Empirical Essays on the Economics of Healthcare and Education

A dissertation submitted for the degree of

Doctor of Philosophy in Economics



Ines Lee

St. Antony's College

University of Oxford

Michaelmas Term

Academic Year 2019–2020

Declaration

I certify that this thesis, submitted for the degree of Doctor of Philosophy in Economics at the University of Oxford, is my own work except where otherwise clearly stated.

I confirm that Chapters 1 and 2 are solely my own work. Chapter 3 is co-authored with Sami Miaari. I contributed to developing the empirical strategy for this project, producing the empirical results, and writing the draft for this chapter.

This thesis is approximately 80,000 words.

Acknowledgements

I have many people to thank for helping me reach this point of my research journey. I am very grateful to have had the invaluable advice, support, and encouragement of my supervisors Professors Steve Bond and Climent Quintana-Domeque. They have challenged me to think deeply and critically about my research topics and to make my research questions matter to society. I hope parts of this thesis meet this challenge.

The Applied Microeconomics Group at Oxford is a wonderful community where I've received a lot of support. Johannes Abeler gave me much guidance during my undergraduate years including my first research assistant job. Brian Bell introduced me to research when I was a Masters student and made the academic work look less daunting. Abi Adams provided me with valuable advice on research and career choices.

My doctoral research has been generously supported by the Swire Scholarship Programme. I thank the Swire Group for welcoming me into a community of incredible scholars. I also thank the George Webb Medley fund and the St. Antony's research travel fund for enabling me to pursue research topics that I am passionate about and to travel internationally to present them.

Lastly, I am eternally grateful to my family and friends who have been my emotional and spiritual advisors through all these years. Their loving support and ceaseless encouragement has made this thesis possible.

Title: *Empirical Essays on the Economics of Healthcare and Education*
Name: Ines Lee **Degree:** D.Phil in Economics **College:** St. Antony's

Abstract

This thesis consists of two chapters on healthcare and one chapter on education. Scholars and policy-makers have long recognised that health and education are important components of human capital, which in turn affect individual welfare and economic growth. This thesis aim to enhance our understanding of the drivers of healthcare quality, health outcomes, and educational performance.

The first two chapters focus on supply- and demand-side factors that affect healthcare quality and outcomes. In Chapter 1, I explore the impacts of a pay-for-performance scheme that creates asymmetric incentives for hospitals to provide quality care cross different patient groups. I find that these incentives improve outcomes for targeted patients and have positive spillover effects on patients for whom hospitals are not explicitly incentivised to improve quality, suggesting that there are complementarities in quality provision within hospitals. These results suggest that asymmetric incentives across groups need not disadvantage the group with "low-powered" incentives.

In Chapter 2, I explore the impacts of health insurance on hospital readmissions among heart attack patients, focusing on demand-side responses to changes in insurance characteristics. Using the age-65 eligibility threshold for Medicare as a source of variation in insurance coverage, I estimate discontinuities in 30-day readmissions and treatment intensity. Among men, there is evidence that eligibility for Medicare insurance increases 30-day readmission rates and that treatment intensity is smooth across the age threshold. These findings suggest that in addition to provider-side factors, the focus of the existing literature, patient-side factors also matter for post-discharge outcomes.

In Chapter 3, co-authored with Sami Miaari, we examine the impact of conflict-related mobility restrictions on educational outcomes in the West Bank between 2000 to 2006 during the Israeli-Palestinian conflict. A prominent feature of this conflict is the system of mobility restrictions enforced through various physical barriers such as checkpoints. We find strong evidence that mobility restrictions in the form of checkpoints have adverse impacts on educational performance. There is evidence that three mechanisms could explain this effect: students lose time when crossing checkpoints on journeys to school, students' psychological wellbeing is negatively impacted by exposure to conflict, and physical resources at schools are negatively affected by mobility restrictions.

Contents

Thesis Introduction **8**

I Complementarities in Quality Provision within Hospitals: Evidence from a Pay-for-Performance Program **17**

1 Introduction **20**

2 Background **27**

3 Model **31**

3.1 Setup 31

3.2 Model Predictions 37

4 Empirical Strategy **41**

4.1 Matching Procedure 42

4.2 Estimating Equation 43

5 Data, Sample, and Descriptive Statistics **47**

5.1 Data Sources 47

5.2 Sample Selection 49

5.3 Descriptive Statistics 50

6 Results **56**

6.1 Impact on Readmission Rates 56

6.2 Impact on Treatment Intensity 59

6.3 Impact on Gaming Behaviour 62

7 Robustness Checks and Confounding Factors **65**

7.1 Robustness Checks 65

7.2 Other Confounding Factors 68

8 Conclusion **70**

II The Impacts of Health Insurance on Readmissions: Evidence from Heart Attack Patients on Medicare **71**

1 Introduction **74**

2 Conceptual Framework **81**

3 Background and Data Sources **87**

3.1 Background 87

3.2 Data 91

3.3 Variable Definition and Sample Selection 92

3.4 Descriptive Statistics 95

4 Empirical Strategy **99**

4.1 Reduced-Form Model 99

4.2 Identification Assumptions 101

5 Results: Impacts on 30-day readmissions **107**

5.1 Men: Discontinuities in readmissions 107

5.2 Women: Discontinuities in readmissions 115

6 Mechanisms: Impacts on treatment intensity **118**

6.1 Men: Discontinuities in treatment intensity 118

6.2 Women: Discontinuities in treatment intensity 121

7 Discussion and Conclusion **125**

III Obstacles on the Road to School: The Impact of Mobility Restrictions on Educational Performance **127**

1 Introduction **129**

2 Background **137**

2.1 The Israeli-Palestinian conflict and Mobility Restrictions . . . 137

2.2 The Education System in the West Bank . . . 142

2.3 How do mobility restrictions affect educational performance? . . . 144

3 Data **147**

3.1 Education Data . . . 147

3.2 Conflict Data . . . 148

3.3 Other Data Sources . . . 151

3.4 Descriptive Statistics . . . 151

4 Empirical Strategy **156**

4.1 Estimating Equations . . . 156

4.2 Identification Assumptions . . . 161

5 Results **163**

5.1 Main Results . . . 163

5.2 Robustness Checks and Threats to Identification . . . 172

6 Mechanisms **178**

6.1 Impact via Time Loss . . . 178

6.2 Impact via Psychological Well-being . . . 182

6.3 Impact via School-learning Environment . . . 187

6.4 Impact via Teachers . . . 189

7 Conclusion **192**

Thesis Conclusion 194

Bibliography 209

Appendix 210

A Appendix to Chapter 1 211

B Appendix to Chapter 2 227

C Appendix to Chapter 3 246

Introduction to Thesis

This thesis consists of three chapters, the first two on healthcare and the last one on education. Numerous studies have shown that investments in health and education can have positive benefits at an individual and aggregate level (e.g. Bleakley, 2010) and international organisations place heavy emphasis on the accumulation of knowledge and improved health as routes to development (e.g. World Bank, 2018). Therefore, well-functioning healthcare and education systems that enable individuals to make such investments play an important role in society. Designing and implementing such systems require an understanding of the determinants of healthcare and educational quality as well as health and educational outcomes. The chapters in this thesis aim to enhance our understanding of the drivers of healthcare quality, health outcomes, and educational performance.

Chapters on Healthcare

The first two chapters of this thesis focus on the healthcare system of the United States. Two features of the American healthcare system make it unique among industrialised nations. The first is that the US spends more as a percentage of GDP on healthcare than similarly advanced nations. In 2018, the US spent 18% of GDP on healthcare (Sisko et al., 2019). This is equivalent to $3.65 trillion or over $10,000 per capita. In comparison, other OECD countries with similar GDP per capita levels to the US spend roughly half as much per person on healthcare (Sawyer and Cox, 2018). For example, in 2017, health consumption expenditures per capita in the UK were under $4,200 (US dollars, PPP adjusted). The high healthcare costs in the US are often explained by higher prices for services, higher costs to administer the system, more utilisation of these services, or a combination of these elements. Despite the relatively high expenditure on healthcare, the US lags behind similarly wealthy OECD countries on measures such as all-cause mortality and mortality considered preventable by timely and effective care (Sawyer and McDermott, 2019).

The second prominent feature of the American healthcare system is that the US remains the only industrialised nation without universal healthcare coverage. Health insurance coverage in the US is provided by several private and public sources. In 2018, roughly 67% of the population had

some type of private insurance while 34% of the population had some form of public insurance (Berchick et al., 2019).[1] 8.5% of the population, or 27.5 million individuals, did not have health insurance at any point during the year.

The two main public insurance sources in the US are Medicare and Medicaid. Medicare is the main federal public health insurance program for the elderly, primarily providing health insurance coverage for individuals aged 65 and above. Nationally, around 75% of Medicare beneficiaries are enrolled in Traditional Medicare (TM) where they can freely choose their providers. The remainder are enrolled in Medicare Advantage (MA), which is an alternative Medicare plan run by private insurance companies but regulated by the government.[2] Medicaid is the federal social protection program for the poor. The remaining public coverage source is military healthcare coverage offered to retired service members. This includes CHAMPVA (Civilian Health and Medical Program of the Department of Veterans Affairs), as well as other care provided by the Department of Veterans Affairs (VA) and the military. In 2018, the proportions of the population that had Medicare, Medicaid, and VA/CHAMPVA coverage were 17.8%, 17.9%, and 1.0% respectively (Berchick et al., 2019).[3] The main source of private insurance in the US is employment-based coverage. In 2018, 55.1% of individuals had some form of employment-based insurance coverage. Some individuals also purchase private insurance coverage directly (10.8% of the population).

