearby hospitals receive an influx of patients. To do so, I estimate equation 2 from the preceding section.

From the table, we can see that if a hospital shuts down, the closest two hospitals generally have an increase in the number of patients, though the third closest has no statistically significant change. Because the dummy for a closed down hospital nearby was interacted

Table 2.4: Increase in Patients at Nearby Hospitals

	Closest	2nd Closest	3rd Closest
Heart Failure	0.10*	-0.08	0.06
	[-0.01, 0.20]	[-0.19, 0.04]	[-0.06, 0.17]
Pneumonia	-0.01	0.23***	0.04
	[-0.10, 0.87]	[0.14, 0.33]	[-0.05, 0.13]
AMI	0.32***	0.21	-0.05
	[0.08,0.56]	[-0.14, 0.55]	[-0.4, 0.29]

95% confidence interval in brackets
* $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$

with the size of that hospital, the interpretation of the coefficient is that if a hospital that has had an average of 100 AMI patients over the last three years, the closest hospital has an associated increase of 32 patients on average. Overall, the estimates are somewhat small, with AMI showing the most evidence that the nearest hospitals take in a large portion of patients who likely would have gone to the hospital that has closed. This might be because AMI is usually more time sensitive than heart failure and pneumonia for most patients. Though this specification requires more investigation, it is not the main regression of interest, and it still shows some evidence that the closest two hospitals saw an increase in patients following a hospital shutdown. I will discuss ways I could improve upon this test in the conclusion.

Last, I move on to the demographic regression of interest. Table 5 shows my estimation of equation 3. Because my earlier regression showed that the third closest hospital showed no increase in patients after a hospital closing, I only included the first two closest hospitals. The column “poorer” refers to the indicator for having a hospital that has a higher ratio of SSI patients close down nearby.

Table 2.5: Effect of Hospital Closures on Nearby Readmission Rates

Heart Failure	Poorer	Richer	Similar SSI
Closest Hospital	0.13	0.27	0.03
	[-0.38,0.65]	[-0.35, 0.90]	[-0.42, 0.48]
Second Closest Hospital	-0.15	-0.02	0.06
	[-0.65, 0.35]	[-0.67, 0.62]	[-0.40, 0.52]
Pneumonia			
Closest Hospital	0.42	0.32	0.17
	[-0.06, 0.9]	[-0.30, 0.95]	[-0.25, 0.59]
Second Closest Hospital	-0.47	-0.31	-0.09
	[-0.96, 0.03]	[-0.91, 0.29]	[-0.53, 0.34]
AMI			
Closest Hospital	-0.07	-0.04	0.16
	[-0.73,0.58]	[-0.76, 0.68]	[-0.37, 0.70]
Second Closest Hospital	-0.18	0.07	-0.04
	[-0.92, 0.55]	[-0.61, 0.75]	[-0.62, 0.55]

From the results, I am unable to reject the null that an increase of poor patients has no effect on a hospital's readmission rate. None of the estimates are significant at the 95% confidence level. While this provides some evidence that socioeconomic status is not to blame for higher readmission rates, there are a number of ways this study could be improved, which I discuss in the conclusion.

2.5 Conclusion

In the years since the ACA was passed with a provision to reduce hospital readmissions, readmission rates around the country have fallen. Because these rates have fallen, the CMS has continued to revise and expand the Hospital Readmission Reduction Program. CMS has been considering adding the critical access hospitals, and they are continuing to add more conditions to the program. However, it is important to know its true effects. Though rates have fallen, the CMS should keep in mind that other hospitals not associated with the program saw similar declines during the same time period. It is worth investigating what may have actually caused these declines. If another program is in fact responsible, then the CMS should focus on expanding that program instead. There were a slew of provisions in the ACA that also could have affected readmissions rates, such as the Community-based Care Transitions Program, the Partnership for Patients, or the Accountable Care Organizations, just to name a few.

At the same time, the CMS should be wary of changing their risk-adjustment process. At the beginning of the 2016 fiscal year, CMS announced that it is working with the National Quality Forum to run a two-year trial to test adding a socioeconomic factor to the risk adjustment. Numerous academics and papers have been published documenting the correlation between the treatment of more poor patients and higher rates of readmission, though none have shown any causal evidence. In my study, I was unable to reject the null that an increase in poor patients at a hospital increased the readmissions rate.

