Title
Essays on Health, Education, and Consumer Information

Author
Heck, Brandon William

Publication Date
2016

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ESSAYS ON HEALTH, EDUCATION, AND CONSUMER INFORMATION

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Brandon Heck

June 2016

The Dissertation of Brandon Heck is approved:

Professor Carlos Dobkin, Chair

Professor Jonathan Robinson

Professor George Bulman

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Contents

List of Figures . . . vi
List of Tables . . . viii
Abstract . . . x
Acknowledgements . . . xii

1 Firm and Consumer Response to Quality Disclosure: Evidence from Hospital Grades 1

1.1 Introduction . . . 1
1.2 Conceptual Framework . . . 8
1.3 Background . . . 11
1.3.1 Hospital Safety Score Methodology . . . 12
1.3.2 Media Exposure . . . 13
1.4 Data . . . 17
1.5 Empirical Design . . . 20
1.6 Results . . . 25
1.6.1 Patient Response . . . 25
1.6.2 Hospital Response . . . 28
1.7 Conclusion . . . 33
1.A Appendix A . . . 36
1.A.1 Detailed Hospital Safety Score Methodology . . . 36
1.A.2 Exploration of Response in Survey Submission . . . 38
1.A.3 Frictions in Hospital Choice . . . 40
1.A.4 Leapfrog Group Internal Auditing . . . 42
1.B Appendix B . . . 51
1.B.1 Robustness Checks . . . 52
1.B.2 Survey Resubmission . . . 54
1.B.3 Survey Gaming . . . 56

1.B.4 Analysis of Price Changes 58
1.B.5 Analysis of the C-B Margin 59

2 The Causal Effect of Education on Health: Evidence from British Administrative Hospital Records 74

2.1 Introduction 74
2.2 Background 77
2.2.1 Compulsory Schooling Reforms in Britain 77
2.2.2 Education-Health Gradients 78
2.3 Data 81
2.4 Empirical Methodology 85
2.5 Results 87
2.5.1 Effects of the Reforms on Educational Attainment 87
2.5.2 Effects of the Reforms on Self-Reported Health - Census . 89
2.5.3 Effects of the Reforms on Hospitalizations 92
2.6 Conclusion 96
2.A Appendix A 99
2.B Appendix B 107

3 The Effect of Community Water Fluoridation on Health and Education 111

3.1 Introduction 111
3.2 Literature Review 116
3.3 Background 122
3.4 Data 125
3.5 Econometric Framework 128
3.6 Results 132
3.6.1 Checking Balance of Covariates of Public Water System Fluoridation 132

3.6.2 1986-1987 National Survey of Oral Health of US Schoolchildren 134
3.6.3 National Health and Nutrition Examination Survey III . . 143
3.6.4 Educational Achievement 146
3.7 Conclusion 148

References **151**

List of Figures

1.1 Compliance with Treatment Assignment (First Stage) 14

1.2 Hospital Safety Score Density 15

1.3 The Effect of Grade Assignment on Inpatient Admissions 27

1.4 The Effect of Grade Assignment on Healthcare-Associated Infections 30

1.5 The Effect of Grade Assignment on Scores 33

1.A.1 Transition Plot 49

1.A.2 Changes in Patient Composition 50

1.A.3 Probability of Survey Submission 51

1.B.1 Bandwidth Robustness - Main Results 65

1.B.2 Changes in Survey Score by Competition and Hospital Size 66

1.B.3 Timeline of Leapfrog Hospital Survey 67

1.B.4 Resubmitting and Non-Resubmitting Hospitals 68

1.B.5 Assessment of Outcome Data Misreporting 69

1.B.6 Changes in Hospital Costs 70

1.B.7 Changes in Inpatient Admissions - C-B Margin 71

1.B.8 Changes in Healthcare-Associated Infections - C-B Margin 72

1.B.9 Changes in Scores - C-B Margin 73

2.1 Effect of the Compulsory Schooling Reforms on Educational Attainment 88

2.2 Effect of Compulsory Schooling Reforms on All Hospitalizations and Mortality 93

2.3 Effect of Compulsory Schooling Reforms on Broad Hospitalization Categories 94

2.4 Effect of Compulsory Schooling Reforms on Specific Morbidities . . 96

2.A.1 Effect of Reforms on Health - Survey Data 105

2.A.2 Education-Health Gradient 106

3.1 Relationship Between Fluoride Level and Deciduous Cavities . . . 138
3.2 Relationship Between Fluoride Level and Permanent Cavities . . . 139
3.3 Linear Probability Model Permanent 139
3.4 Linear Probability Model Deciduous 140
3.5 Relationship Between Fluoride Level and Very Mild or Worse Fluorosis 143
3.6 Linear Probability Model Very Mild or Worse Fluorosis 144

List of Tables

1.1 Hospital Safety Score Media Exposure 16

1.2 Regression Discontinuity Estimates for Patient Sorting 28

1.3 Regression Discontinuity Estimates for Hospital Response 34

1.A.1 Hospital Safety Score Measures 44

1.A.2 Hospital Safety Score Grading Criteria 45

1.A.3 Transition Matrix 46

1.A.4 Centers for Medicare and Medicaid Services Data Reporting Periods 47

1.A.5 Leapfrog Hospital Survey Probabilities of Submission 48

1.B.1 Regression Discontinuity Robustness Estimates of Main Results . . 60

1.B.2 Robustness to Choice of Market Definition 61

1.B.3 Grade Distributions of Survey Resubmitters 62

1.B.4 Regression Discontinuity Estimates for Resubmitting and Non-Resubmitting Hospitals 63

1.B.5 Regression Discontinuity Estimates for Hospital Costs 64

2.1 Correlations Between Education and Health 80

2.2 Differences of Individual Characteristics Across Education Levels . . 82

2.3 Effects of Compulsory Schooling Reforms on Education 90

2.4 Effect of Compulsory Schooling Reforms on Self-Reported Health . . 91

2.5 Effect of Compulsory Schooling Reforms on Inpatient Admissions . 97

2.A.1 Sequential Adding of Covariates Into Education-Overall Health Gradient 100

2.A.2 Alternative Self-Reported Health Definitions 101

2.A.3 Bandwidth Sensitivity - Effects of Compulsory Schooling Reforms on Education 102

2.A.4 Bandwidth Sensitivity - Effect of Reforms on Self-Reported Health . 103

2.A.5 ICD Codes of Analyzed Outcomes from Hospital Episode Statistics . 104

2.B.1 Data Dropped in Cleaning Process 110
3.1 Balance Check - Rural 133
3.2 Balance Check - Urban 134
3.3 Effect of Community Water Fluoridation on Cavities 136
3.4 Effect of Community Water Fluoridation on Fluorosis 141
3.5 NHANES Oral Health Results 145
3.6 NHANES General Health Results 146
3.7 Effect of Community Water Fluoridation on Educational Achievement 147

Abstract

ESSAYS ON HEALTH, EDUCATION, AND CONSUMER INFORMATION

Brandon Heck

This dissertation focuses on using causal inference techniques to assess the impact that public policies have on health outcomes, and the behavioral response of consumers and firms to consumer information. In two of the three chapters of this dissertation, I rely on quasi-experimental variation, namely regression discontinuity designs, to find causal estimates of the determinants of health and the behavioral responses to public health policies and ratings systems.

The first chapter examines the behavioral responses of patients and hospitals to public information disclosure about hospital quality. Specifically, I analyze the assignment of letter grades to hospitals by the non-profit consortium known as the Leapfrog Group. Letter grades are assigned based on sharp cutoffs of a continuous score, so I implement a regression discontinuity design which provides causal estimates of whether patients sort into hospitals with a better grade and if lower-rated hospitals improve quality relative to higher-rated hospitals in order to achieve a higher grade in the next period. Using inpatient discharge records from 7 states, I find evidence that lower-rated hospitals reduce their rate of healthcare-associated infections and patients sort into hospitals that receive the highest grade. These responses occur nearly entirely in competitive hospital markets where patients

have more choice and hospitals have a higher incentive to improve, and in large hospitals, where hospitals have a higher capacity to improve.

The second chapter exploits two compulsory schooling reforms in England to find estimates of the causal effect of education on health. We use a regression discontinuity design to compare individuals that were just barely subject or not subject to the reforms, and examine the differences in their long term health outcomes. Using hospital records from the National Health Service of England, we do not find any causal effect of an additional year of high school education on mortality or morbidity.

The third chapter analyzes the effect of one of the largest public health interventions in the United States, the artificial fluoridation of community water systems, on dental health, general health, and education. In this chapter, I merge several surveys and administrative data sets to obtain a rich set of controls, a fixed effects model to examine the effect of fluoride on educational achievement, and non-parametric specifications to assess how fluoride affects dental health at different levels of health and fluoride levels. I find consistent evidence that fluoridated water has a significant impact on the prevention of dental cavities in baby teeth, but no robust effect on permanent teeth or on general health or educational attainment. I do find however that fluoridated water significantly increases mild or severe cases of fluorosis.

Acknowledgements

I would not have finished my dissertation without the support of my friends, family, and committee members. I am grateful to my committee members: Carlos Dobkin, Jonathan Robinson, and George Bulman for the constant encouragement, support, and guidance throughout the entire process of carrying out these projects. I would like to extend a special thanks to George Bulman for the lengthy discussions and support, and spending a considerable amount of time reading and re-reading my paper, and providing invaluable detailed feedback (not to mention presentation and interview preparation!).

I would also like to thank my girlfriend, Cherie, who was not involved until the final stages of my dissertation, but was invaluable to my progress both personally and professionally. Thank you for editing my paper, having endless discussions at all times of the day and night, and supporting not only my presentation and writing skills but my mental health throughout this process. I am sure that I would not be where I am without you, thank you!

Finally, thank you to my parents, Laura and Tim Heck and brother Josh Heck for your endless support. I would never have made it to this point without your endless love and support of me and my endeavors.

Chapter 1

1 Firm and Consumer Response to Quality Disclosure: Evidence from Hospital Grades

1.1 Introduction

Each year, between 100,000 and 400,000 Americans die due to preventable medical errors, making this the third leading cause of death in the United States (James 2013 and IOM 1999). The cost to a patient of low quality care is potentially very high, and the complexity of healthcare results in high levels of information asymmetry.[1] Theoretical work has shown that evaluation and disclosure of firm quality can mitigate information asymmetry, inducing quality improvements and allowing consumers to identify high quality firms.[2] As a result, information disclosure has been implemented in a variety of industries, including environmental rankings and emission reports, food labeling, restaurant hygiene grades, school grades, and healthcare report cards. Existing empirical work provides mixed evidence about the costs and benefits of public disclosure.[3]

Patients may face substantial uncertainty with respect to the risks they face

[1]Information asymmetries are particularly pronounced in healthcare for several reasons. The complexity of healthcare and medical diagnoses make it difficult for consumers to understand the quality of their treatment relative to the treatment they would have received with a different physician or at a different facility. Additionally, the majority of patients only receive treatment for a particular condition once, compounding the problem of identifying the quality of the treatment. These constraints make it very difficult for consumers to "shop around" for the highest quality as they do in other industries.

[2]See Arrow (1963), Akerlof (1970), Nelson (1970), Shapiro (1982), Chan and Leland (1982, 1986), Dranove and Satterthwaite (1992), and Fishman and Hagerty (2003).

[3]See Marshall et al. 2000, Werner and Asch 2005, and Dranove and Jin 2010 for literature reviews.

from treatment at different hospitals. Thus, patients may not choose hospitals based on risk and hospitals may have a limited profit motive to improve quality since demand is inelastic. In this setting, the introduction of a rating system may induce patients to sort to firms with perceived higher quality and, as a result, hospitals may improve their performance in order to earn a better rating and attract patients. However, the causal effect of ratings are difficult to identify due to endogeneity in the cross-section and mean-reversion in panel data. To obtain causal estimates, I note that the rating system analyzed in this study categorizes hospitals using sharp cutoffs in a continuous measure. Thus the context is ideal for a regression discontinuity design. If categories provide distinct signals of quality which differ from prior beliefs and update consumers' information sets, we should expect that patients will choose to receive treatment at the hospital with higher perceived quality. If hospitals seek to establish or maintain a good reputation through achieving a favorable rating, lower rated hospitals will improve quality measures to improve their standing.

This paper examines the response of hospitals and patients to the assignment of letter grades using the first program of its kind in the United States. The Hospital Safety Score, introduced in June 2012, was implemented by a non-profit coalition of large employers known as the Leapfrog Group, and combines medical data from the Centers for Medicare and Medicaid Services (CMS) with a hospital survey in order to generate a letter grade for most acute care hospitals in the United States.[4]

[4]The Leapfrog Hospital Survey is an annual voluntary survey, which asks a range of questions

This information is used to generate an overall performance index that is updated semiannually and from which grades are assigned based on distinct score cutoffs. I exploit the sharp nature of grade assignments to examine the causal response of hospitals and patients to letter grades using a regression discontinuity design.[5] I combine data on all 28 measures used to construct the score with Medicare Cost Report data, CMS outcome and process measures, and inpatient discharge records to examine hospital and patient response.[6] These linked data make it possible to examine if hospitals respond to a lower grade by improving performance and if patients choose to receive treatment at hospitals with better grades.

Hospital demand with respect to quality information will be relatively inelastic for patients living in rural areas with limited hospital choice. As a result, hospitals facing little competition will have less incentive to improve than those in competitive markets. Therefore, I consider patient sorting in markets with differing levels of competition as measured by the Herfindahl-Hirschman Index (HHI), a widely used measure of market concentration.[7] I also separately examine hospital response for large urban and small rural hospitals as they face very different economic environments and have fundamentally different management

including hospital demographics, hygiene, leadership, staffing, and readmission rates.

[5]Similar designs have been used to analyze the effects of publicly available firm quality information on firm and consumer behavior by Anderson and Magruder (2012), Ramanarayanan and Snyder (2012), and Darden and McCarthy (2015).

[6]Process measures are how often a hospital gives patients recommended treatment for a given medical condition or procedure. Outcome measures are what happens to a patient during care.

[7]See Melnick, Zwanziger 1988, Kessler and McClellan 2000, Cooper et al. 2011 for uses of the HHI as a measure of hospital competition. See for example, Luft et al. 1990, Adams et al. 1991, Adams and Wright 2008, Sivey 2011 for distance as an important factor in hospital choice.

structures.[8] Most large hospitals are located in competitive urban markets, so large hospitals may have a higher capacity and incentive to improve quality as a response to their grade.[9]

In order for hospitals to have a strong incentive to improve quality, patients must respond to quality information by shifting to hospitals with higher perceived quality. To assess how grades affect patient behavior, I analyze whether nearly identical hospitals with different grades experience differential changes in patient volume. Notably, I differentiate the results between elective and non-elective admissions, as patients are able to use quality information in their hospital choice only when their visit is not an emergency. I find an 8 percent increase in elective admissions for "A" hospitals relative to nearly identical "B" hospitals in competitive markets, but no significant sorting response in low competition markets. So, patients respond to receiving the strongest quality signal about a hospital in areas with greater hospital choice.

An important question in assessing the impacts of public disclosure policies is whether there are asymmetric effects depending on whether the information about the firm is positive or negative. I construct a novel data set to show that media exposure for "A" hospitals is greater than for any other grade, creating a stronger incentive for hospitals to cross the threshold from "B" to "A" than from

[8]See Bloom et al. (2013) for a discussion of the variation in hospital management practices.

[9]About 71 percent of hospitals with above the median number of inpatient beds are located in markets with above the median level of competition. Additionally, about 88 percent of hospitals in the top quartile of number of inpatient beds are located in the markets in the top quartile of competitiveness.

“C” to “B,” and for patients to choose to receive treatment at “A” rather than “B” hospitals. Unfortunately, the very low density of “F” and “D” hospitals does not allow precise estimation of the effect of negative news about hospitals.

My results are consistent with studies that find evidence of consumers shifting from healthcare institutions perceived to be of low quality to those perceived to be of higher quality (Romano and Zhou 2004, Dranove et al. 2008, Scanlon et al. 2008, Pope 2009, Schut et al. 2012). However, several studies find no evidence of an increase in overall patient volume for higher rated facilities (Mennemeyer et al. 1997, Howard and Kaplan 2006, Ramanarayanan and Snyder 2012), and others find evidence of only a decrease in market share for the lowest rated facilities (Cutler et al. 2004, Dranove and Sfekas 2008, Dranove et al. 2012). Public disclosure of quality information has also increased demand for higher quality firms across a variety of industries and products in non-healthcare settings.[10] Using a variety of techniques, the literature on health plan choice finds that the dissemination of report cards about health plans updates consumers’ information sets and impacts consumer choice, although the effects are generally small (Tai-Seale and Wedig 2001, Beaulieu 2002, Chernew, McLaughlin, Scanlon, and Solon 2002, Jin and Sorensen 2006, Chernew, Gowrisankaran, Scanlon 2007, Dafny and

[10]Public disclosure has increased demand in industries such as cereal (Ippolito and Mathias 1990), wine (Hilger et al. 2011), fish (Hallstein and Villas-Boas 2010), fertility clinics (Bundorf et al. 2009), restaurants (Jin and Leslie 2003, Anderson and Magruder 2012), and education (Figlio and Lucas 2004). Notably, Anderson and Magruder (2012) use a regression discontinuity design to assess whether Yelp ratings have an impact on restaurant demand, and find that a higher star rating does decrease reservation availability at peak dinner hours.

Dranove 2008, Darden and McCarthy 2015).

I find evidence that lower-rated hospitals responded by improving quality relative to higher-rated hospitals. Specifically, using inpatient discharge records, I find that “B” hospitals significantly reduced their rate of healthcare-associated infections relative to nearly identical “A” hospitals. These improvements in quality are driven entirely by large hospitals which may have a higher capacity for improvement and tend to be located in competitive markets.

I find that the response of hospitals is consistent with a large body of literature which finds that disclosure of quality information drives improvement in low-quality firms.[11] There is some evidence that this also holds true for low-rated healthcare institutions (Hannan et al. 1994, Hauck and Street 2007, Propper et al. 2010, Lu 2012, Ramanarayanan and Snyder 2012). However, the majority of these studies use difference-in-differences strategies, which must be interpreted with caution since mean reversion is highly prevalent in this context and a parallel trends assumption is required.[12] Techniques that address mean reversion and

[11]Empirical work has examined disclosure of quality information about firms and government entities in a variety of non-healthcare settings, including schools, finance, manufacturing, and restaurants. A number of papers have found that disclosure policies have prompted response by firms to improve quality (Ippolito and Mathias 1990, Cohen and Konar 1997, Afsah et al. 2000, Carnoy and Loeb 2002, Jin and Leslie 2003, Hanushek and Raymond 2005, Scorse 2007, Powers et al. 2008, Bennear and Olmstead 2008, Chatterji and Toffel 2010), although some have argued that disclosure can be harmful, particularly to public organizations, because firms receiving poor ratings look less attractive to external funders and high quality employees and students (Espeland and Sauder 2007). Additionally, many studies that examine consumer reviews and evaluations for companies such as Ebay, Amazon, and Yelp find that consumers do respond to online reviews and evaluations (Alm and Melnik 2002, Cabral and Hortacsu 2004, Chevalier and Mayzlin 2006, Jin and Kato 2006, Lockwood et al. 2006, Bryan et al. 2007, Moretti 2011, Anderson and Magruder 2012).

[12]The overstatement of effects in difference-in-difference designs with mean reversion was first discussed by Chay et al. (2005).

require weaker assumptions that are more likely to hold, such as a regression discontinuity design, provide a valuable alternative.

A common concern is that firms will engage in gaming in response to the publication of quality measures. The gaming can take the form of shifting resources or manipulating scores to improve in the areas that are publicly reported, possibly at the expense of quality in other areas.[13] Studies have found gaming to be prominent in the healthcare setting, with hospitals often shifting resources toward rating-relevant measures and away from other important tasks that do not affect the rating (Omoigui 1996, Schneider and Epstein 1996, Mannion et al. 2005, Bevan and Hood 2006, Hood 2006, Dranove et al. 2008, Besley et al. 2009, Lu 2012).[14] I assess explicit gaming of the survey by comparing central line-associated blood stream infection (CLABSI) rates reported by hospitals on the survey with those reported to the CMS. Interestingly, I do not find evidence of misreporting outcome data on the survey.

The paper proceeds as follows. Section 2 provides the context for understand-

[13]In the non-healthcare setting, gaming is particularly prevalent in school accountability programs (Haney 2000, 2001, Figlio and Gertzler 2002, Jacob and Levitt 2003, Jacob 2005, Cullen and Reback 2006). These papers find that although schools do respond to accountability systems by improving their scores on the measures they are being graded on, they often show no improvement or even regression on unobserved outcomes.

[14]Lu (2012) finds that due to a program in the nursing home industry which reported some aspects of quality and not others, nursing homes improved along the reported dimensions, but only at the detriment of the unreported dimensions. A rating system in England which assigned between 0 and 4 stars to each hospital trust and was based primarily on waiting time targets was found to improve waiting times, but evidence of waiting list manipulation and deterioration of other aspects of quality were discovered (Mannion et al. 2005, Bevan and Hood 2006, Hood 2006, Besley et al. 2009). There is some evidence that a public release of mortality data in New York in 1989 led to hospitals and physicians purposely denying higher risk patients as a result of the disclosure (Omoigui 1996, Schneider and Epstein 1996, Dranove et al. 2008).

ing the behavior of patients and hospitals. Section 3 discusses the institutional background of the Leapfrog Group, the Leapfrog Hospital Survey, and the media exposure of the Hospital Safety Score. Section 4 describes each of the data sources used in the paper. Section 5 details the identification strategy. Section 6 presents the results of the effect on firm and consumer response. Section 7 discusses the results and conclusions.

1.2 Conceptual Framework

I model patients as initially having substantial uncertainty about the quality of hospitals and the risks that they will face from receiving treatment. Particularly, patients likely know the extremes of the quality distribution in their local area but are not aware of the relative quality of mid-tier hosiptals. For example, patients may know that the local Kaiser hospital in their area is higher quality than the public county hospital, but are uncertain about the relative quality of other local hospitals. Compounding this, due to the complexity of healthcare it is difficult for patients to assess the quality of their own treatment, therefore even when they have experiences at these hospitals they are unable to assess the conterfactual of their treatment. In this study, patients receive information via the assignment of letter grades "A," "B," "C," "D," or "F" to hospitals. If grades provide distinct signals of quality and these signals differ from prior beliefs, patients' information sets are updated, reducing uncertainty about hospital quality and inducing patients to receive treatment at the hospital which they perceive to be of higher quality.

Hospitals may respond to their own grade assignment by improving quality for several reasons. Primarily, if hospitals believe that patients are choosing where to receive treatment based on grades, competitive pressures will lead to improvements in quality by hospitals that receive a lower grade. It is also possible that even without the knowledge that patients are sorting based on grade, administrators and healthcare workers may wish to receive a higher grade to establish or maintain a good reputation. Due to enhanced media exposure and esteem for the highest grade, the differential incentive to move from a “B” to an “A” is larger than between any other adjacent grades.

This study identifies the causal effect of letter grades by leveraging differences in changes in outcomes between hospitals of nearly identical quality but which receive adjacent letter grades. In order to generate differential changes in outcomes, it must be the case that hospitals with adjacent grades face differential incentives to improve. If hospitals strive to receive an “A” in each period and know their location in the distribution of the score, then there would be no differential incentive for “A” and “B” hospitals to improve. However, there is some uncertainty among hospitals about where exactly they lie in the distribution because the overall score is not publicly reported and the grade thresholds move each period, requiring either a manual calculation or small payment to obtain their score every six months. This uncertainty leads to some hospitals that barely received an "A" believing their score is worse than it actually is and hospitals that barely did not receive an "A" believing that their score is better than it actually is, driving a differential incentive between "A" and "B" or any two other adjacent grades.[15] For the hospital response results where this point is salient, if hospitals actually do have perfect or close to perfect information, this means that my estimates are attenuated, so my estimates can be thought of as a lower bound on the true effect.

An additional point to consider is that all of the information used to construct the score was publicly available information even before the program began, so any effect of letter grades must be due to limited attention (also known as a comparison friction) by patients.[16] If patients respond to letter grades, then this implies that at least some consumers are not able or willing to use the voluminous and complex information presented to them in a disaggregated format, which is contrary to the assumptions of many traditional economic models.

The response of patients and hospitals are likely heterogeneous based on the competitive environment that the hospitals face. Patients in a rural environment with minimal hospital choice have little ability to respond to ratings. However, patients in urban areas with competitive hospital markets have substantially more

[15]Suppose that the threshold for an "A" is a 90 and hospitals received a "B" if their score is between 80 and 89 and an "A" if their score is between 90 and 99. Then, for example, hospitals receiving a 90 may believe that they are actually a 95, thereby putting less effort into improving because their grade next period is "safe." Additionally, a hospital receiving an 89 may believe they are actually an 85, which will likely lead them to increase effort to improve.

[16]See for example, Hirshleifer and Hong Teoh (2003) and Kling et al. (2012).

choice and will have the opportunity to review quality measures and choose where they receive treatment accordingly. Therefore, the incentive for hospitals to improve their grade is increasing with the number of competitors. More competitive markets are comprised not only of many hospitals, but also contain relatively larger hospitals, since hospital density is correlated with size.

While competitive markets incentivize hospitals to improve quality, hospital size impacts their capacity to improve. There are several reasons why large hospitals might react more strongly to grades than smaller hospitals. Bloom et al. (2014) show that larger hospitals have higher management quality, which they suggest could be due to economies of scale since management practices have a fixed cost element. Additionally, larger hospitals require more formalized management practices since informal management practices decrease in effectiveness as the size of the hospital grows. Therefore, hospitals that are larger and in competitive markets will have both an increased incentive and capacity to improve their grade.

1.3 Background

In 1998, a group of large employers came together to discuss how they could improve healthcare quality and affordability. A report by the Institute of Medicine in 2000 found that up to 98,000 Americans die every year from preventable medical errors in hospitals, giving the group a primary purpose (IOM 2000). In November

2000, the Leapfrog Group was officially launched with a focus on patient safety. In 2001, the Leapfrog Group launched the Leapfrog Hospital Survey which is filled out by hospitals voluntarily, is not tied to any monetary rewards or penalties, and is designed to assess both the quality and safety of hospitals.

In 2012, the Leapfrog Group, guided by a panel of nine experts, implemented a new system using 28 measures of hospital safety and quality, where hospitals receive a single score every six months based on a weighted sum of the 28 measures, called the Hospital Safety Score.[17] Then, based on the number of standard deviations from the national mean in a given six month rating period, hospitals are assigned a letter grade, "A," "B," "C," "D," or "F." The methodology and development of the Hospital Safety Score is described in Austin et al. (2014). A summary of the assignment of grades is described in the following subsection.

1.3.1 Hospital Safety Score Methodology

The Hospital Safety Score is a weighted sum of 28 different measures, 12 of which are taken from the survey and the other 16 of which are measures constructed from publicly available data, primarily from the Centers for Medicare and Medicaid Services (CMS).[18] Of these, 15 are process of care or structural measures and 13 are outcome measures. Ten of the process measures and two of the outcome

[17] The Hospital Safety Score actually began with 26 measures. In October 2013, measures for Catheter-Associated Urinary Tract Infection and Surgical Site Infection for Major Colon Surgery were added to the Hospital Safety Score, changing the total number of measures to 28.

[18] A complete explanation of the methodology used by Leapfrog to calculate the Hospital Safety Score can be found in Austin et al. (2014). Appendix Section A1 contains a brief summary of the methodology used to construct scores and assign grades.

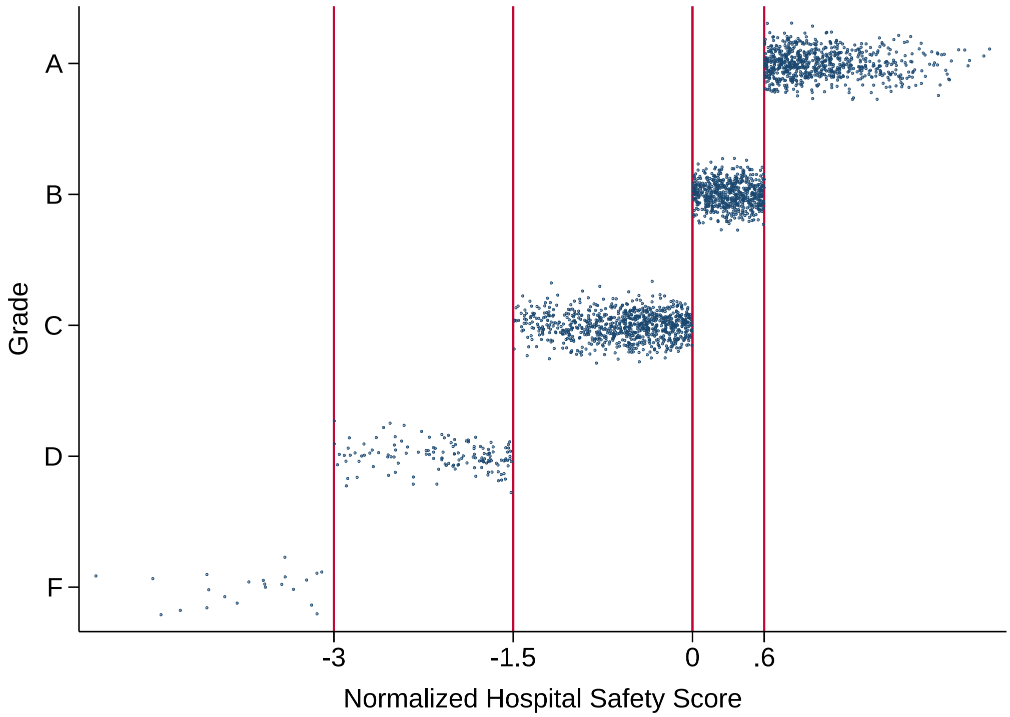
measures are from the survey. A list of all measures used to compute the Hospital Safety Score is provided in Table A1. An example of the criteria for the assignment of grades from Spring 2014 is shown in Table A2, and Figure 1 shows that there is perfect compliance with the assignment rule. Panel A of Figure 2 presents the distribution of scores, showing where the thresholds were in Spring 2014 and that the overall score is a continuous measure. Panels B and C of Figure 2 show that there is no bunching at either the B-A or C-B margins, the only margins examined in this study due to the very low density of hospitals that received grades of "D" and "F."[19] This is strong evidence that the regression discontinuity assumption (that hospitals cannot precisely manipulate which grade they receive) is satisfied.

1.3.2 Media Exposure

The first order question when considering the effects of a rating system is whether firms and consumers are aware of the program at all. I find that in fact, the Hospital Safety Score is widely cited and discussed by both mainstream news outlets as well as many individual hospitals and local news organizations. To assess the media exposure of the Hospital Safety Score, I constructed a data set by searching the websites of hospitals and news outlets to discover any articles or news reports related to the grades of hospitals. Table 1 shows summary statistics

[19]It is very unlikely that there would be precise manipulation over a threshold for two reasons. First, hospitals do not know what their total outcome score will be, nor can they exactly manipulate it to achieve a particular grade. Second, the threshold moves each rating period, so a hospital does not know the exact threshold ex-ante.

Figure 1.1: Compliance with Treatment Assignment (First Stage)



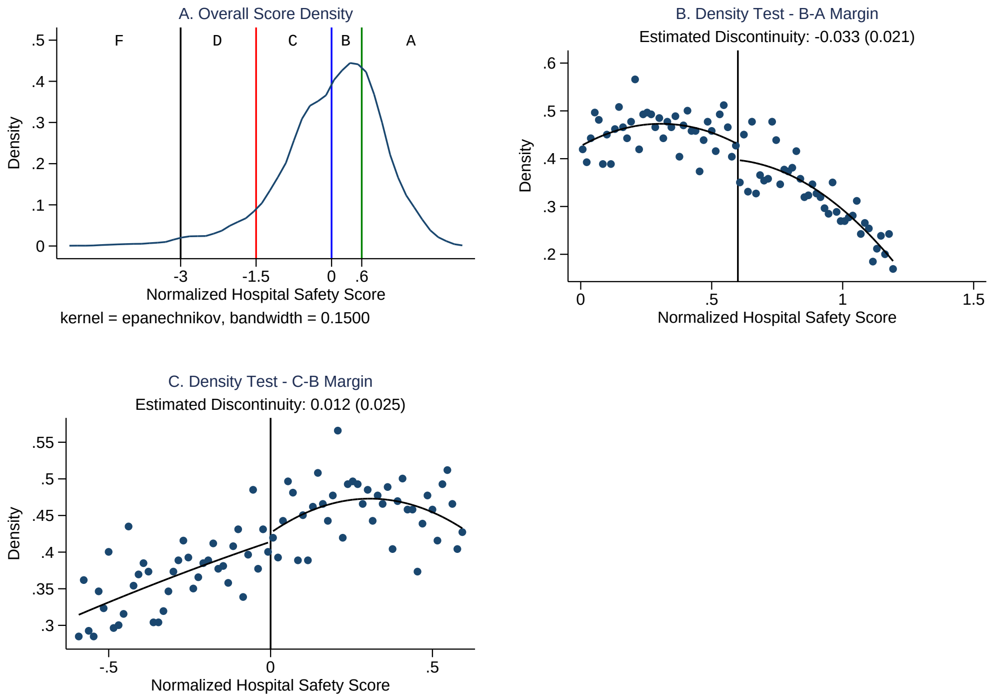
Notes: This graph presents the compliance with treatment assignment. Points represent one hospital and are jittered vertically by adding noise drawn from a normal distribution with mean 0 and standard deviation of 0.1. The Hospital Safety Score is normalized so that the values correspond to standard deviations from the national mean in Spring 2014. Data comes from the Leapfrog Group for Spring 2014.

of the media exposure of a selection of 208 hospitals.[20]

"A" hospitals receive a large amount of attention in the media. Specifically, as shown in column (1), 77 percent of "A" hospitals have some mention of their grade either by themselves or in the local media. About 60 percent of all hospitals that receive an "A" advertise this achievement, usually through articles displayed on their website. Additionally, 53 percent of hospitals had their grade reported in an article by a local newspaper or magazine discussing the grades of all local hospitals. About 30 percent of hospitals receiving an "A" also had articles written

[20]Hospitals were selected based solely on being close to the B-A threshold in Spring 2014.