Within the US healthcare sector, the majority of spending goes to the hospital sector. The level and growth of spending in this sector are notable. Hospital care in the US represents almost 6% of gross domestic product, comprising roughly one-third of total spending on healthcare (Commonwealth Fund, 2014). This makes the hospital industry larger than construction (4.1%), mining (1.6%), and agriculture (0.8%) (Duffin, 2019; Gaynor et al., 2015). Furthermore, the costs of hospital services have grown faster than costs in other parts of the US healthcare system. For example, from 1997 to 2012, the cost of hospital services grew by 149% while the cost of physician services grew by 55% (Rosenthal, 2017).

The increasing costs of the US hospital sector have spurred policy-makers to seek solutions

[1] Some people may have more than one coverage type during the calendar year.
[2] Medicare Advantage is also known as Medicare Part C.
[3] Individuals may have multiple sources of public insurance coverage.

to reduce costs while maintaining or improving healthcare quality. In the past decade, increasing attention has been placed on post-discharge outcomes as an avenue to reduce costs and improve healthcare quality. The post-discharge outcome that has received most focus is hospital readmissions, an event that occurs when a patient who has been discharged from hospital is admitted again within a specified time frame (usually 30-days).

There are several reasons why readmissions have received attention from policy-makers. First, they are costly to the healthcare system. The estimated annual cost of readmissions among Medicare patients alone is $26 billion per year with $17 billion associated with avoidable readmissions (Hines et al., 2014). Second, readmissions may be an indication of quality of care. This is motivated by the fact that there is large variation in readmission rates across hospitals even when controlling for patient characteristics such as age and co-morbidities. Furthermore, it is estimated that a large proportion of readmissions are caused by potentially avoidable events such as improper administration of medication at home or complications with surgical procedures. Third, readmissions can be a major source of patient and family stress and may contribute substantially to loss of functional ability, particularly in older patients (Horwitz et al., 2011).

Given these properties of readmission rates, hospital-level readmission rates now often feature in healthcare policies in several ways. First, it is increasingly common to use readmissions as a measure of provider quality. For example, a key quality metric on the Centers for Medicare & Medicaid Services' (CMS) *Hospital Compare* website is 30-day readmission rates for several diagnoses. Second, recent payment reforms incentivise hospitals to improve quality by tying payments to readmission outcomes. An example of such payment reforms is the Hospital Readmissions Reduction Program, implemented under the Patient Protection and Affordable Care Act (2010) (ACA), that financially penalises hospitals for above-average 30-day readmission rates in several diagnoses domains for patients with Traditional Medicare insurance.

The healthcare chapters of this thesis focus on the hospital sector in the US and use data from New York State (NYS). In these chapters, focus is placed on hospital readmission rates as both a measure of hospital quality and health outcome. Relative to national standards, NYS has high healthcare spending per capita at roughly $12,000 per annum (Centers for Medicare &

Medicaid Services, 2017c), 20% higher than the national average. NYS has also been one of the worst-performing states in terms of hospital readmissions (Kaiser Health News, 2014).

The ***first chapter*** exploits a recent payment reform implemented under the ACA, the Hospital Readmissions Reduction Program (HRRP), to study how hospitals respond to asymmetric incentives. In many healthcare systems, multiple actors (e.g. governments and private insurance companies) pay providers, leading them to face a mix of incentives if payment schemes are misaligned. Understanding how providers respond to mixed incentives is important for improving the design of payment systems.

The HRRP financially penalises hospitals for above-average readmission rates among Traditional Medicare patients admitted for specific conditions (heart attack, heart failure, and pneumonia). Readmission outcomes of patients insured by other policies do not affect hospital performance metrics that are linked to payments under the HRRP. This feature of the program thereby creates asymmetric incentives for hospitals to improve quality across different patient groups. In particular, the program implies that hospitals' incentive to provide quality care depends on the following features: whom the patient is insured by, what condition the patient is admitted for, and whether the patient was admitted before or after the program was implemented. Using these features of the program, I study the following questions: (1) *What is the impact of these financial incentives on health outcomes and care quality provided to targeted (TM) patients?* (2) *What is the impact of this program on patients who are not directly targeted by the program?*

To answer these questions, this chapter begins by outlining a model that generates quality spillovers across insurance groups via complementarities in a hospital's cost function. Using a matched difference-in-differences strategy and exploiting within-hospital variation in incentives to reduce readmissions, I estimate the effects of the program on targeted TM and non-targeted non-TM patients (defined here as Medicare Advantage and privately-insured individuals) treated at the same hospitals in NYS.

The results indicate that hospitals that expect to be penalised by the program respond strongly by reducing readmission rates among targeted TM patients admitted for diagnoses monitored by the program. Among these hospitals, readmission rates among non-targeted patients admitted for some

monitored diagnoses also fell, indicating that there are cost complementarities in quality provision. For example, readmission rates fell by over 3 percentage points (pp) and 2.5 pp respectively among targeted and non-targeted patients admitted for heart attack.

Overall, the results of this chapter shed light on the nature of quality provision within hospitals. They suggest that the types of process adjustments incurred by hospitals to reduce readmissions are difficult to direct toward particular patients and not others, generating spillovers across patient groups. Given the policy goal of containing healthcare costs while improving healthcare quality, understanding any potential sources of cost savings from complementaries could be useful.

Despite the popularity of readmissions as a quality metric and the effectiveness of recent policies targeting this metric, the focus on readmissions has been controversial for several reasons. First, it is unclear whether hospitals are the appropriate entity to be held accountable (Joynt and Jha, 2012). Factors outside the hospital's control are likely to matter for readmissions. Patient behaviour at home after discharge (e.g. adherence to medication) is likely to affect the patient's likelihood of being hospitalised again. Patient characteristics such as socio-economic status are strongly correlated with readmission probabilities. Current risk-adjustment methods do not account for how socio-economic factors influence readmissions; this means that hospitals that receive a high proportion of low socio-economic status patients (e.g. "safety-net" hospitals) will appear to be low quality according to the readmission performance metric. Second, unlike mortality, the link between readmissions and patient welfare is unclear. Reduction in readmissions are not necessarily welfare enhancing if they imply that patients are denied the post-discharge care that they would derive utility from receiving.

The ***second chapter*** of this thesis considers these criticisms by asking whether there is evidence that demand-side factors might also impact post-discharge outcomes. In particular, I ask: *What is the impact of demand-side responses to eligibility for public insurance on readmissions?* There are at least two ways through which insurance eligibility might affect post-discharge outcomes such as readmissions. The first way is via patient-side factors. On the one hand, insurance lowers out-of-pocket costs and encourages the utilisation of healthcare facilities after a hospital discharge. On the other hand, insurance may increase access to affordable medication after discharge, which

could lower the risk of readmissions. The second way is via provider-side factors. As the first chapter shows, different insurance policies may reimburse hospitals in different ways for both the initial stay and any subsequent readmission. Hospitals may therefore be incentivised to tailor treatment choices during the patient's initial stay based on their insurer, which may in turn affect post-discharge outcomes.

Understanding the impact of health insurance on readmissions therefore requires variation in insurance coverage and a feasible way to measure treatment quality. The empirical analysis of this chapter exploits two institutional and medical features to achieve this. Regarding insurance coverage, I use the age-65 eligibility threshold for Medicare as a source of variation in insurance coverage. Regarding treatment quality, I focus on a group of patients admitted for heart attack and use the fact that heart attack treatments are broadly characterised by two competing management approaches. The first is an invasive approach marked by cardiac catheterisation shortly after admission followed by surgery; the second is a conservative approach in which the patient is given drugs and monitored. It has been documented that patients who receive the intensive treatment early tend to have better health outcomes (e.g. Bhatt et al., 2004; Currie et al., 2016). Drawing on these features, I estimate regression discontinuity models for 30-day readmissions and treatment quality.

The results indicate that among male heart attack patients, eligibility for Medicare insurance significantly increases the probability of being readmitted within 30 days of discharge by over 2.5 pp. This increase is driven by patients returning for non-major vascular events such chest pain. Since insurance is likely to have a stronger effect on patients returning with mild conditions compared to those with critical conditions, the heterogeneous impacts across reasons for readmissions support the interpretation that this increase captures the effect of insurance. There are no discontinuities in observed treatment intensity at the age threshold. Insofar as observable treatment procedures primarily capture supply-side responses, this finding suggests that demand-side factors also affect the probability of readmission. For female heart attack patients, in contrast, there is no evidence that eligibility for Medicare insurance increases the probability of being readmitted within 30 days of discharge. There is some evidence of changes in treatment intensity, which may offset any demand-side impact of eligibility for Medicare insurance on readmission rates for female patients.

Overall, by highlighting the potential role of patient-side factors in shaping post-discharge outcomes, this chapter sheds light on a potentially important determinant of readmissions that has so far been overlooked in the literature and in policy debates.

Chapter on Education

The third chapter of this thesis turns to the subject of education in the West Bank. By serving as important stepping stones for good future labour market outcomes, educational achievement has important implications for individual welfare. Several studies have shown that investments in schooling yield higher marginal returns in less developed countries than in developed ones (Psacharopoulos, 1972; Psacharopoulos and Patrinos, 2018). Unfortunately, students attending schools in less developed countries often do so under conditions of conflict. Exposure to various aspects of conflict might affect their educational attainment and performance.