While I did not find any evidence that more poor patients at a hospital lead to an increased readmission rates, I would be interested in using disaggregated Medicare claims data to better investigate this question. Using the disaggregated Medicare claims data would improve the analysis in a number of ways. With the claims data, it would be possible to look at the period immediately following the closing as opposed to having to use a start date that was up to six months off. Second, Medicare claims data has patient zip codes, so I could get a better sense of patient socioeconomic characteristics. Third, many of the critics of the risk adjustment cite both race and socioeconomic status. I did not have any data on race, though I could look at both using Medicare claims data. Last, while a number of hospitals have closed since 2008, using a longer time series would allow for a larger number of hospitals to be used. A longer time series would also help when examining the effect of the HRRP as well, since Hospital Compare only had a few years of data before the ACA was signed.

APPENDIX A

ROBUSTNESS CHECKS

Figure A.1: Manhattan Census Blocks

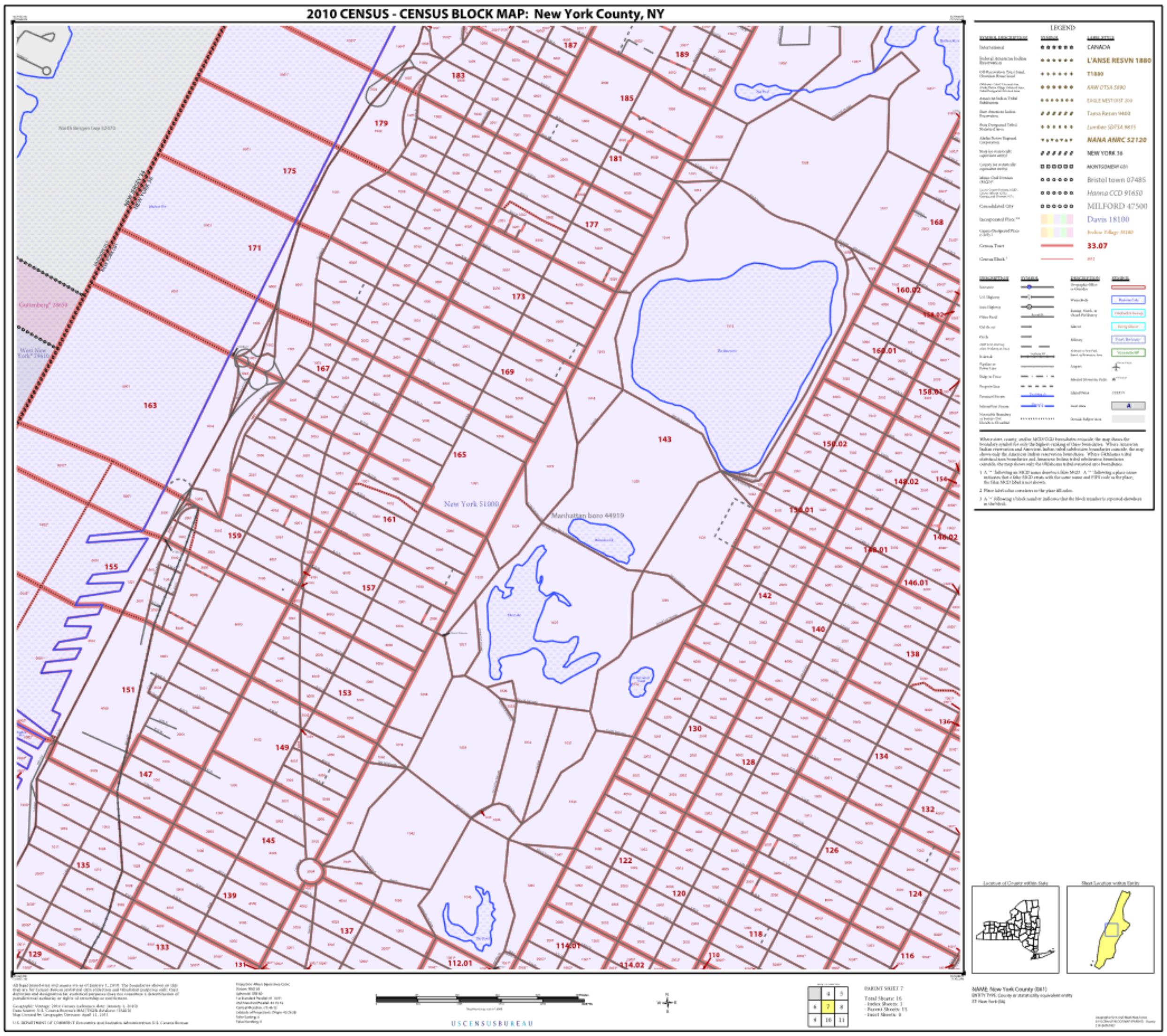
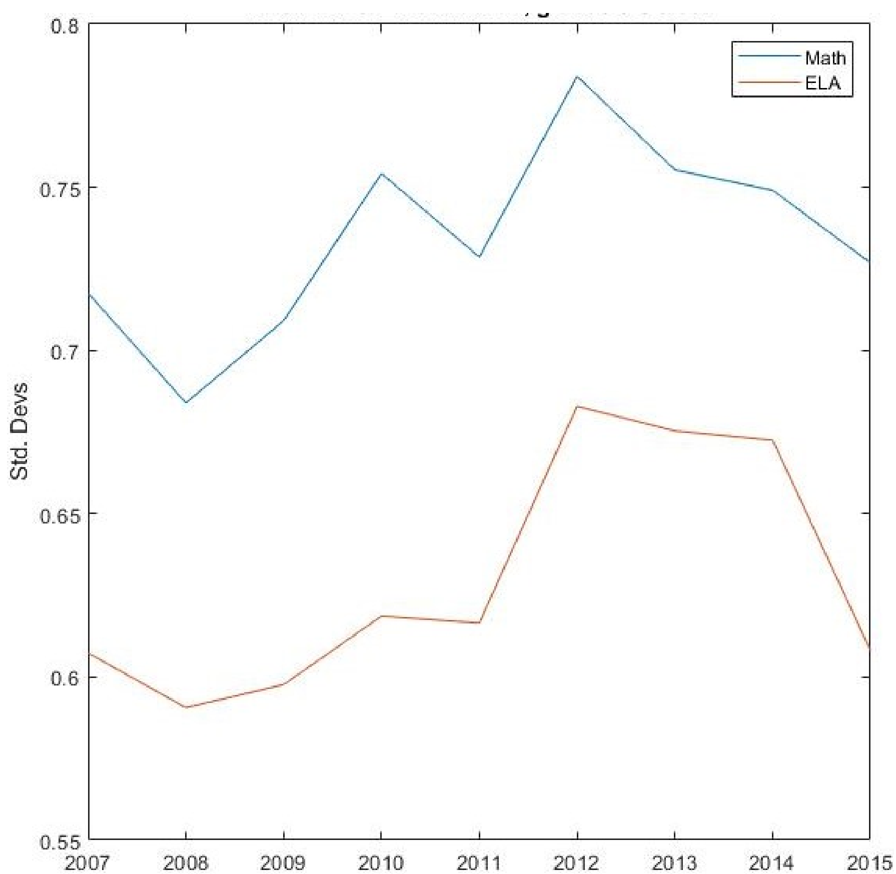