Figure 1.2: Hospital Safety Score Density



Notes: The kernel density estimates in Graph A show the density of Hospital Safety Scores and is estimated with a bandwidth of 0.15 and an epanechnikov kernel. Graphs B and C show McCrary density tests at the B-A and C-B margins, respectively. The Hospital Safety Score is normalized so that the values correspond to standard deviations from the national mean in a given rating period. The lines are fitted values from regressions that include a second-order polynomial in distance fully interacted with a dummy for being above the threshold. Data comes from the Leapfrog Group for Spring 2014.

from local newspapers specifically about their hospital receiving an “A.”

Column (2) shows that “B” hospitals never produced media reports or articles specifically about the grade of their hospital or received local media attention specifically about the grade of their hospital. They receive only 1 of the 3 types of media coverage (articles about the grades of all local hospitals), and by definition, receive the same amount of coverage as “A” hospitals from that source. Therefore, “B” hospitals receive any kind of media attention about 54 percent of the time, whereas “A” hospitals receive any media attention close to 80 percent of the time.

Table 1.1: Hospital Safety Score Media Exposure

	"A"	"B"
	(1)	(2)
Percentage of Hospitals with any mention in the media	77.09 percent	53.80 percent
Percentage of hospitals advertising their grade in their own media	59.84 percent	0 percent
Percentage of hospitals with a local news source with an article about all local grades	53.80 percent	53.80 percent
Percentage of hospitals with a local news source with an article about that specific hospital	30.87 percent	0 percent
Mean number of local news articles with discussion of all local grades	0.94	0.94
Mean number of local news articles specifically about a hospital	0.62	0

Notes: This table displays summary statistics for an evaluation of media exposure of 208 hospitals. Column (1) presents summary statistics for hospitals that received an "A" in Spring 2014. Column (2) presents summary statistics for hospitals that received a "B" in Spring 2014.

However, this does not represent the entire difference in media coverage, since "A" hospitals have not only a higher probability of receiving any media attention, but are able to receive more than one type. Specifically, 38 and 14 percent of hospitals have 2 and 3 different types of media attention, respectively, but "B" hospitals have only 1 at most. A final point is that "A" hospitals may additionally advertise their achievement through television or other non-online media which I have not recorded. This gives credence to the idea that hospitals not only have an incentive to get an "A," but have little incentive to transition between any other two grades.

In addition to local news sources, Leapfrog has been featured in articles by major news organizations such as MSNBC, The New York Times, the Wall Street Journal, AARP Magazine, Forbes, and national public health websites such as

Health Leaders Media. Given the widespread attention by hospitals themselves, local newspapers, and mainstream media outlets, it is plausible that a non-trivial proportion of patients are aware of the Leapfrog Group and the Hospital Safety Score.

1.4 Data

I use data from the Leapfrog Group on all 28 measures that are used to compute the Hospital Safety Score for each hospital that was rated in each of the seven rating periods. In order to identify the hospitals with a larger capacity and incentive to respond to the grades, I run separate analyses for "large" and "small" hospitals, which I determine using the number of inpatient beds in the hospital, obtained from the Center for Medicare and Medicaid Service's (CMS) Medicare Cost Report data. To assess explicit gaming of the survey, I obtain data on central line-associated blood stream infection (CLABSI) rates from CMS Hospital Compare. Finally, to examine patient sorting and the hospital response to healthcare-associated infections, I obtain hospital records from Arizona, California, Florida, Kentucky, New Jersey, New York, and Washington.[21]

The primary data set comes directly from the Leapfrog Group, and contains data for the seven waves between Spring 2012 and Spring 2015. Specifically,

[21]These states represent 33 percent of the population of the United States. Individual level records aggregated by hospital, admission type, and admission month and year were obtained for California was obtained because quarter of admission is masked in California's public files starting in 2012. Individual level data for all other states was aggregated to the same level.

the data contains the CMS number (which is used to link all of the data sets), hospital name, address, city, state, zip code, and geographic coordinates of each rated hospital, along with their overall score, grade, and score for each of the 28 measures that are used to make up the Hospital Safety Score.[22]

I use the Medicare Cost Report Data to obtain the number of beds of each hospital, and total costs and charges in a given fiscal year. The total costs and charges are used to compute cost-to-charge ratios for each hospital, which are then used to compute hospital costs per month using inpatient discharge records.[23] Medicare-certified institutional providers are required to submit an annual cost report to a Medicare Administrative Contractor and this information is collected and made publicly available by CMS. The cost report contains provider information such as facility characteristics, utilization data, cost and charges by cost center (in total and for Medicare), Medicare settlement data, and financial statement data. CMS maintains the cost report data in the Healthcare Provider Cost Reporting Information System (HCRIS).

To assess the accuracy of outcome reporting on the survey, I compare CLABSI rates as reported to CMS through the National Healthcare Safety Network System (NHSN) to the rate reported on the survey. Outcome data for the Hospital

[22]The Hospital Safety Score actually began with 26 measures. In October 2013, measures for Catheter-Associated Urinary Tract Infection and Surgical Site Infection for Major Colon Surgery were added to the Hospital Safety Score, changing the total number of measures to 28.

[23]I choose to analyze deflated charges rather than listed charges because charges are often several orders of magnitude larger than the price hospitals actually charge to insurance companies or individuals, so deflating the charges more accurately reflects the revenue of the hospital.

Safety Score is taken from CMS Hospital Compare, which obtains its data on healthcare-associated infections from the NHSN tool. NHSN is an online tool that allows hospitals to report, track, and compare instances of healthcare-associated infections.

Finally, I utilize state-specific records on inpatient admissions, collected invidually from the health departments directly or indirectly through the Healthcare Cost and Utilization Project's (HCUP) State Inpatient Databases.[24] Individual level hospital records were collected, which allow me to calculate the total number of admissions by admission type in each month as well as the total rate of healthcare-associated infections.[25] The records were provided by the Arizona Department of Health Services, California Office of Statewide Health Planning and Development, Florida Agency for Healthcare Administration, Kentucky Department for Public Health, New Jersey Department of Health, (via the Statewide Planning and Research Cooperative System), New York State Department of Health, and the Washington State Department of Health (via the Comprehensive Hospital Abstract Reporting System). These files represent admissions at licensed general acute care hospitals. The data sets exclude individuals that are admitted to federally regulated hospitals such as Veterans Health Administration hospitals.

[24] Arizona, Kentucky, New Jersey, and New York data were collected through HCUP, and all other states were obtained from their respective state health departments.

[25] The only exception is California which does not contain quarter or month of admission in the public use files starting in 2012. Therefore, I obtained a file aggregated by hospital, admission type, and admission month and year, which was used for the patient sorting analysis.

1.5 Empirical Design

Many studies examining the effect of public firm information disclosure have used cross-sectional regressions (Luft et al. 1990, Elsbach and Kramer 1994, Carnoy and Loeb 2002, Jacobs 2006, Schut et al. 2012), difference-in-differences (DiD) designs (Afsah et al. 2000, Powers et al. 2008, Chatterji and Toffel 2010), or before and after comparisons (Ippolito and Mathias 1990, Hannan et al. 1994, Cohen and Konar 1997, Peterson et al. 1998, Romano and Zhou 2004). Cross-sectional estimation does not lead to causal estimates because it cannot control for unobserved time-invariant characteristics, and thus cannot credibly measure the impact of disclosure. For example, Schut et al. (2012) find that publicly reported lower readmission rates are associated with an increase in demand, but this correlation may simply reflect that better hospitals receive a higher market share, not necessarily as a result of disclosure. Additionally, before and after comparison of means may over or understate the true effect of disclosure if firm quality or market share would have changed similarly in the absence of disclosure. Thus, much of the literature on disclosure uses DiD designs. However, mean reversion is problematic because firms receiving a bad rating may have received a bad draw of patients and will naturally revert back to their true quality in subsequent periods. As shown by Chay et al. (2005), DiD estimates can result in very biased estimates in the presence of mean reversion, but this problem can be overcome by using a regression discontinuity (RD) design.

Mean reversion is caused by a correlation between treatment assignment and the difference between post-intervention and pre-intervention error terms, which manifests as a trend with a slope approaching -1 as this correlation increases. In this study, the slope approaches -1 as the variance due to patient heterogeneity increases, the variance due to hospital-level shocks increases, and the hospital gets smaller. This is intuitive because a hospital with only a very small number of patients is more likely to have an error which will cause them to have a particularly high error rate, resulting in a low score. RD designs fix this problem by controlling for the trend in the underlying assignment variable and analyzing only discrete changes at the cutoff.

I exploit the assignment of grades via sharp cutoffs of the Hospital Safety Score by using an RD design to find the causal effects of grade assignment on patient sorting and changes in hospital quality over time.[26] I examine whether nearly identical hospitals that receive adjacent grades experience differential changes in patient volume and if hospitals that receive a lower grade improve quality measures relative to hospitals that receive the next highest grade. The model is specified as follows:

[26]RD designs have been used to assess the causal impact of disclosure on demand and firm quality improvements by Anderson and Magruder (2012), who looked at the change in restaurant demand due to Yelp ratings, Ramanarayanan and Snyder (2012) who looked at the change in demand and change in rankings for dialysis facilities after they are placed into mortality categories, and Darden and McCarthy (2015) who looked at star ratings for Medicare Advantage plans.

$$Y_{h(t+1)} - Y_{ht} = \beta_0 + \beta_1 GRADE_{ht} + f\left(S_{ht}\right) + \epsilon_{ht} \tag{1}$$

where S_{ht} is the index score, centered at the threshold between "C" and "B" or "B" and "A" for hospital h in rating period t. The dependent variable $Y_{h(t+1)} - Y_{ht}$ is the difference between either the overall score, survey score, or healthcare-associated infection rate between period $t+1$ and period t, or the percentage change in inpatient admissions between period t and period $t+1$.[27] $GRADE_{ht}$ takes on a value of one if $S_{ht} > 0$, and zero otherwise. Then, β_1 is the estimate of the discontinuity, the causal effect of receiving a score just over the cutoff. The function $f(.)$ captures the underlying relationship between the index score and the dependent variable, and consists of a polynomial in S_{ht} fully interacted with $GRADE_{ht}$. Some variables are normalized to the mean and standard deviation of the variable so that discontinuity estimates can be interpreted in standard deviations.[28]

To examine clinical outcomes, I use inpatient discharge data to examine changes in healthcare-associated infections (HAI) as a response to grade assignment.[29] I

[27] The "Overall Score" or "Index Score" is the Hospital Safety Score, that is, a weighted sum of all 28 measures. The "Survey Score" is defined as the weighted sum of all 12 variables on the survey. The weights are the weights to be used for a hospital with no missing measures (as shown in the Hospital Safety Score "Scoring Methodology" document). However, applying different weights to different hospitals or applying no weights gives very similar estimates.

[28] The normalized variables are healthcare-associated infections, overall score, survey score, Charlson comorbidity index, and average age.

[29] I define a patient as having an HAI if they received any of the following during their inpatient stay: catheter-associated urinary tract infection (CAUTI), central line-associated bloodstream infection (CLABSI), surgical site infection (SSI), ventilator-associated pneumonia (VAP), or a methicillin-resistant staphylococcus aureus infection (MRSA).

construct HAI rates as the total number of HAIs in a given hospital-month divided by the total number of admissions in that hospital-month, and examine the change in these rates between six month rating periods. The primary concern with comparing infection rates across hospitals is that the composition of patients is changing as a response to grades. I do not find evidence that the average health of patients as measured by the Charlson comorbidity index, the average age of patients, and a particularly high risk patient population, the number of individuals over 70 years old, are changing at the B-A margin, as shown by Figure A2. This provides evidence that patient composition is not changing as a response to grades.[30]

I consider two dimensions of heterogeneity that are likely to influence hospital and patient response: hospital size and competitive environment. There are several reasons to believe that large hospitals might react more strongly to grades than smaller hospitals. Bloom et al. (2014) show that larger hospitals have higher management quality, which they suggest could be due to economies of scale since management practices have a fixed cost element. Additionally, larger hospitals require more formalized management practices since informal management practices decrease in effectiveness as the size of the hospital grows. Specifically, larger

[30]Hypothetically, patient composition could be altered through two mechanisms that act in opposite directions. First, patients that are sicker and more vulnerable to infections may be more likely than more healthy patients to pay attention to ratings and to choose to receive treatment at "A" hospitals. Acting in the reverse direction, it is also possible that more conscientious individuals that pay particular attention to ratings and may be more likely to choose an "A" hospital may also pay more attention than the average individual to ensuring that they do not receive an infection. They may do this through personal hygiene as well as ensuring that doctors and nurses are performing the tasks necessary to keep them safe.

hospitals are more likely to have an administrative team with the capacity to improve quality measures by either making quality improvements or by gaming (manipulating survey answers without real quality improvements). Finally, hospital size is strongly positively correlated with hospital density, so large hospitals have an increased incentive to respond as a result of increased local competition. I define large and small hospitals as having above or below the median number of beds, respectively.

To test whether patient sorting is stronger in areas with higher levels of competition, I compute the Herfindahl–Hirschman Index (HHI), where each hospital's market is defined as a 20 mile fixed radius around each hospital.[31] The HHI is an established measure of the size of firms relative to the industry, and therefore an indicator of the level of competition in the market. The HHI for hospital i is computed as the sum of the squared market shares of all hospitals in hospital i's market. I define "more competitive" and "less competitive" hospitals as having below or above the median HHI, respectively.[32]

[31] I have detailed geographic data for patients for only seven states, so variable radius or patient flow methodologies would only be feasible for a subset of outcomes and states. Mutter, Wong, and Zhan (2005) compute correlations between the various market definitions, including predicted patient flows which Kessler and McClennan (2000) use as an instrument for HHI, and find very high correlations between them. Table B2 shows that my primary result is robust to choice of market definition.

[32] For the states in which I perform patient sorting analysis, the median "less competitive" hospital has 4 hospitals in a 20 mile radius and 13 hospitals within 50 miles, while the median "more competitive" hospital has 23 hospitals in a 20 mile radius and 72 hospitals in a 50 mile radius.

1.6 Results

1.6.1 Patient Response

I explore the response of patients to receiving quality information about hospitals by assessing the percentage increase in elective and non-elective inpatient admissions on both the C-B and B-A margins. If patients are responding to hospital grade assignments, we would expect that we would observe this response in elective admissions, where patients have the ability to choose where they receive treatment. I assess changes in all admissions, non-elective and elective admissions separately for hospitals in less competitive and more competitive markets (as defined in Section 5).

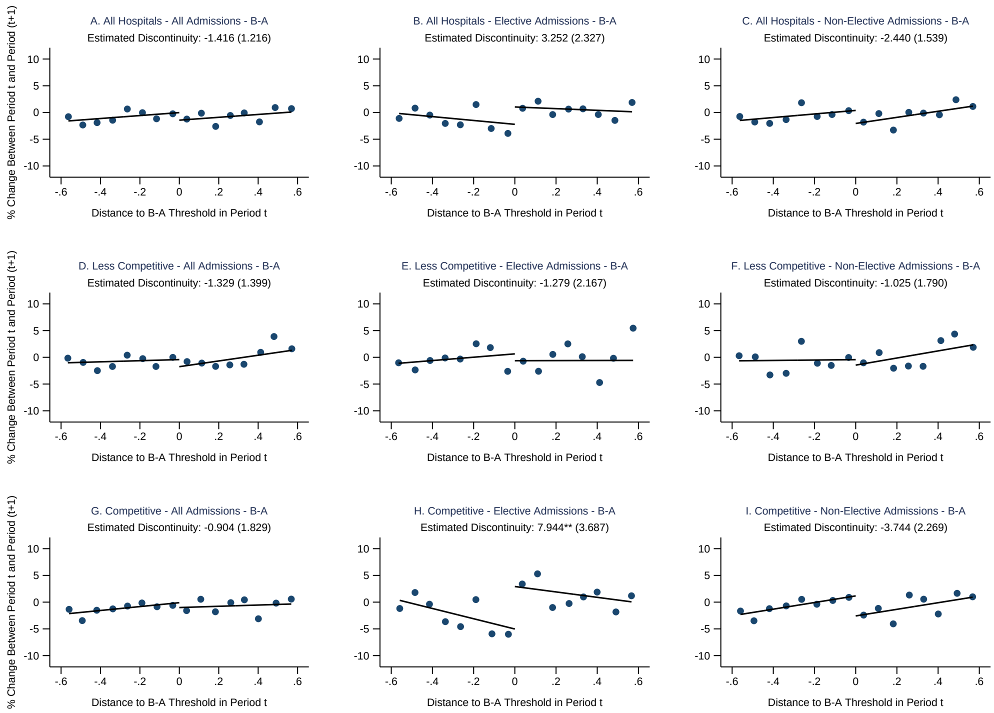
Figure 3 examines the response of patients to the grade assignment of hospitals. Figure 3B provides some evidence of patient sorting because "A" hospitals are increasing their elective admissions by about 3 percent relative to "B" hospitals on the B-A margin, although the estimate is not statistically significant. Figure 3C shows that "B" hospitals are increasing their non-elective admissions by about 2 percent relative to "A" hospitals, but this estimate is not statistically significant. Figure 3A examines the change in the sum of all admissions, excluding pregnancies, which shows that "B" hospitals are slightly increasing their overall admissions relative to "A" hospitals. This occurs because the discontinuity for non-elective admissions is negative and non-elective admissions represent about 75 percent of elective or non-elective admissions. Because hospitals advertise their achievement and receive additional media attention when receiving an “A,” we would expect to see patient sorting only on the B-A margin and only for elective admissions. Consistent with this idea, Figure B7 does not show any significant patient sorting on the C-B margin.[33] Taken together, these results are suggestive of some patient sorting among all hospitals.

Quality information will have a minimal effect on patient sorting if patients are in areas with very few hospitals, as patients in those markets have very limited choice. Therefore I refine my analysis of patient sorting by categorizing local hospital markets based on competitiveness. I define hospitals as “more competitive” or “less competitive” and analyze patient sorting for each category. All discontinuity estimates indicate that there is no statistically significant difference in market share between hospitals of adjacent grades for hospitals in less competitve markets. However, in more competitive markets, elective admissions for “A” hospitals increase by about 8 percent relative to “B” hospitals, which is statistically significant at the 5 percent level (Figure 3H). There is a 4 percent increase in non-elective admissions in “B” hospitals relative to “A” hospitals, although this is not a statistically significant estimate (Figure 3I).[34] Because non-elective admissions represent approximately 75 percent of all admissions, it follows that the

[33] All figures in the main text display results only for the B-A margin. Figures displaying all results on the C-B margin and a more detailed discussion of these results can be found in Appendix B5.

[34] One possible explanation is that many non-elective patients are admitted through the emergency department, so if there are a large number of beds being used by elective admissions, the hospital may choose to transfer the patient to a different facility.

Figure 1.3: The Effect of Grade Assignment on Inpatient Admissions



Notes: These graphs present changes in inpatient admissions on the B-A margin. The dependent variable for all models is the percentage change in inpatient admissions between six month rating periods. All admissions are defined as all elective or non-elective admissions not including deliveries. Elective admissions are admissions scheduled more than 24 hours in advance. Non-elective admissions are admissions not scheduled more than 24 hours in advance. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable and are weighted by the number of beds in each hospital. Admission counts are based on hospital discharge records from Arizona, California, Florida, Kentucky, New Jersey, New York, and Washington. Sample is trimmed at the 99th percentile of the dependent variable. "All Hospitals" includes all hospitals in the sample. "Less Competitive" includes only hospitals with above the median value of the Herfindahl-Hirschman Index. "Competitive" includes only hospitals with below the median value of the Herfindahl-Hirschman Index. The running variable distance to threshold comes from the Leapfrog Group for Spring 2012 to Spring 2015.

opposite discontinuity estimates of elective and non-elective admissions results in an estimate very near 0 for all admissions (Figure 3G).

Table 2 displays all patient sorting results and confirms that the only significant patient sorting is where we would most expect, for elective admissions and in competitive hospital markets, where patients have the capacity for choice, and on

Table 1.2: Regression Discontinuity Estimates for Patient Sorting

Competition Sample	B-A			C-B		
	All	More	Less	All	More	Less
	(1)	(2)	(3)	(4)	(5)	(6)
All Admissions	-1.416	-0.904	-1.329	-1.083	0.254	-3.093*
	(1.216)	(1.829)	(1.399)	(1.351)	(1.868)	(1.816)
Observations	700	336	363	695	309	384
Hospitals	395	207	215	420	210	208
Elective Admissions	3.252	7.944**	-1.279	-2.376	-2.296	-3.898
	(2.327)	(3.687)	(2.167)	(2.216)	(2.968)	(3.295)
Observations	696	334	362	690	308	382
Hospitals	390	204	213	417	208	236
Non-Elective Admissions	-2.440	-3.744	-1.025	-0.784	0.639	-2.954
	(1.539)	(2.269)	(1.790)	(1.722)	(2.442)	(2.220)
Observations	700	335	365	692	308	384
Hospitals	393	205	215	419	208	239

Notes: The rows "All Admissions," "Elective Admissions," and "Non-Elective Admissions" present the discontinuity estimates for the given admission type. All admissions are defined as all elective or non-elective admissions not including deliveries. Elective admissions are admissions scheduled more than 24 hours in advance. Non-elective admissions are admissions not scheduled more than 24 hours in advance. The dependent variable for all models is the percentage change in inpatient admissions between six month rating periods. More Competitive and less competitive hospitals are defined as below and above the median Herfindahl-Hirschman Index, respectively. All models include a first-order polynomial in distance fully interacted with a dummy for being over the grade threshold. Admission counts are based on hospital discharge records from Arizona, California, Florida, Kentucky, New Jersey, New York, and Washington. Sample is trimmed at the 99th percentile of the dependent variable. All regressions are weighted by the number of beds in the hospital. The running variable distance to threshold comes from the Leapfrog Group for Spring 2012 to Spring 2015. Standard errors clustered by hospital in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

the B-A margin, where patients receive the strongest quality signal. So, despite the frictions in hospital choice as described in Appendix Section A3, there is some evidence of patient sorting, which is concentrated nearly entirely in competitive hospital markets.

1.6.2 Hospital Response

In this section, I assess whether hospitals receiving a lower grade improve quality relative to nearly identical hospitals that receive the next highest grade. The previous results have shown that some patients are responding by choosing to receive treatment at "A" hospitals rather than lower grade hospitals, so I expect that competitive pressures will lead "B" hospitals to improve quality in order to receive an "A" in the next period. To evaluate this hypothesis, I examine the change in healthcare-associated infections as well as the change in the overall and survey scores.

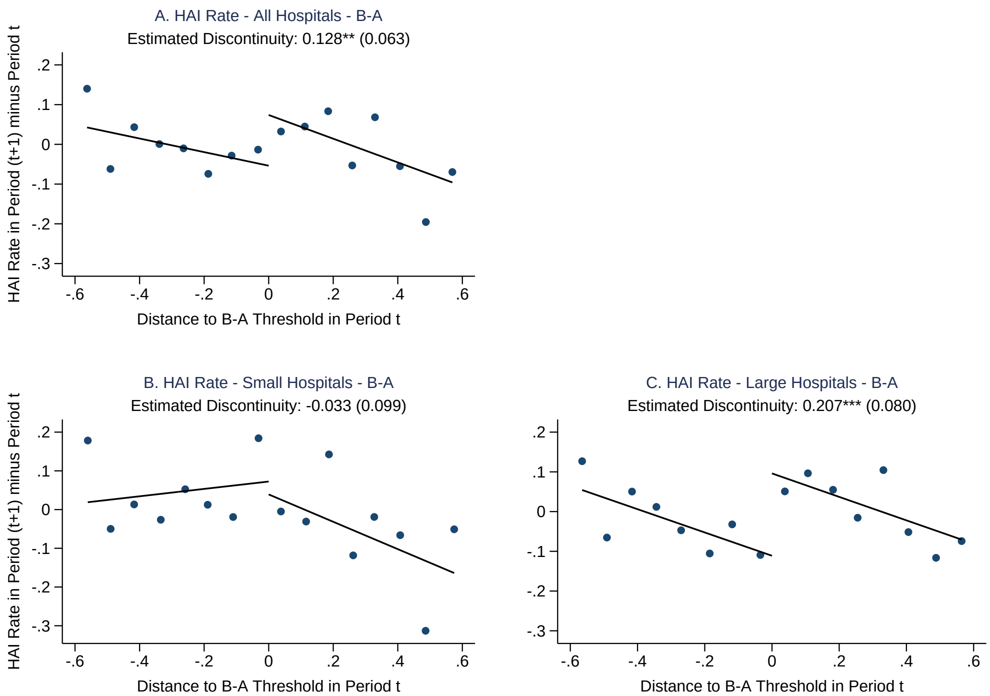
1.6.2.1 Healthcare-Associated Infections

Healthcare-associated infections (HAI) are easily preventable, responsive to intervention, and can be dramatically lowered in a short time period.[35] I use inpatient discharge records to construct the rate of HAIs in each hospital and examine the change in the HAI rate between rating periods.

Figure 4 presents changes in HAI rates between six month rating periods. Figure 4A shows that when pooling large and small hospitals, "B" hospitals are reducing their HAI rate by a statistically significant 0.128 standard deviations relative to nearly identical "A" hospitals, providing evidence that "B" hospitals are responding to their grade by decreasing infections. Figure 4C shows that large

[35] Healthcare-associated infections (HAI) are alarmingly prevalent in many US hospitals, and yet are the most easily preventable healthcare-acquired conditions. There are close to 100,000 deaths (Klevens et al. 2007) per year due to HAIs, which cost the healthcare system \$5.7-\$31.5 billion (Scott 2009). Over the past several decades however, great progress has been made in reducing HAIs due to interventions to improve handwashing (Pfoh et al. 2013, WHO 2009, Gould et al. 2010, Boyce 2011), checklists (Gawande 2009), and safety culture and education (Marschall et al. 2008). Process interventions aimed at HAI-prevention are often highly effective and show results in as little as 3-4 months. For example, Edmond et al. (2010) find that nurses wearing electronic badges that sensed whether they used alcohol to wash their hands increased compliance from 66 percent to 92 percent in an intervention period of only 10 days, and Vernaz et al. (2008) found that a handwashing campaign involving education on the frequent and proper use of alcohol-based hand rubs significantly and immediately reduced the number of methicillin-resistant Staphylococcus aureus (MRSA) infections.

Figure 1.4: The Effect of Grade Assignment on Healthcare-Associated Infections



Notes: These graphs present changes in the rate of healthcare-associated infections on the B-A margin. The numerator of "HAI Rate" is defined as 1 if the patient received any healthcare-associated infection during their inpatient stay. The denominator of "HAI Rate" is the total number of admissions to the hospital in a given month. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. HAI rates are based on hospital discharge records from Arizona, Florida, Kentucky, New Jersey, New York, and Washington. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. The running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015.

hospitals are entirely driving this response. Specifically, large hospitals that just missed an "A" are reducing HAI rates by 0.207 standard deviations relative to those that just received an "A," which is significant at the 1 percent level. On the other hand, I do not find evidence that small "B" hospitals respond through infection reduction (Figure 4B). Consistent with the notion that hospitals would respond only on the B-A margin, Figure B8 shows that hospitals are not responding through HAI improvements on the C-B margin.

Using inpatient discharge records, this study has shown that some patients chose to receive treatment at "A" hospitals rather than a lower grade hospital as a result of grade assignment. Due to competitive pressures and the desire to maintain a good reputation, "B" hospitals have improved quality by reducing healthcare-associated infections to attempt to get an "A" in the subsequent period. In the next section, I examine whether hospitals that receive a lower grade are explicitly attempting to receive a higher grade by improving their overall score and score on the survey.

1.6.2.2 Overall Score and Survey Score

Figure 5 presents changes in the overall score and survey score. Figure 5A shows evidence of a small and marginally significant response to grades in the overall score on the B-A margin. Figure 5B shows that the "B" hospitals are increasing their survey score by 0.063 standard deviations more than "A" hospitals, which is significant at the 5 percent level and indicates that the response in overall score is driven by increases in the survey.[36] As shown in previous results, some hospitals may have responded more strongly to grade assignment, but the noise from non-responding hospitals is obscuring these results. In order to identify the response for the subset of hospitals with a sufficient incentive and capacity to respond, Panels C-F examine only hospitals which completed the 2011 survey (which applies to the first rating period; Spring 2012) and separately examine

[36] The other measures that are used to calculate the overall score are from the CMS and there is no discontinuity on either margin because the reporting periods are not aligned with Leapfrog rating periods, as discussed in Section 5.

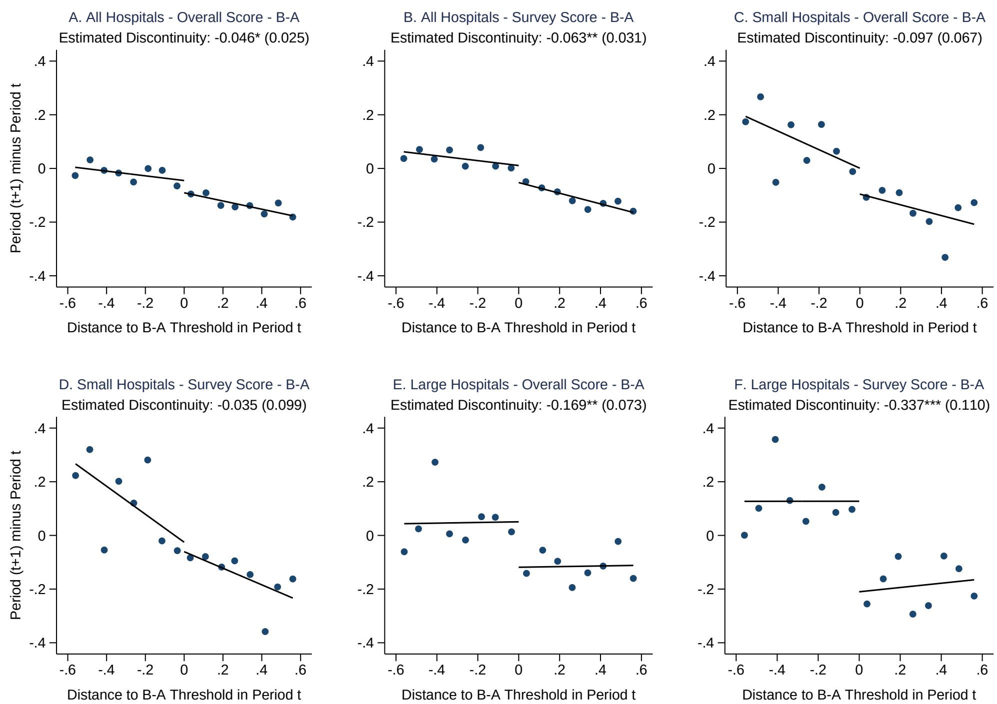
large and small hospitals.[37]

Consistent with the results for healthcare-associated infections, responses in scores are driven by large hospitals. Figure 5E shows that large hospitals that just missed an "A" are increasing their overall score by 0.169 standard deviations more than those that just received an "A," which is significant at the 5 percent level. This response in overall score for large hospitals is entirely driven by increases in the survey score (Figure 5F). Specifically, large hospitals that just missed an "A" are increasing their survey score by 0.337 standard deviations more than those that just received an "A," which is significant at the 1 percent level. However, Figure 5 (C-D) show that small hospitals do not respond by improving their scores. Consistent with previous results, Figure B9 does not provide any evidence of a response in overall or survey scores on the C-B margin.

Table 3 displays the point estimates for the hospital responses in overall score, survey score, and HAI rates. The results indicate that "B" hospitals improve both their score on the survey and objectively measured HAIs relative to "A" hospitals. For both outcomes, the response is driven entirely by large hospitals, which have both an increased capacity and incentive to respond to grades. Consistent with the notion that there is a differential incentive to improve only on the B-A margin, there is no response on the C-B margin for any group of hospitals or outcomes.

[37]See explanation and analysis of response in survey submission in Appendix Section A2.

Figure 1.5: The Effect of Grade Assignment on Scores



Notes: These graphs present changes in the Hospital Safety Score and the Leapfrog Hospital Survey total score on the B-A margin. The dependent variable in each model is the change in the given variable between six month rating periods. "Overall Score" is the Hospital Safety Score, that is, a weighted sum of 26 or 28 measures. "Survey Score" is a weighted sum of all 12 measures on the Leapfrog Hospital Survey. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. "All Hospitals" includes all hospitals rated by the Leapfrog Group. "Large Hospitals" and "Small Hospitals" includes only hospitals with above or below the median number of beds, respectively, and includes only hospitals that took the survey reported in the first release of the Hospital Safety Score in June 2012. Overall score, survey score, and the running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015.

1.7 Conclusion

This paper has assessed the behavioral responses of patients and hospitals to quality information disclosure, specifically the assignment of letter grades to hospitals by the Leapfrog Group. To obtain the causal response to grade assignment, I exploit the method in which grades are assigned based off of a continuous score using a regression discontinuity design.

Table 1.3: Regression Discontinuity Estimates for Hospital Response

	B-A			C-B		
	All	Small	Large	All	Small	Large
	(1)	(2)	(3)	(4)	(5)	(6)
Overall Score	-0.046*	-0.097	-0.169**	0.025	0.039	-0.099
	(0.025)	(0.067)	(0.073)	(0.027)	(0.091)	(0.088)
Observations	4,766	761	747	5,227	630	623
Hospitals	1792	319	318	1932	290	270
Survey Score	-0.063**	-0.035	-0.337***	0.004	0.141	-0.136
	(0.031)	(0.099)	(0.110)	(0.032)	(0.137)	(0.148)
Observations	4,766	761	747	5,227	630	623
Hospitals	1792	319	318	1932	290	270
HAI Rate	0.128**	-0.033	0.207**	-0.042	-0.056	-0.037
	(0.063)	(0.099)	(0.080)	(0.051)	(0.104)	(0.059)
Observations	515	274	241	512	259	253
Hospitals	274	138	136	300	149	151

Notes: The rows "Overall Score," "Survey Score," and "HAI Rate" present the discontinuity estimates for the given variable. The dependent variable in each model is the change in the given variable between six month rating periods. "Overall Score" is the Hospital Safety Score, that is, a weighted sum of 26 or 28 measures. "Survey Score" is a weighted sum of all 12 measures on the Leapfrog Hospital Survey. The numerator of "HAI Rate" is defined as 1 if the patient received any healthcare-associated infection during their inpatient stay. The denominator of "HAI Rate" is the total number of admissions to the hospital in a given month. Large and small hospitals are defined as having above or below the median number of beds, respectively. All models include a first-order polynomial in distance interacted with a dummy for being over the grade threshold. HAI rates are based on hospital discharge records from Arizona, Florida, Kentucky, New Jersey, New York, and Washington. The running variable distance to threshold, overall score, and survey score come from the Leapfrog Group for Spring 2012 to Spring 2015. Standard errors clustered by hospital in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

I assess whether patients are responding to the grade assignment of hospitals by choosing to be treated at nearly identical hospitals with different grades. I find evidence that patients sort to higher grade hospitals in competitive hospital markets but there is no response in areas with minimal hospital choice. Specifically, in competitive markets, I find that "A" hospitals increase their volume of elective admissions by 8 percent more than "B" hospitals, but there is no change in patient volume between "C" and "B" hospitals. This result occurs because hospitals receive much more media attention when they receive an "A" than when they receive any other grade, creating the strongest differential signal and the largest differential incentive to improve between "A" and "B" hospitals.