Existing literature has shown that conflict-related violence (e.g. fatalities) has a detrimental impact on educational attainment and performance. Another prominent but often overlooked feature of contemporary conflicts is infrastructures such as walls, barriers, and buffer zones. These conflict infrastructures often represent a source of security for those constructing these infrastructures while they represent a source of mobility restrictions for those living in close proximity. Recent examples of such conflict infrastructures include the separation wall in Jerusalem, the Berlin Wall, and peace walls in Northern Ireland. Despite the important role played by conflict infrastructure in contested areas, little is known about their impact on the educational performance of students.

In the ***third chapter*** of this thesis, co-authored with Sami Miaari, we ask: *What is the impact of mobility restrictions on the educational performance of high-school students?* We study this question in the context of the West Bank during the Israeli-Palestinian (IP) conflict between 2000 and 2006, a period of heightened conflict known as the Second Intifada. A prominent feature of the IP conflict is the system of mobility restrictions enforced in the West Bank through various physical barriers. The most prominent of these barriers are checkpoints which Palestinians often cross to commute between Palestinian villages and between Israel and the West Bank. Although

the experience of crossing a checkpoint depends on many factors such as when and where one is crossing, the crossing process usually involves showing identification cards to security personnel. Delays incurred at checkpoints over this period ranged from 15 minutes to 3 hours, and depended on various factors such as the traffic volumes at checkpoints, the number of personnel at the checkpoint, and even the mood of those guarding the checkpoint.

To answer this question, we combine several novel data sources that allow us to measure educational performance and multiple dimensions of conflict, including exposure to violence as well as exposure to physical barriers that restrict mobility. Using information on the home and school location of a student, we construct measures that capture whether that student crosses a checkpoint on the shortest route to school in the year that they sit their final high school exam (the *Tawjihi General Examination* or TGE). Using within school-location and within home-location variation in exposure to violence and mobility restrictions, we estimate the impact of encountering a checkpoint on the road to school on performance in the TGE. We find that having to cross a checkpoint on the road to school reduces the probability of passing the final exam by 3 pp and the overall test score by 0.07 standard deviations. The magnitudes of these effects are larger than our estimates of the impact of exposure to an additional fatality in the home or school location. For example, the estimated impact of an additional fatality in the school locality reduces the probability of passing by 0.07 pp and the overall score by 0.0012 standard deviations. The estimated effects are also comparable to those of other inputs into the education production function, such as teacher quality and school management quality, examined in the existing literature.

We provide evidence that the following three mechanisms are important for explaining the negative relationship between academic performance and exposure to mobility restrictions. First, students who cross checkpoints on the road to school lose time in the morning and early evening, reducing available study time. Second, exposure to mobility restrictions has adverse impacts on the psychological wellbeing of students which may leave them in a worse state to learn and study. Lastly, there is also evidence that barriers surrounding the school area have negative impacts on the physical resources available at school. In particular, checkpoints within 10 km of the school locality center reduce the probability of schools having a library, science lab, or computer lab.

Overall, the results of this chapter suggest that multiple dimensions of conflict have detrimental impacts on educational performance. To the best of our knowledge, this is the first analysis of the impact of mobility restrictions on educational performance. These results suggest that safeguards should be taken to minimise exposure of students to physical barriers, for example by reaching a consensus that barriers should not be built near school areas.

Chapter I

Complementarities in Quality Provision within Hospitals: Evidence from a Pay-for-Performance Program

Chapter Abstract

This chapter provides evidence of cost complementarities in quality provision within hospitals. I study the impacts of the Hospital Readmissions Reduction Program (HRRP), a pay-for-performance scheme that financially penalises hospitals for above-average readmission rates among Traditional Medicare (TM) patients but not among patients insured by other insurance policies, thereby creating asymmetric incentives to provide quality across different insurance groups. I develop a model that generates quality spillovers across different insurance groups via cost complementarities in quality provision. Using a matched difference-in-differences strategy and exploiting within-hospital variation in incentives, I estimate the effect of the program on targeted TM and non-targeted non-TM patients treated at the same hospitals in New York State. Hospitals that expect to be penalised by the program respond strongly by reducing readmission rates among targeted TM patients admitted for diagnoses monitored by the program. Among these hospitals, readmission rates among non-targeted patients admitted for some monitored diagnoses also fall, indicating that there are cost complementarities in quality provision. For example, readmission rates fell by over 3 percentage points (pp) and 2.5 pp respectively among targeted and non-targeted patients admitted for heart attack. There is some evidence that these reduction are driven by increases in treatment intensity and little evidence that they are driven by hospitals finding ways to game the system.

Chapter Notation

- h = Hospital
- d = Diagnosis
- $j = \{m, n\}$ = Insurance type
- κ = Hospital efficiency parameter
- q = Quality
- Π = Revenue; π = Profits
- $\phi(\cdot)$ = Hospital's costs
- τ = Penalty
- x = Number of patients
- $V(\cdot)$ = Hospital utility function

1 | Introduction

Many policies target certain subgroups and not others, creating asymmetric incentives for providers to exert effort across subgroups. For example, healthcare policies are often implemented at the insurance-group level and target patients with particular forms of that insurance coverage.[1] Education policies often target under-privileged children in schools.[2] A concern with these subgroup-specific policies is that since providers such as hospital and schools often serve multiple groups at once, asymmetric incentives across groups may cause providers to shift resources from the non-targeted to the targeted groups in order to meet performance targets. Theory suggests that the impact of asymmetric incentives on the targeted and non-targeted groups depends on the underlying production technology of the output under consideration. For example, applying the logic of standard multi-tasking models (e.g. Holmstrom and Milgrom, 1991), the impact of asymmetric incentives across subgroups depends on whether the tasks of serving multiple groups are substitutes or complements. The effect of such policies is therefore an empirical question.

This goal of this chapter is to examine the impacts of a policy that creates asymmetric incentives across subgroups in a healthcare setting. I use the Hospital Readmissions Reduction Program (HRRP), a pay-for-performance scheme implemented under the Affordable Care Act in 2010, as a source of variation in the financial incentives that hospitals face to provide quality. This program reduces payments to hospitals if their 30-day readmission rate among Traditional Medicare (TM) patients admitted for any one of several conditions (heart attack, heart failure, and pneumonia) is above the corresponding national average. Since this program measures hospital quality based on readmission outcomes of TM patients admitted for particular conditions, it only explicitly incentivises hospitals to increase quality of care among this group of patients. The design of this program and the absence of similar changes affecting other insurance groups who may be treated at the same hospital (e.g. privately insured) for the same condition offers a natural setting to test the impact of asymmetric incentives on the outcomes of targeted and non-targeted patients.

[1] For example, the Center for Medicare and Medicaid Services (CMS) implements policies that target patients covered by their insurance schemes.

[2] For example, under the *No Child Left Behind* Act, districts and schools must break down test scores by subgroups to determine whether all students are making adequate yearly progress. Groups are usually categorised based on socio-economic criteria.

Methodology. I begin by developing a model in which hospitals treat patients with different insurance policies and payments for hospital services differ by insurance groups. Following the program, the model uses an outcome-based definition of quality: the proportion of patients not readmitted within 30 days. Hospitals differ in an efficiency parameter which generates variation in pre-policy quality levels. The model allows for cost complementarities in quality provision across insurance groups: increasing quality for one insurance group may reduce the marginal cost of increasing quality for other insurance groups. One interpretation of these cost complementarities is that some quality-enhancing activities (e.g. improving ward hygiene) cannot be assigned to or narrowly targeted towards particular patients. Thus, improvements for one group may benefit others treated at the same hospital, generating spillovers.

This model delivers two main predictions about how hospitals will respond to the pay-for-performance component in the Medicare payment scheme. First, hospitals will respond by improving quality for TM (i.e. targeted) patients if they expect that their pre-policy readmission rate would be above the national average and thus they would face a penalty. Second, the impact of the program on non-targeted patients depends on the degree of cost complementarities in quality provision between different insurance groups. In particular, quality will increase for non-targeted patients if quality provision across insurance sectors are complements while quality will decrease if they are substitutes.

To test the predictions of the model, I combine matching and difference-in-differences techniques to estimate the impacts of the program on targeted and non-targeted patients treated at hospitals in New York State. Since the penalties are based on hospital-level readmission rates among TM patients admitted for specific conditions, the program creates variation in financial incentives based on the following factors: (i) the diagnosis the patient is admitted for (monitored or not), (ii) the period of admission (pre- or post-policy), (iii) the insurance policy of the patient (targeted, TM or non-targeted, non-TM), (iv) the hospital's pre-policy quality which determines its expected penalty. Drawing on the program design, I match patients admitted for monitored conditions with patients admitted for a subset of non-monitored conditions (gastrointestinal) at the same hospital to form a comparison patient group. I further match on observable characteristics that may influence readmission outcomes (age bins, gender, minority status), their admission period (pre-

and post-policy), and primary expected payer. I then exploit within-hospital and across-diagnosis and across-time variation in financial incentives to estimate the impacts of the policy. I provide evidence supporting the assumptions required for this strategy to identify the causal impacts of the program. For example, within insurance and hospital-performance groups, readmission trends of monitored and non-monitored diagnoses appear parallel before the introduction of the policy.

Main results. In line with the first prediction of the model, I find strong evidence that under-performing hospitals respond by reducing readmission rates among TM patients. Patients treated at hospitals that expect to be penalised see large and statistically significant reductions in readmission probabilities for monitored conditions. These responses vary by diagnoses. Among heart attack TM patients, readmission rates fell by 3.36 percentage points (pp) relative to a pre-policy mean of 22.88% (a 12.6% reduction in the number readmitted). Among heart failure TM patients, the program reduced readmission rates by 0.73 pp, relative to a pre-policy rate of 28.28% (a 1.98% reduction). Among pneumonia TM patients, the program reduced readmission rates by 0.94 pp, relative to a pre-policy rate of 21.21% (a 4.4% reduction).