Figure A.2: Black-White Test Gap



Note: This graph shows the difference between the average of all white students' math and English test scores and the average of all black students' math and English test scores.

Figure A.3: Distribution of Census Tract Median Household Income for Charter and Tradtional Public School Students, 2015

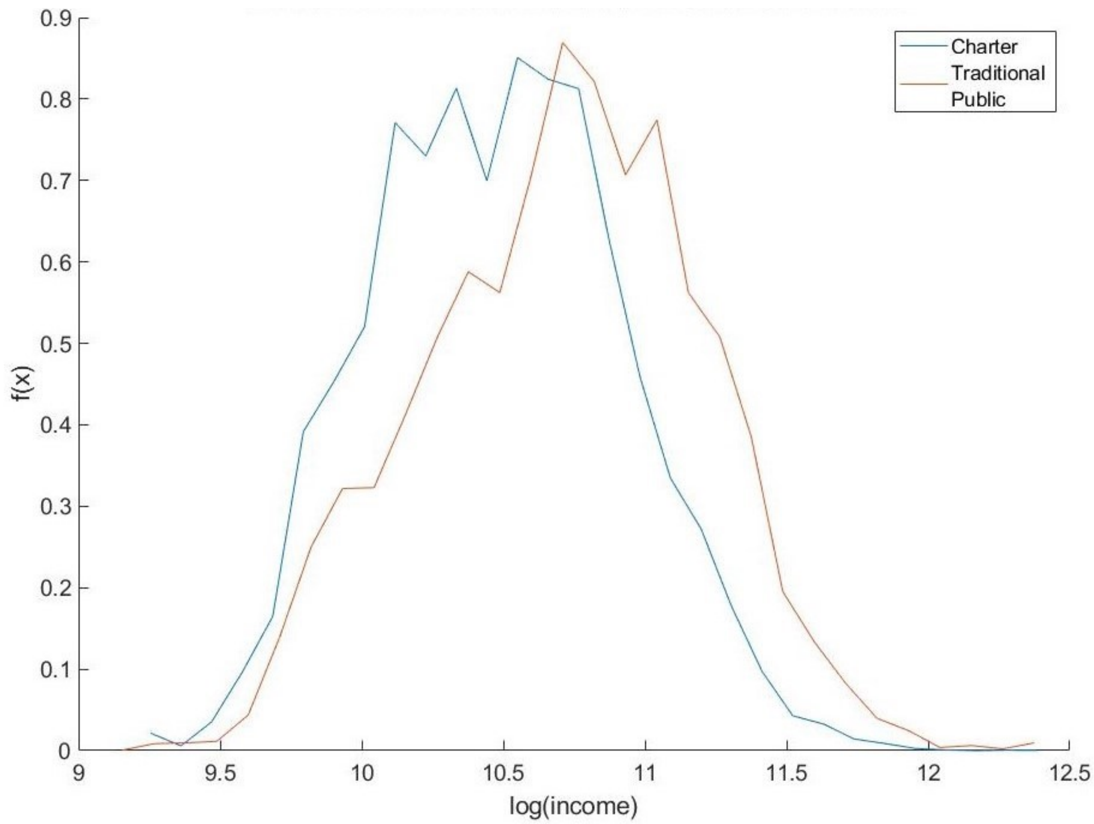


Table A.1: Within Network Changes in Value-Added, Elementary & Middle Schools Only

	English	Math
Age	-0.00790*	-0.00987*
	(0.00403)	(0.00583)
I{#2-3}	-0.0167	-0.0113
	(0.0209)	(0.0302)
I{#4-5}	-0.0422	-0.0651
	(0.0357)	(0.0516)
I{#¿=6}	-0.0499	-0.0800
	(0.0419)	(0.0607)
Network-Year FE	Yes	Yes
N	498	498

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Figure A.4: Coefficients from Nonparametric Age Specification, English