I also find evidence that hospitals respond to receiving a grade of "B" by significantly improving quality relative to nearly identical "A" hospitals. Using inpatient discharge records, I find that "B" hospitals decreased their rate of healthcare-associated infections relative to "A" hospitals. Additionally, "B" hospitals significantly improved their score relative to "A" hospitals, which implies that lower-ranked hospitals are directly trying to improve their standing in the ranking system. These responses are driven by large hospitals, which have a higher capacity for improvement, as well as a higher incentive for improvement because most large hospitals are in competitive markets.

This paper has shown that some patients sort into hospitals with the strongest quality signal, and hospitals receiving the second-best rating improved rating measures and patient outcomes to attempt to receive the highest rating. These responses are much larger in more competitive markets, which is consistent with recent literature that has found that increased competition has improved patient outcomes.[38] Generally, these results imply that rating systems have the capacity to improve consumer welfare by sorting consumers into higher quality firms and by incentivizing firms to improve quality.

[38] See Kessler and McClellan (2000), Cooper et al. (2011), Gaynor et al. (2013).

Appendix A. Additional Background

Appendix A1. Detailed Hospital Safety Score Methodology

This section describes the methodology and the measures that make up the Hospital Safety Score, as well as a description of the reporting periods of the Centers for Medicare and Medicaid Services (CMS) measures used in the Hospital Safety Score, and a description of the grade transitions of hospitals.

The Hospital Safety Score is a weighted sum of 28 different measures, 12 of which are taken from the survey and the other 16 of which are measures constructed from publicly available data, primarily from the Centers for Medicare and Medicaid Services (CMS). Of these, 15 are process of care or structural measures and 13 are outcome measures. Ten of the process measures and two of the outcome measures are from the survey. A list of all measures used to compute the Hospital Safety Score is provided in Table A1. An example of the criteria for the assignment of grades from Spring 2014 is shown in Table A2.

The following is a brief summary of the methodology used to construct the Hospital Safety Score and how the underlying score is used to assign letter grades to hospitals. Step 1: Each of the 15 process and 13 outcome measures are given a score. Step 2: Data from the Centers for Medicare and Medicaid Services are trimmed to the 99th percentile (any value over the 99th percentile is set equal to the 99th percentile value). Step 3: Z-scores are calculated for each measure. Step

4: Any Z-score below -5 is truncated at -5. Step 5: If the hospital has data on each measure, use the standard weights as given in the methodology, but if the hospital is missing data on any outcome measure or did not complete the survey, weights must be recalculated for them. Step 6: Obtain weighted process and outcome scores by multiplying the Z-score of each measure by the appropriate weight and summing them to obtain a weighted process score and a weighted outcome score. Step 7: Sum the weighted process score and outcome score (since both are worth 50 percent of the overall score) and add 3 to obtain the final Hospital Safety Score. Step 8: Letter grades are assigned based on thresholds of the overall score determined by the number of standard deviations from the national mean in that rating period.

There are several ways in which Leapfrog deals with missing data. If the hospital fails to complete the survey, a score is computed based on the CMS measures. Additionally, Computerized Prescriber Order Entry and ICU Physician Staffing, two of the process measures on the survey, receive scores from the American Hospital Association (AHA) Annual Survey if the hospital did not fill out the survey. Finally, outcome measures must be based on at least 30 patients or the value for that measure will be considered missing, which is relevant for smaller hospitals.

Figure A1 and Table A3 present two alternative representations of the variation in scores over time. This paper focuses on analyzing the changes in scores over time, so a potential concern is that hospitals are not shifting their score over time enough to be able to identify effects. Although most hospitals do not shift their grade from one period to the next, a substantial portion are shifting each period. Particularly, 17-20 percent of hospitals are shifting between an "A" and a "B" grade between rating periods, and about 13-22 percent are shifting between a "B" and a "C," providing plenty of variation to identify effects.

About half of the Hospital Safety Score is made up of measures from the CMS, so a natural question is whether these objective outcome and process measures are changing as a result of grade assignment. However, as reported in Table A4, the CMS measures that contribute to the calculation of the Hospital Safety Score have different reporting periods. The majority of the measures are reported as two year rolling averages. Additionally, the most current reporting periods for many measures end before the first grades were assigned in June 2012. Therefore, it is not feasible to examine the measures reported by the CMS as a response to grades.

Appendix A2. Exploration of Response in Survey Submission

This section explores the concern that there is a discontinuous jump in the probability of survey submission between letter grades, which would confound any responses found in changes in scores.

I wish to examine only those periods where hospitals took the survey because these are the periods where the hospitals are paying the most attention to their grade and are therefore the most likely to respond. However, since taking the survey could potentially be a response to grade assignment, it would confound the results to examine only periods in which a hospital took the survey.

Table A5 gives an indication that hospitals receiving lower grades tended to stop taking the survey more than hospitals receiving higher grades, although this discrepancy narrowed significantly after the first year of the program. However, Figure A3 shows that there is not a discontinuity in the change in survey submission at the B-A margin. Specifically, Panels A-C does not provide evidence that hospitals are more likely to switch from not taking the survey to taking the survey if they received a “B” rather than an “A.” Likewise, Panels D-F does not provide evidence that hospitals are more likely to stop taking the survey if they received a “B” rather than an “A.” Therefore, there is some evidence of differential attrition of survey submission depending on grade, but there is not a differential change in survey submission status across the B-A threshold, so this should not confound my results.

I restrict my attention to hospitals that took the 2011 survey which applied to the first rating period; Spring 2012. Taking this survey could not be a response to their grade, but indicates that the hospital was aware of Leapfrog prior to the implementation of the Hospital Safety Score, so these hospitals have an increased capacity to respond to their grade assignment.

Appendix A3. Frictions in Hospital Choice

Unlike restaurants or supermarket products, choosing the hospital at which to receive treatment is constrained by several factors. The first consideration regarding whether or not grades affect hospital choice is whether hospitals are at full capacity and whether higher-grade hospitals would be able to accomodate a higher volume. Throughout the 1980's and 1990's, increases in technology led to substantial excess bed capacity, so most studies have found minimal overutilization and constraints in inpatient bed capacity.[39]

The primary friction in hospital choice is health insurance. Health maintenance organization (HMO) plans require patients to pick a primary care physician (PCP) and all health services must go through that PCP. With an HMO plan, only a small network of hospitals is available and a referral from a patient's PCP is needed to see any other healthcare professional. A preferred provider organization (PPO) allows more flexibility than an HMO (and no PCP), but seeing a healthcare professional outside the network means that the patient will not receive full coverage. High deductible health plans (HDHP) generally have the same network restrictions as PPOs, but with much higher deductibles and lower premiums. If a patient has an HDHP, then they are eligible for a tax-advantaged health savings account (HSA) to save for emergencies. Point of service (POS) plans vary, but are often a hybrid of an HMO and PPO, where sometimes referral from a PCP is needed to see

[39] See Gaynor and Anderson 1995, Keeler and Ying 1996, Madden 1999, Ennis, Schoenbaum, and Keeler 2000, Green 2002, DeLia 2006, Green and Liu 2015.

a specialist, but members may also be partially covered for out-of-network care. According to the 2014 Health Benefits Survey, 58 percent of workers with employer provided health coverage enrolled in a PPO, 13 percent in an HMO, 20 percent in a HDHP, and 8 percent in a POS plan. Therefore, 13 percent of workers are relatively unlikely to be able to respond to any kind of consumer information, while 87 percent are also constrained by coverage, but generally have at least some coverage outside of their primary network. Additionally, 16 percent of the United States population is enrolled in Medicaid, while 15 percent is enrolled in Medicare, and these individuals can receive health services from any hospital or doctor that accepts Medicaid or Medicare, respectively.[40]

Although health insurance makes hospital choice more constrained than firm choice in many other industries, Chandra et al. (2015) have recently explored the possibility that hospital choice is not as sticky as commonly believed. They show that higher quality hospitals, where quality is measured by many of the same measures used to compute the Hospital Safety Score, tend to have higher market shares at a point in time and increase their market share as their quality increases. They conclude that the healthcare sector may be closer to traditional sectors than commonly assumed.

A final consideration is whether it is actually doctors or patients who see the

[40] A report from the Department of Health and Human Services shows that access to Medicaid is constrained and that about half of providers that accepted Medicaid managed care could not actually offer appointments to enrollees (HSS 2014). On the contrary, 96 percent of Medicare beneficiaries report having a usual source of care and about 90 percent are able to schedule timely appointments (Boccuti 2013).

quality information and respond to the grades. It is possible that while patients are not aware of the grades, doctors are, and directing their patients accordingly. It is not possible to distinguish to what extent it is doctors or patients that are responding to the information. However, if there is a hospital-choice response to grades, regardless of whether patients or doctors are using the information, the implication that people are using the quality information to make a hospital-choice decision is the same.

Appendix A4. Leapfrog Group Internal Auditing

To ensure hospitals are accurately reporting responses to the survey, the Leapfrog Group implements several steps. First, the hospital CEO or their designee must complete and sign an "Affirmation of Accuracy" after each section of the survey, and Leapfrog exercises the right to check documentation at random before publishing any results. After submission of the survey, "Quantitative responses are assessed using empirically driven, normative data quality standards." Each response of a hospital is divided into three categories depending on the plausibility of their response based on data quality thresholds. If the response is deemed plausible but outside of the data quality thresholds or implausible and with the intent to mislead, follow-up with the hospital is conducted. Data quality thresholds for each question are determined from both external data sources (e.g., state quality reports, other national performance measurement entities) and historical survey data. This review is done monthly during the survey cycle (April 1 - December 31).

Table A1: Hospital Safety Score Measures

	Process Measures	Adverse Outcome Measures
Leapfrog Hospital Survey Measures	Computerized Prescriber Order Entry ICU Physician Staffing Leadership Structures and Systems Culture Measurement, Feedback and Intervention Teamwork Training and Skill Building Identification and Mitigation of Risks and Hazards Nursing Workforce Medication Reconciliation Hand Hygiene Care of the Ventilated Patient	Central Line-Associated Blood Stream Infections Cather-Associated Urinary Tract Infections
Centers for Medicare and Medicaid Services Measures	Patients Received Antiobiotic within 1 Hour Prior to Surgery Patients Received the Right Antiobiotic for Surgery Antiobiotic Discontinued After 24 Hours Urinary Catheter was Removed on Postoperative Day 1 or 2 Surgery Patients Received Appropriate Blood Clot Treatment	Foreign Object Retained After Surgery Air Embolism Pressure Ulcer - Stages 3 and 4 Falls and Trauma Surgical Site Infection Colon Death Due to Complications Collapsed Lung Breathing Failure After Surgery Postoperative PE/DVT Wounds Split Open After Surgery Accidental Cuts or Tears

Notes: This table presents the 28 measures used to compute the Hospital Safety Score after and including Fall 2013. Before Fall 2013, 26 measures were used, where catheter-associated urinary tract infection and surgical site infection for major colon surgery were the two measures added in Fall 2013. This table represents only the primary source of each measure.

Table A2: Hospital Safety Score Grading Criteria

Grade	Safety Score Criteria	Standard Deviations	Count of Hospitals	Percentage of Hospitals
A	≥ 3.151	0.6 SDs Above Mean	804	31.9%
B	≥ 2.961	0.0 SDs Above Mean	668	26.5%
C	≥ 2.485	1.5 SDs Below Mean	878	34.8%
D	≥ 2.009	3.0 SDs Below Mean	150	5.9%
F	< 2.009	$>$ 3.0 SDs Below Mean	22	0.9%
Totals			2,522	

Notes: This table presents the criteria for each grade and the distribution of hospitals across grades. The values in the "Standard Deviations" column correspond to the lower threshold for each grade. Source: "Explanation of Safety Score Grades" published on www.hospitalsafetyscore.org in April 2014.

Table A3: Transition Matrix

		Fall 2013				
	Grade	A	B	C	D	F
	A	79.56	19.38	2.96	0	0
	B	17.71	57.82	13.70	2.106	0
Spring 2014	C	2.18	21.82	78.64	23.74	5.88
	D	0.14	0.33	4.32	68.35	23.53
	F	0	0	0	5.760	70.59

Notes: This table presents the transition probabilities from Fall 2013 to Spring 2014. Each cell represents the percentage of hospitals that received the specified grade in Spring 2014 after receiving the specified grade in Fall 2013. Data comes from the Leapfrog Group for Fall 2013 and Spring 2014.

Table A4: Centers for Medicare and Medicaid Services Data Reporting Periods

	Spring 2012	Fall 2012	Spring 2013	Fall 2013	Spring 2014	Fall 2014
CLABSI	06-08	1/1/11-9/30/11	4/1/11-3/31/12	10/1/11-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
CAUTI				1/1/11-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
SSI Colon				1/1/12-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
PSI 4	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12
PSI 6	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12
PSI 11	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12	7/1/10-6/30/12
PSI 12	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12
PSI 14	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12
PSI 15	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12
SCIPINF1	4/1/10-3/31/11	10/1/10-9/30/11	4/1/11-3/31/12	10/1/11-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
SCIPINF2	4/1/10-3/31/11	10/1/10-9/30/11	4/1/11-3/31/12	10/1/11-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
SCIPINF3	4/1/10-3/31/11	10/1/10-9/30/11	4/1/11-3/31/12	10/1/11-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
SCIPINF9	4/1/10-3/31/11	10/1/10-9/30/11	4/1/11-3/31/12	10/1/11-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
SCIPVTE2	4/1/10-3/31/11	10/1/10-9/30/11	4/1/11-3/31/12	10/1/11-9/30/12	4/1/12-3/31/13	10/1/12-9/30/13
Foreign Object Retained	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12	7/1/10-6/30/12
Air Embolism	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12	7/1/10-6/30/12
Pressure Ulcers	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12	7/1/10-6/30/12
Falls and Trauma	10/1/08-6/30/10	7/1/09-6/30/11	7/1/09-6/30/11	7/1/10-6/30/12	7/1/10-6/30/12	7/1/10-6/30/12

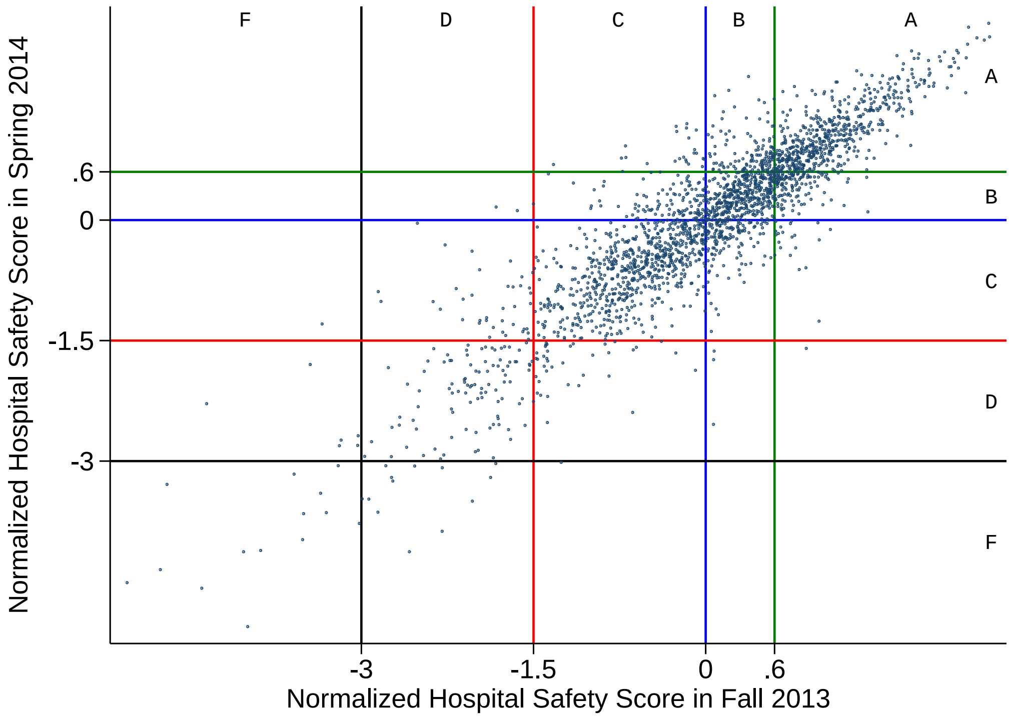
Notes: This table presents the reporting periods for each Centers for Medicare and Medicaid Service measure used in the Hospital Safety Score. Each cell displays the reporting period that corresponds to the Hospital Safety Score period in the first row.

Table A5: Leapfrog Hospital Survey Probabilities of Submission

Grade in Period (t-1)	Percentage Submitted Survey Conditional on Submitted Survey in Period (t-1)					
	Fall 2012	Spring 2013	Fall 2013	Spring 2014	Fall 2014	Spring 2015
A	89.98	100	95.71	100	98.85	100
B	92.07	98.77	96.30	100	97.86	100
C	75.77	98.86	90.00	100	89.47	100
D	72.22	100	60.00	100	80.00	100

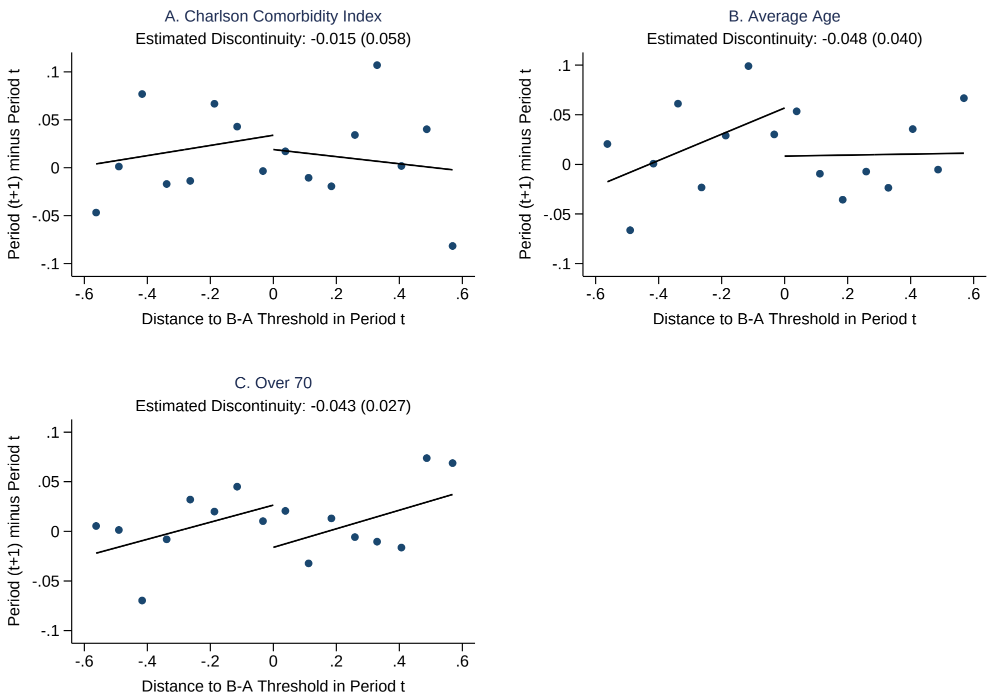
Notes: This table presents the probabilities of submission of the Leapfrog Hospital Survey conditional on the grade of the hospital in the previous period. Data comes from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure A1: Transition Plot



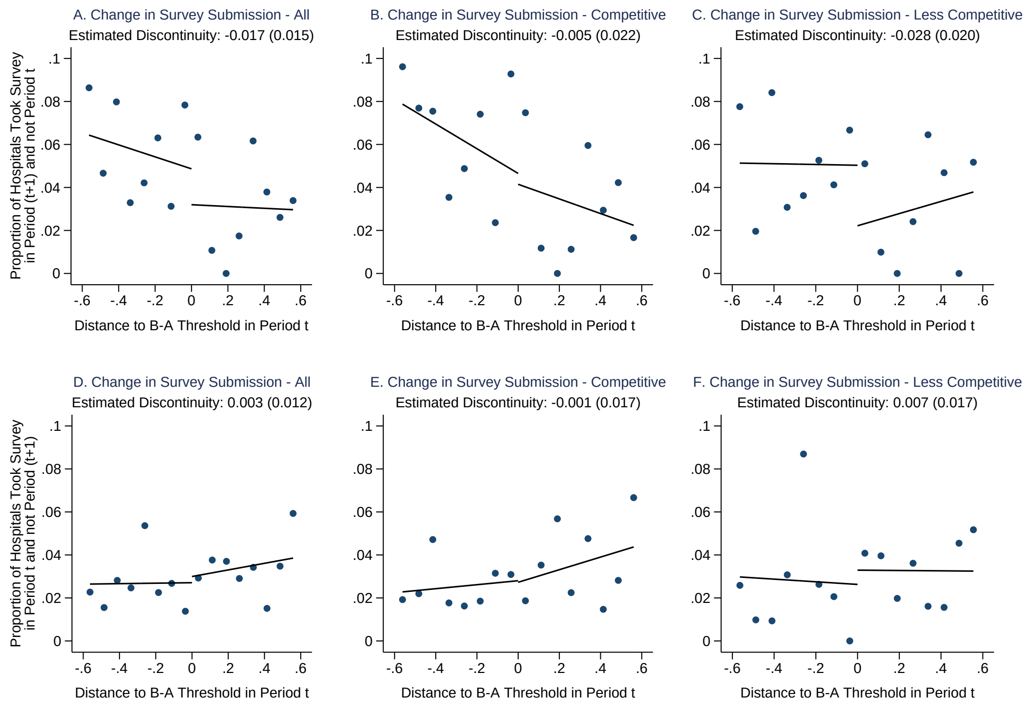
Notes: This graph presents the transitions in the Hospital Safety Score between Fall 2013 and Spring 2014. The Hospital Safety Score is normalized so that the values correspond to standard deviations from the national mean in a given rating period. Each point represents a single hospital. Data comes from the Leapfrog Group for Fall 2013 and Spring 2014.

Figure A2: Changes in Patient Composition



Notes: This graph presents the change in patient composition at the B-A margin. Graphs A, B, and C present the changes in the Charlson comorbidity index, average age, and number of individuals over 70, respectively, admitted as an inpatient between six month rating periods. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. Dependent variables are computed from the age of the patient (in years) and the diagnoses of the patient which come from hospital discharge records from Arizona, Florida, Kentucky, New Jersey, New York, and Washington. The running variable distance to threshold comes from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure A3: Probability of Survey Submission



Notes: This graph presents the probability of submitting the Leapfrog Hospital Survey conditional on survey submission in the previous period. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the Hospital Safety Score. Data on survey submission comes from the Leapfrog Group for Spring 2012 to Spring 2015.

Appendix B. Robustness and Additional Checks

This section presents robustness checks of the main results of this paper, as well as a check for explicit gaming of the Leapfrog Hospital Survey and an exploration of the change in prices as a result of grade assignment.

Appendix B1. Robustness Checks

Table B1 tests the sensitivity of the major results to a variety of different specifications such as bandwidth choice, polynomial order, and kernel. Specifically, columns (1), (2), and (3) parametrically control for trends using linear, quadratic, and cubic polynomials, respectively. Columns (4), (5), and (6) provide estimates using local linear regression and a triangular kernel, restricting bandwidth choice to optimal bandwidths proposed by Calonico et al. (2014), Imbens and Kalyanaraman (2011), and a cross-validation procedure by Ludwig and Miller (2007). Patients and hospitals respond only on the B-A margin, so all robustness checks are only for responses on the B-A margin.

The row labeled "Survey Score" provides discontinuity estimates for large hospitals that took the first survey, which shows that the estimates are relatively stable and nearly always significant at the 5 percent level despite bandwidth or polynomial choice. The only exception is when controlling for trends using a quadratic polynomial because using higher order polynomials are overfitting since the mean reversion in ratings is linear and the sample size is small. The row "Elective Admissions" presents discontinuity estimates for patient sorting among hospitals in competitive markets, which shows that estimates range from about 7 to 11 percent depending on the polynomial and bandwidth choice. All estimates are statistically significant at the 5 percent level except when controlling for trends using a cubic polynomial or when examining a very small bandwidth, both of which result in large standard errors. The row “HAI Rate” presents discontinuity estimates for the difference in HAI rates for large hospitals, which shows that due to the small sample size, higher order polynomials and small bandwidths result in insignificant estimates due to large standard errors.

Figure B1 plots the discontinuity estimates against 9 bandwidths of increasing size, and shows that each of the main results have a very consistent magnitude across bandwidths and the discontinuity estimates for survey scores and elective admissions are statistically significant across nearly all bandwidths. Due to large standard errors because of a small sample size, the discontinuity estimates for the HAI rate are not significant below a bandwidth of 0.6. Table B2 shows the robustness of patient sorting in elective admissions on the B-A margin to choice of market definition. Specifically, each column provides the estimates from the regression discontinuity model when the market of the hospital is defined as a circle around the hospital with the given radius (as the crow flies). The estimates range from about 7-8.5 percent and are all statistically significant at least at the 10 percent level, and many are significant at the 5 percent level. So, the results are robust to choice of radius for the definition of hospital markets.

In order to assess the relative importance of competition versus hospital size in my results, I compare the response of hospitals in their HAI rate and the survey score for large and more competitive hospitals. Figure B2 (A-B) show that both groups respond by improving their survey score significantly, but this response is larger for the set of large hospitals. Similarly, Figure B2 (C-D) show that this same pattern holds for changes in HAIs, although the difference between size and competition is less pronounced. Location in a competitive market and size of hospital are strongly positively correlated, so it is not possible to tell which is driving the increased response, but the difference in effect sizes is suggestive that the size of the hospital is more strongly associated with the hospital's response to information than whether it is located in a competitive market.

Appendix B2. Survey Resubmission

The Leapfrog Hospital Survey is issued annually, and as shown in Figure B3, hospitals must submit the survey by June 30 to have their survey score reflected in their fall score. If hospitals do not resubmit the survey, the same score for all survey measures is used again for the spring of the following year, but their score will be updated in the spring if the hospital does choose to resubmit.[41] Table B3 shows that the lower the grade received, the higher the probability that the hospital will resubmit the survey, suggesting that hospitals are resubmitting only in order to improve their score. Additionally, both submissions and resubmissions have increased drastically between Spring 2012 and Spring 2015, an indication

[41] Hospitals may resubmit their survey with an update to any of their survey scores in accordance with the following statement from the Leapfrog guidelines: "Updated Submissions: Hospitals may update and resubmit their surveys as often as needed to reflect actual progress achieved or additional commitments undertaken in these patient safety areas. Hospitals submitting new information will have new results replace the posted results from the prior submission to reflect this progress, consistent with Leapfrog's monthly update of survey results." If a hospital resubmits their survey between the August 31 snapshot and the closing of the survey on December 31, their spring survey score will represent the most recently submitted score.

that some hospitals have realized the benefit both of submitting and resubmitting the survey.

Figure B4 (A-B) show the probability of resubmission on the B-A and C-B margins, respectively. Figure B4 (Panel A) shows that “B” hospitals at the B-A margin are 0.12 percentage points (81 percent) more likely to resubmit than “A” hospitals at the margin, confirming that some hospitals are strategically resubmitting in the hopes of improving their grade. Figure B4 (Panel B) shows that “C” hospitals at the C-B margin are 0.084 percentage points more likely to resubmit than “B” hospitals at the margin, although this estimate is not statistically significant. So, the probability of resubmission jumps at both thresholds, but the differential incentive is strongest at the B-A margin.

Figure B4 (Panel C) plots the change in the survey score only for resubmitters between their original survey score in the fall and their resubmitted score in the spring. This figure shows that nearly all resubmissions are improvements in the survey score, indicating that hospitals are submitting purely to improve their grade. However, there is no discontinuous change between “A” and “B” hospitals, which could be for three reasons. First, hospitals that resubmit are particularly motivated to achieve a higher grade, so hospitals that barely received an “A” are aware that they are very close to the threshold and may receive a “B” in the next period if they do not resubmit. Second, if hospitals are making real quality improvements, six months may be too short of a time to make these improvements. Providing some evidence to this hypothesis, Figure B4 (Panel D) shows that resubmitting hospitals that received a “B” improve their survey score significantly relative to “A” hospitals when examining changes between annual surveys. So, “B” hospitals are improving relative to “A” hospitals between annual waves, but not between six month periods. Third, hospitals are able to resubmit for about one month before the fall score is released, so some resubmissions may not be a response to the grade at all.

Because there is no discontinuous response in resubmission scores, Figure B4 (E-F) show that there is almost no difference in the size of the response in the survey score for large hospitals regardless of whether I include resubmitters. Taken together, the evidence shows that resubmitters are resubmitting strategically to improve their grade, but the response in the survey score for large hospitals is not affected by resubmissions.

Appendix B3. Survey Gaming

When evaluating the efficacy of any rating program, it is crucial to understand whether participants are taking part in gaming of the ratings. In this setting, gaming would most explicitly take the form of reporting untruthful survey answers. To check this explicitly, I note that hospitals report central line-associated blood stream infection (CLABSI) rates on the Leapfrog Hospital Survey and to the Centers for Medicare and Medicaid Services (CMS) using an online tool from the National Healthcare Safety Network (NHSN). The CMS provides guidelines for both internal and external validation which involves on-site review of medical records and auditing of electronic medical records. Therefore, to check whether hospitals are accurately reporting CLABSI rates on the survey, I compute the difference between the standardized infection ratio (SIR) for CLABSI from the CMS and the survey. Specifically, I compare the CLABSI rate reported on the survey to the CLABSI rate from the CMS that the hospital would have received if they had not taken the survey. The reporting periods of the measure from each source are skewed by several months so they will not be identical, but should form a mean 0 distribution.

Figure B5 displays the distributions of the difference between CLABSI rates. Figure B5 (Panel A) shows the distribution for all hospitals. Figure B5 (Panel B) shows the distribution for all hospitals that have recently received a "B," which have an enhanced incentive to increase their score. Figure B5 (Panel C) shows the distribution only for resubmitting hospitals, which are particularly motivated to increase their grade and are thus the most likely to game the survey. All of the graphs show that the distribution is mean 0 and most values are very close to 0. So, there does not appear to be any systematic misreporting of outcome rates on the survey, even for the hospitals with the highest incentive to manipulate.

Appendix B4. Analysis of Price Changes

Consumers make their decision to give their patronage to a firm based on two attributes, price and quality. In this paper, I have focused on quality rather than price because it is unlikely that hospitals alter their prices in response to their grade assignment. Due to the opaque nature of hospital charges and the mediation of costs for the patient through health insurance, most patients are unaware of the actual price that they will face at any given hospital. Although hospital charges are now public information after the passage of the Affordable Care Act, charges are often several times larger than the prices hospitals are actually paid by insurance companies. In 2004 for example, hospitals were only paid 38 percent of their charges by patients or their insurers (Reinhardt 2006). Since patients do not know to what extent charges listed on a hospital's chargemaster represent the cost they will actually have to pay, it would be very difficult for patients to "shop around" and choose a hospital based on price. Therefore, it would be very unlikely that hospitals would change prices to attract consumers.

To check whether hospitals are adjusting their prices in response to their grade assignment, I examine the percent change in costs by multiplying the monthly charges for each hospital, from the hospital discharge records, by the cost-to-charge ratio for each hospital. We can think of costs defined in this manner as charges, deflated so that they more accurately reflect the prices that patients or their insurance companies would actually pay. I do not find any evidence that hospitals are adjusting their deflated charges in response to their grade assignment on either margin or even for elective admissions, where patients have choice over which hospital to be treated at (Figure B6). Table B5 confirms these point estimates.

Appendix B5. Analysis of the C-B Margin

This section presents changes in inpatient admissions, healthcare-associated infections, and overall and survey scores on the C-B margin. Specifically, Appendix Figures B7, B8, and B9 show changes in inpatient admissions, healthcare-associated infections, and overall and survey scores, respectively, on the C-B margin. As noted in the text, both patients and hospitals view the differential signal between "B" and "A" grades to be much more pronounced than between "C" and "B" grades. This is likely because individuals view extreme ratings as more meaningful than middle ratings, and thus, there is a substantially higher level of advertising for hospitals that receive an "A" grade. Therefore, as shown in Appendix Figures B7, B8, and B9, there is no response of either patients or hospitals on the C-B margin, even for subsets where a response is likely and seen on the B-A margin. These results confirm the hypothesis that patients and hospitals view the distinction between "B" and "A" grades as more meaningful and a stronger signal than the distinction between "C" and "B" grades.