Regarding the second prediction of the model, the results indicate that the asymmetric incentives generated by the program did not lead hospitals to divert resources away from non-targeted patients. In fact, the program also had beneficial effects on non-targeted heart attack patients treated at hospitals that expect to be penalised. Specifically, among heart attack non-TM patients, the program reduced readmission rates by 2.38 pp, relative to a pre-policy mean of 16.08% (a 14.8% reduction). Among heart failure non-TM patients, the program reduced readmission rates by 0.98 pp relative to a baseline of 24.08%. Among pneumonia non-TM patients, the program did not have significant impacts on readmission rates. Since the readmission time trends for non-monitored conditions appear fairly flat pre- and post-policy, this suggests that these results are not driven by hospitals substituting resources into the treatment of monitored conditions and away from the treatment of non-monitored conditions.

The reductions in readmissions could be driven by hospitals improving quality of treatment or by hospitals changing their readmission strategies to avoid penalties. To explore the mechanisms driving the reduction, I first examine the effects of the program on several measures of medical

treatment intensity. In line with existing studies, I find that the program had no significant impact on the cost of the initial hospitalisation. This is consistent with the idea that some quality-enhancing activities may not be assignable to particular patients and therefore not reflected in billed charges of these hospitalisations. There is some evidence that hospitals responded by increasing the number of treatment procedures provided to TM patients admitted for heart attack and heart failure. For example, the average number of treatment procedures received by heart attack TM patients increased by 5% (from a baseline of 5 procedures); the number of treatment procedures for heart attack non-TM patients did not change. Lastly, the program reduced length of stay for heart attack cases, increased it for heart failure cases, and had no impact on length of stay of pneumonia patients.

Second, I examine the effects of the program on hospitals' readmission strategies. There is little evidence that hospitals are "gaming" the system by delaying readmissions outside the 30-day window. There is also no evidence that the program increased the probability that a patient returns to the emergency department (ED) within 30 days but is not readmitted; thus the lower rates of readmissions do not appear to be showing up as ED returns. Lastly, there is no evidence that hospitals are increasing the number of transfers so that readmissions are counted against the hospital that the patient is transferred to (rather than the hospital the patient is transferred from). The limited evidence of gaming suggests that reductions in readmissions reflect improvements in quality.

Related Literature and Contribution. This chapter is related to several strands of literature. First, it is related to a recent literature that examines the early impact of the HRRP. Generally, these studies find that the HRRP lowered 30-day readmissions among TM beneficiaries, the targeted patient group, admitted for monitored conditions. These studies use various empirical strategies to estimate the effects of the program. The standard empirical strategy is a difference-in-differences strategy comparing changes in readmission rates over time between monitored and non-monitored conditions (Carey and Lin, 2015; Zuckerman et al., 2016). Other studies also exploit the fact that penalties depend on past readmission rates (Gupta, 2017; Mellor et al., 2017): hospitals with readmissions above the national average have a greater incentive to reduce readmissions compared

to those below.

A shortcoming of these existing analyses of the HRRP is that they neglect the potential effects of the program on the non-targeted patient group. On the one hand, studies that use Medicare claim files (Gupta, 2017; Zuckerman et al., 2016) by default cannot test the impact of the policy on the non-targeted patients (admitted for monitored conditions) since the outcomes for these patients are not observed. On the other hand, studies that use all-payer databases often assume that the impact is restricted to the targeted group. Using a similar dataset as this chapter, one specification in Carey and Lin (2015) uses non-TM patients as a control group and finds no effects of the policy. If there are (positive) spillover effects, estimated changes for TM and non-TM patients would tend to cancel out and thus this specification would yield no significant estimated treatment effect.

This chapter contributes to this literature by allowing for and estimating spillover effects of the HRRP to non-targeted groups, for example through cost complementarities. In this respect, this chapter is most similar to Chen and Grabowski (2017) who test for the unintended consequences of the HRRP among Medicare advantage and privately insured patients and Batt et al. (2018) who examine the effect of the program on non-targeted patients and non-monitored conditions. In contrast to Chen and Grabowski (2017) and Batt et al. (2018), this analysis uses within-hospital variation and explores how responses vary across hospitals based on pre-policy performance (which in turn determine the hospitals' exposure to the program penalties). Ziedan (2018) and Batt et al. (2018) also explore spillovers across diagnoses.

Second, this chapter is related to a large theoretical literature dating back to Holmstrom and Milgrom (1991) that studies how contracts affect effort exerted in multitasking environments and what this implies about the optimal contract in such environments. The empirical literature on the impact of payment schemes on health outcomes and provider behaviour in healthcare settings is small. Dumont et al. (2008) find a trade-off between quality and quantity of care. In particular, Canadian physicians who voluntarily signed up to contracts that paid less for a specific quantity of consultations increased quality, as measured by the average time per consultation and other activities (such as teaching) that are not remunerated. Feng Lu (2012) studies the impact of a mandatory quality disclosure policy and finds evidence of cost substitutability: nursing homes

improved quality on measures that are targeted but deteriorated in regards to non-targeted ones. Rodríguez-Lesmes and Vera-Hernández (2018) use the Quality and Outcome Framework in the UK, the largest primary-care pay-for-performance programme worldwide (Roland and Olesen, 2016), to empirically test whether rewarded tasks are cost complements or substitutes. Consistent with the findings of this chapter, they find evidence of complementarities in the clinical areas of cardiovascular disease.

This chapter contributes to this literature by providing new evidence on the impact of a large nationwide pay-for-performance program which created asymmetric incentives across insurance groups. I provide new evidence that the asymmetric incentives across groups did not disadvantage the group with "low-powered" incentives. In fact, non-targeted patients treated at hospitals that improved quality for targeted patients also saw reductions in readmissions.

Third, this chapter is related to a general literature on spillover effects. The general term "spillover" is an umbrella term for any effect (usually, of a program) that impacts economic units other than the treated.[3] In the healthcare literature, the term "spillover" usually refers to how changes in one insurance sector may affect outcomes in another insurance sector (see Chernew et al. (2010) for a review). Two types of affected outcomes are commonly analysed. First, utilisation spillovers examine whether better management of care among privately-insured patients results in lower costs/different utilisation patterns for other patients who also benefit from better management techniques. Second, price spillovers (also known as "cost shifting") examine whether changes in public reimbursement rates influence prices among privately-insured patients.[4]

This chapter contributes to this literature by providing new evidence that a program targeting the public-insurance sector affects outcomes in the private-insurance sector. In contrast, most papers in this literature look at effects that work in the opposite direction. Furthermore, the institutional features of the program considered in this chapter allow for a research design that overcomes some of the selection issues faced by previous studies. For example, studies that analyse

[3] Spillover effects include externalities (e.g. Crépon et al., 2013), general equilibrium effects (e.g. Crépon et al., 2013), and peer effects (e.g. Neidell and Waldfogel, 2010).

[4] Darden et al. (2018) test whether the HRRP causes hospitals to adjust prices charged to private insurers in response to potential cuts in public funding. They find that hospitals that faced net payment reductions as a result of the HRRP negotiated higher average private payments.

the impact of private health plan penetration have to deal with patient selection into insurance plans: changes in the costs of publicly-insured patients when the share of private health-insurance plans increases reflect both changes in expenditure per patient and changes in the mix of patients left in the publicly-insured patient pool. In my setting, the changes in reimbursement rates generated by the HRRP are arguably more exogenous than private health plan penetration.

The rest of this chapter is structured as follows. In Section 2, I describe the institutional details relevant to the analysis. In Section 3, I describe the model and explain its key predictions. In Section 4, I outline the empirical strategy used to estimate the impacts of the program. In Section 5, I describe the main data sources, sample selection procedure, and provide descriptive statistics for the main sample. In Section 6, I present the main results. In Section 7, I discuss the robustness of the baseline results. Finally, Section 8 concludes.

2 | Background

Nationally, the level and growth of spending on Medicare and particularly amongst Traditional Medicare beneficiaries has been a persistent policy concern.[5] This issue is a particular concern in New York State (NYS), the setting of this chapter. Relative to national standards, NYS has high Medicare spending per enrollee at over $12,000 per annum (Centers for Medicare & Medicaid Services, 2017c). It is increasingly recognised that expenditure on post-discharge care such as readmissions is contributing to healthcare costs. Over the sample period (2008–2014), NYS has been among the five worst performing states in terms of all-cause hospital readmissions (Kaiser Health News, 2014).

The Patient Protection and Affordable Care Act (ACA), implemented in March 2010, introduced three pay-for-performance programs that aim to improve health outcomes while reducing healthcare costs. Each program targets different quality metrics. This chapter focuses on the Hospital Readmissions Reduction Program which uses readmission rates as a performance and quality measure. This program is directed towards TM patients in that only outcomes of TM patients affect the performance measure. Over the same period, no other main insurance provider implemented similar payment reforms targeting readmissions.

Under the Hospital Readmissions Reduction Program (HRRP), a hospital h is penalised in year t if its performance in years $t-4$ to $t-2$ is worse than the national average hospital with the same observable patient mix (i.e. the demographics of the patient population based on factors such as disease severity and socioeconomic characteristics).[6] Performance is measured by risk-standardised 30-day readmission rates among TM patients admitted for diagnoses monitored by the program. In the first round of penalties, October 2012 to September 2013, three conditions were monitored: acute myocardial infarction (AMI, commonly known as heart attacks), heart failure (HF), and pneumonia (PN). The CMS revises the policy continuously, adding new diagnoses and procedures

[5] As described in the introductory chapter, Medicare is the main federal public health insurance program in the United States. Nationally, around 75% of Medicare beneficiaries are enrolled in Traditional Medicare while the remaining 25% are enrolled in Medicare Advantage.