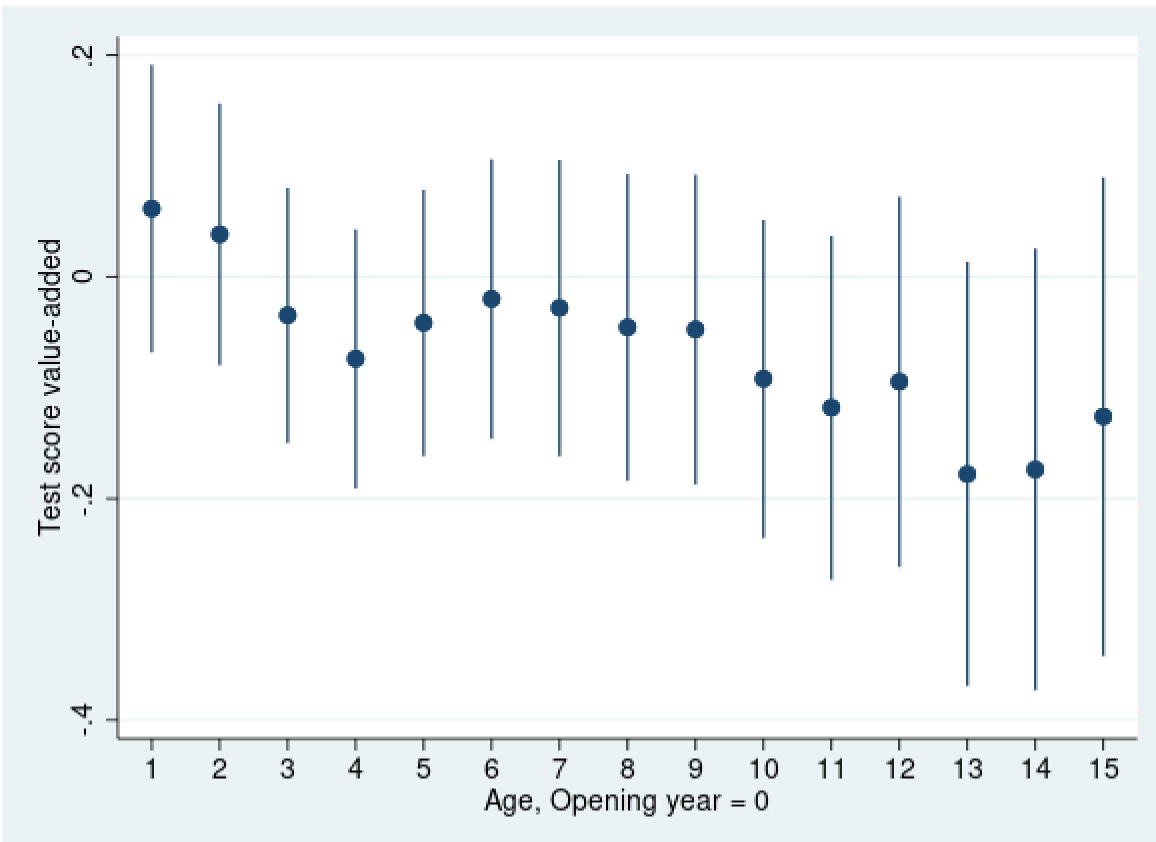


Figure A.5: Coefficients from Nonparametric Age Specification, Math

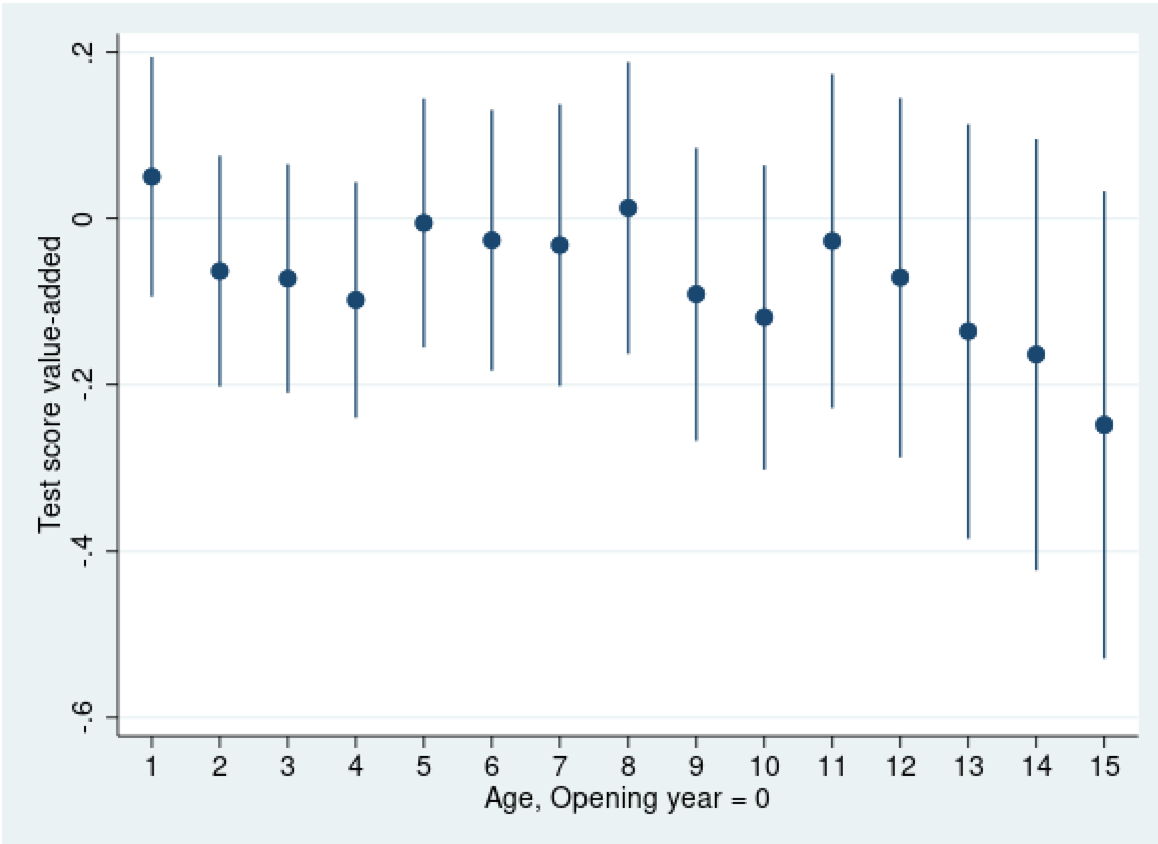


Table A.2: Span of Control – Only Elementary & Middle Schools

	English			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Age	-0.0034	-0.0082*	-0.0065	-0.0038	-0.0154**	-0.0095
	(0.0039)	(0.0048)	(0.0049)	(0.0055)	(0.0068)	(0.0069)
# Schools Open in Netw.	0.0001			0.0002		
	(0.0055)			(0.0078)		
Total Netw. Enroll (in 1000s)		0.0140			0.0340**	
		(0.0105)			(0.0147)	
# Grades Served			0.0007			0.0013
			(0.0008)			(0.0011)
School FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	304	304	304	304	304	304

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A.3: Span of Control – Year Fixed Effects

	English			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
# Schools Open in Network	0.0027			0.0005		
	(0.0062)			(0.0085)		
Total Netw. Enroll. (1000s)		0.0119			0.0168	
		(0.0116)			(0.0161)	
# Grades Served			0.0006			0.0006
			(0.0009)			(0.0012)
School FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	372	372	372	359	359	359

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A.4: Span of Control with Year Fixed Effects, Only Elem. & Middle Schools