Table B1: Regression Discontinuity Robustness Estimates of Main Results

	Parametric			Local Linear Order 1		
	Linear	Quadratic	Cubic	CCT	IK	CV
	(1)	(2)	(3)	(4)	(5)	(6)
Survey Score	-0.337***	-0.248	-0.394**	-0.360**	-0.315**	-0.301***
	(0.110)	(0.157)	(0.193)	(0.176)	(0.151)	(0.116)
Bandwidth	0.6	0.6	0.6	0.206	0.342	0.598
Observations	749	749	749	299	499	747
Elective Admissions	7.944**	11.023*	9.898	11.017*	11.173*	9.257**
	(3.687)	(5.719)	(7.224)	(6.198)	(5.805)	(3.879)
Bandwidth	0.6	0.6	0.6	0.245	0.274	0.594
Observations	334	334	334	149	162	332
HAI Rate	0.207**	0.157	0.114	0.118	0.119	0.186**
	(0.080)	(0.117)	(0.151)	(0.134)	(0.134)	(0.089)
Bandwidth	0.6	0.6	0.6	0.242	0.243	0.591
Observations	241	241	241	159	159	252

Notes: The "Survey Score" row provides discontinuity estimates in the change in survey total score for hospitals that took the first Leapfrog survey and have above the median number of beds. The "Elective Admissions" row provides discontinuity estimates in the percentage change in elective admissions for hospitals in relatively competitive markets, that is, hospitals with lower than the median HHI. The numerator of "HAI Rate" is defined as 1 if the patient received any healthcare-associated infection during their inpatient stay. The denominator of "HAI Rate" is the total number of admissions to the hospital in a given month. CCT represents the optimal bandwidth using the optimal bandwidth procedure proposed by Calonico et al. (2014), IK uses the procedure proposed by Imbens and Kalyanaraman (2011), and CV uses the cross-validation procedure proposed by Ludwig and Miller (2007). Columns (1), (2), and (3) present models including a first, second, and third order polynomial in distance, respectively, fully interacted with a dummy for being over the grade threshold. Columns (4), (5), and (6) present estimates using local linear regression with a triangular kernel as proposed by Calonico et al. (2014). HAI counts are based on hospital discharge records from Arizona, Florida, Kentucky, New Jersey, New York, and Washington. Survey scores and the running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015. Standard errors clustered by hospital in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table B2: Robustness to Choice of Market Definition

Market Radius	15	20	25	30	35	40	45	50
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discontinuity Estimates	6.824*	7.944**	8.263**	8.825**	8.301**	6.824*	7.337*	6.838*
	(3.578)	(3.687)	(3.707)	(3.709)	(3.699)	(3.684)	(3.739)	(3.705)
Observations	338	334	337	333	334	339	341	339
Hospitals	208	202	203	201	202	204	205	209

Notes: This table presents robustness of the response in elective admissions on the B-A margin for hospitals in competitive markets to different market definitions. The dependent variable for all models is the percentage change in elective inpatient admissions. All models include a first-order polynomial in distance interacted with a dummy for being over the grade threshold. Admission counts are based on hospital discharge records from Arizona, California, Florida, Kentucky, New Jersey, New York, and Washington. Sample is trimmed at the 99th percentile of the dependent variable. The sample includes only hospitals with below the median value of the Herfindahl-Hirschman Index. The running variable distance to threshold comes from the Leapfrog Group for Spring 2012 to Spring 2015. Standard errors clustered by hospital in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table B3: Grade Distributions of Survey Resubmitters

	Resubmitters			Total Survey Submitters		
Fall Grade	Spring 2013	Spring 2014	Spring 2015	Spring 2013	Spring 2014	Spring 2015
A	39 (10.08 percent)	51 (11.21 percent)	57 (12.95 percent)	387	455	440
B	46 (23.12 percent)	87 (32.46 percent)	78 (29.21 percent)	199	268	267
C	82 (35.19 percent)	100 (42.37 percent)	126 (45.65 percent)	233	236	276
D	18 (47.37 percent)	29 (59.18 percent)	31 (60.78 percent)	38	49	51
F	6 (85.71 percent)	3 (60.00 percent)	11 (73.33 percent)	7	5	15

Notes: This table presents the number and percentage of hospitals that submitted and resubmitted the Leapfrog Hospital Survey by their grade in the fall of that year. Resubmission percentages are conditional on taking the survey. Survey submission and resubmission counts come from the Leapfrog Group for Spring 2013 to Spring 2015.

Table B4: Regression Discontinuity Estimates for Resubmitting and Non-Resubmitting Hospitals

	Survey Score				Proportion Resubmitted	
With Resubmitters	Only	Only-Annual	Yes	No	B-A	C-B
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity Estimates	-0.057	-0.262**	-0.337***	-0.341**	-0.121***	-0.084
	(0.105)	(0.111)	(0.110)	(0.137)	(0.042)	(0.059)
Observations	273	778	747	566	1,490	1,131
Hospitals	226	448	318	297	799	713

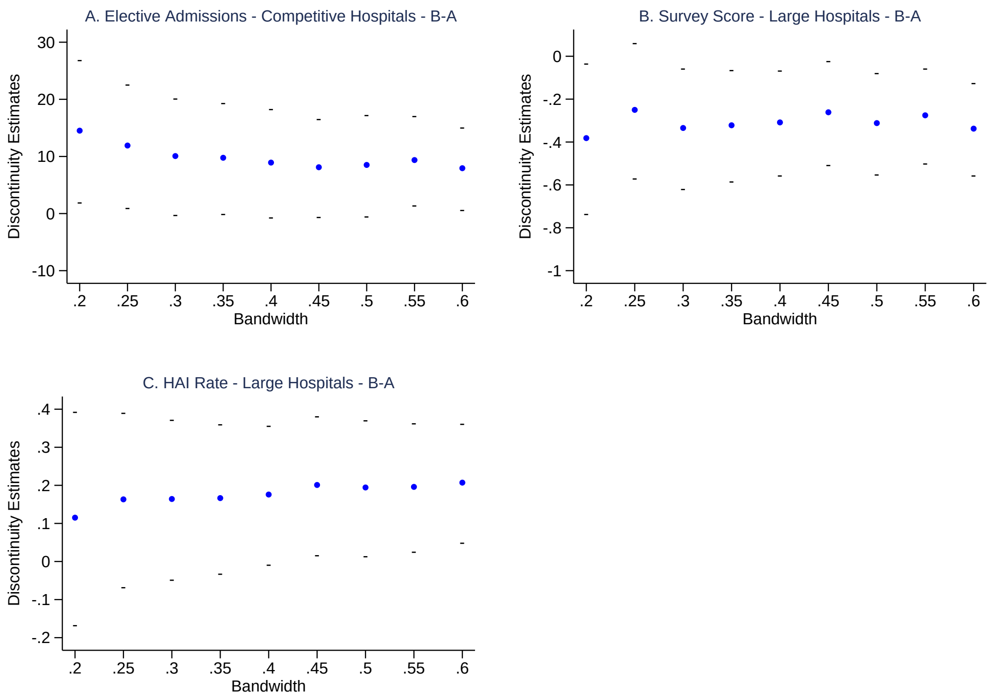
Notes: Column (1) presents the discontinuity estimate in survey score for resubmitters between the original Fall survey and the resubmitted Spring survey. Column (2) presents the discontinuity estimate in survey score for resubmitters between adjacent waves of the survey. Columns (3) and (4) present the discontinuity estimates in survey scores for large hospitals when including and not including resubmitters, respectively. Columns (5) and (6) present the discontinuity in the probability of resubmitting conditional on their Fall grade on the B-A margin and C-B margins, respectively. Large hospitals are defined as having above the median number of beds and small hospitals as having below the median number of beds. All variables are normalized so that the values correspond to standard deviations. All models include a first-order polynomial in distance fully interacted with a dummy for being over the grade threshold. The sample includes only hospitals that took the survey reported in the first release of the Hospital Safety Score in June 2012 for columns (3) and (4), only resubmitters for columns (1) and (2), and only hospitals which took the survey in the Fall for columns (5) and (6). Survey scores and the running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015. Standard errors clustered by hospital in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table B5: Regression Discontinuity Estimates for Hospital Costs

	B-A Grade Margin			C-B Grade Margin		
Admission Type	All	Elective	Non-Elective	All	Elective	Non-Elective
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity Estimates	-2.385	-2.756	-0.336	-0.280	-2.607	0.589
	(1.499)	(1.674)	(1.235)	(1.463)	(2.251)	(1.306)
Observations	333	276	326	275	328	272
Hospitals	173	129	177	165	124	162

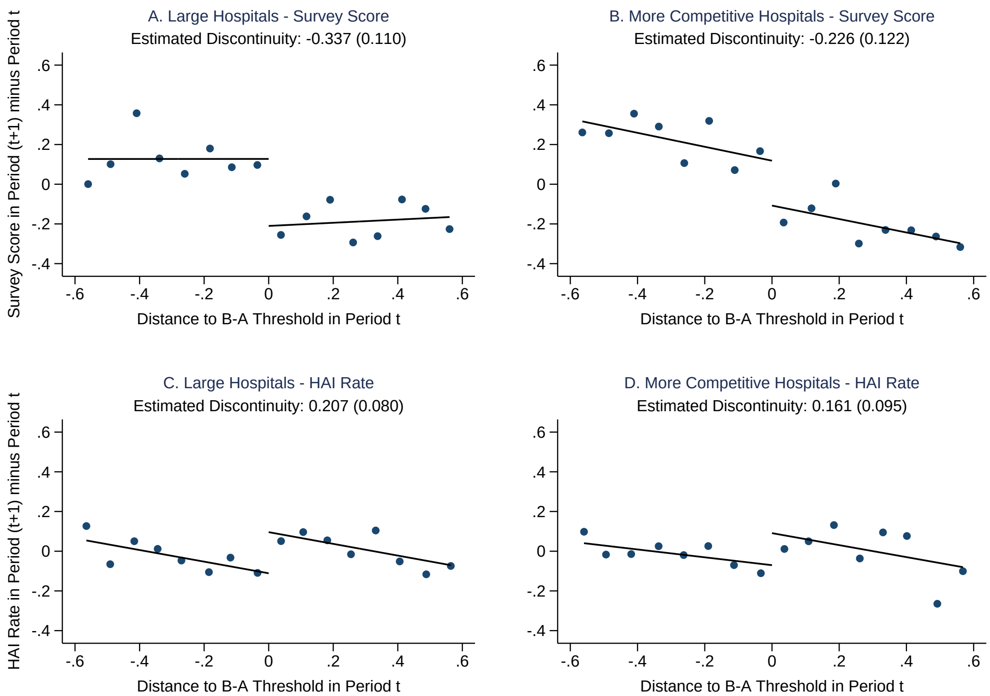
Notes: The dependent variable for all models is the percentage change in inpatient admissions between six month rating periods. All admissions are defined as all elective or non-elective admissions not including deliveries. Elective admissions are admissions scheduled more than 24 hours in advance. Non-elective admissions are admissions not scheduled more than 24 hours in advance. All models include a first-order polynomial in distance fully interacted with a dummy for being over the grade threshold. Costs are computed by multiplying the mean list charges for a given hospital-month by the cost-to-charge ratio for that hospital. Cost esimates are based on hospital discharge records from Arizona, Florida, Kentucky, New Jersey, New York, and Washington and cost-to-charge ratios derived from Medicare Cost Reports. The running variable distance to threshold comes from the Leapfrog Group for Spring 2012 to Spring 2015. Standard errors clustered by hospital in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Figure B1: Bandwidth Robustness - Main Results



Notes: These graphs present robustness to bandwidth choice for the main results. Graph A presents discontinuity estimates for the percentage change in elective admissions for hospitals below the median Herfindahl-Hirschman Index on the B-A margin. Graph B presents discontinuity estimates for the change in survey score for hospitals above the median number of beds on the B-A margin. Graph C presents discontinuity estimates for the change in the rate of healthcare-associated infections for hospitals above the median number of beds on the B-A margin. Dashes represent 95 percent confidence bounds. HAI rates and inpatient admissions are based on hospital discharge records from Arizona, California, Florida, Kentucky, New Jersey, New York, and Washington. Survey scores and the running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure B2: Changes in Survey Score by Competition and Hospital Size



Notes: These graphs present changes in the Leapfrog Hospital Survey and healthcare-associated infection (HAI) count for large and more competitive hospitals. Large hospitals and more competitive hospitals have above the median number of beds and below the median Herfindahl-Hirschman Index, respectively. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. The sample includes only hospitals that took the survey reported in the first release of the Hospital Safety Score in June 2012. Survey score and the running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure B3: Timeline of Leapfrog Hospital Survey

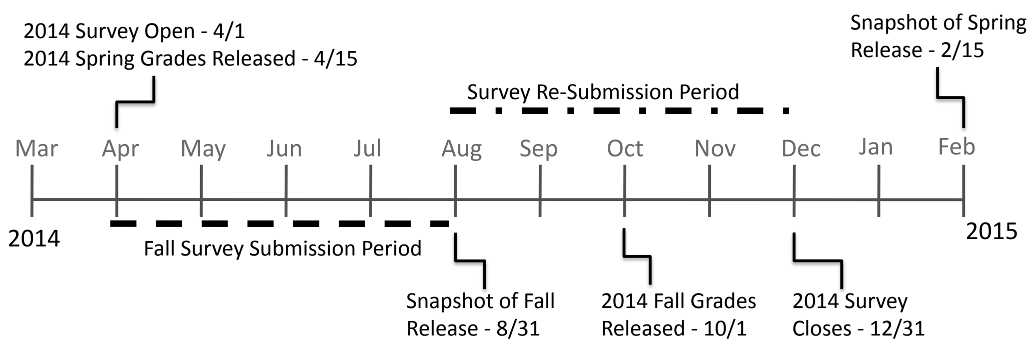
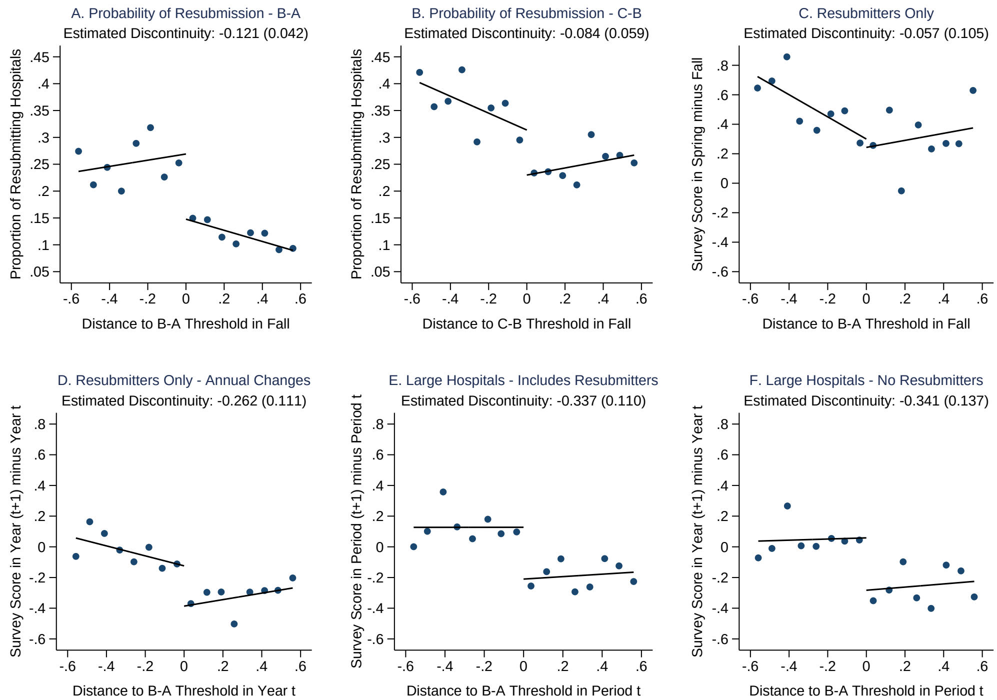
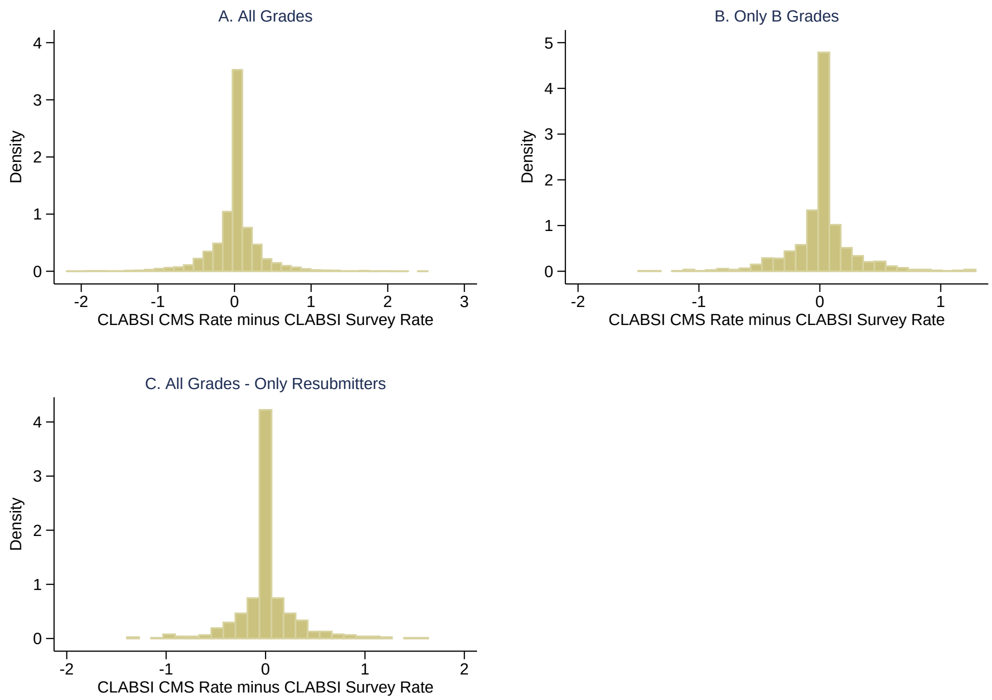


Figure B4: Resubmitting and Non-Resubmitting Hospitals



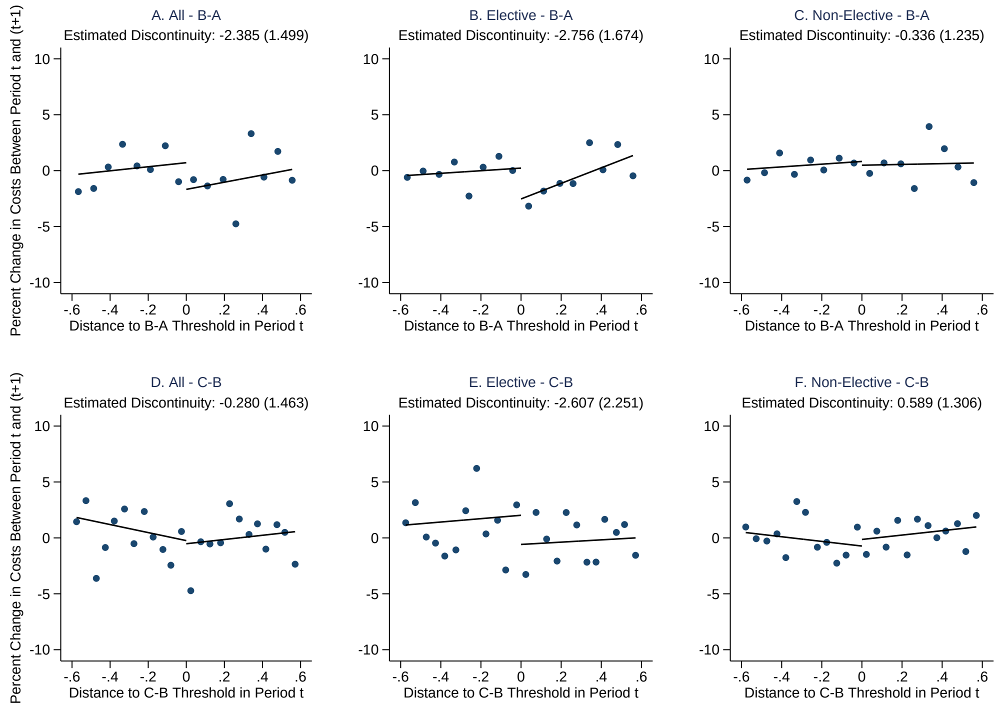
Notes: These graphs examine the behavior of hospitals that resubmit the survey and to what extent these resubmitters drive the response to the survey. Graph A presents the change in the survey score for resubmitters between the original Fall survey and the resubmitted Spring survey. Graph B presents the change in survey score for resubmitters between adjacent waves of the survey. Graphs C and D present changes in survey scores for large hospitals when including and not including resubmitters, respectively. Graphs E and F present the probability of resubmitting conditional on their Fall grade on the B-A margin and C-B margins, respectively. Large and small hospitals are defined as having above and below the median number of beds, respectively. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. The sample includes only hospitals that took the survey reported in the first release of the Hospital Safety Score in June 2012 for Graphs C and D, only resubmitters for Graphs A and B, and only hospitals which took the survey in the Fall for Graphs E and F. Survey scores and the running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure B5: Assessment of Outcome Data Misreporting



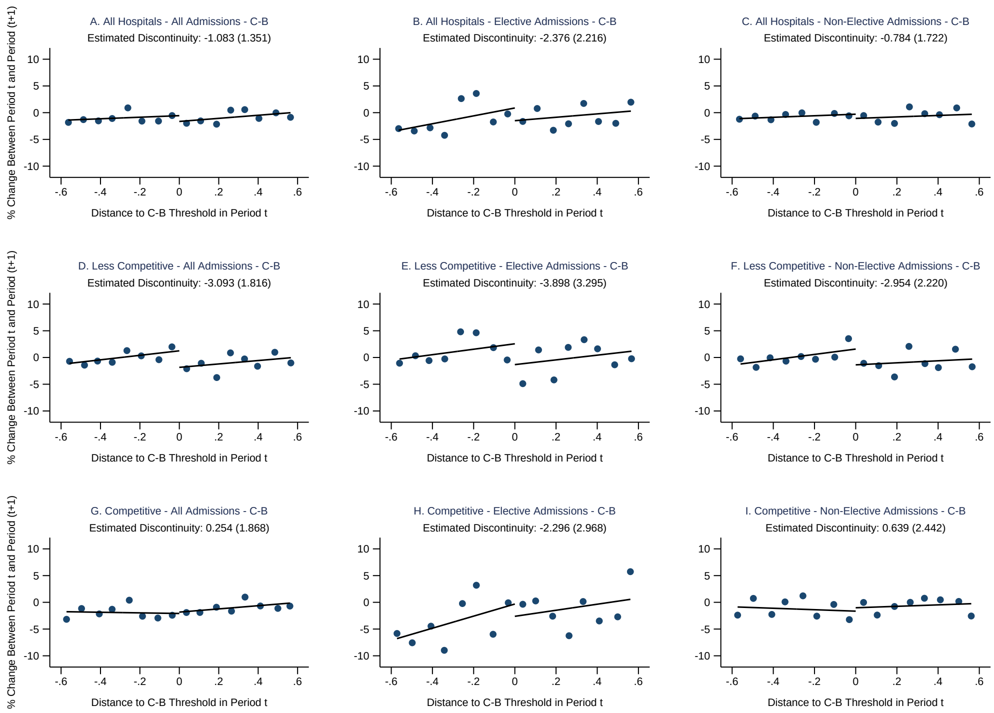
Notes: These graphs present the estimated distributions of the difference between Central Line-Associated Cather Infections (CLABSI) rates reported by the Centers for Medicare and Medicaid Services and on the Leapfrog Hospital Survey. Graph A presents the estimated distribution of the difference in CLABSI rates for hospitals that had recently received any grade. Graph B presents the estimated distribution of the difference in CLABSI rates only for hospitals that had recently received a "B." Graph C presents the difference in CLABSI rates only for surveys that were resubmitted.

Figure B6: Changes in Hospital Costs



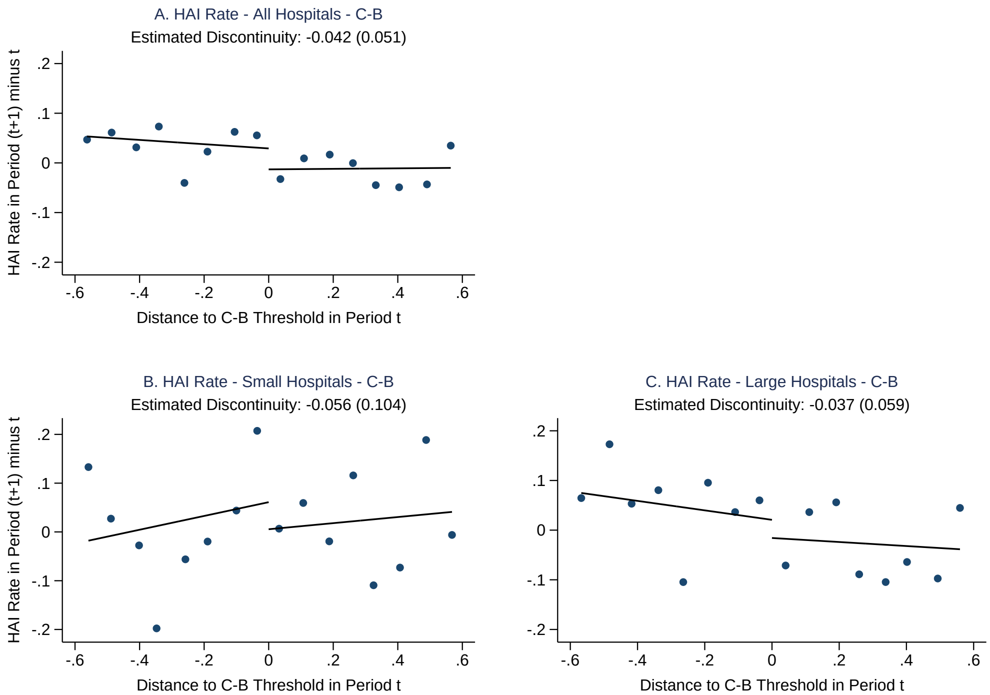
Notes: These graphs present changes in overall costs for hospitals. The dependent variable for all models is the percent change in costs between six month rating periods. All admissions are defined as all elective or non-elective admissions not including deliveries. Elective admissions are admissions scheduled more than 24 hours in advance. Non-elective admissions are admissions not scheduled more than 24 hours in advance. All graphs are weighted by the number of beds in the hospital. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. Costs are computed by multiplying the mean list charges for a given hospital-month by the cost-to-charge ratio for that hospital. List charges are based on hospital discharge records from Arizona, Florida, New Jersey, New York, and Washington. Cost-to-charge ratios are by fiscal year and are derived from Medicare Cost Reports. The running variable distance to threshold comes from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure B7: Changes in Inpatient Admissions - C-B Margin



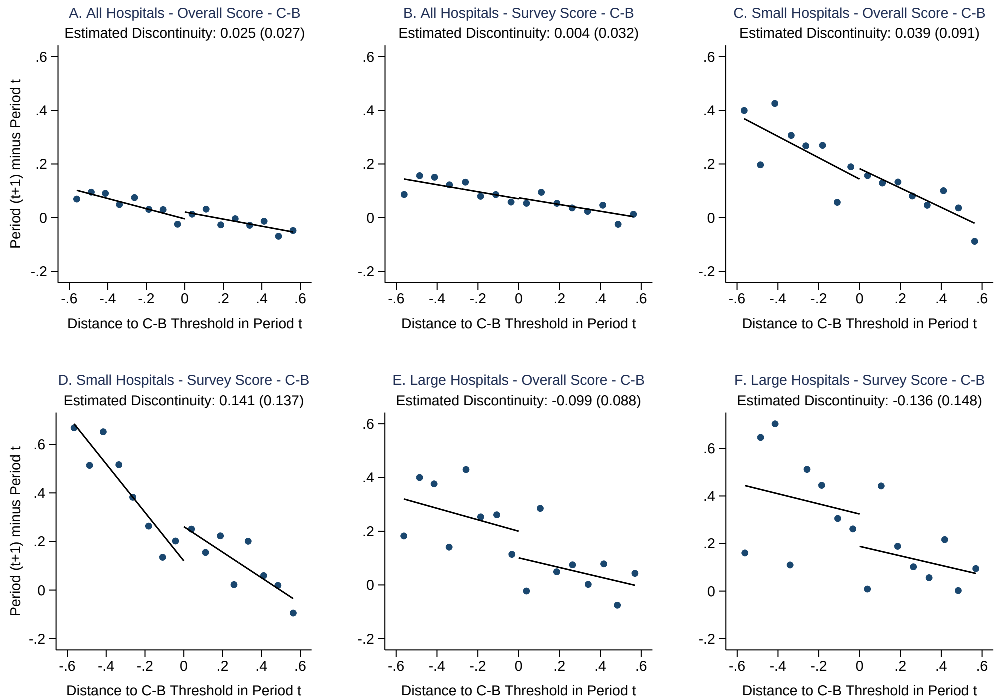
Notes: These graphs present changes in inpatient admissions on the C-B margin. The dependent variable for all models is the percentage change in inpatient admissions between six month rating periods. All admissions are defined as all elective or non-elective admissions not including deliveries. Elective admissions are admissions scheduled more than 24 hours in advance. Non-elective admissions are admissions not scheduled more than 24 hours in advance. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable and are weighted by the number of beds in each hospital. Admission counts are based on hospital discharge records from Arizona, California, Florida, Kentucky, New Jersey, New York, and Washington. Sample is trimmed at the 99th percentile of the dependent variable. "All Hospitals" includes all hospitals in the sample. "Less Competitive" includes only hospitals with above the median value of the Herfindahl-Hirschman Index. "Competitive" includes only hospitals with below the median value of the Herfindahl-Hirschman Index. The running variable distance to threshold comes from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure B8: Changes in Healthcare-Associated Infections - C-B Margin



Notes: These graphs present changes in the rate of healthcare-associated infections on the C-B margin. The numerator of "HAI Rate" is defined as 1 if the patient received any healthcare-associated infection during their inpatient stay. The denominator of "HAI Rate" is the total number of admissions to the hospital in a given month. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. HAI rates are based on hospital discharge records from Arizona, Florida, Kentucky, New Jersey, New York, and Washington. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. The running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015.

Figure B9: Changes in Scores - C-B Margin



Notes: These graphs present changes in the Hospital Safety Score and the Leapfrog Hospital Survey total score on the C-B margin. The dependent variable in each model is the change in the given variable between six month rating periods. "Overall Score" is the Hospital Safety Score, that is, a weighted sum of 26 or 28 measures. "Survey Score" is a weighted sum of all 12 measures on the Leapfrog Hospital Survey. All variables are normalized so that the values correspond to standard deviations. The lines are fitted values from regressions that include a first-order polynomial in distance interacted with a dummy for being above the threshold. The points represent means of the dependent variable for cells of 0.075 standard deviations of the running variable. "All Hospitals" includes all hospitals rated by the Leapfrog Group. "Large Hospitals" and "Small Hospitals" includes only hospitals with above or below the median number of beds, respectively, and includes only hospitals that took the survey reported in the first release of the Hospital Safety Score in June 2012. Overall score, survey score, and the running variable distance to threshold come from the Leapfrog Group for Spring 2012 to Spring 2015.

Chapter 2

2 The Causal Effect of Education on Health: Evidence from British Administrative Hospital Records

2.1 Introduction

There exists a robust and well-documented positive correlation between education and health, regardless of the education level, time period, or country being studied.[42] However, the recent evidence on the causal effect of education on health, focusing on quasi-experimental designs exploiting compulsory schooling reforms, is largely inconclusive and most studies rely on imprecise survey data.[43] In the most cited study in this literature, Lleras-Muney (2005) exploits variation in compulsory schooling laws in the United States across states and over time, providing instrumental variables estimates indicating reductions in mortality of 3-6 percentage points.[44] Since then, many studies have exploited compulsory schooling reforms in Europe to estimate the causal effect of education, but the evidence is still inconclusive. Particularly, some studies find no evidence of an effect (Albouy and Lequien 2009, Clark and Royer 2013, Johnston et al. 2015), some find significant causal effects (Oreopoulos 2006, Arendt 2008, Silles 2009, Lager and Torssander

[42]Goldman (2001) provides an expansive review of the literature.

[43]See Eide and Showalter (2011) for a thorough review of recent literature on both pathways of the causal relationship between education and health. Appendix Figure A1 shows regression discontinuity estimates of overall health and several morbidities using survey data, which prominently displays that estimates using only survey data are often very imprecise.

[44]However, Mazumder (2008) reanalyzes the study of Lieras-Muney (2005) using a broader sample of census data, and discovers that the effects of education on mortality induced by the compulsory schooling law changes are not robust to the inclusion of state-specific time trends.

2012, Fischer et al. 2013), while others find inconclusive results (Arendt 2005, Jürges et al. 2010, Powdthavee 2010, Gathman et al. 2014).[45]

This paper supplements the previous literature by exploiting quasi-experimental variation in education induced by two changes to compulsory schooling laws in Britain. The first reform in 1947 increased the minimum school leaving age from 14 to 15 while the second reform in 1972 increased it from 15 to 16. These British reforms are an ideal scenario for the assessment of outcomes across cohorts because the first reform affected roughly 50 percent of the population, and the second reform affected 25 percent of the population, compared with only 5 percent of relevant cohorts in the US (Lleras-Muney 2005). The reforms affected individuals based on a strict date of birth cutoff, so we are able to use a regression discontinuity (RD) design to estimate the causal effect of education on health.

This design uses the fact that the reforms led to sharp increases in educational attainment. Since it is reasonable that individuals born just on either side of the threshold are similar in both observable and unobservable characteristics, the RD design allows us to estimate the causal effect of education on health without the concern of unobservable factors biasing our estimates. Several other studies, namely Clark and Royer (2013), have used the same design to analyze the effect of these compulsory schooling reforms on health. However, this is the first study to use a very large data set of administrative hospital records to produce precise

[45]Several other studies claim causal effects of education on health using mixed structural and instrumental variables approaches (Heckman et al. 2015) and cross-Europe fixed effects approaches (Brunello et al. 2015).

estimates of the English compulsory schooling reforms on morbidities that are known to be closely associated with education.

We assess health outcomes using the expansive Hospital Episode Statistics data, which contains records of every hospital admission and outpatient visit at National Health Service hospitals in England. By pooling 22 years of inpatient records and 8 years of outpatient records, we are able to precisely assess the effect of education on mortality and many objectively measured health outcomes. Additionally, to produce estimates of the first stage relationship, examine the health-education gradient, and illustrate the imprecision of survey data, we use the Health Survey For England and the General Household Survey. This is also the first study to estimate the relationship between education and self-reported health using the 2011 census, so we are able to estimate the effect of the reforms for individuals ten years older than previous studies.

Consistent with Clark and Royer (2013), we do not find evidence of a causal effect of education on self-reported health or mortality, even for individuals ten years older than those in their sample. Additionally, we do not find an effect even for morbidities which are known to be associated with education, such as diabetes and lung cancer.

The remainder of the paper proceeds as follows. Section I describes the data sets used in this study. Section II details our empirical methodology. Section III presents first-stage estimates, RD estimates, and correlations between education and health from survey data, as well as our main results, causal estimates from HES administrative data, and Section IV concludes.

2.2 Background

2.2.1 Compulsory Schooling Reforms in Britain

An important distinction between American and British compulsory education laws is that British laws are national while compulsory schooling in the US is state-regulated. These laws specify the maximum age by which children must start school and the minimum age at which children can leave school. This paper examines two reforms which increased the minimum age which children are allowed to leave school by one year.