[6] Precisely, the program defines the penalty year t as October t to September $t+1$ and the penalties in that year are based on performance from July $t-4$ to June $t-1$.

to the list of monitored conditions.[7] Intuitively, a hospital is financially penalised if it performs worse than how the CMS expects it to perform given the types of patients it treats.

The penalties in year t are imposed on the total Medicare payments from admissions that year under Medicare's inpatient prospective payment system (IPPS).[8] The total penalty is $\tau \in [0\%, P^{cap}\%]$ of total Medicare payments in year t where P^{cap} is the maximum penalty cap. The cap started from a base of 1% in 2012 and increased to 3% in 2014. In the first year of the program, 64% (of 3,500) acute care hospitals across the US were penalised and the average penalty was 0.42% of IPPS payments for a total of $290 million (Boccuti and Casillas, 2015). In NYS, over 88% of hospitals were penalised in the first year and the average penalty was 0.59%.[9] In October 2014, when the maximum penalties increased to 3%, total penalties rose to $420 million (Rau, 2015).

To calculate the penalty rate τ faced by hospital h in year t, the CMS calculates two variables based on hospital discharge data from $t-4$ to $t-2$: (i) the risk-adjusted predicted readmission rate for diagnosis d ($\hat{r}_{dh}$) and (ii) the risk-adjusted expected readmission rate ($\tilde{r}_{dh}$). These two variables are computed using a hierarchical logistic regression model which adjusts the hospital raw readmission rate by patient risk factors (e.g. age and co-morbidities).[10] The risk-adjusted predicted readmission rate ($\hat{r}_{dh}$) predicts the readmission rate using estimates from the model, conditional on patient mix remaining constant. The risk-adjusted expected readmission rate ($\tilde{r}_{dh}$) roughly corresponds to the readmission rate of the national average hospital with the same patient mix as the hospital. Hospital h is penalised if the predicted rate is greater than the expected rate in *any* of the monitored diagnoses ($\frac{\hat{r}_{dh}}{\tilde{r}_{dh}} > 1$); there is no penalty or bonus payment otherwise. The penalty for diagnosis d is defined as $\max\left\{\frac{\hat{r}_{dh}}{\tilde{r}_{dh}} - 1, 0\right\}$ multiplied by the fraction of all Medicare revenue that comes from Medicare patients admitted for diagnosis d. To arrive at the total penalty rate, the CMS then aggregates penalties for excess readmissions across all monitored conditions. The final penalty rate is the lower of (i) the sum of the penalties across monitored diagnoses and (ii) $P^{cap}\%$ of total Medicare payments from all conditions. Appendix A.1 contains more details about

[7] E.g. Beginning in October 2014, total hip or knee replacement (THA/KA) and chronic obstructive pulmonary disease (COPD) were added to the list of monitored procedures.

[8] Under the IPPS, hospitals are reimbursed for their operating costs for Traditional Medicare (TM) patients based on prospectively set rates. Each diagnosis related group (DRG) has a payment weight assigned to it, based on the average resources used to treat Medicare patients in that DRG.

[9] Own calculations based on CMS data described in Section 5.

[10] See Appendix A.1 for further details of the CMS risk adjustment model.

the penalty formula.

Figure 1.1 illustrates the penalties faced by New York State hospitals in the first year of the program (October 2012–September 2013) as a function of their excess readmission ratio for AMI cases (calculated using discharge data from July 2008 to June 2011). Dots represent actual hospitals and the bold line is the penalty schedule. The bold line shows that if the excess readmission ratio is greater than one and the total penalty is below the penalty cap, then the marginal penalty per readmission is constant. The marginal penalty per readmission due to the program only changes based on whether the risk-adjusted readmission rate exceeds the national average (and is below the value that generates the maximum penalty cap), but it does not matter how close or far away the hospital is from the threshold.[11]

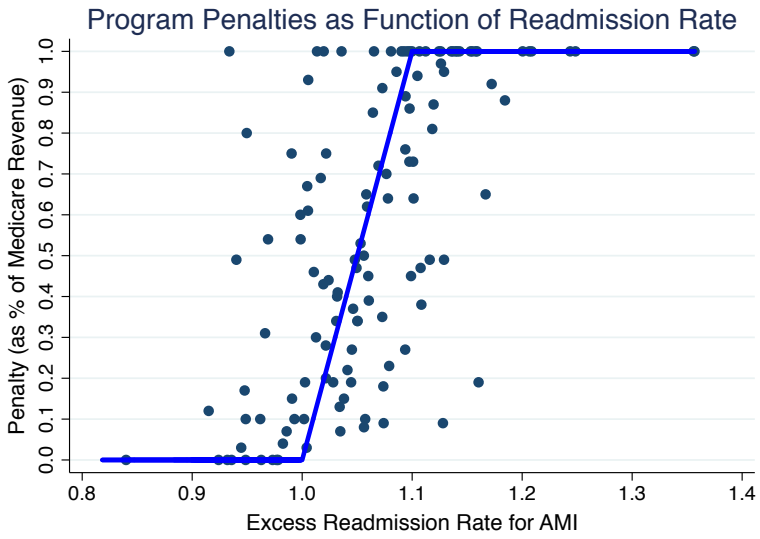


Figure 1.1: Actual HRRP penalties as function of excess readmission rate (for AMI) for hospitals in sample. Dots represent hospitals in New York State, the bold line represents the penalty schedule. Excess readmission rate is the ratio of risk-adjusted predicted to expected readmission rates.

The key features of the program which create variation in the incentives that hospitals face are summarised below and are used in the empirical analysis to identify the impacts of the program. Since there are multiple sources of variation, it is worth clarifying that the term "monitored"

[11] Note that the overall penalty is determined by excess readmissions in three diagnoses while this diagram plots penalties as a function of excess readmissions in one diagnosis only, thus there is quite a lot of dispersion around the bold line. When the penalty rate is plotted as a function of readmission rates in other conditions, the pattern looks similar. For example, hospitals at the penalty cap (penalty rate = 1) with excess readmission rates for AMI below one have excess readmission rates above one for other monitored conditions.

describes whether a *diagnosis* affects the penalties while the term "targeted" describes whether the readmission outcome of a *patient* affects the penalty.

- **Timing.** The HRRP was announced as part of the ACA in March 2010, although rules about the penalty structure were not released until August 2011 by the CMS. Since penalties in year t are based on past performance in $t-4$ to $t-2$, hospitals are expected to react immediately once details about the program are announced. Hospital incentives therefore change after the announcement of the program.
- **Monitored vs. Non-monitored diagnoses.** In the first round of penalties, penalties are based on readmission outcomes of TM patients admitted for three diagnoses: AMI, HF, and PN. Furthermore, since below-average performance in any one of the monitored diagnoses results in a penalty, hospitals cannot compensate for poor performance in one diagnosis by performing better in another. Therefore, hospitals' incentive to provide quality depends on the diagnosis the patient is admitted for.
- **Targeted vs. Non-targeted patients.** The penalties are calculated using the outcomes of TM patients.[12] Outcomes of patients insured by other insurance schemes such as Medicare Advantage and private insurance do not affect the penalty. There are no similar payment reforms for private insurance plans over the same period. Therefore, hospitals' incentives to provide quality depends on the insurance coverage of the patient.
- **Pre-policy quality relative to national average.** Since the penalties are based on past performance relative to the national average, hospitals that have previously performed poorly are expected to respond more to the program than hospitals that performed well. Therefore, a hospital's incentives to alter quality depends on whether its pre-policy readmission rate is above or below the national average.

[12] Legally, Chapter IV Subchapter B of the Social Security Act, under which the HRRP is outlined, applies to the Medicare program only. Furthermore, the CMS website specifies that readmissions rates are calculated using Medicare Claim Files. These claims data do not include claims data for beneficiaries who enroll in Medicare Advantage (Medicare Part C) plans or other privately insured patients. Hospitals are expected to know this.

3 | Model

The following model analyses how hospitals will alter quality for TM and non-TM patients in response to the program. The model is similar in spirit to Zhang et al. (2016); its main novelty is to allow for cost complementarities in quality provision which generates spillovers across insurance groups. The model delivers two predictions which are tested in the empirical analysis. Mathematical details for these predictions can be found in Appendix A.2.

I assume that each hospital is made up of multiple departments specialising in the treatment of different diagnoses.[13] The main decision-maker in the model is a hospital department specialising in a particular monitored diagnosis. Therefore, all variables below are to be understood to be diagnosis- and hospital-specific. This notational simplification is consistent with the empirical strategy used below which estimates the impacts of the policy for each monitored diagnosis separately (AMI/HF/PN) using within-hospital variation in financial incentives.

3.1 Setup

Quality. Quality of care is a multi-faceted notion. Donabedian (1988) proposes a three-component approach to assess quality: (i) structure (attributes of hospitals such as the nurse-to-patient ratio), (ii) process (adherence to well-established clinical guidelines such as giving AMI patients beta-blockers on arrival), and (iii) outcome/clinical (health outcomes of patients such as readmissions). This model uses readmission outcomes as a measure of quality (in line with the program's design) and assumes that the underlying structure- or process-based quality changes that may be important for outcome-based changes are imperfectly observed, for example, by policymakers in charge of setting the hospital payment scheme.