	English			Math		
# Schools Open in Network	-0.0007 (0.0054)			0.0008 (0.0079)		
Total Netw. Enroll (1000s)		0.0138 (0.0104)			0.0355** (0.0150)	
# Grades Served			0.0006 (0.0008)			0.0014 (0.0011)
School FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	304	304	304	304	304	304

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

REFERENCES

Abdulkadiroğlu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. (2019) "Do Parents Value School Effectiveness?." *Working Paper*

Abdulkadiroglu, Atila, Joshua Angrist, Sarah Cohodes, Susan Dynarski, Jon Fullerton, Thomas Kane, and Parag Pathak. 2009. *Informing the Debate: Comparing Boston's Charter, Pilot, and Traditional Schools.* The Boston Foundation (TBF). Boston, January

Abdulkadiroglu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak. 2011. *Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots.* Quarterly Journal of Economics 126 (2): 699–748.

Abdulkadiroglu, Atila, et al. *Charters without lotteries: Testing takeovers in New Orleans and Boston.* The American Economic Review 106.7 (2016): 1878-1920.

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind. "What can we learn from experiments? Understanding the threats to the scalability of experimental results." American Economic Review 107, no. 5 (2017): 282-86.

Amarasingham, R., Moore, B. J., Tabak, Y. P., Drazner, M. H., Clark, C. A., Zhang, S., et al. (2010). An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Medical care, 48(11), 981-988.

Anderson, G. F., & Steinberg, E. P. (1984). Hospital readmissions in the Medicare population. New England Journal of Medicine, 311(21), 1349-1353.

Andrews, I. and Oster, E. (2017). Weighting for External Validity. NBER Working Paper Series, 23826.

Angrist, J. D., Hull, P. D., Pathak, P. A.,& Walters, C. R. (2017). Leveraging Lotteries For School Value-Added: Testing And Estimation. The Quarterly Journal of Economics, 871–919

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. *Explaining charter school effectiveness.* American Economic Journal: Applied Economics 5.4 (2013): 1-27.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2010. *Inputs and Impacts in Charter Schools: KIPP Lynn.* American Economic Review 100(2): 239–43.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2012. *Who Benefits from KIPP?* Journal of Policy Analysis and Management 31 (4): 837–60.

Angrist, Joshua D., et al. *Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice.* Journal of Labor Economics 34.2 (2016): 275-318.

Axon, R. N., & Williams, M. V. (2011). Hospital readmission as an accountability measure. Jama, 305(5), 504-505.

Baude, Patrick L., Marcus Casey, Eric A. Hanushek, Gregory R. Phelan, and Steven G. Rivkin. "The evolution of charter school quality." Economica (2014).

Benbassat, J., & Taragin, M. (2000). Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of internal medicine, 160(8), 1074-1081.

Booker, Kevin, Brian Gill, Tim Sass, and Ron Zimmer. "Charter high schools' effects on educational attainment and earnings." No. 8004. Mathematica Policy Research, 2014.

Boozary AS, Manchin J, III, Wicker RF. The Medicare Hospital Readmissions Reduction Program: Time for Reform. JAMA. 2015;314(4):347-348.

Campbell, D. and Stanley, J. (1963). Experimental and quasi-experimental designs for research. Houghton Mifflin Company, Boston, MA.

Center for Research on Education Outcomes (CREDO) (2013, June). "National Charter School Study." Palo Alto: CREDO, Stanford University.

Center for Research on Education Outcomes (CREDO), Woodworth, J., Raymond, M., Han, C., Negassi, Y., Richardson, P., & Snow, W. (2017). "Charter Management Organizations 2017." Palo Alto, CA: Center for Research on Education Outcomes.

Chabrier, Julia, Sarah Cohodes, and Philip Oreopoulos. *What Can We Learn from Charter School Lotteries?.* The Journal of Economic Perspectives 30.3 (2016): 57-84.

Chandra, A., Finkelstein, A., Sacarny, A., & Syverson, C. (2016). Healthcare Exceptionalism? Performance and Allocation in the US Healthcare Sector. National Bureau of Economic Research.

Chaput-Toupin, E et al. "[Are early unforseen rehospitalizations inevitable?]." Revue d'epidemiologie et de sante publique 44.3 (1996): 221-227.

Chingos, Matthew M."The impact of a universal class-size reduction policy: Evidence from Florida's statewide mandate." Economics of Education Review 31, no. 5 (2012): 543-562.