The 1944 Education Act raised the minimum age at which children are allowed to leave school from 14 to 15, specifying that the change will take effect on April 1, 1947. The 1944 Act also gave the Minister of Education the authority to raise the minimum school leaving age to 16, which he did in 1972. This second increase in the minimum school leaving age to 16 took effect on September 1, 1972. In practice, the 1947 change meant that students could not leave school until part way through ninth grade and the 1972 change meant they could not leave school until part way through tenth grade.

It is important to understand what specifically the education consisted of when examining the effect of education on health. For example, if the students staying one additional year were exposed to a health class that they otherwise would not have, this would likely have a larger effect than if the additional education was simply an extension of their current curriculum. In this setting, the key elements of the school system did not change, but the extra years were instead used to simply expand the already existing curriculum. Ministry of Education reports and commentary from the time suggest that the extra year created by the 1947 change was used to introduce some students to more advanced materials and help other students master more basic material.

2.2.2 Education-Health Gradients

A strong correlation between education and health has been documented in many different settings and at different levels of education, beginning with Kitagawa and Hauser (1968). Appendix Figure A2 presents a visual representation of the health-education gradient for overall health, depression, and the probability of smoking. Each point in the figure represents the mean of the outcome variable at the given education level. The panels show that there is a steep gradient across education levels for overall health, and the probability of having depression and smoking, often with large differences between the education levels affected by the reforms.[46]

Column (1) of Table 1 displays univariate correlations between two outcomes from the Health Survey for England: the proportion of individuals that reported

[46]In results not shown, there are similarly steep gradients for many other outcomes as well.

being in "Bad Health or Worse" and those reporting having "Limited Activity," between individuals that left school at 15 versus 14 for the 1947 reform and 16 versus 15 for the 1972 reform.[47] Column (1) shows that a strong education-health gradient does exist even at the low levels of education that we are examining in this study and for individuals that are very close to the same age as they are in the 2011 census.[48] Specifically, for individuals born between April 1, 1928 and April 1, 1938, individuals that left school at 15 had an 8.2 percentage point lower probability of reporting being in "Bad Health or Worse" than those that left school at 14. Similarly, for individuals born between September 1, 1952 and September 1, 1962, individuals that left school at 16 had an 8.8 percentage point lower probability of reporting being in "Bad Health or Worse" than those that left school at 15. On bases of 17 percent for the 1947 reform and 12.8 percent for the 1972 reform, these are extremely large differences. It appears that there are very large differences in health status between individuals that were induced to stay one more year by the reforms and those that were not.

Column (2) adds a small set of observable covariates to the regression which were previously omitted. For each outcome in each reform, the estimate drops dramatically when adding the observable covariates, indicating that part of the

[47] "Bad Health or Worse" corresponds to reporting "Bad" or "Very Bad" for their general health status. "Limited Activity" corresponds to reporting that a long-term illness or disability limits day-to-day activities either "a lot" or "a little," and that this has lasted or was expected to last at least 12 months. Appendix Table A2 shows that the same qualitative conclusions are reached regardless of which level of health we choose as the outcome.

[48] We specifically chose to use data from the Health Survey for England for 2009-2014 to produce the correlations in Table 1 in order to provide correlations for individuals the same age as those in the 2011 census.

Table 2.1: Correlations Between Education and Health

	Bad Health or Worse		Limited Activity	
	(1)	(2)	(3)	(4)
Panel A. 1947 Reform				
Age left education = 15	-0.082***	-0.064***	-0.142***	-0.108***
	(0.015)	(0.023)	(0.024)	(0.041)
Outcome Mean	0.170	0.170	0.517	0.517
Observations	2,599	1,449	1,776	945
Panel B. 1972 Reform				
Age left education = 16	-0.088***	-0.048***	-0.099***	-0.043*
	(0.013)	(0.014)	(0.020)	(0.024)
Outcome Mean	0.128	0.128	0.322	0.322
Observations	3,834	2,716	2,742	1,928
Includes Covariates	No	Yes	No	Yes

Notes: Sample comes from the Health Survey for England 2009-2013. The sample for the 1947 reforms consists of individuals born five years on either side of April 1, 1933. The sample for the 1972 reform consists of individuals born five years on either side of September 1, 1957. The included covariates are age, sex, race, income, whether the individual is a frequent drinker, and if the individual has ever smoked. Robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

education-health gradient is explained by these covariates. We only have a small set of observable characteristics, so it is feasible that there are many additional unobservable factors explaining the education-health relationship, and as we add additional covariates, the relationship could continue to diminish, potentially to 0.[49]

Table 2 presents the difference between many observable characteristics and outcomes across the education levels affected by the reforms. Table 2 shows that many characteristics and outcomes are significantly different, and nearly all are economically significant, between the education levels affected by the reforms.

[49] Appendix Table A1 presents how the estimates of the relationship change for the outcome "Bad Health or Worse" when adding each additional covariate.

Notably, individuals that left school at 16 that were affected by the 1972 reform have about 25 percent higher income than those that left school only one year earlier.[50] Moreover, the individuals receiving an additional year of education had lower inpatient and outpatient visits for both reforms, including a 26 percent reduction in outpatient visits for those individuals affected by the 1972 reform, which implies an improved health status for those more educated individuals.

These large differences in individual characteristics across education levels will lead to omitted variable bias in the education-health gradient if these factors are not taken into account. The statistical and economic significance of the differences of many of these characteristics across education levels leads us to strongly believe that there are likely large differences in unobservable characteristics as well, conveying the need for a regression discontinuity design to control for differences in observable and unobservable characteristics, and provide us with causal estimates of this relationship.

2.3 Data

To obtain first-stage estimates, as well as preliminary regression discontinuity (RD) estimates and correlations between health and education, we pool survey data from the Health Survey of England (HSE) for all waves from 1991-2014, along with all waves of the General Household Survey from 1986-1996. For our main

[50]Oreopoulos (2006) and Clark and Royer (2013) do find causal effects of these same reforms on income, but they estimate less than a ten percent effect.

Table 2.2: Differences of Individual Characteristics Across Education Levels

	1947 Reform			1972 Reform		
Age left education	14	15	Difference	15	16	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.423	0.468	0.045**	0.466	0.452	-0.014
	[0.494]	[0.499]	(0.020)	[0.499]	0.498	(0.017)
Age	79.60	76.26	-3.344***	55.56	53.40	-2.158***
	[2.773]	[2.745]	(0.109)	[3.051]	[3.185]	(0.108)
White	0.932	0.970	0.038***	0.941	0.934	-0.007
	[0.251]	[0.170]	(0.009)	[0.236]	[0.249]	(0.008)
Income	17943	19257	1,313.766	29024	36398	7,374.066***
	[20835]	[19128]	(940.177)	[26385]	[31798]	(1,106.415)
Frequent Drinker	0.192	0.201	0.009	0.184	0.167	-0.016
	[0.394]	[0.401]	(0.018)	[0.387]	[0.373]	(0.014)
Ever Smoke	0.598	0.622	0.023	0.693	0.622	-0.071***
	[0.490]	[0.485]	(0.019)	[0.461]	[0.485]	(0.016)
Diabetes	0.182	0.178	-0.005	0.0923	0.0775	-0.015
	[0.386[	[0.382]	(0.015)	[0.290]	[0.267]	(0.010)
High Blood Pressure	0.647	0.586	-0.061***	0.357	0.309	-0.048***
	[0.478]	[0.493]	(0.019)	[0.479]	[0.462]	(0.017)
Inpatient Count	0.138	0.128	-0.010	0.137	0.124	-0.013
	[0.507]	[0.503]	(0.007)	[0.787]	[0.449]	(0.008)
Outpatient Count	0.441	0.407	-0.035	0.343	0.255	-0.087***
	[2.024]	[1.681	(0.025)	[1.834]	[1.295]	(0.020)

Notes: Sample comes from the Health Survey for England 2009-2013 and the General Household Survey 1998-2007. The sample for the 1947 reforms consists of individuals born five years on either side of April 1, 1933. The sample for the 1972 reform consists of individuals born five years on either side of September 1, 1957. Columns (1), (2), (4), and (5) represent means of the specified variable for all individuals that left school at the specified age. Columns (3) and (6) represent the difference in means between individuals that left school at 15 versus 14 for the 1947 reform and 16 versus 15 for the 1972 reform. Standard deviations in brackets and robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

results, we use the National Health Service's (NHS) Hospital Episode Statistics (HES). Specifically, we pool all inpatient admissions from 1989-2010, and separately, all outpatient appointments from 2003-2010. Additionally, we combine the HES with the English census to obtain hospitalization rates. Finally, we perform the first analysis of the reforms on self-reported health from the 2011 census.[51]

[51]These years of data represent all available inpatient and outpatient records available for the HES as of 2013.

Starting in 1991 and continuing to the present, the HSE is an annual survey that combines answers from a questionnaire with objective information obtained by a nurse visit. The HSE contains demographic information such as gender and age, as well as information on what age each individual last attended school, self-reported questions about health and health behaviors, and a limited amount of objective health measures. We use the HSE for three purposes. First, we provide estimates of the correlation between education and health by assessing the relationship between the age an individual left full time education and a variety of health behaviors at key levels of education. Second, we provide first-stage estimates by assessing the increase in overall educational attainment. Third, we construct RD and instrumental variables (IV) estimates for the effect of the reforms on a variety of self-reported health measures.

In order to increase the precision of our estimates, we combine the HSE with data from the GHS. The GHS is an annual repeated cross-sectional survey of about 10,000 households annually in Great Britain. Although the GHS does not contain information on the majority of the health measures in the HSE, it does contain a question on general health, as well as age left full time education, in addition to basic demographic information including month and year of birth for the 1986-1996 waves.[52] These surveys are used to construct estimates of the first-

[52]Since the GHS only contains month of birth until 1996, we use the 1986-1996 waves for the first-stage estimates. Additionally, the GHS did not include the number of inpatient and outpatient visits until the 1998 wave and the survey ended in 2007, so we use the 1998-2007 waves to find the mean of inpatient and outpatient visits at education levels affected by the reforms.

stage relationship between the reforms and educational attainment, as well as to motivate our analysis by constructing health-education gradients for the number of inpatient and outpatient visits.

Our primary data sources are the inpatient and outpatient records of the HES. Originally conceived in 1987, HES increased the recording of national admission records for England from only 10 percent to 100 percent after the development of HES. HES collects data from all inpatient admissions, outpatient visits, and emergency department visits, with data available from 1989, 2003, and 2007, respectively. Since we were unable to obtain access to the month of birth for emergency department patients, we assess outcomes using all inpatient and outpatient data available in the HES. Although the HES contains information on many demographic and geographic characteristics of patients and hospitals, we use only the month and year of birth, primary and subsequent diagnoses, and sex of each patient for our analysis.[53]

Additionally, we analyze self-reported health outcomes and construct hospitalization rates from the 2001 and 2011 censuses.[54] The census asks only a small

[53]The HES is an aggregation of submissions by heterogeneous health providers across the country, and the submissions to the HES of some providers contain measurement error of various magnitudes in their dates of birth, admission dates, and appointment dates. To address this systematic measurement error, we undertook an extensive process of cleaning the data to remove providers from our data sets in the years in which there is significant measurement error. A full discussion of this process, along with how much data is dropped in each file year, is contained in Appendix B. Due to the large amount of measurement error across providers, we use the cleaned data set for all results presented in this study, although our results are very similar if we use the original data.

[54]To construct hospitalization rates, we subtract off the mortality count for each month and create the population for each day by assuming that the mortality rate is uniform throughout any given month. Then, the denominator for a hospitalization rate for any given month of birth is the mean of the population for each day of that month. Rates are then scaled by dividing by

number of health-related questions, but covers all of England and Wales, and thus generates very precise estimates. The 2001 census is used to construct hospitalization rates and the 2011 census is used to estimate the effect of the reforms on self-reported health.

2.4 Empirical Methodology

To assess the causal effect of education on health outcomes, we use a regression discontinuity framework. First, we estimate the effects of the compulsory schooling reforms on educational attainment, our first-stage relationship, separately for each reform, as follows:

$$E_{ict} = \alpha + \beta_D D_{ic} + f(C_{ic}) + \epsilon_{ict} \quad (2)$$

The dependent variable E_{ict} is a measure of educational attainment. The subscripts indicate the individual i, cohort c, and time t. The explanatory variable of interest, the instrument D_{ic}, takes on a value of one if an individual was born after the relevant cutoff (April 1933 for the 1947 reform and September 1957 for the 1972 change), and 0 otherwise. C_{ic} is an individual's birth cohort relative to the relevant cutoff. The function $f(.)$ captures the underlying relationship between birth cohort and educational attainment, and consists of a quadratic polynomial in C_{ic} fully interacted with D_{ic}.

22, the number of years in which we have data, and multiplying by 10,000, to create a monthly rate per 10,000 individuals.

After documenting the strong first-stage relationship between the reforms and educational attainment, we estimate the relationship between health and education using the 2011 census as well as the Hospital Episode Statistics (HES) administrative hospital records. To generate credible causal estimates of the effect of education on health, we use compulsory school law changes as an instrument for education. The model is specified as follows:

$$Y_{ict} = \gamma + \alpha_E E_{ic} + g(C_{ic}) + e_{ict} \tag{3}$$

where Y_{ict} is a health outcome (i.e., hospitalization rate) for individual i in cohort c at time t. We use the compulsory schooling reforms as an instrument for educational attainment E_{ic}. The underlying relationship between the dependent variable Y_{ict} and the birth cohort is captured by the flexible function $g(.)$. Again, $g(.)$ is made up of a quadratic polynomial in C_{ic} fully interacted with the instrument D_{ic}.

Since our primary data source, the HES, does not contain any information on educational attainment, we estimate our first-stage from the pooled Household Survey for England (HSE) and General Household Survey (GHS) data. Since the surveys contain information on educational attainment and the instrument (month and year of birth), and the HES contains information on health outcomes and the instrument, we are able to perform two-sample instrumental variables (TSIV) estimation. Our IV estimate is the ratio of the reduced-form estimate of the effect of the policy change on an outcome of interest and the first-stage estimate of the effect of the policy change on education. We use the delta method to compute the standard error for two-sample IV estimation.[55]

2.5 Results

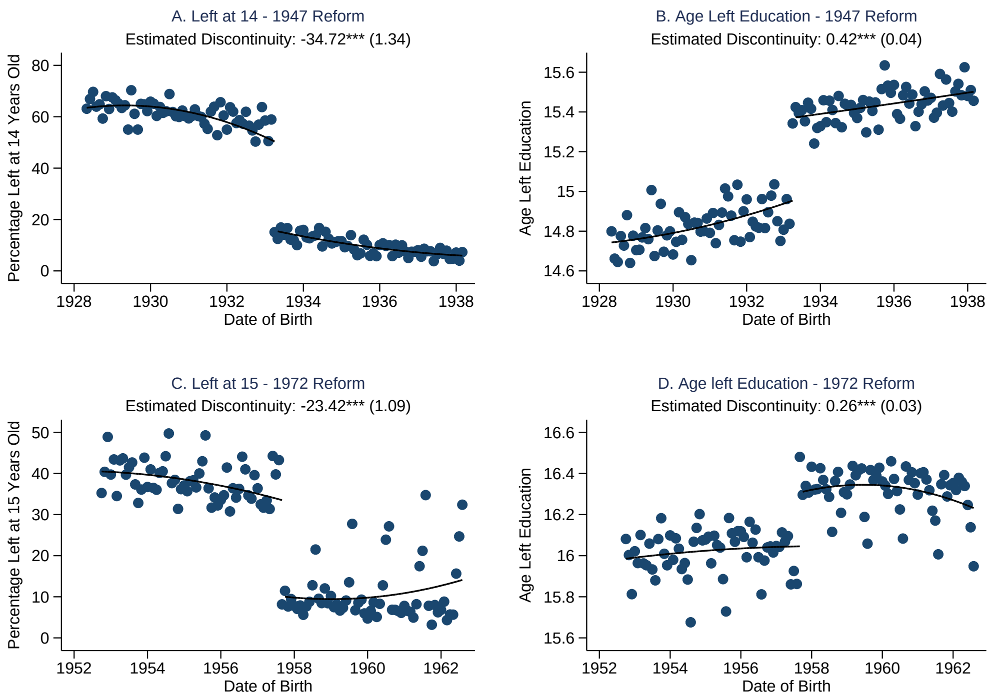
Our results are split into three subsections. First, we describe the impacts of the schooling reforms on educational attainment. Next, we present causal estimates of education on self-reported health from the 2011 census. Finally, we present the effect of the law changes on health outcomes from the Hospital Episode Statistics (HES).

2.5.1 Effects of the Reforms on Educational Attainment

Figure 1 presents the relationship between birth cohort and the percentage of each cohort that left school at 14 for the 1947 reform and left school at 15 for the 1972 reform, as well as the relationship between birth cohort and the age an individual left full time education. Each panel contains superimposed fitted quadratic trends, where the data is fitted separately on each side of the cutoff, as well as estimated discontinuities obtained from the estimation of equation (1). Panels A and B of Figure 2 show that the 1947 reform reduced the percentage of individuals leaving education at 14 years old by roughly 34 percent and increased

[55]Specifically, we use the formula $Var(\frac{\pi}{\alpha}) = \frac{Var(\hat{\pi})}{\alpha^2} + \frac{\pi^2 Var(\hat{\alpha})}{\alpha^4}$ where π represents the point estimate from the reduced form model, α represents the point estimate from the first stage estimation, that is, β_D in equation (1), and $Var(x)$ represents the variance of x.

Figure 2.1: Effect of the Compulsory Schooling Reforms on Educational Attainment



Notes: Sample comes from the Health Survey for England 1991-2014 and the General Household Survey 1986-1996. The lines are fitted values from regressions that include a second-order polynomial in month-year-of-birth fully interacted with a dummy for being above the relevant cutoff. The points represent means of the dependent variable for each month-year of birth cell. Robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

the mean age at which individuals left full time education by about 0.4 years, respectively. Additionally, the 1972 reform reduced the percentage of individuals receiving less than 10 years of education by about 21 percent and increased the mean age at which individuals left full time education by about 0.2 years.

Table 3 quantifies these estimates using a five year bandwidth as used in Figure 1, and includes month of birth and survey year dummies to absorb idiosyncratic noise.[56] There are several important points to note about these estimates. This

[56]Appendix Table A3 presents first-stage estimates of the effect of the reforms on education

suggests that a large fraction of individuals were "compliers," especially relative to the proportion affected in the United States. Moreover, the first stage estimates are very precisely estimated, with t-statistics exceeding twenty for both reforms. Additionally, there is very little spillover from either reform, which implies that the reforms exactly performed the job they were meant to, causing students who otherwise would have dropped out to remain in school for just one more year for each reform, but no more. Especially for the 1972 reform, it appears that for many months, students receive less education than they are required to. This occurs because the taking of the O-level satisfies the compulsory schooling requirement. Since these exams occur in May and June, students who are born in June, July, and August can actually finish school before they turn 16. This is discussed in much more detail in the Online Appendix of Clark and Royer (2013), and due to its systematic nature, this pattern can be controlled for using month-of-birth dummies, the inclusion of which has very little effect on our estimates.

2.5.2 Effects of the Reforms on Self-Reported Health - Census

After documenting the strong-first relationship between the reforms and educational attainment, we turn to the impacts of the reforms on self-reported health. To examine self-reported health, we use administrative data from the 2011 census which covers all individuals in England and Wales, allowing us to generate

using three different optimal bandwidth procedures. Regardless of the bandwidth used, the estimates are generally stable, bounded between the estimates in Table 3, and similar to those of Clark and Royer (2013).

Table 2.3: Effects of Compulsory Schooling Reforms on Education

	Years of education	≤9 years	≤10 years	≤11 years	≤12 years
Panel A. 1947 Reform					
Estimate	0.429***	-34.194***	-9.005*	0.130	0.155
	(0.109)	(4.653)	(4.945)	(2.938)	(2.254)
Outcome Mean	15.13	35.08	70.06	87.40	93.99
Observations	2,808	2,808	2,808	2,808	2,808
Panel B. 1972 Reform					
Estimate	0.269***	0.113	-23.295***	0.850	-4.586
	(0.095)	(1.281)	(3.719)	(4.567)	(3.597)
Outcome Mean	16.17	2.311	23.69	72.40	84.11
Observations	3,329	3,329	3,329	3,329	3,329

Notes: Sample comes from the Health Survey for England 1991-2014 and the General Household Survey 1986-1996. All estimates based on regressions that include a quadratic function of month-year of birth centered at the relevant cutoff, a quadratic interaction of month-year of birth and the relevant reform dummy, and dummy variables for year of survey and month of birth. "Outcome Mean" represents the mean of the outcome at the threshold. The bandwidth for all estimates is 60 months. Robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

very precise estimates. Clark and Royer (2013) have produced similar estimates using the 2001 census, but we examine the relationship between the reforms and self-reported health when individuals are 10 years older.

Table 4 provides estimates of the effect of the reforms on two self-reported outcomes: Being in "Bad Health or Worse" or having "Limited Activity."[57] A bandwidth of three years is used for all estimates, as this is close to the bandwidth obtained by two out of three optimal bandwidth procedures (presented in Appendix Table A4). We find that none of the estimates are statistically significant at the five percent level and are precisely estimated. For example, we find

[57] "Bad or Worse Health" corresponds to reporting "Bad" or "Very Bad" for their general health status. "Limited Activity" corresponds to reporting that a long-term illness or disability limits day-to-day activities either "a lot" or "a little," and that this has lasted or was expected to last at least 12 months.

Table 2.4: Effect of Compulsory Schooling Reforms on Self-Reported Health

	Mean	RF	IV
Panel A. 1947 Reform			
Bad Health or Worse	0.153	0.0012	0.0031
		(0.0024)	(0.0062)
Limited Activity	0.572	0.0054*	0.014*
		(0.0028)	(0.0074)
Panel B. 1972 Reform			
Bad Health or Worse	0.0743	0.0005	0.0022
		(0.0007)	(0.0033)
Limited Activity	0.192	0.0014	0.0064
		(0.0014)	(0.0066)

Notes: These estimates come from 2011 census data aggregated to the month-year-of-birth level. The column labeled "Mean" presents the value of the dependent variable at the threshold for the specified bandwidth. The column labeled "RF" presents reduced-form estimates which include a quadratic function of month-year of birth fully interacted with the relevant reform dummy, and an indicator for the month of birth. The column labeled "IV" presents instrumental variables estimates, which are the ratio of the reduced-form estimate and the first-stage estimate. The standard errors for these estimates are calculated using the delta method. The bandwidth for all estimates is 36 months. Robust standard errors in parantheses.

that the instrumental variables (IV) estimate for being in "Bad Health or Worse" for the individuals affected by the 1947 reform is 0.31 percentage points. On a base of 15.3 percent, this implies that we can rule out an effect size (that is, those affected by the reform had better health) larger than 5.92 percent. For the 1972 reform, we find an IV estimate of 0.22 percentage points. On a base of 7.4 percent, this implies that we can rule out an effect size larger than 5.74 percent. In Table 1, we found that looking at the same individuals at about the same age, the correlations between education level and these same self-reported outcomes revealed very large correlations, up to a 50 percent effect of education. However, using the reforms as an instrument, we find that the causal effect of education on self-reported health is no larger than 6 percent.

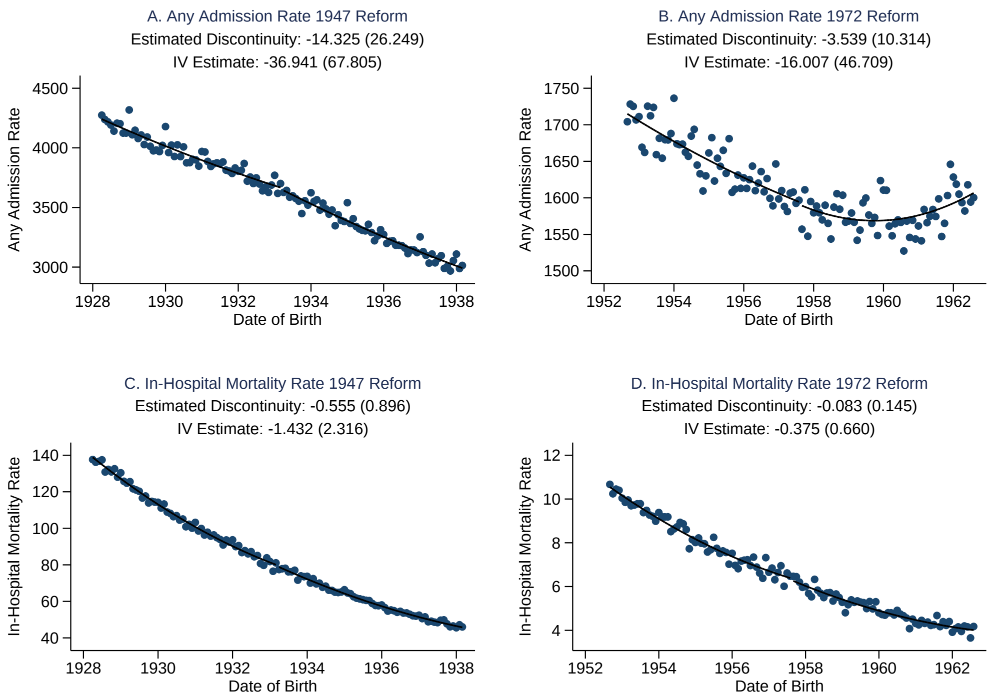
2.5.3 Effects of the Reforms on Hospitalizations

Using survey data, we did not find significant estimates of the reforms on self-reported health for a variety of measures that should be associated with education, but they are imprecisely estimated. Using the 2011 census, we were able to precisely show that the reforms did not affect two general measures of health. However, this does not rule out the possibility that there is an effect of the reforms on morbidities which are closely associated with education. To accomplish this, we turn to the Hospital Episode Statistics (HES).

Figure 2 displays the month-of-birth profile of the overall admission rate and the rate of in-hospital deaths. For example, “Any Rate” is interpreted as the number of inpatient admissions with any diagnosis per 10,000 individuals per month. The month-of-birth profile has a steep negative slope, but there is no evidence of any effect of the reforms for either in-hospital mortality or all inpatient admissions. For example, we are able to rule out an effect size larger than 4.6 percent for the overall admission rate for the 1947 reform and 6.7 percent for the 1972 reform. Not surprisingly, the effect sizes for any admission are quite similar to that of overall health from the census but much smaller than the estimates obtained from surveys.

We have shown that there is no effect on in-hospital mortality by any cause or the overall admission rate, but our data allows us to examine the effect of the reforms on specific diagnoses, some of which are highly related to education. First,

Figure 2.2: Effect of Compulsory Schooling Reforms on All Hospitalizations and Mortality

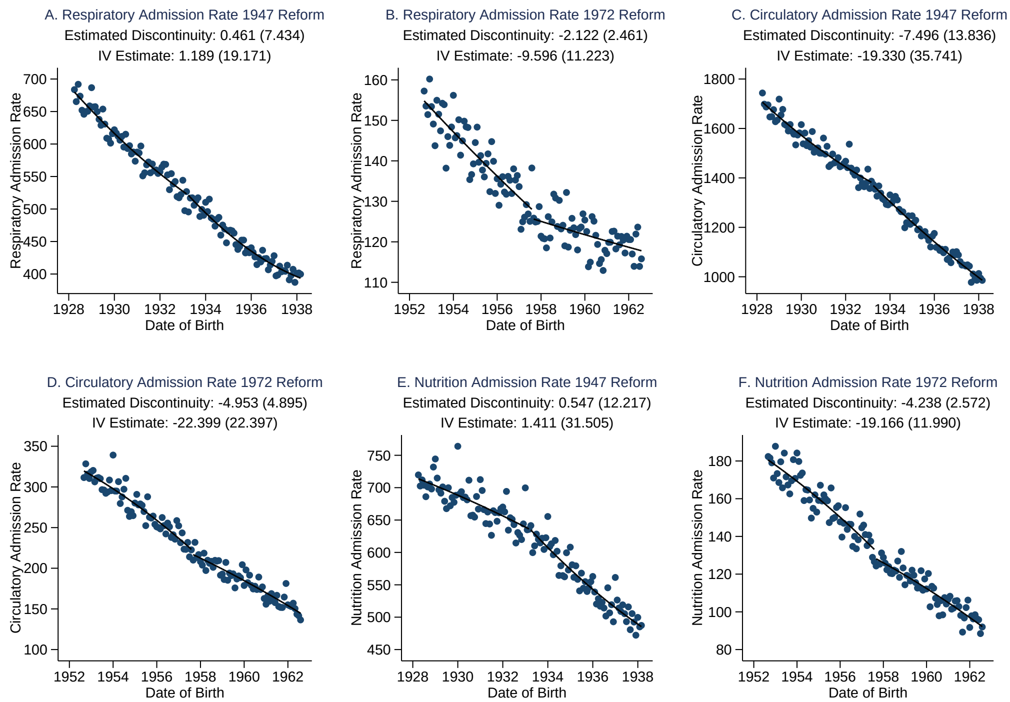


Notes: Sample comes from the Hospital Episode Statistics Inpatient Sample 1989-2010. The lines are fitted values from regressions that include a second-order polynomial in month-year-of-birth fully interacted with a dummy for being above the relevant cutoff. The points represent means of the dependent variable for each month-year of birth cell. Robust standard errors in parantheses.

we look at three broad categories of hospitalizations: respiratory, circulatory, and nutrition-related diseases.[58] These categories contain many medical conditions which are likely to be associated with education. Specifically, nutrition-related diseases consist of diabetes and other problems associated with diet and weight, so education is particularly likely to affect this set of diagnoses. However, there is no significant difference between those affected by either reform and those that

[58]Appendix Table A5 lists the ICD codes for each of the outcomes analyzed in this study. Specifically, the categories are "Diseases of the respiratory system," "Diseases of the Circulatory System," and "Endocrine, nutritional and metabolic diseases, and immunity disorders."

Figure 2.3: Effect of Compulsory Schooling Reforms on Broad Hospitalization Categories



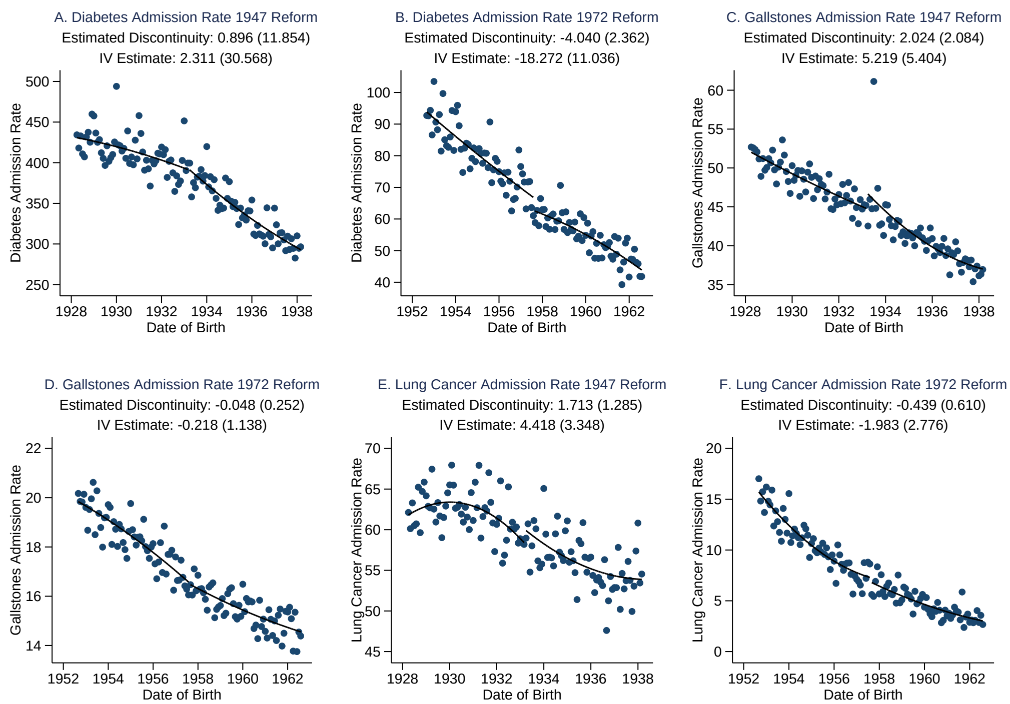
Notes: Sample comes from the Hospital Episode Statistics Inpatient Sample 1989-2010. The lines are fitted values from regressions that include a second-order polynomial in month-year of birth fully interacted with a dummy for being above the relevant cutoff. The points represent means of the dependent variable for each month-year of birth cell. Robust standard errors in parantheses.

were not for any of the three diagnosis categories. As expected, the effects are somewhat less precisely estimated than the effect on the overall admission rate, but are still small relative to the estimates obtained using survey data. For example, as shown in Panel E, the IV estimate for the 1947 reform for nutrition-related diseases is 1.411 on a base of 635.488, so we can rule out an effect larger than 9.5 percent.

Figure 4 presents the month-of-birth profile for three conditions which are known to be mostly preventable: diabetes, gallstones, and lung cancer. The panels show that even for three specific conditions with relatively low rates, there is no significant difference between those affected and not affected by the reforms. As expected, the variance of the profile is higher than for categories with a higher number of admissions, which leads to estimates with higher standard errors. However, visual inspection suggests that the variance of the month-of-birth cells is much lower even for these specific conditions than for any outcome using survey data. This is confirmed by the point estimates presented in Appendix Figure A1 and Figure 4. For example, using the HES data, we can rule out an effect of the 1947 reform on diabetes of greater than 14.73 percent, but using the survey, we can only rule out effects larger than 120.89 percent of the mean.

Using administrative hospital records, we have been able to examine the effect of the reforms on a multitude of specific outcomes that we believe may be more closely affected by education than the previous literature has been able to analyze. We have found that even for morbidities which are closely associated with education, there is no evidence of any effect of an additional year of high-school education on long-term health. Table 5 includes estimates of the reforms on all hospitalization outcomes presented in Figures 2-4, along with the means of each outcome.

Figure 2.4: Effect of Compulsory Schooling Reforms on Specific Morbidities



Notes: Sample comes from the Hospital Episode Statistics Inpatient Sample 1989-2010. The lines are fitted values from regressions that include a second-order polynomial in month-year of birth fully interacted with a dummy for being above the relevant cutoff. The points represent means of the dependent variable for each month-year of birth cell. Robust standard errors in parantheses.