Hospital departments. Each hospital department treats various patients with different insurance policies j. This analysis considers the case where $j \in \{m, n\}$ so that the patient has Traditional Medicare (TM) insurance if $j = m$ and some other type of insurance (e.g. private insurance) if

[13] For example, AMI patients are treated in the AMI-wing while PN patients are treated in the PN-wing.

$j = n$. Given a particular diagnosis, hospital departments differ in the following respects: (i) an exogenous and time-invariant efficiency parameter $\kappa > 0$ which is constant across insurance groups j and (ii) the exogenous number of patients they receive.

Patient flows. The hospital department receives an exogenous number of patients (x_j) for each insurance type j. Let $0 \leq r_j \leq 1$ be the observed all-cause 30-day readmission rate for insurance group j in the hospital department. The quality provided for that insurance group is defined as the proportion of patients who are not readmitted: $q_j := 1 - r_j$. I assume a simple one-period model so that admissions and readmissions occur within the same period. Since the total number of patients for each insurance group equals the number of exogenous patients and patients readmitted to the same hospital, the total number of patients treated (x_j^T) can be written as $x_j^T = x_j + r_j x_j$.[14]

Revenue. Revenues (denoted by upper-case Π) are given by the sum of revenue from admissions and readmissions. Let p_j denote the average payment the hospital department receives for each initial hospitalisation and p_j^r denote the average payment received for a readmission, which may differ from the payment received for the initial admission. Prices are assumed to be set exogenously.[15] The hospital department's revenues can then be written as:

$$\Pi_m + \Pi_n = \sum_{j \in \{m,n\}} (p_j x_j + p_j^r x_j r_j) = \sum_{j \in \{m,n\}} (p_j x_j + p_j^r x_j \times (1 - q_j)) \tag{2}$$

where the second equality follows from the definition of quality as the proportion of patients who are not readmitted.

[14] For simplicity, this model assumes that if a patient is readmitted, they are readmitted to the same hospital and for the same diagnosis. In reality, this need not be the case as a patient who has a repeated hospitalisation may be readmitted at a different hospital and/or for a different condition. The assumption that a patient is readmitted to the same hospital can be relaxed by allowing a hospital-specific readmission rate for insurance group j to be $r_j(1 - \gamma_h)$ where γ_h is the probability that a patient who is readmitted goes to another hospital. The assumption that the patient is readmitted for the same diagnosis simplifies the analysis by ensuring that the payment is the same at the initial admission and readmission. This assumption can be relaxed by allowing the readmission payment to vary by hospital episode for the patient, depending on which diagnosis group they are readmitted for.

[15] For administratively set Medicare payments, this assumption is natural. For private payments which are usually determined by bargaining between providers and insurers, this assumption is natural if payment determination occurs after quality is determined (as is standard in existing models, see Gaynor et al. (2015).)

Costs. Let the function $\phi(q_m, q_n; x_m, x_n, \kappa)$ denote costs faced by the hospital department. This function returns the cost of treating a given number of TM and non-TM patients at the specified quality levels for a given diagnosis. This function depends on the quality provided q, the number of patients x, and the hospital efficiency parameter κ. I assume that the cost function has the following properties. First, it is continuous and increasing in q_j and x_j and at least twice-differentiable with respect to q_j for all hospitals departments and where $j := \{m, n\}$. Second, following Rodríguez-Lesmes and Vera-Hernández (2018), I assume that $\phi(q_m, q_n; x_m, x_n, \kappa) = \frac{1}{\kappa}\phi(q_m, q_n; x_m, x_n)$. This means each hospital department's cost is equal to the cost of a reference hospital department with $\kappa = 1$, $\phi(q_m, q_n; x_m, x_n)$, weighted by the inverse of the hospital department efficiency parameter κ. Less efficient hospitals departments (lower κ) therefore have higher quality costs for a given quality and output level.

Spillovers across insurance groups may arise in this model via the costs function if there are cost complementarities in quality provision. More formally, if $\partial^2\phi/\partial q_m \partial q_n < 0$ then there are cost complementarities; if $\partial^2\phi/\partial q_m \partial q_n > 0$ then there is cost substitutability; and if $\partial^2\phi/\partial q_m \partial q_n = 0$ then there is neither.[16]

Cost complementarities in quality provision may reflect various mechanisms at play.[17] One example is that the introduction of a targeted incentive program like the HRRP may lead to changes in treatment choices for both TM and non-TM patients by inducing physicians to adopt new practice procedures that result in lower readmissions. Once a physician has gained the know-how of how to provide better quality of care for TM patients, the marginal cost of also improving quality of care for non-TM patients may be lower. Another example is that the policy may lead hospitals to invest in better staff. Hospitals may decide to hire more experienced physicians to treat patients admitted for monitored diagnoses. Since physicians are likely to treat both TM and non-TM patient groups, both groups would benefit from the expertise of these new recruits.

These examples suggest that cost complementarities shed light on the nature of the quality-enhancing activity. In particular, the existence of cost complementaries suggest that in this context

[16] An example of a cost function with cost complementarities is $\phi(q_m, q_n) = c_m q_m^2 + c_n q_n^2 - \omega q_m q_n$ where $\omega > 0$.

[17] Cost complementarities are sufficient but not necessary for economies of scale in quality provision. Economies of scale may arise from two sources: complementarities or fixed costs (Butler, 1995; Clark and Huckman, 2012). This chapter focuses on complementarities.

quality-enhancing activities cannot be assigned to or be narrowly targeted towards particular patients. Hospital services that have this unassignable nature such as the quality of staff, medical expertise, and ward hygiene fall under this category. For such activities, improvements for one patient group would benefit other patient groups, generating spillovers.

Note that some types of these unassignable activities may be unverifiable (e.g. attitude of nurses) while others might be verifiable (e.g. ward hygiene). Therefore, some of the costs associated with unassignable activities may not be captured in the billed charges of a particular hospitalisation. This is consistent with recent studies of the impacts of the HHRP showing little evidence of increased billed charges or other measures of treatment intensity. For example, Gupta (2017) and Mellor et al. (2017) both find that readmission rates fell but find no evidence of increased charges, length of stay, or treatment intensity.

Modelling the HRRP. To translate the program structure into the framework of the model, I make the following simplifications. First, given the static one-period setup of the model, I assume that performance in period t affects the penalty in period t. Although the actual penalty is dynamic in the sense that readmissions today affect profits tomorrow, the fact that the penalties are based on three years of past data implies that within a period that spans three years or more, performance at the start of the period affects profits towards the end of the period. For example, in the analysis below, the "post-policy" period is September 2011 to December 2014, so quality chosen at the start of this period (2011) feeds into penalties applied from October 2013 onwards. Therefore, this simplification still allows one to capture the main features of the program.

Second, following Zhang et al. (2016), I assume that quality improvement decisions are made at the hospital department level. Each hospital department treating patients with monitored diagnoses acts by assigning a penalty cap to their monitored diagnosis to meet the overall hospital-level target. This is a reasonable assumption given that physicians with different specialities tend to work in different departments and thus decisions are likely to be made "locally" at the diagnosis level. The structure of the program implies that the penalty rate for a hospital department treating a

monitored diagnosis can be written as follows (see Appendix A.1 for derivation):

$$\tau_m = \min\left\{ \max\left\{ \frac{r_m}{\tilde{r}_m} - 1, 0 \right\}, \frac{P^{cap}}{\psi_m} \right\} \tag{3}$$

where P^{cap} is the penalty cap assigned by the policy, ψ_m is the ratio of TM revenues from all monitored diagnosis to TM revenues across all diagnoses (monitored and non-monitored), and other terms are defined as before. Since there is no heterogeneity between patients in the model, there is no distinction between observed and risk-adjusted readmission rates. Thus, the numerator in this penalty formula refers to observed rates for simplicity whereas the actual CMS penalty uses risk-adjusted rates in both the numerator and denominator (see Section 2 of this chapter).

To see how equation (3) can be used to calculate penalties for a particular monitored diagnosis, suppose that a hospital receives \$1 million from Medicare overall. Of the total revenue, \$50,000 is from Medicare AMI patients (with an excess readmissions ratio of 1.03), \$90,000 is from Medicare HF patients (with an excess readmissions ratio of 1.07), \$60,000 is from Medicare PN patients(with an excess readmissions ratio of 0.7) and the remaining is from other (non-monitored) conditions. The penalty cap is 1%. In this case, the penalty rate for the AMI department is $\min\left\{0.03, \frac{0.01}{0.2}\right\} = 0.03$; the penalty rate for the HF department is $\min\left\{0.07, \frac{0.01}{0.2}\right\} = 0.05$, and the penalty rate for the PN department is 0. To calculate the penalty in dollar terms for a particular monitored diagnosis, one multiplies the revenue for that monitored diagnosis by the penalty rate for that diagnosis: $\Pi_m \times \tau_m$. This gives penalties of \$1,500, \$4,500, and \$0 for AMI, HF, and PN in this example.

Third, since the actual penalty faced by a hospital is a function of their own performance relative to other hospitals, hospitals need to form expectations about their penalty. I assume that when forming expectations about τ, the hospital sets the expected readmission rate ($\tilde{r}_m$, the denominator of the penalty formula) equal to some exogenously given rate, such as the national risk-adjusted average readmission rates published in CMS's reports or website.