Cohodes, S., Setren, E., and Walters, C.R. (2019). *Can Successful Schools Replicate? Scaling Up Boston's Charter School Sector*, NBER Working Paper No.25796

Cook, T. and Campbell, D. (1979). Quasi-Experimentation: Design & Analysis Issues for Field Settings. Houghton Mifflin

Cronbach, L. and Shapiro, K. (1982). Designing evaluations of educational and social programs. Jossey-Bass, San Francisco, CA.

Davis, Jonathan, Jonathan Guryan, Kelly Hallberg, and Jens Ludwig. “The economics of scale-up.” No. w23925. National Bureau of Economic Research, 2017.

Dee, T. S., Dobbie, W., Jacob, B. A., & Rockoff, J. (2019). “The causes and consequences of test score manipulation: Evidence from the new york regents examinations.” *American Economic Journal: Applied Economics*, 11(3), 382-423.

Dinerstein, Michael and Troy Smith. 2016 *Quantifying the Supply Response of Private Schools to Public Policies*, working paper.

Dobbie, Will S., and Roland G. Fryer Jr. *Charter Schools and Labor Market Outcomes.* No. w22502. National Bureau of Economic Research, 2016.

Dobbie, Will, and Roland G. Fryer Jr. *Getting beneath the veil of effective schools: Evidence from New York City.* American Economic Journal: Applied Economics 5.4 (2013): 28-60.

Dobbie, Will, and Roland G. Fryer Jr. “The medium-term impacts of high-achieving charter schools.” *Journal of Political Economy* 123.5 (2015): 985-1037.

Dobbie, Will, and Roland G. Fryer, Jr. *Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children’s Zone.* American Economic Journal: Applied Economics 3.3 (2011): 158-187.

Dobbie, Will. *Teacher characteristics and student achievement: Evidence from Teach For America.* Unpublished manuscript, Harvard University (2011).

DuGoff, E, S Bishop, and P Rawal. ”Hospital readmission reduction program reignites debate over risk adjusting quality measures.” Health Aff (2014).

Dynarski, S. , Hyman, J. and Schanzenbach, D. W. (2013), “Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion.” Journal of Policy Analysis and Management, 32: 692-717.

Frankl, S. E., Breeling, J. L., & Goldman, L. (1991). Preventability of emergent hospital readmission. The American journal of medicine, 90(1), 667-674.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Natalya Verbitsky-Savitz et al. “Charter-school management organizations: Diverse strategies and diverse student impacts.” Mathematica Policy Research, Inc. (2012).

Gautam, P., Macduff, C., Brown, I., & Squair, J. (1996). Unplanned readmissions of elderly patients. Health bulletin, 54(6), 449-457.

Gaynor, M. & Vogt W, .B. ”Antitrust and Competition in Health Care Markets.”In A.J. Culyer and J.P. Newhouse, eds. Handbook of Health Economics, Vol. 1B. New York:North-Holland,2000.

Gaynor, M., & Vogt, W. B. (2003). Competition Among Hospitals, 34 RAND J. Econ, 764, 764.

Gerhardt, G., Yemane, A., Hickman, P., Oelschlaeger, A., Rollins, E., & Brennan, N. (2013). Medicare readmission rates showed meaningful decline in 2012. Medicare Medicaid Res Rev, 3(2), E1-E12.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, Emily Dwoyer, and Marsha Silverberg. 2010. *The Evaluation of Charter School Impacts: Final Report.* NCEE 2010-4029. Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. http://ies.ed.gov/ncee/pubs/20104029/.

Graham, H., & Livesley, B. (1983). Can readmissions to a geriatric medical unit be prevented? The Lancet, 321(8321), 404-406.

Hasnain-Wynia, R., Baker, D. W., Nerenz, D., Feinglass, J., Beal, A. C., Landrum, M. B., et al. (2007). Disparities in health care are driven by where minority patients seek care: examination of the hospital quality alliance measures. Archives of Internal Medicine, 167(12), 1233-1239.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments. In Handbook of Econometrics, volume 6B, pages 4875–5143.

Heckman, J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. Econometrica, 73(3):669–738.

Hedges, L. V. and O'Muircheartaigh, C. (2011). Improving generalizations from designed experiments. Northwestern University manuscript.

Hoxby, Caroline M., Sonali Murarka, and Jenny Kang. *How New York City's charter schools affect achievement.* Cambridge, MA: New York City Charter Schools Evaluation Project (2009): 1-85.

Hoxby, Caroline Minter, and Jonah E. Rockoff. *The impact of charter schools on student achievement.* Cambridge, MA: Department of Economics, Harvard University, 2004.

Hu, J., Gonsahn, M. D., & Nerenz, D. R. (2014). Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health Affairs, 33(5), 778-785.

James, J. (2013). Medicare hospital readmissions reduction program. Health Affairs, 34(2), 1-5.

Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the Medicare fee-for-service program. New England Journal of Medicine, 360(14), 1418-1428.

Joynt, K. E., & Jha, A. K. (2012). Thirty-day readmissions—truth and consequences. New England Journal of Medicine, 366(15), 1366-1369.

Joynt, K. E., & Jha, A. K. (2013). A path forward on Medicare readmissions. New England Journal of Medicine, 368(13), 1175-1177.

Joynt, K. E., & Jha, A. K. (2013). Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. Jama, 309(4), 342-343.

Joynt, K. E., Chatterjee, P., Orav, E. J., & Jha, A. K. (2015). Hospital closures had no measurable impact on local hospitalization rates or mortality rates, 2003–11. Health Affairs, 34(5), 765-772.

Joynt, K. E., Orav, E. J., & Jha, A. K. (2011). Thirty-day readmission rates for Medicare beneficiaries by race and site of care. Jama, 305(7), 675-681.

Kelly, J., McDowell, H., Crawford, V., & Stout, R. (1992). Readmissions to a geriatric medical unit: is prevention possible? Aging Clinical and Experimental Research, 4(1), 61-67.

Krueger, A. B. (2003). "Economic Considerations and Class Size." The Economic Journal,113(February):F34–F63

Krueger, A. B. and Whitmore, D. M. (2001), "The Effect of Attending a Small Class in the Early Grades on College-test Taking and Middle School Test Results: Evidence from Project Star." The Economic Journal, 111: 1-28.

Krueger, Alan B. “Experimental Estimates of Education Production Functions.” *The Quarterly Journal of Economics* 114, no. 2 (1999): 497-532.

Ladd, Helen F., Charles T. Clotfelter, and John B. Holbein. *The growing segmentation of the charter school sector in North Carolina*. Education Finance and Policy (2016).

Lucas Jr, R. E. (1978). “On the size distribution of business firms.” *The Bell Journal of Economics*, 508-523.

National Alliance for Public Charter Schools, “A Growing Movement: America’s Largest Charter School Communities,” (2015).

Place, Kate, and Philip Gleason. “Do Charter Middle Schools Improve Students’ College Outcomes? Evaluation Brief. NCEE 2019-4005.” National Center for Education Evaluation and Regional Assistance (2019).

Raj Chetty, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, Danny Yagan, “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star” , *The Quarterly Journal of Economics*, Volume 126, Issue 4, November 2011, Pages 1593–1660

Romley, J. A., & Goldman, D. P. (2011). How Costly is Hospital Quality? A Revealed-Preference Approach. The Journal of industrial economics, 59(4), 578-608.

Rosen, Sherwin. (1982) “Authority, control, and the distribution of earnings.” *The Bell Journal of Economics* : 311-323.

Shih, T., Ryan, A. M., Gonzalez, A. A., & Dimick, J. B. (2015). Medicare’s Hospital Readmissions Reduction Program in Surgery May Disproportionately Affect Minority-serving Hospitals. Annals of surgery, 261(6), 1027-1031.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. Journal of the Royal Statistical Society, Series A(Part 2):369–386.

Success Academy (2016), *Success Academy Receives $25 Million Gift from the Robertson Foundation to Expand System of Top-Tier Public Charter Schools*, https:// www.successacademies.org/press-releases/success-academy-receives-25-million-gift-from-the-robertson-foundation-to-expand-system-of-top-tier-public-charter-schools/, accessed on: March 9, 2020.

The Board of Trustees of the State University of New York. *Policies for the Renewal of Not-for-Profit Charter School Education Corporations and Charter Schools Authorized by the Board of Trustees of the State University of New York*. September 4, 2013.

Tipton, E. (2013). Improving generalizations from experiments using propensity score subclassification: Assumptions, properties, and contexts. Journal of Educational and Behavioral Statistics, 38:239–266.

Town, R., & Vistnes, G. (2001). Hospital competition in HMO networks. Journal of health economics, 20(5), 733-753.

U.S. Department of Education, National Center for Education Statistics, Common Core of Data, "Public Elementary/Secondary School Universe Survey," 2008-2009

Walters, Christopher R. "The demand for effective charter schools." *Journal of Political Economy* 126, no. 6 (2018): 2179-2223.

Zhang, Dennis J. , Itai Gurvich, Jan A. Van Mieghem, Eric Park, Robert S. Young, & Mark V. Williams (2016) Hospital Readmissions Reduction Program: An Economic and Operational Analysis. Management Science

Zuckerman, R. B., Sheingold, S. H., Orav, E. J., Ruhter, J., & Epstein, A. M. (2016). Readmissions, observation, and the hospital readmissions reduction program. New England Journal of Medicine, 374(16), 1543-1551.