2.6 Conclusion

In this study, we have re-examined the causal effect of education on health by examining the effect of two compulsory schooling reforms in England on the long-term health of the affected individuals. Clark and Royer (2013) and several other studies also examined these reforms and did not find substantial effects of the reforms on mortality, self-reported overall health, or using survey data, self-reported morbidity. However, previous studies have not been able to precisely address the extent to which the reforms affected morbidities that are known to be closely

Table 5: Effect of Compulsory Schooling Reforms on Inpatient Admissions

	Mean	RF	IV
	(1)	(2)	(3)
Panel A. 1947 Reform			
Any	3662.896	-14.325	-36.941
		(26.249)	67.805
Mortality	79.707	-0.555	-1.432
		(0.896)	(2.316)
Respiratory	522.162	0.461	1.189
		(7.434)	(19.171)
Circulatory	1379.863	-7.496	-19.330
		(13.836)	(35.741)
Nutrition	635.488	0.547	1.411
		(12.217)	(31.505)
Diabetes	391.136	0.896	2.311
		(11.854)	(30.568)
Gallstones	44.754	2.024	5.219
		(2.084)	(5.404)
Lung Cancer	58.265	1.713	4.418
		(1.285)	(3.348)
Panel B. 1972 Reform			
Any	1593.616	-3.539	-16.007
		(10.314)	(46.709)
Mortality	6.366	-0.083	-0.375
		(0.145)	(0.660)
Respiratory	127.808	-2.122	-9.596
		(2.461)	(11.223)
Circulatory	222.183	-4.953	-22.399
		(4.895)	(22.397)
Nutrition	132.598	-4.238	-19.166
		(2.572)	(11.990)
Diabetes	66.615	-4.040	-18.272
		(2.362)	(11.036)
Gallstones	16.538	-0.048	-0.218
		(0.252)	(1.138)
Lung Cancer	7.274	-0.439	-1.983
		(0.610)	(2.776)

Notes: Sample comes from the Hospital Episode Statistics Inpatient Sample 1989-2010. The column labeled "Mean" presents the value of the dependent variable at the threshold. The column labeled "RF" presents reduced-form estimates which include a quadratic function of month-year-of-birth fully interacted with the relevant reform dummy. All regressions are weighted by the total number of inpatient admissions in the month-year-of-birth cell. The column labeled "IV" presents the instrumental variables estimates, which are the ratio of the reduced-form estimate and the first-stage estimate. The standard errors for these estimates are calculated using the delta method. The bandwidth for all estimates is 60 months. Robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

associated with educational attainment.

This study fills this gap by using 20 years of administrative hospital records for all of England from the Hospital Episode Statistics (HES). By analyzing hundreds of millions of hospital admissions, this study has been able to precisely estimate the effect of an additional year of high school education on morbidities that we believe are closely associated with education such as diabetes and lung cancer. There is little reason to believe that these reforms are unique in their lack of effect on health. Rather, it is more plausible that reforms of this nature have only very small effects on health. If this is indeed the case, then this study complements studies such as Clark and Royer (2013) by showing that the health returns of education are at best, very small, even for morbidities that are thought to be closely associated with educational attainment.

Appendix A. Additional Figures and Tables

In this appendix, we present a number of additional figures and tables that are described in the main text.

Table A1: Sequential Adding of Covariates Into Education-Overall Health Gradient

	Bad Health or Worse						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. 1947 Reform							
Age left education = 15	-0.082***	-0.083***	-0.076***	-0.068***	-0.081***	-0.065***	-0.064***
	(0.015)	(0.015)	(0.018)	(0.018)	(0.022)	(0.023)	(0.023)
Male		0.026*	0.027*	0.026*	0.052***	0.044**	0.033*
		(0.015)	(0.015)	(0.015)	(0.017)	(0.018)	(0.019)
Age			0.002	0.003	0.004	0.006*	0.006*
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
White				-0.142***	-0.176***	-0.241***	-0.240***
				(0.043)	(0.057)	(0.073)	(0.072)
Income					-0.000	-0.000***	-0.000***
					(0.000)	(0.000)	(0.000)
Frequent Drinker						0.005	-0.002
						(0.024)	(0.024)
Ever Smoked							0.045**
							(0.019)
Panel B. 1972 Reform							
Age left education = 16	-0.088***	-0.088***	-0.087***	-0.088***	-0.059***	-0.051***	-0.048***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Male		0.005	0.005	0.004	0.020*	0.017	0.015
		(0.011)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)
Age			0.000	0.000	0.001	-0.001	-0.001
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
White				-0.048*	-0.024	0.005	0.003
				(0.025)	(0.030)	(0.029)	(0.029)
Income					-0.000***	-0.000***	-0.000***
					(0.000)	(0.000)	(0.000)
Frequent Drinker						0.013	0.006
						(0.015)	(0.015)
Ever Smoked							0.059***
							(0.011)

Notes: Sample comes from the Health Survey for England 2009-2013. Robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table A2: Alternative Self-Reported Health Definitions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. 1947 Reform				Age left education = 15			
Bad Health or Worse	-0.082***	-0.083***	-0.076***	-0.068***	-0.081***	-0.065***	-0.064***
	(0.015)	(0.015)	(0.018)	(0.018)	(0.022)	(0.023)	(0.023)
Fair Health or Worse	-0.135***	-0.134***	-0.105***	-0.096***	-0.116***	-0.122***	-0.119***
	(0.019)	(0.020)	(0.023)	(0.023)	(0.027)	(0.032)	(0.031)
Good Health or Worse	-0.035***	-0.034***	-0.027*	-0.024	-0.020	-0.020	-0.020
	(0.013)	(0.013)	(0.015)	(0.015)	(0.019)	(0.023)	(0.023)
Panel B. 1972 Reform				Age left education = 16			
Bad Health or Worse	-0.088***	-0.088***	-0.087***	-0.088***	-0.059***	-0.051***	-0.048***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Fair Health or Worse	-0.112***	-0.112***	-0.108***	-0.108***	-0.083***	-0.072***	-0.067***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.020)	(0.021)	(0.020)
Good Health or Worse	-0.074***	-0.074***	-0.076***	-0.076***	-0.058***	-0.058***	-0.053***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.017)	(0.019)	(0.018)
Covariates Included							
Male		X	X	X	X	X	X
Age			X	X	X	X	X
White				X	X	X	X
Income					X	X	X
Frequent Drinker						X	X
Ever Smoked							X

Notes: Sample includes individuals in the Health Survey for England: 2009-2014. Robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table A3: Bandwidth Sensitivity - Effects of Compulsory Schooling Reforms on Education

	Years of education	≤9 years	≤10 years	≤11 years	≤12 years
Panel A. 1947 Reform					
	Calonico, Cattaneo, and Titiunik (2014)				
Estimate	0.290**	-28.277***	2.567	1.922	1.095
	(0.120)	(5.467)	(6.289)	(7.264)	(2.936)
Outcome Mean	15.13	35.25	69.67	72.20	93.93
Bandwidth	49	43	38	26	35
	Imbens and Kalyanaraman (2011)				
Estimate	0.464***	-25.541***	-10.200**	2.645	0.643
	(0.103)	(5.721)	(4.916)	(5.384)	(2.724)
Outcome Mean	15.13	35.21	70.15	72.10	93.98
Bandwidth	70	40	61	44	41
	Ludwig and Miller (2007) - Cross Validation				
Estimate	0.482***	-28.100***	-12.875***	0.104	0.155
	(0.102)	(5.223)	(4.370)	(6.024)	(2.254)
Outcome Mean	15.13	35.19	70.26	71.95	94.05
Bandwidth	72	48	78	36	60
Panel B. 1972 Reform					
	Calonico, Cattaneo, and Titiunik (2014)				
Estimate	0.225	1.012	-25.246***	1.922	-4.511
	(0.152)	(2.334)	(5.502)	(7.264)	(4.916)
Outcome Mean	16.18	2.505	23.20	72.20	84
Bandwidth	25	26	28	26	34
	Imbens and Kalyanaraman (2011)				
Estimate	0.278**	0.548	-24.674***	2.645	-4.364
	(0.117)	(1.529)	(3.803)	(5.384)	(4.707)
Outcome Mean	16.18	2.377	23.54	72.10	83.93
Bandwidth	41	49	57	44	37
	Ludwig and Miller (2007) - Cross Validation				
Estimate	0.289**	-0.128	12.922	0.104	-5.457
	(0.126)	(0.896)	(14.426)	(6.024)	(4.058)
Outcome Mean	16.18	2.467	22.40	71.95	84.06
Bandwidth	36	120	6	36	48

Notes: Sample includes individuals in the Health Survey for England: 2009-2014. All estimates based on regressions that include a quadratic function of month-year of birth centered at the relevant cutoff, a quadratic interaction of month-year-of-birth and the relevant reform dummy, and dummy variables for year of survey and month of birth. Bandwidths are in months. "Outcome Mean" represents the mean of the outcome at the threshold. Robust standard errors in parantheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table A4: Bandwidth Sensitivity - Effect of Reforms on Self-Reported Health

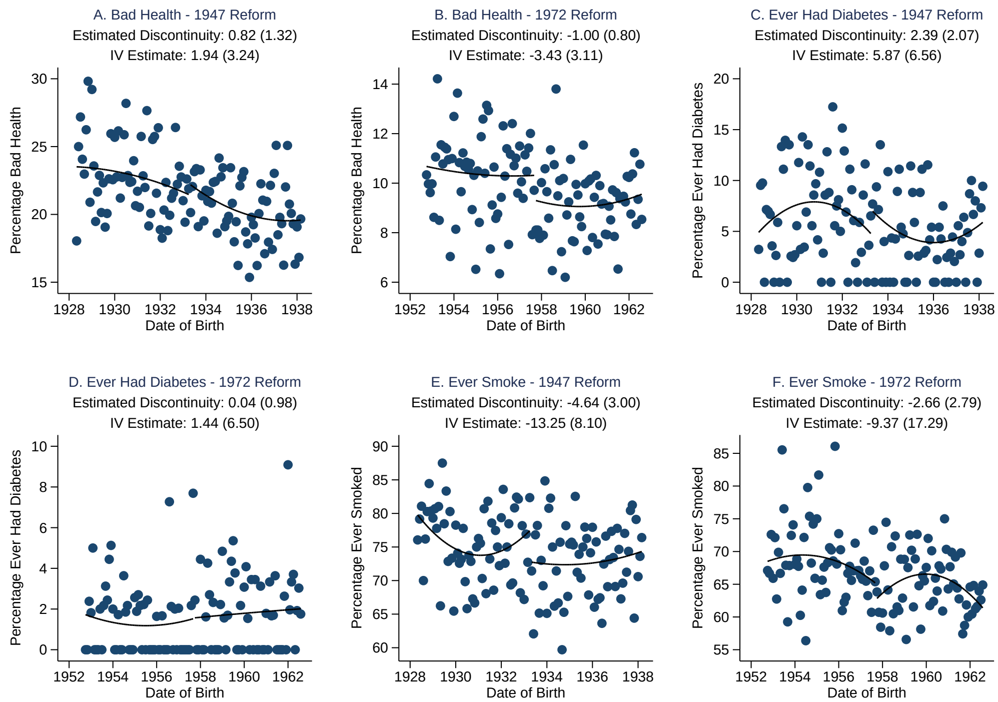
	Mean	RF	IV	Bandwidth
Panel A. 1947 Reform				
Health Fair or Bad				
CCT	0.54	0.0021	0.0054	42
		(0.0031)	(0.0081)	
IK	0.54	0.0001	0.0003	132
		(0.0018)	(0.0047)	
CV	0.54	0.0017	0.0043	33
		(0.0036)	(0.0092)	
Limited Activity				
CCT	0.57	0.0056*	0.0144	33
		(0.0031)	(0.0082)	
IK	0.57	0.0031	0.00810	125
		(0.0020)	(0.0052)	
CV	0.57	0.0057*	0.0147	33
		(0.0031)	(0.0080)	
Panel B. 1972 Reform				
Health Fair or Bad				
CCT	0.23	-0.0008	-0.00370	48
		(0.0013)	(0.0059)	
IK	0.23	-0.0034***	-0.0155	166
		(0.0008)	(0.0043)	
CV	0.23	-0.0010	-0.00440	48
		(0.0013)	(0.0059)	
Limited Activity				
CCT	0.19	0.0010	0.0045	44
		(0.0013)	(0.0061)	
IK	0.19	-0.0013*	-0.0060	145
		(0.0008)	(0.0036)	
CV	0.19	-0.0002	-0.0009	48
		(0.0012)	(0.0055)	

Notes: These estimates come from 2011 census data aggregated to the month-year-of-birth level. The column labeled “Mean” presents the value of the dependent variable at the threshold for the specified bandwidth. The column labeled “RF” presents reduced-form estimates which include a quadratic function of month-year of birth fully interacted with the relevant reform dummy, and an indicator for the month of birth. The column labeled “IV” presents the instrumental variables estimates, which are the ratio of the reduced-form estimate and the first-stage estimate. The standard errors for these estimates are calculated using the delta method. Tbe column labeled “Bandwidth” presents the bandwidth of the estimates in months, calculated using the optimal bandwidth selection process in the first column. CCT represents the optimal bandwidth using the optimal bandwidth procedure proposed by Calonico et al. (2014), IK uses the procedure proposed by Imbens and Kalyanaraman (2011), and CV uses the cross-validation procedure proposed by Ludwig and Miller (2007). Robust standard errors in parantheses.

Table A5: ICD Codes of Analyzed Outcomes from Hospital Episode Statistics

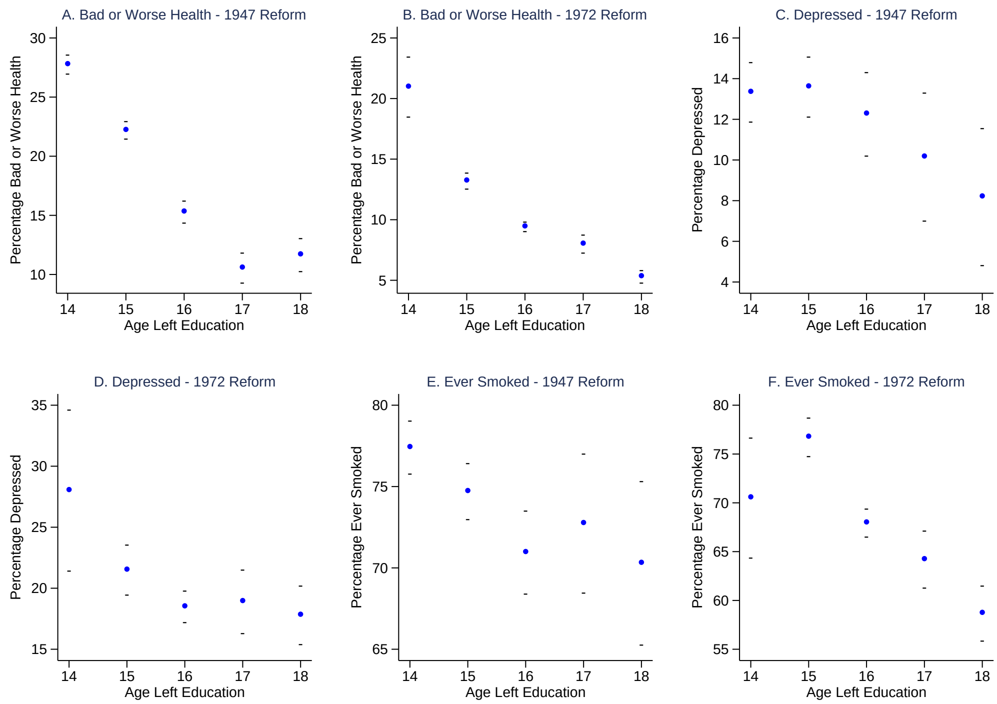
Diagnosis Category	ICD 9 Codes
Any	Any Code
Respiratory	460-510
Circulatory	390-450
Nutrition	240-270
Diabetes	250
Gallstones	574
Lung Cancer	162

Figure A1: Effect of Reforms on Health - Survey Data



Notes: Sample comes from the Health Survey for England 1991-2014 and the General Household Survey 1986-1996. The lines are fitted values from regressions that include a second-order polynomial in month-year of birth fully interacted with a dummy for being above the relevant cutoff. The points represent means of the dependent variable for each month-year of birth cell. Robust standard errors in parantheses. "IV Estimate" displays the treatment effect from equation (2) using two-stage least squares.

Figure A2: Education-Health Gradient



Notes: Sample comes from the Health Survey for England 2009-2013 and the General Household Survey 1986-1996. Each point represents the mean of the variable on the y-axis at the given education level. Bars represent 95 percent confidence bounds.

B. Description of Hospital Episode Statistics Data Cleaning Methods

The Hospital Episode Statistics data sets contain frequent and systematic measurement error where patients are not assigned the correct month or day of birth, admission, appointment, or arrival. In some cases, health care providers entered just one month or day or for all, or a very large proportion of, dates. For many other providers, the proportion of a given month is only marginally larger than the proportion of other months, but is consistently high across multiple years. Often, this occurs for very large providers, particularly for date of birth, and in almost all problematic cases, the proportion of dates of birth being assigned to January are due to measurement error and are too large relative to the true distribution.

In order to address these issues, we implemented several multi-stage procedures using the provider codes available in the data sets. The first stage of the first procedure was to calculate the proportion of each date of birth, appointment date, admission date, or arrival date corresponding to each month, along with the mean of the proportions of each month across all providers. Then, for each provider in each file year, we calculated a t-statistic where the null hypothesis is that the proportion for any given month for a given provider and file year is equal to the mean of proportions for that month. Using this t-statistic, we calculated a p-value indicating the likelihood that the event would occur in any given month. By raising this p-value to a power equal to the number of providers in a given file year, multiplied by the total number of months across all file years, we obtain the probability that the event occurred over and over again for a certain provider.

Subsequently, we remove all provider year pairs with a p-value of greater than 0.01, so that the next stage of our assessment focuses only on provider year pairs which have a proportion of at least one month with a very low probability of occurring by random chance. Next, we create a figure for each provider year pair with the proportion of each month in order to manually look for patterns within providers and across file years. By manual inspection of the figures, we create a list of particularly problematic provider year pairs, which constitute the dropped observations of our data sets. This procedure alone was used to analyze dates of birth, but given the particularly problematic nature of admission, appointment, and arrival dates, and that these variables contain the day, month, and year of admission, appointment, or arrival, we implement an additional procedure for more accurate analysis.

Our second procedure involves constructing a separate data set with the total number of admissions, appointments, or arrivals for each day of the sample and for each provider. We then construct an additional set of figures with the total count of admissions, appointments, or arrivals on each day. These figures allow us to look at the full distribution of dates across the sample and assess whether the distribution evolves smoothly over the sample period. Specifically, we focused on providers in which one day or only a few days had a much larger number of hospital visits than any other day, so that it was clear that all patient admissions were being assigned only to those days, regardless of the true day and month in which they visited the hospital. By utilizing both sets of figures, we can accurately assess whether the distribution of a certain provider is due in large part to measurement error or accurately reflects their true distribution.

The following table describes the proportion of each data set that is flagged and subsequently dropped in our primary data sets.

Table B1: Data Dropped in Cleaning Process

	Inpatient			Outpatient		
File Year	DOB	Admi	Total	DOB	Appo	Total
1989	3.80	0	3.80			
1990	3.56	2.91	6.46			
1991	3.52	0.41	3.94			
1992	4.50	1.10	5.58			
1993	3.68	1.89	5.57			
1994	2.97	1.13	4.11			
1995	3.03	0.03	3.06			
1996	3.82	0.78	4.60			
1997	5.30	3.92	9.22			
1998	7.51	1.19	8.70			
1999	8.82	2.72	11.55			
2000	9.13	0.51	9.65			
2001	9.20	1.42	10.66			
2002	8.83	2.02	10.85			
2003	8.55	0.07	8.61	2.81	1.82	4.63
2004	8.21	0.09	8.34	2.51	0.19	2.70
2005	9.10	0.05	9.15	2.41	2.25	4.67
2006	10.23	0.23	10.46	2.89	0.36	3.24
2007	10.35	1.36	11.71	3.63	1.72	5.40
2008	9.59	0.03	9.62	3.13	0.25	3.38
2009	9.11	0.23	9.35	3.31	0.06	3.36
2010	9.74	0.00	9.74	3.31	0.17	3.48

Notes: Each cell presents the percentage of data dropped from each file-year and type. “DOB” refers to hospitals that were dropped because they had an unusually high proportion for a particular day of birth. “Admi” refers to hospitals that were dropped because they had an unusually high proportion for a particular day of admission.

Chapter 3

3 The Effect of Community Water Fluoridation on Health and Education

3.1 Introduction

Since the first artificial fluoridation of water in the United States in 1945 in Grand Rapids, Michigan, there has been and continues to be a fervorous debate over the effects of water fluoridation on oral health as well as more general outcomes such as bone disease, cancer, and intelligence. Although there exists many products and dental treatments which contain or are specifically designed to provide fluoride to their consumers, the fluoridation of public water systems is by far the most prevalent and controversial means of fluoride intake. Several studies (Burt 1989; Allen et al. 1992) have estimated the mean cost of water fluoridation in the United States to be roughly $0.96 per capita per year in 2013 dollars. With an estimated 226 million individuals drinking fluoridated water in the United States as of 2012, this is an approximate cost of $217 million per year to artificially fluoridate public water systems. Furthermore, the cost of a filling to repair a cavity in 2013 generally ranges from $75 to $200. Given estimates of the effect of water fluoridation on the prevention of cavities, ignoring the cost of the known adverse outcome fluorosis, Burt (1989) and Griffin et al. (2001) estimate significant cost savings of water fluoridation of upwards of 50 percent. Despite the results of these studies of extreme cost-savings, there remains stark controversy over the magnitude of fluoride's effects on oral health, as well as the existence of effects on general health and mental development.

Since the inception of water fluoridation, there has been a plethora of research into its effects on oral health, and more recently, general adverse health outcomes. McDonagh et al. (2000) and Higgins et al. (2009a, 2009b, 2009c, 2009d, 2009e) perform meta-analyses of over 100 non-experimental studies of water fluoridation as well as clinical studies of other fluoride supplements, and find mean reductions in cavities of roughly 25 percent. The primary data source of our analysis and the source of several other studies due to being the only data source in the United States with detailed information on residential history, fluoride exposure, cavities, and fluorosis, is the 1986-1987 National Survey of Oral Health of US Schoolchildren (NSOHS). Performing a comparison of means using differing methodology with the NSOHS data set, Yiamouyiannis (1990) finds no significant difference in oral health between fluoridated and unfluoridated individuals, while Brunelle and Carlos (1990) find a significant and large effect of 18-25% depending on sample restrictions. A significant improvement in quality came from Burt et al. (1997) which utilize linear regression analysis to control for individual demographic characteristics. They determine that water fluoridation is correlated with significant reductions in both deciduous and permanent cavities, but also with significant increases in the adverse outcome fluorosis.

Additionally, Liu et al. (2000) along with many other Chinese studies using comparisons of near by villages with differing fluoride levels, and a meta-analysis of these studies by Choi et al. (2012), find that villages with larger amounts of fluoride in the water are associated with lower intelligence. Moreover, Bassin et al. (2001) find a significant increase in risk of osteosarcoma cancer in young boys drinking fluoridated water, but no such increase in females. However, Cicuttini et al. (2001) perform a meta-analysis of 33 studies on bone density and strength and conclude that these adverse effects may occur only at concentrations roughly four times greater than that recommended in the United States. Overall, the majority of studies have found that higher levels of fluoride in public water supplies are associated with significantly reduced cavities and increased levels of mild fluorosis, but not with significant effects on general health.

This paper contributes to the literature in a number of important ways. Using a variety of data sets, I address many branches of the literature, specifically the effect of water fluoridation on cavities and fluorosis, as well as the effect on several adverse health outcomes. Additionally, I introduce a richer set of controls than previous studies via additional data on regional characteristics, specifically median county income. Moreover, I employ more advanced econometric techniques than previous studies by relaxing the assumptions of the linear regression model and flexibly estimating the relationship between oral health and water fluoridation by utilizing non-parametric analysis and linear probability models. Finally, this paper is the first to place water fluoridation in a panel context and assess the impact of water fluoridation on educational achievement across more than 1000 counties across the United States, dating back to the 1950's when artificial water fluoridation first began.

My primary data source is the 1986-1987 National Survey of Oral Health of US Schoolchildren (NSOHS), which I supplement with median county income from the US Census Bureau. In order to assess whether political leaning has any bearing on whether a community is more or less likely to adopt water fluoridation, I obtain information on the proportion of individuals that voted for a particular party of the presidential candidate for each county in the sample for the years 1968-1984. Additionally, I combine the Fluoridation Census 1992 with the National Health and Nutrition Examination Survey III (NHANES III) in order to perform a robustness check on the regression results from the NSOHS data as well as examine the effect of water fluoridation on several general health outcomes. Finally, I combine information on fluoridation status and the date when water fluoridation first began from the Center for Disease Control's (CDC) "My Water's Fluoride" with educational outcomes from seven waves of the decennial census. This allows me to estimate a fixed-effects model over a 60 year period of the effect of water fluoridation on high school and college graduation rates.

In addition to using a fixed-effects model to estimate educational outcomes, I also employ econometric techniques that have not been utilized in this literature. Specifically, since the NSOHS contains information on the parts per million (ppm) level of fluoride for each child, I am able to perform non-parametric analysis in order to assess the relationship between fluoride and oral health at all levels of fluoride without the functional form assumptions of the linear regression model. Additionally, I construct plots of a series of linear probability models, which provide not only the cumulative density function of the oral health outcome, but the specific relationship between water fluoridation and oral health outcomes at all levels of oral health.

I find that children that drink optimally fluoridated water have on average 0.9 less of a carious surface than children that do not drink optimally fluoridated water, which is consistently significant across multiple data sets and specifications. However, although the point estimate of water fluoridation's effect on permanent surfaces is significant in the NSOHS, showing a significant effect of 0.638 of a surface, this effect is not robust across data sets and moreover, non-parametric analysis reveals that the relationship is flat across all ppm levels. Furthermore, children drinking fluoridated water have on average 1.9 additional surfaces of very mild or worse fluorosis and about 0.4 additional surfaces of mild fluorosis compared to children not drinking fluoridated water, while severe fluorosis occurs only in those individuals drinking naturally fluoridated water far exceeding the recommended optimum. Finally, I find no evidence of an effect of water fluoridation on general health, trouble working for children or adults, retardation in children, or educational achievement in the form of high school and graduation rates over the entire period of fluoridation in the United States.

The paper proceeds as follows: Section II provides a detailed literature review, Section III discusses information on the background of fluoridation in the United States and some brief technical details, Section IV describes the data sets used in the analysis, Section V outlines the econometric framework, Section VI presents the results of the effect of water fluoridation on oral health, adverse health outcomes, and educational achievement, and Section VII concludes.

3.2 Literature Review

To assess the impacts of various fluoride applicators, an extensive battery of clinical studies have been carried out. Higgins et al. (2009a) perform a meta-analysis of 144 clinical studies looking at all types of topical fluoride and find a pooled effect of using fluoridated materials of 26 percent, although there was significant heterogeneity in the estimates based on the type of control group used. Higgins et al. (2009b, 2009c, 2009d, 2009e) perform meta-analyses of clinical trials for fluoride toothpastes, gels, varnishes, and mouthrinses, and find pooled mean effects on decayed, missing, or filled surfaces (dmfs) of 24 percent, 28 percent, 46 percent, and 26 percent, respectively. The authors note that in all meta-analyses, there was significant heterogeneity in methodology and results across studies. The authors emphasize that most studies were of low quality, however, the mean effects are quite consistent across methods of topical fluoride application as well as close to many estimates of the effects of water fluoridation.

In 1997, the British Government commissioned a team of researchers at the NHS Centre for Reviews and Dissemination at the University of York to conduct a systematic review of water fluoridation. McDonagh et al. (2000), named the York Review, analyzed 214 studies of various design and quality. Using meta-analysis techniques, the authors developed four main conclusions. First, the median effect of fluoridation across studies showed a mean decrease in decayed, missing, or filled teeth (dmft) of 2.25. Second, the incidence of fluorosis of aesthetic concern increases from 6.3 percent at a water fluoride level of 0.1 parts per million (ppm) to 12.5 percent at 1.0 ppm. Third, there is no clear association between water fluoridation and bone developments, bone fractures, or cancer. Fourth, all studies were considered to be of medium or low quality, and the authors stressed that additional work with appropriate statistical analysis and proper attention to confounding variables should be undertaken to obtain more precise impacts of water fluoridation.

Using the 1986-1987 National Survey of Oral Health of US Schoolchildren (NSOHS), Yiamouyiannis (1990) split up the surveyed children into 84 geographical areas and classified each as being artificially fluoridated for more than 17 years, less than 17 years but have been artificially fluoridated, or have never had artificial fluoridation. He finds that there is no statistically significant difference between the mean dmft or dft of children in each area. In contrast, Brunelle and Carlos (1990) used residence histories to separate children into two groups, one which had continuous fluoride exposure over their whole life and one with no fluoride exposure. A comparison of means of the two groups shows a mean 18 percent difference in DMFS across all ages. Both papers use simple comparison of means without controlling for confounding variables, and furthermore, Brunelle and Carlos (1990) do not exploit the variation across children with partial lifetime exposure to water fluoridation.

A significant improvement in quality using the NSOHS data set came in the form of Burt, Eklund, and Heller (1997), who use linear regression analysis to determine the effect of water fluoridation on children for both deciduous (baby) surfaces and permanent surfaces. They perform regressions of the number of dmfs or dfs for each child on the ppm level of fluoride in the water at a child's school as well as demographic covariates and survey questions about use of other fluoride materials. They find that there is a significant negative correlation between the ppm level of fluoride in the water collected from school samples, and dmfs or dfs. Specifically, they find a coefficient of -1.08 for dfs and -0.59 for dmfs, which I can interpret as saying that a 1 ppm increase in fluoride in the water decreases decay on average for deciduous teeth by 1.08 surfaces and for permanent teeth by 0.59 surfaces. Additionally, the authors find that fluorosis increases monotonically with the ppm level of fluoride in water, which for mild fluorosis, increases from a prevalance of 13.5 percent at less than 0.3 ppm to 41.4 percent at greater than

1.2 ppm. The NSOHS section of the present paper follows most closely to Burt, Eklund, and Heller (1997) and can in some ways be thought of as an extension and additional analysis of their results.

In order to evaluate the cost effectiveness of water fluoridation, Griffin et al. (2001) used estimates of the effects of water fluoridation on oral health from several published studies as well as the National Survey of Oral Health of US Schoolchildren, National Survey of Oral Health of U.S. Employed Adults and Seniors, and National Health and Nutrition Examination Survey I (NHANES I) and National Health and Nutrition Examination Survey III (NHANES III) to impute baseline estimates of mean dmfs and dfs, as well as the effectiveness of water fluoridation in reducing decay. The authors then computed the lifetime cost of decayed surfaces at various discount rates and compared them to the per capita costs of water fluoridation, where the latter is computed from a study of fixed costs and operating costs of water fluoridation for 44 Florida communities. Using their estimates for the effectiveness of water fluoridation, the authors compared the lifetime costs of decay with the costs of fluoridation and found that even in the worst case, water fluoridation was still significantly cost saving compared to the costs of repairing decay. It is important to note that this study did not address fluorosis and the costs that it may impose.

Glied and Neidel (2008) take the notion that water fluoridation significantly benefits oral health as given, and use access to fluoridation in childhood as a proxy for oral health status, which they then use to test the effect of oral health status on labor market outcomes. The authors used the National Longitudinal Study of Youth, which followed 12000 individuals for several decades to obtain county of birth and demographic characteristics. They then combined this with the 1992 Fluoridation Census, which provided estimates of the population that has optimal fluoridation for each county, and this ratio was used as the fluoridation status. This is the same method that I pursue for the NHANES III data set, which will be discussed in more detail later. They find that women who resided in communities with fluoridated water as children earn roughly 4 percent more as adults, but there is no effect on men. Additionally, they provide evidence that at least until the amendment of the Safe Drinking Water Act that required the disclosure of fluoride content in drinking water, most decisions for fluoridation were made by various government administrators, and more often than not was not a majority public decision. This is especially true since water fluoridation policies are determined at the water district level, which often do not correspond to municipal boundaries that govern other decisions, so the administrators that make these decisions are often not held accountable. Additionally, they show that out of 27 demographic characteristics, only urban status is significantly related to fluoride exposure.

Aside from analyzing the effects of fluoride on oral health, several studies have assessed the adverse impacts of fluoridation on outcomes other than fluorosis.

Bassin et al. (2006) matched individuals from the same hospital with and without osteosarcoma cancer from 11 different teaching hospitals across the US, and conducted telephone interviews about demographic characteristics as well as use of fluoride supplements. The authors find that for males, especially between ages 4 and 12, there is a significant increase of risk of osteosarcoma cancer, although the standard errors are quite large. However, they find no significant relationship for females or males in other age ranges. The authors note that there are several other studies looking at the same hospitals which did not find significant effects, but they note that it may be because they did not look at age-specific effects.

Liu et al. (2000) compared the IQ scores of 60 children in China from a village with a high level of water fluoridation to 59 children from a neighboring village with a low level of water fluoridation. The authors argue that the characteristics of the children are very similar across the villages and find that the mean IQ of the high fluoridation children was 92.27 ± 20.45, whereas the mean for the low fluoridation children was 103.05 ± 13.86, so they conclude that fluoridation has a significant detoriating effect on cognitive functions, although statistically it does not appear to be the case. Liu et al. (2000) is one of the 33 studies performed mostly in China, with two in Iran, that look at the deleterious effects of fluoride on intelligence. Choi et al. (2012) perform a meta-analysis of these studies and find that fluoridated communities have on average a significantly smaller IQ of -0.45. It is important to note however that there is significant heterogeneity across studies, many of the studies had significantly higher fluoride levels than recommended optimal levels, and all of the studies are of low quality.

Cicuttini et al. (2001) reviewed 33 studies on the effects of water fluoridation on bone density and strength, and found that adverse effects of fluoride on bone density and strength occurred in both animals and humans only at concentrations four times greater than the optimal recommended level for community water fluoridation, or not at all. Across the literature, there is not strong statistical support for any adverse effects of fluoride other than fluorosis, but both the quantity and quality of work in this area is quite low so more robust statistical analyses need to be made to make any strong conclusions.