Hospital department's objective. Putting all these elements together, the hospital department's problem can be described as follows. At the start of the period, hospital department

chooses quality (q_m, q_n) to maximise payoffs. Following the existing literature on the behaviour of not-for-profit hospitals (McGuire, 1985; Newhouse, 1970; Sloan, 2000), I allow hospital payoff to be an increasing function of quality q and profits π, given the number of patients x and efficiency parameter κ. The hospital department's problem can be written as

$$\max_{q_m, q_n} \quad V(q_m, q_n, \pi_m + \pi_n; x_m, x_n, \kappa) \tag{4}$$

where profits are revenues minus the penalty imposed and quality costs:

$$\pi_m + \pi_n = \sum_{j \in \{m,n\}} \left[\left(p_j x_j + p_j^r r_j x_j \right) \times \left(1 - \tau_j \right) \right] - \frac{1}{\kappa} \phi(q_m, q_n; x_m, x_n) \tag{5}$$

I make the following assumptions about the payoff function: (i) profits $\pi := \pi_m + \pi_n$ enter linearly into the payoff function and thus $\frac{\partial V}{\partial \pi} = 1$ and (ii) the payoff function is concave in the quality arguments (q_m, q_n). In equation (5), p_j and p_j^r are payments for admissions and readmissions. τ_j is the penalty applied to insurance group j, where $\tau_m > 0$ for hospitals penalised under the HRRP and $\tau_n = 0$ as there are no penalties for readmissions among non-Medicare patients. Lastly, $\frac{1}{\kappa}\phi(q_m, q_n; x_m, x_n)$ are hospital costs where κ is the hospital department efficiency parameter.

The presence of quality q in the payoff function is often referred to as hospital benevolence or altruism (Chalkley and Malcomson, 1998; Chang and Jacobson, 2011). An alternative interpretation is that hospitals care about their reputation, which may depend on the quality of care provided to patients.[18] Recent empirical studies provide evidence that not-for-profit hospitals consider a mix of quality, output, and profits (Chang and Jacobson, 2011; Deneffe and Masson, 2002; Horwitz and Nichols, 2009). Since over 85% of hospitals in the sample are not-for-profit (discussed further in Table 1.2 of Section 4 of this chapter), it is appropriate to allow these arguments to enter the payoff function.

[18] Inclusion of quality into the payoff function does not imply that the patient's well-being fully enters the hospital's payoff function. This is because the hospital may value quality in a different way than patients.

Model simplifications. In addition to the assumptions mentioned previously, it is worth noting that this model makes several further simplifications. First, it assumes that demand for hospital services is exogenous and does not respond to quality of care. This assumption may be appropriate for urgent conditions such as the ones targeted by the program. For time-critical cases, patients are likely to go to the nearest hospital. However, for non-emergency elective conditions that require inpatient services (e.g. hip replacements), demand may respond to better quality of care. If demand responds to quality, then reduction in readmission rates may result in long term benefits for the hospital via greater future demand.[19] Second, it assumes that there are no strategic interactions between hospitals. Specifically, each hospital takes the expected readmission rate (the "national average") as given and does not anticipate how its quality changes will affect this rate. This assumption is suitable in this context given that each hospital is likely to act independently and the total number of hospitals under the HRRP is large (∼3,500).

3.2 Model Predictions

In the *pre-policy* period, there is no pay-for-performance component for either insurance group ($\tau_m = \tau_n = 0$). At an interior optimum, the hospital department will choose quality levels q_j such that the following first order condition holds:[20]

$$\frac{\partial V}{\partial q_j} = p_j^r x_j + \frac{1}{\kappa}\frac{\partial \phi}{\partial q_j} \tag{6}$$

Equation (6) illustrates the trade-off faced by hospitals when deciding on the level of quality provision. The left hand side is the marginal payoff from increasing quality due to hospital altruism. The right-hand side captures the marginal loss in profit from increasing quality. The marginal loss in profit is composed of two parts. First, the profit loss due to readmission reductions: if hospitals

[19] This may be of relevance in the US setting because, since 2008, CMS publishes hospital performance scores on their *Hospital Compare* website. 30-day readmission rates are published as a measure of quality. The goal of this website is to help consumers make informed decisions about where to go for hospital care.

[20] Given that the hospital incurs some operational costs for readmissions and that payoffs are increasing in q, this guarantees a non-zero level of quality.

are reimbursed for readmissions on a per-case basis, readmissions may account for a non-negligible proportion of the hospital's profit margin. Given equation (2), revenue is decreasing in quality since the hospital loses payments from rehospitalisation cases. Second, profits fall due to the marginal cost of increasing quality captured by the function ϕ. Equation (6) also implies that more efficient hospitals (hospitals with higher κ) have higher optimal q and thus the pre-policy distribution of q is governed by κ.[21]

In the *post-policy* period, $\tau_m \geq 0$ for hospital departments affected by the program. Therefore, such hospital departments will choose quality q_m such that the following first-order condition holds:

$$\frac{\partial V}{\partial q_j} - \frac{\partial \tau_m}{\partial q_m}\left[p_m x_m + p_m^r r_m x_m\right] = p_m^r x_m(1-\tau_m) + \frac{1}{\kappa}\frac{\partial \phi}{\partial q_m} \tag{7}$$

Equation (7) illustrates how the trade-off changes as a result of a pay-for-performance program. The program has two effects. First, the marginal benefit of quality provision increases since this will reduce penalties. Second, the marginal loss in revenue due to fewer readmissions is cushioned by the penalty. The impact of quality improvements on the penalties depends on which part of the penalty schedule the hospital is situated at (see Figure 1.1 in Section 2 for reference), leading to the following prediction.

Prediction 1 (Direct impact of penalties on TM patients)**.** *Hospital departments will respond to the penalties by increasing quality for TM patients admitted for monitored diagnoses if the hospital expects that its pre-policy readmission rate is greater than the national average risk-adjusted readmission rate $\tilde{r}_m$. There will be no change otherwise.*

Prediction 1 explains when a hospital has financial incentives to reduce its readmissions. If the hospital expects that its excess readmission ratio is less than one, then it faces no penalties and has no incentive to change its readmission rates. However, if the hospital expects its excess readmission

[21] Implicitly differentiating equation (6) with respect to κ gives: $\frac{\partial q_j}{\partial \kappa} = -\left[\frac{\partial^2 V}{\partial q_j} - \frac{1}{\kappa}\frac{\partial^2 \phi}{\partial q_j^2}\right]^{-1} \frac{1}{\kappa^2}\frac{\partial \phi}{\partial q_j}$. This is positive since $\frac{\partial^2 V}{\partial q_j} - \frac{1}{\kappa}\frac{\partial^2 \phi}{\partial q_j^2} < 0$ if the second order condition is satisfied.

ratio to be greater than one, then the pre-policy readmission rate cannot be optimal. This is because increasing quality will reduce expected penalties and therefore increase profits. Furthermore, the marginal loss in revenue from fewer readmissions is reduced because of the penalties applied to payments. If the excess readmission ratio is so high that the penalty is binding, increasing quality doesn't reduce expected penalties. However, the marginal loss in revenues from fewer readmissions is lower due to the total payments being reduced by the HRRP.

To examine the impact of the program on non-TM patients, the group of patients indirectly targeted by the program, differentiate equation (6) for $j = n$ with respect to τ_m. The resulting equation illustrates how the indirect impact of the program on non-TM patients depends on the degree of complementarities between q_m and q_n:

$$\frac{\partial q_n/\partial \tau_m}{\partial q_m/\partial \tau_m} = \frac{\partial r_n/\partial \tau_m}{\partial r_m/\partial \tau_m} = -\underbrace{\frac{\partial^2 \phi}{\partial q_m \partial q_n}}_{=:\phi''_{mn}} \left[\frac{\partial^2 \phi}{\partial q_n^2} - \frac{\partial^2 V}{\partial q_n^2}\right]^{-1} \tag{8}$$

Prediction 2 (Indirect impact of penalties on non-TM patients.)**.** *(A.) If the penalties had no effect on TM patients, then the hospital department will not alter quality for non-TM patients. (B.) If the penalties reduced readmissions for TM patients, then the impact of the penalties on readmission rates among non-TM patients admitted for monitored diagnoses depends on the sign of ϕ''_{mn}. In particular, if q_m and q_n are complements ($\phi''_{mn} < 0$), then $\frac{\partial q_n}{\partial \tau_m} > 0$. If q_m and q_n are substitutes ($\phi''_{mn} > 0$), then $\frac{\partial q_n}{\partial \tau_m} < 0$. If there are no dependencies between q_m and q_n ($\phi''_{mn} = 0$), then $\frac{\partial q_n}{\partial \tau_m} = 0$.*

Prediction 2 states the conditions under which non-TM patients will be affected by the program. First, if the penalties did not incentivise the hospital to reduce readmission rates among TM patients, then it will also have no effect on non-TM patients. Second, among hospitals that respond to the policy, some hospitals may also reduce readmissions for non-TM patients. In particular, there will be reductions for non-TM patients if there are cost complementarities in quality provision for TM and non-TM patients. Intuitively, since the penalties do not apply to payments for non-TM patients, the program has no direct effect on the marginal cost or benefits of improving quality for

non-TM patients. If the incentives are not effective in increasing q_m, then they will also not be effective in increasing q_n. If the incentives are effective in increasing q_m, then any effects of the program on non-TM patients must be due to complementarities in quality provision across patient insurance groups.

4 | Empirical Strategy

In light of Predictions 1 and 2, the goal of the empirical strategy is to estimate how the financial incentives created by the program impacted quality of care for TM and non-TM patients depending on whether the hospital they are treated at is likely to be penalised by the program. To do so, I combine matching with general difference-in-differences techniques and exploit within-hospital variation in financial incentives across patient diagnoses (monitored or non-monitored) and across time (pre- and post-policy).