3.3 Background

In 1901, a young dental school graduate named Frederick McKay left for Colorado Springs to open a dental practice, and when he arrived, he realized that the residents had severe brown stains on their teeth. After several years of studying this phenomenon with Dr. G.V. Black, they came to realize that the teeth of the Colorado Springs residents, although heavily stained, were unusually resistant to decay. They also came to realize that all residents drank from the same local water supply, which they suspected may have an ingredient which is causing this phenomenon. As a result, in 1923, McKay visited his parents in Idaho in a town with similar stains, and he advised them to change the source of their water supply.

Amazingly, within a few years, the younger children of the town were sprouting healthy, unmottled teeth. After this result, along with several other researchers, McKay tested other water sources where this phenomenon was occurring and eventually discovered that very high levels of fluoride was in fact the culprit. The mottling of teeth due to high levels of fluoride was thus called fluorosis.

The same year, Dr. H. Trendley Dean, the first director of the U.S. National Institute of Dental Research (NIDR) began investigating how high fluoride levels in water could go before fluorosis occurred. After several years, Dean's colleague Dr. Elias Elvove, a senior chemist, developed a more accurate method for measuring fluoride at the parts per million (ppm) level. With this new method, Dean set out across the country with his team measuring fluoride levels. By the late 1930's, Dean and his team had discovered that fluoride levels of less than 1 ppm did not cause enamel fluorosis in most people and caused only mild enamel fluorosis in a small percentage of people. For years after, Dean hypothesized that adding fluoride to water within a certain range may fight tooth decay while limiting fluorosis. In 1945, Grand Rapids, Michigan became the first city in the world to fluoridate its water supply. After 11 years, the cavities rate among schoolchildren after the fluoride was added dropped by more than 60 percent compared to children before the fluoride was added.

Due to the success of this experiment, water fluoridation in the United States has been spreading throughout the country ever since. The recommended rate of fluoridation decided on by Dean and his team varies by temperature since individuals living in hotter climates tend to drink more water and thus require less fluoride in their water for the same effect. The Center for Disease Control (CDC) has recommended 0.7 - 1.2 ppm of fluoride in the water supply as the optimal range throughout the entire period I am examing, so for the rest of this paper I refer to this as the optimal range. On January 7, 2011, the U.S. Department of Health and Human Services proposed a new level of 0.7 ppm for all climates. The main reason cited for the change is that since nearly all households now have air conditioning, the discrepancy between the amount of water drank in cold and warm climates is drastically reduced compared to earlier days before widely available air conditioning.

Modern research on the effects of fluoride suggest that the benefit is predominantly topical, and is attributable to the inhibition of demineralization, enhancement of remineralization, and inhibition of bacterial activity. Specifically, the mineral content of the hard tissues of the tooth is primarily hydroxyapatite, and during development, a small fraction of weakly bound carbonate impurity ions are incorporated into the hydroxyapatite matrix. These impure ions are then vulnerable to even small amounts of acid which creates weak spots allowing for cavities to be created. However, when applied to the tooth surface, fluoride ions can chemically react with the hydroxyapatite to form fluorapatite, a mineral that is much less soluble in water and more resistant to acidity. A much more thorough and technical discussion is found in Featherstone (1999) and other dental literature.

3.4 Data

The primary data source used in my analysis is the 1986-1987 National Survey of Oral Health of US Schoolchildren (NSOHS), performed by the National Institute of Dental Research. This survey is the largest and most extensive dental survey for children that has been performed in the United States. The survey performed examinations and questionnaires for 40,693 children, ages 4-22, in over 200 counties throughout the United States. Questionnaires filled out by the parents or guardians of the children answered questions about demographic characteristics and whether their children have used fluoride drops, tablets, fluoride treatments at school, and professional fluoride treatments. Additionally, a 500 mL water sample was obtained from each sampled school providing the ppm level of fluoride in the water in each child's school, as well as a binary indicator for the fluoridation status of each child's residence.[59] The examination portion of the survey assessed the level of cavities as well as the severity of fluorosis for each child. In my analysis, I choose to look at only children which have lived at a single residence throughout their entire lives since the school sample from their current school is

[59]The school sample measure serves as the primary variable of interest in my non-parametric analyses and linear probability models, since the measure of home fluoridation is a binary measure indicating whether the residence receives its water from an optimally fluoridated water system or not.

used as the fluoridation status of the individual throughout their entire lives, thus I suspect that the observed effect of water fluoridation would be attenuated for those children which have resided at more than a single residence.[60]

The NSOHS does not contain estimates for the income level of individuals, so in order to obtain an estimate of the income level, I obtained the median income level of the first county in which a child lived, from the Small Area Income and Poverty Estimates 1989. Since I look at only single-resident children in my primary analysis, the median income level of the first county in which a child lived is the median income level of the county in which the child has lived for their entire life.

Additionally, I suspect that the political leaning of individuals may systematically affect the adoption of water fluoridation, so in order to assess whether water fluoridation is balanced on the political leaning of communities, I obtain data on presidential voting records by county for five elections, which covers the period 1968-1984, from McGillivray and Scammon (1988). This data allows me to compute for each county in the sample, the percentage vote for Democrat, Republicans, and other. Since I take the average of the five presidential elections prior to the survey, I suspect that this gives an adequate representation of the political leaning of the individuals of a particular county.

I also provide results from the National Health and Nutrition Examination Survey III (NHANES III), which is the seventh survey from the National Center for

[60]My results are robust however to the attenuation caused by multi-resident individuals, and all results for the full NSOHS data set are available upon request.

Health Statistics (NCHS) and Center for Disease Control and Prevention (CDC). The NHANES III was designed to "provide national estimates of the health and nutritional status of the United States' civilian, noninstitutionalized population aged two months and older." I choose to use the NHANES III as opposed to newer waves of the NHANES series because the NHANES III is the last of the series with detailed information on cavities for each individual.

The NHANES III does not contain any information on fluoride exposure for individuals, so in order to obtain a measure of fluoride exposure, I use the 1992 Fluoridation Census to obtain the approximate number of individuals that are optimally fluoridated in each county by summing the number of individuals served by each optimally fluoridated water system primarily serving a given county. I combined this with the total population for each county in the sample from the 1990 Census from the US Bureau of the Census. So, my measure of fluoride exposure for each county is the proportion of optimally fluoridated individuals given by the Fluoridation Census 1992 relative to the total population of the county given by the 1990 Census. Since the most finely detailed information on the residence of each individual is at the county level and is the current county in which they live at the time of the interview and examination, each individual in any given county is assigned the same fluoridation exposure as every other individual in that county. Also, my sample includes only the 35 counties in the survey with populations of over 500,000 due to confidentiality concerns.

Additionally, I examine the relationship between the fluoridation of water systems and educational achievement in 1118 counties in 18 states throughout the United States. To accomplish this, I combine data from two sources: the CDC's "My Water's Fluoride" and the National Historical Geographic Information System (NHGIS). The CDC's "My Water's Fluoride" is a website maintained by the CDC that contains information on the fluoridation status of water systems as well as the primary county that the water system serves, the date the water system started fluoridating, and how many people the water system serves. The NHGIS is a database which compiles historical census information and makes it available for download. Using the NHGIS, I was able to obtain the percentage of white individuals relative to the total population, the percentage of individuals living in urban areas, family income, the total population, and educational attainment in the form of high school and college completion rates for each county. Data for the aforementioned variables is available for the censuses of 1950, 1970, 1980, 1990, 2000, and 2010. The 1960 census does not contain educational achievement data and artificial fluoridation started in the 1950's, so this is the furthest back that it is necessary to have outcome data for.

3.5 Econometric Framework

Ideally, fluoride exposure would be randomly assigned to communities and a simple comparison of means between fluoridated and unfluoridated communities would allow us to estimate the causal effect of fluoridation while controlling for unobservables. However, fluoride exposure is not randomly assigned and is in fact adopted through a mish-mash of state mandated laws and community-specific political processes. Although the data does not allow me to explicitly control for unobservables, I control for observable characteristics using a variety of methods and I explicitly check the balance of covariates between fluoridated and unfluoridated communities.. Additionally,

Using the NSOHS and NHANES III cross-sectional data sets, I estimate the linear regression model:

$$Y_i = \beta_F F_i + \beta_X X_i + \varepsilon_i \tag{4}$$

The dependent variable Y_i is a health outcome. Subscript i indicates individual i. The coefficient of interest β_F represents the effect of optimal water fluoridation at individual i's residence on the health outcome Y_i, and X_i is a vector of observed covariates.

Since the NSOHS provides the parts per million level of fluoride from each child's school, I estimate a semi-parametric dose-response function, which provides me with point estimates of the effect of water fluoridation on health outcomes at various intervals of fluoride exposure. The specification is as follows:

$$Y_i = \beta_{F1} F1_i + \beta_{F2} F2_i + \beta_{F3} F3_i + \beta_{F4} F4_i + \beta_X X_i + \varepsilon_i \tag{5}$$

The dependent variable Y_i is either the count of surfaces with deciduous or permanent surfaces, or the count of surfaces with very mild or worse, mild or worse, or severe fluorosis. Subscript i indicates individual i. $F1_i$ takes on a value of one if the water at a child's school contains between 0.3-0.7 ppm of fluoride, $F2_i$ indicates the optimal range of 0.7-1.2 ppm, $F3_i$ indicates the range of 1.2-3 ppm, $F4_i$ indicates the range of greater than three ppm, and X_i is a vector of observed covariates.

By combining information on the start date of fluoridation for a water system with outcomes from the decennial census, I am able to construct a panel data set and estimate the following fixed-effects model:

$$E_{it} = \beta_F F_{it} + \beta_X X_{it} + \eta_i + \phi_t + \varepsilon_{it} \tag{6}$$

The dependent variable E_{it} is the high school or college graduation rate. The subscripts indicate the county i and year t. The explanatory variable of interest, F_{it}, takes on a value of one if at least 50 percent of the population of county i is served by optimally fluoridated water systems in year t. X_{it} is a vector of covariates, η_i are fixed effects for county i and ϕ_t are fixed effects for year t.

Additionally, using the NSOHS data set, I relax the assumptions of the linear regression model and perform non-parametric analysis via second degree local polynomials to investigate the relationship at all levels of fluoride between cavities, both deciduous and permanent, and the ppm level of fluoride at each child's school, as well as the relationship between fluorosis and the ppm level of fluoride at each child's school.[61] Instead of estimating the relationship by fitting a line to the entire sample, non-parametric analysis allows the data to determine the functional form of the model. Specifically, the procedure can be viewed as performing a series of weighted least-squares regressions and providing local least-squares estimates, which more accurately allows me to analyze the relationship between the level of fluoride in a child's water at school and oral health outcomes.

Also using the NSOHS data set, in order to explore the effect that fluoride has on oral health at various levels of cavities, I construct linear probability models for each dependent variable. Specifically, an indicator for whether the dependent variable is less than a certain count is regressed on the school fluoride level, and the coefficient is plotted, along with the constant from the same regression.[62] For example, at an indicator of zero where the dependent variable is deciduous cavities, the coefficient at an indicator of 0 indicates the proportion of individuals that "moved" from having one or more deciduous carries to having zero deciduous carries due to a one unit increase in ppm of fluoride in their drinking water.

The panel data set used to estimate equation (3) was created by using the fluoridation start date from the CDC's "My Water's Fluoride" and supposing that a

[61] Each figure presents the scatter plot and second degree local polynomial fit of the dependent variable on the level of fluoride from the school sample, as well as the residuals from a regression of the given dependent variable on all observed covariates, with the mean adjusted to be equal to the mean of the dependent variable. Each data point represents the mean of 200 sequential data points sorted by school fluoride level.

[62] The constant of the linear probability model traces out the CDF of the dependent variable.

water system is fluoridated in a certain year if the water system started artificially fluoridating prior to that date. Moreover, if the water system has natural fluoridation below the optimal level, they are presumed to be not fluoridated for any year, and if the water system has natural fluoridation above the optimal level, they are presumed to be fluoridated for every year. Then, the panel of water systems is aggregated to the county level by supposing that if at least 50 percent of the population of the county is served by optimally fluoridated water systems, then that county is considered to be fluoridated for that year.

3.6 Results

3.6.1 Checking Balance of Covariates of Public Water System Fluoridation

Prior to analyzing the effects of water fluoridation, it is crucial to determine whether the adoption of community water fluoridation is balanced on observable covariates. As discussed by Glied and Neidel (2008), urban communities are much more likely to adopt artificial water fluoridation, primarily due to the large returns to scale of urban areas compared to rural due to much higher population density. As such, I perform balance checks on observable covariates separately for urban and rural counties, so as to compare only within urban and rural areas rather than across them. Tables 1 and 2 show that the only covariate with a significantly different mean across counties with unfluoridated and fluoridated

Table 3.1: Balance Check - Rural

	Overall Mean	Fluoridated	Observations
Median Income	26650	304.68	136
	(6937)	(1,071.76)	
Poverty	18.24	-0.36	137
	(10.09)	(1.59)	
White	0.719	-0.07	142
	(0.357)	(0.07)	
Democrat Vote	38.69	2.22	138
Proportion	(10.76)	(1.62)	
Age	11.22	0.21	142
	(2.638)	(0.46)	

Notes: Each row reports a regression of the given dependent variable on an indicator for whether a county has at least 0.7 ppm of fluoride in the water supply. The indicator was constructed by taking the mean of the school fluoride sample from the NSOHS by county. Standard deviations in parentheses in the first column and robust standard errors in parantheses in the second column. Only counties with at least 80 percent of the NSOHS sample reporting urban status are included in these estimates. * significant at 10%; ** significant at 5%; *** significant at 1%.

status is poverty for the within-urban comparison. This indicates that poorer urban areas are actually more likely to be fluoridated, which is consistent with the notion that a primary goal of water fluoridation is to lessen the oral health gap between communities of varying incomes. In general however, observable covariates are balanced when comparing within urban and rural areas, which lends credibility to our results.

My hypothesis and motivation for obtaining data on the voting records of counties was that counties that voted democratic were more likely to have liberal views and thus be more likely to vote or protest against water fluoridation due to ethical or health concerns. However, Tables 1 and 2 show that there is no significant difference in the proportion of individuals that voted for a democratic presidential candidate between 1968 and 1984 between fluoridated and unfluoridated counties

Table 3.2: Balance Check - Urban

	Overall Mean	Fluoridated	Observations
Median Income	26650	-2,731.63	85
	(6937)	(1,712.85)	
Poverty	18.24	4.64**	85
	(10.09)	(2.16)	
White	0.719	-0.03	91
	(0.357)	(0.08)	
Democrat Vote	38.69	2.28	85
Proportion	(10.76)	(2.43)	
Age	11.22	-0.06	91
	(2.638)	(0.56)	

Notes: Each row reports a regression of the given dependent variable on an indicator for whether a county has at least 0.7 ppm of fluoride in the water supply. The indicator was constructed by taking the mean of the school fluoride sample from the NSOHS by county. Standard deviations in parentheses in the first column and robust standard errors in parantheses in the second column. Only counties with under 80 percent of the NSOHS sample reporting urban status are included in these estimates.* significant at 10%; ** significant at 5%; *** significant at 1%.

when comparing within rural and urban areas.

3.6.2 1986-1987 National Survey of Oral Health of US Schoolchildren

3.6.2.1 Effect on Carious Surfaces

Table 3 presents the point estimates for the effect of fluoride on deciduous cavities in children 12 years old or younger and permanent cavities in children 5 to 17 years old. Since the NSOHS contains information on whether a child has taken fluoride drops or tablets, I can analyze the interaction between water fluoridation, drops, and tablets, to describe the effect of water fluoridation when an individual does or does not use additional fluoride supplements. Column (1) shows that the effect of optimal residential fluoridation for children that have never taken fluoride drops or tablets is a reduction in cavities of 0.969 of a surface, and the effect for a child that has taken fluoride drops or tablets is a change in cavities of -0.969 - 0.177 + 0.557 = -0.589 of a surface, which has a p-value of 0.0856, and is thus significant at the ten percent level. From column (3), the effect of optimal residential fluoridation on permanent cavities for a child that has never taken fluoride drops or tablets is a reduction in cavities of 0.638 of a surface, and the effect for a child that has taken fluoride drops or tablets is a change in cavities of -0.638 - 0.207 + 0.900 = 0.055 of a surface, which has a p-value of 0.8215, and thus is not statistically significant. Moreover, the effect of taking fluoride drops or tablets without consuming optimally fluoridated water is merely -0.177 of a surface for deciduous cavities and -0.207 of a surface for permanent cavities, both of which are insignificant. Thus, my point estimates indicate that taking fluoride drops or tablets are not a suitable substitute for water fluoridation for the prevention of cavities. Moreover, consuming fluoridated water and taking fluoride drops or tablets provides no additional benefit over consuming only fluoridated water.

Additionally, the dose response function shows that the largest effect for deciduous cavities is in the above optimal range of 1.2-3 ppm, which is only slightly larger, but less significant than the response in the optimal range of 0.7-1.2 ppm. For permanent cavities, the largest effect is actually in the range of greater than 3 ppm, but this is only statistically significant at the ten percent level where as fluoride in the optimal range of 0.7-1.2 ppm is only slightly smaller and significant

Table 3.3: Effect of Community Water Fluoridation on Cavities

	Deciduous cavities		Permanent cavities	
	(1)	(2)	(3)	(4)
Residential Fluoridation	-0.969**		-0.638***	
	(0.398)		(0.157)	
ppm(.3-.7)		-0.868***		-0.454**
		(0.294)		(0.219)
ppm(.7-1.2)		-1.261***		-0.561**
		(0.344)		(0.213)
ppm(1.2-3)		-1.341**		0.050
		(0.508)		(0.493)
ppm(>3)		-0.798		-0.652*
		(0.587)		(0.344)
White	-1.111***	-1.092***	-0.452***	-0.489***
	(0.260)	(0.225)	(0.123)	(0.131)
Male	0.216	0.214	-0.460***	-0.462***
	(0.157)	(0.159)	(0.084)	(0.086)
Urban	-0.348	-0.244	-0.122	-0.083
	(0.314)	(0.291)	(0.204)	(0.225)
Fluoride Tablet or Drops	-0.177	-0.173	-0.207	0.074
	(0.282)	(0.234)	(0.146)	(0.144)
Fluoride Tablet or Drop	0.557		0.900***	
x Residential Fluoridation	(0.415)		(0.232)	
Median Income	-0.344*	-0.387**	0.024	0.010
	(0.201)	(0.178)	(0.139)	(0.149)
Constant	4.457***	4.711***	8.960***	8.985***
	(0.482)	(0.426)	(0.615)	(0.658)
Observations	11,182	11,182	18,722	18,722
R-squared	0.025	0.028	0.266	0.266

Notes: Standard errors in parantheses using BRR replicate survey weights. (1) and (2) include indicators for ages 4-12. (3) and (4) include indicators for ages 5-17. * significant at 10%; ** significant at 5%; *** significant at 1%.

at the five percent level.[63]

Figure 1 presents non-parametric analyses for the relationship between the level of fluoride at a child's school and deciduous cavities. First, it is important to note that in each of the non-parametric figures, the local polynomial fits of the

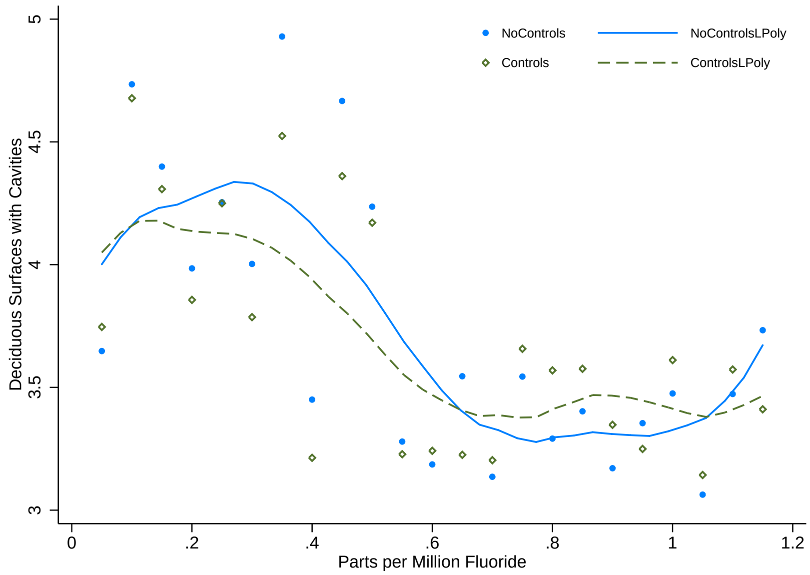
[63]Only 0.7 percent of children in the sample attend schools with greater than 3 ppm of fluoride in the drinking water.

residuals and the original data are very similar, which implies that the relationship shown is in fact due to water fluoridation, and is not changed significantly by controlling for observable characteristics. The local polynomial fit is essentially flat until about 0.3 ppm, but has a steep negative slope until about 0.7 ppm, the beginning of the optimal range and the new optimal quantity for all climates recommended by the CDC in 2011. This relationship implies that the effect of water fluoridation on deciduous cavities is non-existent at very low levels of fluoride, but is quite significant from very low levels of fluoride to the beginning of the optimal range. However, there does not appear to be any additional effect on deciduous cavities at fluoride levels greater than 0.7 ppm. This result was previously discussed by Burt, Eklund, and Heller (1997) and is consistent with the CDC's change to the recommendation for the optimal fluoride level in 2011.

Figure 3 presents a plot of coefficients from a series of linear probability models for deciduous cavities. The effect is strongest at an indicator of 6, which means that having an additional 1 ppm of fluoride in a child's water at school moves 6.3 percent of children from having more than six cavities to having less than or equal to 6 cavities.[64] The magnitude of the effect declines quickly but remains significant until an indicator of 30. The linear probability model in Figure 4 shows that water fluoridation is most effective at preventing deciduous cavities in

[64] Coincidentally, a 1 ppm increase is roughly equivalent to moving from completely unfluoridated to the middle of the optimal fluoridation range.

Figure 3.1: Relationship Between Fluoride Level and Deciduous Cavities



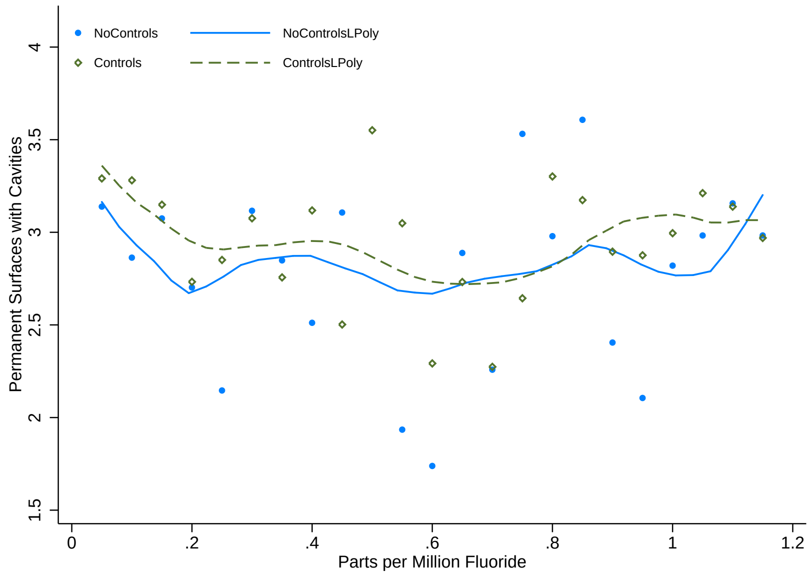
Notes: Each point represents the average number of deciduous cavities in 0.05 ppm bins. All local polynomials are of degree two.

children which already have relatively few cavities.

Surprisingly, the relationship between permanent cavities and fluoride is flat for all ppm levels which indicates that the effect of fluoride on permanent cavities is quite small or non-existent, even though the point estimate is statistically significant (Figure 2). When I relax the rigidity of the linear functional form, the apparent effect of water fluoridation on permanent surfaces seems to vanish.

The linear probability model for permanent cavities in Figure 4 also shows that the effect is strongest at an indicator of 6. However in contrast to the effect on deciduous cavities, the effect of water fluoridation is not significant up to that point, and is significant only for the indicators from 6 to 23. This implies that for permanent teeth, the effect is stronger and only significant for individuals that

Figure 3.2: Relationship Between Fluoride Level and Permanent Cavities



Notes: Each point represents the average number of deciduous cavities in 0.05 ppm bins All local polynomials are of degree two.

Figure 3.3: Linear Probability Model Permanent

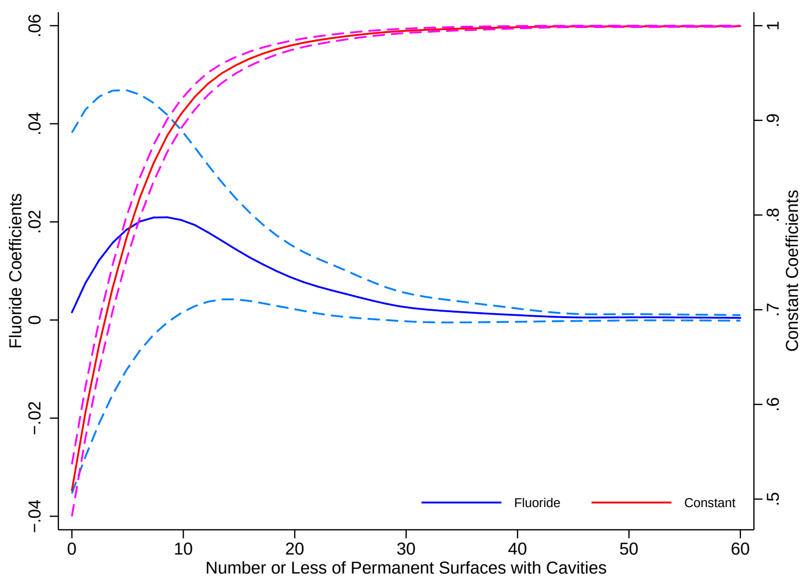
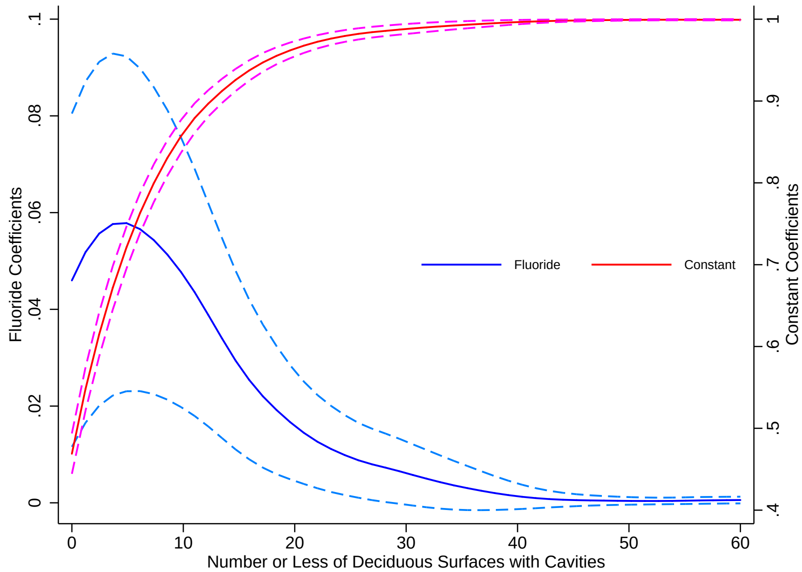


Figure 3.4: Linear Probability Model Deciduous



already have 6 or more carious surfaces. Specifically, the effect at an indicator of 0 is statistically 0, which means that on average, fluoride is not moving anyone from having 1 or more permanent carious surfaces to having 0. So, while water fluoridation seems to have a significant dampening effect on deciduous cavities at all levels of oral health, the effect on permanent cavities is significant only for those children with quite poor oral health, and the effect is still much smaller than for deciduous cavities even for individuals with a large amount of cavities.

3.6.2.2 Effect on Fluorosis

Table 4 presents the point estimates for the effect of fluoride on very mild, mild, or severe fluorosis. The effect of optimal residential fluoridation is an increase of 1.886 surfaces of very mild or worse fluorosis for children who have never taken

Table 3.4: Effect of Community Water Fluoridation on Fluorosis

	Very Mild or Worse Fluorosis		Mild or Worse Fluorosis		Severe Fluorosis	
	(1)	(2)	(3)	(4)	(5)	(6)
Residential Fluoridation	1.886**		0.428***		0.011	
	(0.699)		(0.142)		(0.012)	
ppm(.3-.7)		1.003		0.244		0.006
		(0.660)		(0.163)		(0.010)
ppm(.7-1.2)		2.080***		0.461***		0.014
		(0.662)		(0.165)		(0.011)
ppm(1.2-3)		2.029***		0.823*		0.124
		(0.660)		(0.456)		(0.106)
ppm(>3)		2.578		1.746		0.102
		(1.793)		(1.469)		(0.090)
White	-0.283	-0.324	-0.131	-0.141	-0.012*	-0.015**
	(0.464)	(0.439)	(0.144)	(0.139)	(0.007)	(0.007)
Male	-0.002	0.006	-0.003	0.000	-0.002	-0.001
	(0.142)	(0.142)	(0.043)	(0.043)	(0.005)	(0.005)
Urban	-1.499	-1.579	-0.410	-0.376	-0.005	0.002
	(1.155)	(1.152)	(0.347)	(0.363)	(0.016)	(0.012)
Fluoride Tablet or Drops	0.282	0.332	0.112	0.107*	0.008	0.011**
	(0.172)	(0.201)	(0.071)	(0.059)	(0.008)	(0.005)
Fluoride Tablet or Drop	-0.100		-0.106		0.000	
x Residential Fluoridation	(0.240)		(0.119)		(0.019)	
Median Income	0.649	0.661	0.039	0.046	-0.006	-0.005
	(0.949)	(0.951)	(0.157)	(0.156)	(0.009)	(0.009)
Constant	0.515	0.302	0.435	0.323	0.029	0.018
	(2.886)	(2.856)	(0.569)	(0.560)	(0.019)	(0.019)
Observations	20,308	20,308	20,308	20,308	20,308	20,308
R-squared	0.081	0.083	0.021	0.026	0.001	0.004

Notes: Standard errors in parantheses using BRR replicate survey weights. Includes indicators for ages 5-16. * significant at 10%; ** significant at 5%; *** significant at 1%.

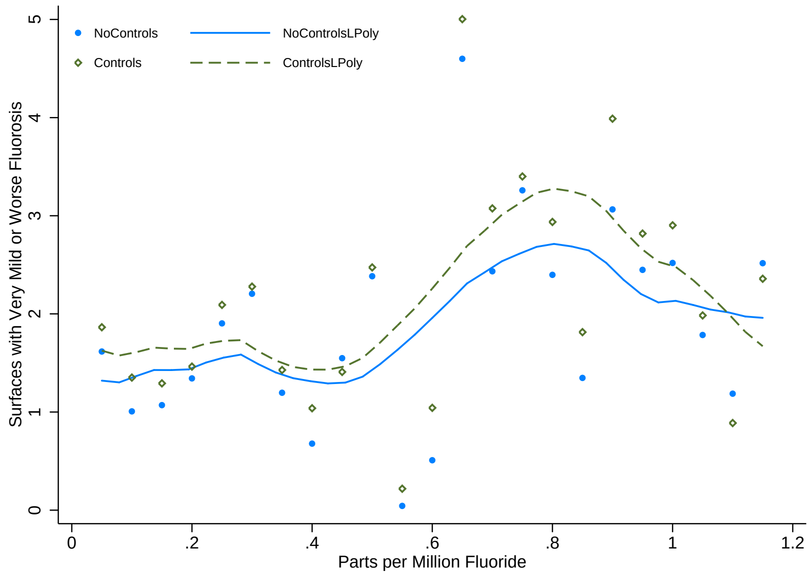
drops or tablets, and an effect of 1.886 + 0.282 - 0.100 = 2.068 for those that have. Taking tablets or drops in addition to water fluoridation seems to have very little impact, and although the magnitude of the point estimate is slightly larger with both types of fluoride intake, the difference of the effect between just water fluoridation and water fluoridation and drops or tablets is not significant.

The effect of optimal residential fluoridation is an increase of 0.428 of a surface with mild or worse fluorosis for children that have never taken drops or tablets, and an increase of 0.428 + 0.112 -0.106 = 0.434 for those that have. As in the previous case, there is not a significant compounding effect on fluorosis by drinking fluoridated water as well as taking drops or tablets.

Unlike in the more minor cases of fluorosis, optimal residential fluoridation does not have a significant impact on severe fluorosis at conventional significance levels. However, the dose response function in column (6) of Table 4 shows that at fluoride levels above the optimal range, the point estimates are about ten times larger than in the optimal range or below, but due to the small sample size in those ranges and the small amount of individuals with severe fluorosis, the estimates are also much more imprecise due to large standard errors, so I cannot reject that there is no effect. The observation that severe fluorosis is much more prevalent in areas with very high concentrations of natural fluoridation is consistent with the initial observations of Dr. McKay in the 1930's.

Figure 5 presents non-parametric analysis for the effect of fluoridation on counts of very mild or worse fluorosis. There is a clear increasing relationship from very low levels of fluoride to about 0.7 ppm, the beginning of the optimal range, but the relationship is essentially flat after that point. From Figure 6, we see that unlike the linear probability models for the effect on cavities, the effect of water fluoridation on fluorosis is actually strongest at an indicator of 0 with a

Figure 3.5: Relationship Between Fluoride Level and Very Mild or Worse Fluorosis



Notes: Each point represents the average number of deciduous cavities in 0.05 ppm bins All local polynomials are of degree two.

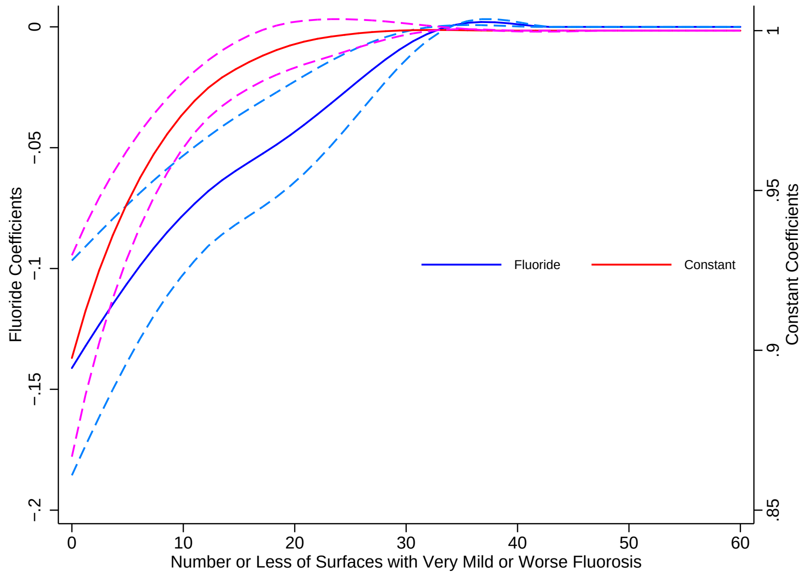
coefficient of -0.138, which means that 13.8 percent of children were moved from having 0 counts of fluorosis to instead having 1 or more counts of fluorosis due to water fluoridation. The effect then declines only gradually, and is still significant until it becomes 0 at an indicator of 27 since there are no children with more than 27 surfaces afflicted with fluorosis.[65]

3.6.3 National Health and Nutrition Examination Survey III

Table 5 presents estimates for the effect of optimal water fluoridation on deciduous cavities, permanent cavities, and self-reported teeth health for both children and adults from the NHANES III cross-sectional examination and survey. The magni-

[65]The results from non-parametric analysis and linear probability models for mild or worse and severe fluorosis are very similar. Results are available upon request.

Figure 3.6: Linear Probability Model Very Mild or Worse Fluorosis



tude of the point estimates for children are quite similar to those from the NSOHS, except that the effect on permanent cavities is not significant, even in children of the same age range, with the NHANES III data. Additionally, individuals living in more fluoridated counties self-report significantly better teeth health. Moreover, there is no significant effect of fluoride on either cavities or self-reported teeth health in adults, although the point estimates have the expected sign. However, the effect of water fluoridation on adults is likely to be severely attenuated since the measure of fluoride is based on the county where the individual currently lives, but the data set does not contain residential history information and thus I cannot discern the fluoridation status of examinees' previous residences and hence their lifetime fluoride exposure. So, although this measure does not show a significant effect of water fluoridation on cavities in adults, my data does not allow me to

Table 3.5: NHANES Oral Health Results

	Children			Adults	
	Deciduous cavities	Permanent cavities (5-17)	Self-Reported Teeth Health	Permanent cavities (17-90)	Self-Reported Teeth Health
	(1)	(2)	(3)	(4)	(5)
Fluoride	-0.799**	-0.516	-0.205***	-0.894	-0.089
	(0.330)	(0.318)	(0.075)	(1.040)	(0.057)
White	-0.547	-0.161	-0.109**	7.214***	-0.094**
	(0.444)	(0.438)	(0.050)	(0.819)	(0.043)
Male	-0.096	-0.062	0.063	-2.446***	0.024
	(0.248)	(0.197)	(0.055)	(0.593)	(0.041)
Urban	-1.026***	0.166	-0.110	-0.972	-0.076*
	(0.299)	(0.237)	(0.097)	(1.422)	(0.044)
Income	-0.531***	-0.148	-0.223***	1.571***	-0.166***
	(0.083)	(0.113)	(0.025)	(0.181)	(0.011)
Constant	4.028***	6.385***	3.094***	0.290	3.309***
	(0.499)	(0.805)	(0.114)	(2.377)	(0.135)
Observations	3,152	2,346	4,249	6,497	6,746
R-squared	0.084	0.251	0.116	0.277	0.117

Notes: Standard errors in parantheses using BRR replicate survey weights. (1) includes indicators for ages 3-12. (2) includes indicators for ages 5-17. (3) includes indicators for ages 3-15. (4) and (5) includes indicators for ages 17-90. The dependent variable in columns (3) and (5) is the self-reported condition of natural teeth, where 1 represents "Excellent," 2 represents "Very Good," 3 represents "Good," 4 represents "Fair," and 5 represents "Good." * significant at 10%; ** significant at 5%; *** significant at 1%.

rule out the possibility that one does exist.

Table 6 presents estimates for the effect of optimal fluoridation on several health outcomes. Fluoride is shown to have no significant effect on self-reported trouble working, whether a doctor has ever reported a child has had retardation, or the general health status of an individual as determined by the examining physician, in either children or adults. In fact, income is the only covariate with a consistent significant effect on these general health outcomes.

Table 3.6: NHANES General Health Results

	Children			Adults	
	Trouble Working	Retardation	General Health	Trouble Working	General Health
	(1)	(2)	(3)	(4)	(5)
Fluoride	0.039	0.001	-0.159	0.041	-0.028
	(0.039)	(0.002)	(0.165)	(0.043)	(0.143)
White	-0.003	-0.003	-0.002	0.003	-0.102*
	(0.031)	(0.002)	(0.051)	(0.021)	(0.057)
Male	0.014	0.001*	0.062***	-0.068***	-0.019
	(0.014)	(0.001)	(0.021)	(0.017)	(0.027)
Urban	0.048*	0.002***	0.304***	0.051	0.157
	(0.025)	(0.001)	(0.104)	(0.051)	(0.145)
Income	-0.020**	-0.001**	-0.051***	-0.047***	-0.080***
	(0.008)	(0.000)	(0.011)	(0.005)	(0.009)
Constant	1.027***	0.001	1.262***	1.089***	1.650***
	(0.037)	(0.001)	(0.140)	(0.064)	(0.183)
Observations	2,583	4,796	4,618	7,100	7,088
R-squared	0.031	0.006	0.057	0.394	0.237

Notes: Standard errors in parantheses using BRR replicate survey weights. (1), (2), and (3) include indicators for ages 4-15. (4) and (5) include indicators for ages 17-90. The dependent variable in columns (1) and (4) is the self-reported level of difficulty performing housework, gardening, exercise, or play, where 1 represents "No difficulty," 2 represents "Some difficulty," 3 represents "Moderate difficulty," and 4 represents "Could not do." The dependent variable in column (2) is a self-reported measure for whether a doctor has ever said the child has mental retardation. The dependent variable in columns (3) and (5) is the general health status of an examinee as determined by the examining physician, where 1 represents "Excellent," 2 represents "Very Good," 3 represents "Good," 4 represents "Fair," and 5 represents "Poor." * significant at 10%; ** significant at 5%; *** significant at 1%.

3.6.4 Educational Achievement

Table 7 presents estimates from panel regressions of high school and college graduation rates on optimal water fluoridation for the years 1950, 1970, 1980, 1990, 2000, and 2010, for 1118 counties across the United States, using data from the decennial census and the CDC's "My Water's Fluoride." Using a fixed effects model weighting each county by the total population served of all of the water systems that primarily serve that county, a fixed effects model without weighting, and a random effects model, I find that there is no significant negative effect of op-

Table 3.7: Effect of Community Water Fluoridation on Educational Achievement

	High School Graduation Rate			College Graduation Rate		
	FE + Weight	FE	RE	FE + Weight	FE	RE
	(1)	(2)	(3)	(4)	(5)	(6)
Fluoridation	0.010*	0.002	-0.002	0.014*	0.004	0.004*
	(0.006)	(0.003)	(0.003)	(0.007)	(0.002)	(0.002)
Urban	0.028*	-0.008	0.106***	0.031***	0.021***	0.041***
	(0.014)	(0.006)	(0.005)	(0.011)	(0.005)	(0.003)
White	0.397***	0.397***	0.244***	-0.108**	-0.028*	-0.001
	(0.042)	(0.020)	(0.012)	(0.043)	(0.016)	(0.007)
Constant	-0.031	-0.089***	-22.266***	0.133***	0.059***	-4.465***
	(0.035)	(0.017)	(0.121)	(0.034)	(0.014)	(0.082)
Observations	6,113	6,113	6,113	6,113	6,113	6,113
R-squared	0.967	0.972		0.836	0.752	
Number of counties	1,118	1,118	1,118	1,118	1,118	1,118

Notes: Standard errors in parantheses clustered by county. Columns (1) and (4) use a fixed effects model and are weighted by the total population served by each of the water systems that primarily serve a given county, as reported by "My Water's Fluoride," which varies by state from 2006-2010. Includes year fixed effects. Columns (2) and (5) use a fixed effects model without weighting, and columns (3) and (6) use a random effects model without weighting. * significant at 10%; ** significant at 5%; *** significant at 1%.

timal fluoridation on high school or college graduation rates in any specification. Educational achievement is of course a rough proxy for intelligence, but is the only proxy for which data exists for such a large number of counties in multiple states and spanning the entire length of time in which artificial fluoridation has been implemented in the United States. Although these outcomes are too rough a proxy to necessarily disprove the claims that fluoride has a deleterious effect on intelligence, it does show that fluoridated counties, even counties that have been fluoridated since the beginning of artificial water fluoridation, have not seen a decline in educational achievement compared to counties which have over time seen 0 or very low levels of water fluoridation.

3.7 Conclusion

Fluoride is a naturally occurring chemical which has been shown to have beneficial effects on oral health through the prevention of tooth decay. Through the use of multiple econometric techniques using two cross-sectional interview and examination data sets from the United States, supplemented with additional controls, as well as a separate panel data set constructed from two distinct sources, I have examined the impact water fluoridation has on oral health, general health, and educational achievement.

The beneficial effect of water fluoridation on deciduous cavities is unarguably significant statistically and qualitatively. With a point estimate of roughly 0.9 of a surface, children drinking optimally fluoridated water have on average 0.9 less of a carious surface than children that do not drink fluoridated water, which translates to a 19 percent decrease in decayed or filled surfaces. Moreover, non-parametric analysis shows that this effect is driven by the difference between those drinking water with very low levels of fluoride and those anywhere within the optimal range, including the floor of the range, 0.7 ppm. Additionally, the effect on deciduous teeth is strongest for those with low levels of cavities, with the effect peaking at 6 cavities and diminishing quickly thereafter.

In contrast however, the effect of water fluoridation on permanent cavities is far from clear-cut. Although the point estimate from the NSOHS is statistically significant and indicates a reduction in cavities of half a surface between fluoridated and non-fluoridated individuals on average, non-parametric analysis reveals a flat relationship at all levels of fluoride, implying that when I relax the assumptions of the linear regression model, there does not appear to be any significant relationship. Furthermore, although the effect on deciduous cavities remains significant using the NHANES III, the effect on permanent cavities is of similar magnitude but no longer significant. Moreover, I find that water fluoridation significantly reduces permanent cavities only for those individuals which already have 6 or more carious surfaces, and thus does not seem to have a beneficial effect on individuals with relatively favorable oral health.

Additionally, my results show that drinking fluoridated water and consuming fluoride drops or tablets does not contain any additional benefit over consuming only fluoridated water. Furthermore, external fluoride supplements such as drops and tablets are shown to have a significantly smaller effect on the prevention of cavities than water fluoridation and thus cannot be viewed as substitutes for water fluoridation for the prevention of cavities. Although the data does not provide information on toothpaste use and my results are in addition to any benefit derived from fluoridated toothpaste or mouth rinse, my results indicate that a threshold may exist across which no benefit is derived from additional fluoride intake.

Besides having a deleterious effect on tooth decay, fluoride causes a cosmetic tooth defect known as fluorosis, which manifests itself in various forms based on the level of fluoride ingested over time. I find that optimally fluoridated individuals on average have about 1.9 surfaces of very mild or worse fluorosis, and about 0.4 surfaces of mild or worse fluorosis, more than non-fluoridated individuals. Although not strictly monotonic, unlike the relationship between water fluoridation and cavities, the effect of water fluoridation on fluorosis continues as the fluoride level in the water supply increases. Although the effect of water fluoridation on severe cases of fluorosis are not shown to be statistically significant, likely due to imprecision as a result of very few individuals being afflicted with severe fluorosis, nearly all cases of severe fluorosis occur due to naturally fluoridated water containing levels far exceeding the recommended optimal range.

Using the NHANES III and a panel data set constructed from the years in which water systems began fluoridating combined with census data, I do not find any evidence of an effect of water fluoridation on general health, retardation in children, trouble working for children or adults, or educational achievement in the form of average high school and college graduation rates for a county. Although my measures are insufficiently precise to preclude the possibility of any adverse health effects or deleterious effects on intelligence, my results indicate no such effects.

References

Adams, Kathleen, Robert Houchens, George Wright, and James Robbins. 1991. "Predicting hospital choice for rural Medicare beneficiaries: the role of severity of illness." *Health Services Research* 26(5): 583-612.

Adams, Kathleen, and George Wright. 2008. "Hospital Choice of Medicare Beneficiaries in a Rural Market: Why Not the Closest?" *The Journal of Rural Health* 7(2): 134-152.

Afsah, Shakeb, Allen Blackman, and Damayanti Ratunanda. 2004. "How do public disclosure pollution control programs work: evidence from Indonesia." *Human Ecology Review* 11(3): 235-246.

Albouy, Valerie, and Laurent Lequien. 2009. "Does Compulsory Education Lower Mortality?" *Journal of Health Economics* 30(3): 515-530.

Allen SJ, LJ Brown, and ML Ringelberg. 1992. "Cost of fluoridation: 44 Florida communities." *Journal of Public Health Dentistry* 52(2):75-80.

Alm, James, and Mikhail Melnik. 2002. "Does a seller's ecommerce reputation matter? Evidence from eBay auctions." *The Journal of Industrial Economics* 50(3): 337-349.

American Hospital Association. 2005. Hospital Statistics. Chicago: HealthForum.

Anderson, Michael, and Jeremy Magruder. "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database." *The Economic Journal* 122(563): 957-989.

Angrist, Joshua, and Alan Krueger. 1992. "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." *Journal of the American Statistical Association* 87(418):328-336.

Arendt, Jacob Nielsen. 2005. "Does Education Cause Better Health? A Panel Data Analysis Using School Reforms for Identification." *Economics of Education Review* 24(2): 149-160.

Austin, Matthew, et al. 2014. "Safety in Numbers: The Development of Leapfrog's Composite Patient Safety Score for U.S. Hospitals." *Journal of Patient Safety*

10(1): 64-71.

Bassin, Elise B., Roger B. Davis, Murray A. Mittleman, and David Wypij. 2006. “Age-specific fluoride exposure in drinking water and osteosarcoma (United States).” *Cancer Causes & Control* 17(4):421-428.

Beaulieu, Nancy Dean. 2002. “Quality information and consumer health plan choices.” *Journal of Health Economics* 21(1): 43-63.

Bennear, Lori, and Sheila Olmstead. 2008. “The impacts of the “right to know”: Information disclosure and the violation of drinking water standards.” *Journal of Environmental Economics and Management* 56(2): 117-130.

Besley, Timothy J., Gwyn Bevan, and Konrad Burchardi. 2009. "Naming & Shaming: The impacts of different regimes on hospital waiting times in England and Wales."

Bevan, Gwyn, and Christopher Hood. 2006. “Have targets improved performance in the English NHS?” *BMJ: British Medical Journal* 332(7538): 419.

Black, Nick, Anthony Brady, David Harrison, and Kathy Rowan. 2004. "Hospitals’ star ratings and clinical outcomes: ecological study." *BMJ: British Medical Journal* 328(7445): 924-925.

Bloom, Nicholas, Rafaella Sadun, and John Van Reenen. 2014. “Does Management Matter in Healthcare?”

Boccuti, Cristina, Christina Swoope, Anthony Damico, and Tricia Neuman. 2013. "Medicare patients’ access to physicians: a synthesis of the evidence." *Kaiser Family Foundation*. http://kff. org/medicare/issue-brief/medicarepatients-access-to-physicians-asynthesis-of-the-evidence/.

Boyce, John. 2011. “Measuring healthcare worker hand hygiene activity: current practices and emerging technologies.” *Infection Control* 32(10):1016–1028.

Brunelle, J. A., J. P. Carlos, and Epidemiology Branch. 1990. “Recent trends in dental caries in US children and the effect of water fluoridation.” *Journal of Dental Research* 69:723-727.

Brunello, Giorgio, Margherita Fort, Nicole Schneeweis, and Rudolf Winter-Ebmer. 2016. “The Causal Effect of Education on Health: What is the Role of Health Behaviors?” *Health Economics* 25(3): 314-336.

Bryan, Doug, David Lucking-Reiley, Naghi Prasad, and Daniel Reeves. 2007. "Pennies from eBay: the determinants of price in online auctions." *The Journal of Industrial Economics* 55(2): 223-233.

Bundorf, Kate, Natalie Chun, Gopi Shah Goda, and D. Kessler. 2009. "Do Markets Respond to Quality Information? The Case of Fertility Clinics." *Journal of Health Economics* 28(3): 718-727.

Burt, Brian A., Stephen A. Eklund, and Keith E. Heller. 1997. "Dental caries and dental fluorosis at varying water fluoride concentrations." *Journal of Public Health Dentistry* 57(3):136-143.

Burt, Brian A. 1989. "Proceedings for the workshop: Cost-effectiveness of caries prevention in dental public health, Ann Arbor, Michigan, May 17–19." *Journal of Public Health Dentistry* (special issue): 331-337.

Cabral, Luis, and Ali Hortacsu. 2004. "The Dynamics of Seller Reputation: Theory and Evidence from eBay." *National Bureau of Economic Research* No. w10363.

Calonico, Sebastian, Matias Cattaneo, and Rocıo Titiunik. 2014. "Robust data-driven inference in the regression-discontinuity design." *Stata Journal* 14(4): 909-946.

Carnoy, Martin, and Susanna Loeb. 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis* 24(4): 305-331.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2015. "Healthcare Exceptionalism? Performance and Allocation in the US Healthcare Sector." No. w21603. National Bureau of Economic Research.

Chatterji, Aaron, and Michael Toffel. 2010. "How firms respond to being rated." *Strategic Management Journal* 31(9): 917-945.

Chay, Kenneth, Patrick McEwan, and Miguel Urquiola. 2005. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *American Economic Review* 95(4): 1237-1258.

Chernew, Michael, Gautam Gowrisankaran, and Dennis Scanlon. 2001. "Learning and the Value of Information : Evidence from Health Plan Report Cards."

National Bureau of Economic Research No. w8589.

Chernew, Michael, Catherine McLaughlin, Dennis Scanlon, and Gary Solon. “The impact of health plan report cards on managed care enrollment.” *Journal of Health Economics* 21(1): 19-41.

Chestnutt, I. G., J. Kleijnen, M. McDonagh, E. T. Treasure, P. Whiting, and P. Wilson. 2002. “The York Review - A systematic review of public water fluoridation: a commentary.” *British Dental Journal* 192:495-497.

Chevalier, Judith, and Dina Mayzlin. 2006. “The Effect of Word of Mouth on Sales: Online Book Reviews.” *Journal of Marketing Research* 43(3): 345-354.

Choi, Anna L., Philippe Grandjean, Guifan Sun, and Ying Zhang. 2012. “Developmental fluoride neurotoxicity: a systematic review and meta-analysis.” *Environmental Health Perspectives* 120(10):1362.

Cicuttini, Flavia M., Lisa L. Demos, Christopher K. Fairley, Hana Kazda, and Martha I. Sinclair. 2001. “Water fluoridation, osteoporosis, fractures-recent developments.” *Australian Dental Journal* 46(2):80-87.

Clark, Damon, and Heather Royer. 2013. “The Effect of Education on Adult Mortality and Health: Evidence from Britain.” *American Economic Review* (TBA).

Cohen, Mark, and Shameek Konar. 1997. “Information As Regulation: The Effect of Community Right to Know Laws on Toxic Emissions.” *Journal of Environmental Economics and Management* 32(1): 109-124.

Cullen, Julie Berry and Randall Reback, 2006. "Tinkering toward accolades: school gaming under a performance accountability system" in T. Gronberg and D. Jansen, eds., Improving School Accountability: Check-Ups or Choice, Advances in Applied Microeconomics 14 (Amsterdam: Elsevier Science).

Cutler, David, Robert Ilckman, and Mary Beth Landrum. 2004. “The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery.” *American Economic Review* 94(2): 342-346.

Dafny, Leemore, and David Dranove. 2008. “Do Report Cards Tell Consumers Anything They Don’t Already Know? The Case of Medicare HMOs.” *RAND Journal of Economics* 39(3): 790-821.

Darden, Michael, and Ian McCarthy. 2015. "The Star Treatment: Estimating the Impact of Star Ratings on Medicare Advantage Enrollments." *The Journal of Human Resources* 50(4): 980-1008.

DeLong, Elizabeth, James Jollis, Daniel Mark, Lawrence Muhlbaier, and Eric Peterson. 1998. "The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly." *Journal of the American College of Cardiology* 32(4): 993-999.

Dafny, Leemore, and David Dranove. 2008. "Do report cards tell consumers anything they don't know already? The case of Medicare HMOs." 39(3): 790-821.

DeLia, David. 2006. "Annual Bed Statistics Give a Misleading Picture of Hospital Surge Capacity." *Annals of Emergency Medicine* 48(4): 384-388.

Department of Health and Human Services. 2014. "ACCESS TO CARE: PROVIDER AVAILABILITY IN MEDICAID MANAGED CARE." OEI-02-13-00670. Retrieved from http://oig.hhs.gov/oei/reports/oei-02-13-00670.pdf.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Sattherthwaite. 2008. "Is More Information Better? The Effects of "Report Cards" on healthcare Providers." *Journal of Political Economy* 111(8): 555-588.

Dranove, David, Subramaniam Ramanarayanan, and Yasutora Watanabe. 2012. "Delivering Bad News: Market Responses to Negligence." *Journal of Law and Economics* 55(1): 1-25.

Dranove, David, and Andrew Sfekas. 2008. "Start spreading the news: A structural estimate of the effects of New York hospital report cards." *Journal of Health Economics* 27: 1201-1207.

Edmond, MB, A Goodell, W Zuelzer, et al. 2010. "Successful use of alcohol sensor technology to monitor and report and hygiene compliance." *Journal of Hospital Infection* 76: 364–365.

Eide, Eric, and Mark Showalter. 2011. "Estimating the relation between health and education: What do we know and what do we need to know?" *Economics of Education Review* 30(5): 778-791.

Elsbach, Kimberly, and Roderick Kramer. 1996. "Members' Responses to Organizational Identity Threats: Encountering and Countering the Business Week

Rankings." *Administrative Science Quarterly* 41(3): 442-476.

Ennis S, Schoenbaum M, Keeler T. 2000. "Optimal Prices and Costs for Hospitals with Excess Bed Capacity." *Applied Economics* 32(9): 1201–1212.

Epstein, Arnold, and Eric Schneider. 1996. "Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care — A Survey of Cardiovascular Specialists." *The New England Journal of Medicine* 335: 251-256.

Espeland, Wendy, and Michael Sauder. 2007. "Rankings and reactivity: how public measures recreate social worlds." *American Journal of Sociology* 113(1): 1-40.

Featherstone, JD. 1999. "Prevention and reversal of dental caries: role of low level fluoride." *Community Dent Oral Epidemiol*iogy 27(1):31-40.

Figlio, David, and Maurice Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review* 94(3): 591-604.

Figlio, David, and Lawrence Getzler. 2002. "Accountability, ability and disability: Gaming the system." National Bureau of Economic Research No. w9307.

Gawande, Atul. 2010. "The checklist manifesto: How to get things right." Vol. 200. New York: Metropolitan Books.

Gaynor M, Anderson GF. 1995. "Uncertain Demand, the Structure of Hospital Costs, and the Cost of Empty Hospital Beds." *Journal of Health Economics* 14(3): 291–317.

Glied, Sherry, and Matthew Neidell (2010). "The economic value of teeth." *Journal of Human Resources* 45(2):468-496.

Goldman, Noreen. 2001. "Social Inequalities in Health." *Annals of the New York Academy of Sciences* 954(1): 118-139.

Gould, Dinah, Donna Moralejo, Nicholas Drey, and Jane Chudleigh. 2010. "Interventions to improve hand hygiene compliance in patient care." *The Cochrane Library.*

Green, Linda. 2002. "How Many Hospital Beds?" *Inquiry* 39(4): 400–412.

Green, Linda, and Nan Liu. 2015. “A Study of New York City Obstetrics Units Demonstrates the Potential for Reducing Hospital Inpatient Capacity.” *Medical Care Research and Review*: 1077558715572388.

Griffin, Susan O., Kari Jones, and Scott L. Tomar. 2001. “An economic evaluation of community water fluoridation.” *Journal of Public Health Dentistry* 61(2):78-86.

Hallstein, Eric, and Sofia Villas-Boas. 2010 "Can Household Consumers Save the Wild Fish? Analysis of a Seafood Sustainability Advisory in a US Supermarket."

Hannan, EL, H Kilbum, M Racz, E Shields, and MR Chassin. 1994. “Improving the outcomes of coronary artery bypass surgery in New York State.” *Journal of the American Medical Association* 271(10): 761-761.

Hanushek, Eric, and Margaret Raymond. 2005. “Does School Accountability Lead to Improved Student Performance?” *Journal of Policy Analysis and Management* 24(2): 297-327.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and estimation of treatment effects with a regression discontinuity design." *Econometrica* 69(1): 201-209.

Hauck, Katharina, and Andrew Street. 2007. “Do targets matter? A comparison of English and Welsh National Health priorities.” *Health Economics* 16(3): 275-290.

Heckman, James, John Humphries, and Gregory Veramendi. 2015. "The causal effects of education on earnings and health." Unpublished manuscript, Department of Economics, University of Chicago.

Higgins, J.P., V. C. Marinho, Stuart Logan, and Aubrey Sheiham. 2003. “Topical fluoride (toothpastes, mouthrinses, gels or varnishes) for preventing dental caries in children and adolescents.” *Cochrane Database Syst Rev 4.*

Hilger, James, Greg Rafert, and Sofia Villas-Boas. 2011. “Expert opinion and the demand for experience goods: an experimental approach in the retail wine market.” *The Review of Economics and Statistics* 93(4): 1289-1296.

Hirshleifer, David, and Hong Teoh, Siew. “Limited attention, information disclosure, and financial reporting.” *Journal of Accounting and Economics* 36(1-3):

337-386.

Hood, Christopher. 2006. "Gaming in Targetworld: The Targets Approach to Managing British Public Services." *Public Administration Review* 66(4): 515-521.

Howard, David, and Bruce Kaplan. 2006. "Do Report Cards Influence Hospital Choice? The Case of Kidney Transplantation." *Inquiry* 43(2): 150-159.

Imbens, Guido, and Karthik Kalyanaraman. 2011. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* rdr043.

Imbens, Guido, and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142: 615-635.

IOM (Institute of Medicine). 2000. To Err is Human—Building a Safer Health System. Washington, DC: The National Academies Press.

Ippolito, Pauline and Alan Mathios. 1990. "Information, Advertising and Health Choices: A Study of the Cereal Market." *The RAND Journal of Economics* 21(3): 459-480.

Jacob, Brian. 2005. "Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics* 89: 761-796.

Jacob, Brian, and Steven Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *The Quarterly Journal of Economics* 118(3): 843-877.

James, John. 2013. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." *The Journal of Patient Safety* 9(3): 122-128.

Jin, Ginger, and Andrew Kato. 2006. "Price, quality, and reputation: evidence from an online field experiment." *RAND Journal of Economics* 37(4): 983-1005.

Jin, Ginger, and Philip Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *The Quarterly Journal of Economics* 118(2): 409-451.

Jin, Ginger, and Alan Sorensen. 2005. "Information and Consumer Choice: The Value of Publicized Health Plan Ratings." *National Bureau of Economic Research* No. w11514.

Johnston, David, Grace Lordon, Michael Shields, and Agne Suziedelyte. 2015. "Education and Health Knowledge: Evidence from UK Compulsory Schooling Reform" *Social Science & Medicine* 127: 92-100.

Jürges, Hendrik, Eberhard Kruk, and Steffen Reinhold. 2013. "The Effect of Compulsory Schooling on Health Evidence from Biomarkers." *Journal of Population Economics* 26(2): 645-672.

Klevens, R. Monina, Jonathan R. Edwards, Chesley L. Richards, Teresa C. Horan, Robert P. Gaynes, Daniel A. Pollock, and Denise M. Cardo. 2007. "Estimating health care-associated infections and deaths in US hospitals, 2002." *Public health reports* 122(2): 160.

Kling, Jeffrey, Sendhil Mullainathan, Eldar Shafir, Lee Vermeulen, and Marian Wrobel. 2012. "Comparison Friction: Experimental Evidence from Medicare Drug Plans." *The Quarterly Journal of Economics* 127(1): 199-235.

Kohn, Linda T., Janet Corrigan, and Molla S. Donaldson. 2000. "To err is human: building a safer health system." Washington, D.C.: National Academy Press.

Lillard, Dean, and Eamon Molloy. 2010. "Live and learn or learn and live: Does education lead to longer lives?" Working paper, Cornell University.

Liu, S., W. Lu, Y. Lu, Z. R. Sun, X. Wang, L. N. Wu. 2000. "Effect of high-fluoride water on intelligence in children." *Fluoride* 33(2):74-78.

Lleras-Muney, Adriana. 2005. "The Relationship Between Education and Adult Mortality in the United States." *Review of Economic Studies* 72: 189-221.

Lockwood, Kate, Paul Resnick, John Swanson, and Richard Zeckhauser. 2006. "The value of reputation on eBay: a controlled experiment." *Experimental Economics* 9(2): 79-101.

Lu, Susan. 2012. "Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes." *Journal of Economics and Management Strategy* 21(3): 673-705.

Luft, Harold, Deborah Garnick, David Mark, Deborah Peltzman, Ciaran Phibbs, Erik Lichtenberg, Stephen J. McPhee. 1990. “Does Quality Influence Choice of Hospital?” *Journal of the American Medical Association* 263(21): 2899-2906.

Ludwig, Jens, and Douglas Miller. 2007. “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design.” *Quarterly Journal of Economics* 122(1): 159-208.

CW Madden. 1999. “Excess Capacity: Markets, Regulation, and Values.” *Health Services Research* 33: 1651–1668.

Marschall, Jonas, Leonard Mermel, David Classen, Kathleen Arias, Kelly Podgorny, Deverick Anderson, Helen Burstin et al. 2008. "Strategies to prevent central line–associated bloodstream infections in acute care hospitals." *Strategies* 29(S1): S22-S30.

Mazumder, Bhash. 2008. “Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws.” *Federal Reserve Bank of Chicago Economic Perspectives* 32(2): 2-16.

McCrary, Justin. 2008. “Manipulation of the running variable in the regression discontinuity design: A density test.” *Journal of Econometrics* 142(2): 698–714.

McGillivray, Alice V., Richard M. Scammon. 1988. “America At the Polls 2: The Vote for President 1986-1984.” Elections Research Center, 1321 Connecticut Ave. N.W., Washington, D.C. 20036.

Mennemeyer, Stephen, Michael Morrisey, and Leslie Howard. 1997. “Death and reputation: how consumers acted upon HCFA mortality information.” *Inquiry* 34(2): 117-128.

Moretti, Enrico. 2011. “Social learning and peer effects in consumption: evidence from movie sales.” *Review of Economic Studies* 78(1): 356-393.

Omoigui NA, Miller DP, Brown KJ, et al. 1996. “Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes.” *Circulation* 93(1): 27–33.

Oreopoulos, Phillip. 2006. “Estimating Average and Local Average Treatment Effects of Education when Compulsory School Laws Really Matter.” *American Economic Review* 96(1): 152-175.

Pfoh, Elizabeth, Sydney Dy, and Cyrus Engineer. 2013. "Interventions to improve hand hygiene compliance: brief update review." *Making healthcare Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices.* Chapter 8.

Pope, Devin. 2009. "Reacting to rankings: Evidence from "America's Best Hospitals." *Journal of Health Economics* 28(6): 1154-1165.

Powdthavee, Nattavudh. 2010. "Does Education Reduce the Risk of Hypertension? Estimating the Biomarker Effect of Compulsory Schooling in England." *Journal of Human Capital* 4(2): 173-202.

Powers, Nicholas, Allan Blackman, Thomas Lyon, and Urvashi Narain. 2008. "Does Disclosure Reduce Pollution? Evidence from India's Green Rating Project. *Resources for the Future.*

Propper, Carol, Matt Sutton, Carolyn Whitnall, and Frank Windmeijer. 2010. "Incentives and targets in hospital care: Evidence from a natural experiment." *Journal of Public Economics* 94(3): 318-335.

Ramanarayanan, Subramaniam, and Jason Snyder. 2012. "Information Disclosure and Firm Performance: Evidence from the Dialysis Industry." *UCLA Anderson.*

Reinhart, Uwe. 2006. "The Pricing Of U.S. Hospital Services: Chaos Behind A Veil Of Secrecy." *Health Affairs* 25(1): 57-69.

Romano, Patrick, and Hong Zhou. 2004. "Do Well-Publicized Risk-Adjusted Outcomes Reports Affect Hospital Volume?" *Medical Care* 42(4): 367-377.

Scanlon, Dennis, Richard Lindrooth, and Jon Christianson. "Steering Patients to Safer Hospitals? The Effect of a Tiered Hospital Network on Hospital Admissions." *Health Services Research* 43 (5 Pt 2): 1849-1868.

Schneider, Eric, and Arnold Epstein. 1996. "Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care — A Survey of Cardiovascular Specialists." *New England Journal of Medicine* 335: 251-256.

Schut, Frederik, Stephanie van der Geest, and Marco Varkevisser. 2012. "Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands." *Journal of Health Economics* 31(2): 371-378.

Scorse, Jason. 2000. "Does being a "Top 10" worst polluter affect facility environmental releases? Evidence from the US toxic release inventory." *Atlantic Monthly.*

Scott, Douglas. 2009. "The direct medical costs of healthcare-associated infections in US hospitals and the benefits of prevention." 1-16.

Silles, Mary. 2009. "The causal effect of education on health: Evidence from the United Kingdom." *Economics of Education Review* 28(1): 122-128.

Sivey, Peter. 2011. "The effect of waiting time and distance on hospital choice for English cataract patients." *Health Economics* 21(4): 444-456.

Snelling, Iain. 2003 "Do star ratings really reflect hospital performance?" *Journal of Health Organization and Management* 17(3): 210-223.

Tai-Seale, M., GJ Wedig. 2002. "The effect of report cards on consumer choice in the health insurance market." *Journal of Health Economics* 21(6): 1031-1048.

Vernaz, Nathalie, Hugo Sax, Didier Pittet, Pascal Bonnabry, Jacques Schrenzel, and Stephan Harbarth. 2008. "Temporal effects of antibiotic use and hand rub consumption on the incidence of MRSA and Clostridium difficile." *Journal of Antimicrobial Chemotherapy* 62(3): 601-607.

World Health Organization. 2009. "WHO Guidelines on Hand Hygiene in healthcare: First Global Patient Safety Challenge."

Yiamouyiannis, John A. 1990. "Water fluoridation and tooth decay: Results from the 1986-1987 national survey of U. S. school children." *Fluoride* 23(2):55-67.