The key identifying assumptions underlying this design are: (a) within insurance and hospital performance groups, readmission trends are parallel across diagnoses ("parallel readmission trends") (b) within insurance and hospital performance groups, non-monitored diagnoses are not affected by the program (stable unit treatment value assumption, "SUTVA") and (c) there are no other confounding factors (e.g. alternative policies) that might affect readmission rates of TM or non-TM patients around the same time the HRRP is implemented. Assumption (a) may be violated if patients admitted for a monitored diagnosis differ from patients admitted for "comparison" diagnoses in unobservable ways. It may also be violated if hospitals select different types of patients to treat before and after the policy in order to reduce readmission rates. Assumption (b) may be violated if physicians who treat monitored patients also treat comparison-group patients. It may also be violated if the policy leads to improvements in treatment procedures that are also used to treat patients admitted for non-monitored conditions. Assumption (c) may be violated if private insurance companies also implement policies around the same time that reduce readmission rates for non-TM patients.

To limit violations of assumption (a), I use a matching procedure that matches patients admitted for monitored diagnoses (AMI/HF/PN) with patients admitted for non-monitored diagnoses at the same hospital. To limit violations of assumption (b), I restrict patients in the "comparison" group to be those admitted for diagnoses that are biologically dissimilar to the monitored diagnoses. Specifically, following several studies, I restrict the set of comparison diagnoses to include only

gastrointestinal conditions.[22] This is intended to help reduce potential spillovers across diagnoses via physician knowledge accumulation or via shared treatment procedures.[23] The following sections describe the matching procedure and then discuss the empirical specification. Graphical evidence in support of the first two identification assumptions is provided in Section 5. While it is difficult to rule out all other factors that might affect readmission rates near the implementation period of the HRRP (assumption c), robustness checks in Section 7 suggest that this is unlikely to be a major confounding factor.

4.1 Matching Procedure

The matching procedure proceeds as follows. First, for each of the three monitored conditions, I construct individual datasets of patients who are admitted for these conditions and who satisfy some sample selection criteria such as having a standard hospital discharge (selection criteria are further discussed in Section 5). Second, I construct one dataset of potential "comparison" patients who are admitted for non-monitored conditions (gastrointestinal). Third, for each patient admitted for a monitored diagnosis, I match them to patients in the potential comparison group based on a set of observable characteristics. For each monitored diagnosis dataset, I match without replacement: each patient in the comparison dataset can only be matched once to one patient within the AMI dataset, and similarly to only one patient within each of the HF and PN datasets. However, the same comparison patient may be matched to a patient admitted for heart attack and a patient admitted for heart failure. I drop unmatched patients from the sample.

I match exactly on the following variables: period of admission (pre- or post policy), hospital ID, primary payer, age bins (5-year intervals), gender, and minority status (white vs. non-white). If there is more than one potential match, the best match based on age, co-morbidity, and race is

[22] Mellor et al. (2017) and Carey and Lin (2015) use gastrointestinal conditions as the control. See Appendix A.3 for the list of ICD9 codes for the comparison diagnoses.

[23] In the sample, there appears to be little overlap in terms of physicians and procedures across monitored and non-monitored diagnoses. In terms of procedural overlap, fewer than 0.2% of gastrointestinal patients are treated with any of the top-3 AMI procedures while fewer than 8% of patients admitted for gastrointestinal conditions are treated with any of the top-3 HF or PN procedures. In terms of physician overlap, among doctors that specialise in AMI cases (defined as having more than 50% of their patients being admitted for AMI), fewer than 6.5% treat comparison patients in the same year. The figures for physicians specialising in HF and PN are even lower: 5.6% and 3.1% respectively.

selected.[24] Matching on patient age, gender, and minority status attempts to minimise the observable differences between patients admitted for monitored and non-monitored diagnoses. Matching on period of admission ensures that relative differences between monitored and non-monitored diagnoses in changes in readmission rates across time are not due to compositional changes in the types of patients admitted for various diagnoses. Matching on the hospital ID means that patients admitted for non-monitored diagnoses are treated at the same hospital as patients admitted for monitored diagnoses. Estimated impacts of the program are therefore not driven by differential hospital responses to the policy (e.g. due to hospital efficiency). Matching on the primary payer means that, within insurance group, provider reimbursement methods are unlikely to differ across patients admitted for monitored and non-monitored conditions.

The post-policy date is defined as admissions after August 2011, following the convention of most existing studies.[25] It is possible that some hospitals started responding immediately post-ACA which was enacted in March 2010. Robustness checks in Section 7 show that the results are robust to various definitions of the post-policy date.

4.2 Estimating Equation

The main estimating equation for the analysis is:

$$\begin{aligned} y_{idht} = \beta_1 P_{dt} + \beta_2 (P_{dt} \cdot E_i) + \beta_3 (P_{dt} \cdot A_{dh}) + \beta_4 (P_{dt} \cdot E_i \cdot A_{dh}) \\ + \alpha_h + \delta_d + \Xi_t + Z'_{idht}\eta + X'_{idht}\zeta + \epsilon_{idht} \end{aligned} \tag{9}$$

The subscripts in equation (9) are defined as follows. The subscript i denotes the patient. The subscript d denotes the diagnosis that the patient is admitted for. For example, in the AMI sample, d denotes AMI or a comparison diagnosis. The subscript h denotes the hospital that the patient is initially admitted to. The subscript t denotes the time period (measured in months) when the patient is admitted.

[24] If there are any further ties, I randomly select the match.
[25] One exception is Zuckerman et al. (2016) who use April 2010 as the post-program date.

The main variables in equation (9) are defined as follows. y_{idht} is the outcome variable of interest for patient i admitted for diagnosis d at hospital h at time t. In the main analysis, this is an indicator variable that equals one if the patient is readmitted to hospital within 30 days and zero otherwise. $P_{dt} := D_d \times T_t = 1$ equals one if diagnosis d is monitored at time t. D_d is an indicator that equals one if diagnosis d is one of the monitored diagnoses. T_t is an indicator that equals 1 for months after the announcement of the policy (after August 2011). For example, if $d = AMI$ and t is August 2012, then $P_{dt} = 1$. E_i takes on a value of one if the patient's hospital stay is covered by TM and zero otherwise.

A_{dh} is an indicator for whether the hospital's observed readmission rate between August 2008 and March 2010 is above the risk-adjusted national average rate published in CMS reports. This variable proxies for whether the hospital expects to be penalised by the program in the post-policy period, which is required to test Prediction 1. Two features about the construction of A_{dh} are worth noting. First, the actual CMS penalty formula defines "above-threshold" hospitals as those with risk-adjusted predicted readmission greater than the risk-adjusted expected readmission rate (see Section 2). Rather than using the CMS statistical model to classify hospitals into "above-threshold" and "below-threshold" categories, this construction of A_{dh} uses a simpler rule: compare the *observed* readmission rate among TM patients to published statistics on national average risk-adjusted readmission rates. This seems reasonable given the complexity of calculating risk-adjusted values. Furthermore, the results in Section 6 provide support for this assumption about hospital behaviour: hospitals that are classified as poor-performers according to this simpler rule respond to the program; hospitals that are classified as good-performers do not respond. Second, the construction of A_{dh} does not use data after March 2010 in case hospitals do respond after the enactment of the ACA (i.e. before August 2011). If this is the case, then a construction of A_{dh} that uses data after March 2010 will capture not only the hospital's exposure to the program but also its response to the program. I examine other methods of calculation A_{dh} in Section 6.

The remaining variables are defined as follows. α_h is a vector of hospital fixed effects. This captures time-invariant hospital characteristics that may affect patients' probability of being readmitted and can be interpreted as some measure of hospital efficiency. δ_d denote diagnosis dummies

that capture any diagnosis-specific variation.[26] Ξ_t is a vector of month-year time dummies that captures seasonal fluctuations and trends in readmission rates that affect all hospitals and diagnoses in the same way. Z'_{idht} is a vector of the other main and interaction effects.[27]. X'_{idht} is a vector of patient characteristics that were coarsely matched (age, co-morbidities index, race) and a vector of hospital characteristics that may affect readmissions (physician-to-bed ratio, nurse-to-bed ratio, and number of beds). Following existing studies (e.g. McGarry et al., 2016; Mellor et al., 2017), I cluster standard errors at the hospital level.

The main coefficients of interest in equation (9) are β_1 to β_4. The sum $\sum_{s=1}^{s=4}\beta_s$ captures the impact of the program on TM patients treated at hospitals that expect to be penalised by the program ("poor-performance" hospitals). $\beta_1+\beta_3$ is the impact of the program on non-TM patients treated at hospitals that expect to be penalised by the programs (the "spillover effect" in these hospitals). $\beta_1+\beta_2$ is the impact of the program on TM patients treated at hospitals that do not expect to be penalised by the program ("good-performance" hospitals) while β_1 is the spillover effect in these hospitals. The β coefficients can be interpreted as (the negative of) the empirical counterparts of $\partial q_n/\partial\tau_m$ and $\partial q_m/\partial\tau_m$ which feature in the model (see Section 3). While the model analyses the impact of the policy as an infinitesimal change in the penalty, the matched differences-in-differences estimator exploits a discrete change.[28] For example, one can therefore interpret $\sum_{s=1}^{s=4}\beta_s$ as (minus) the change in q_m due to a discrete change in financial penalties: $\sum_{s=1}^{s=4}\beta_s = -\Delta q_m/\Delta\tau_m$ where the size $\Delta\tau_m$ equals the average penalty faced by hospitals in the sample. In the sample, this equals 0.55% of total Medicare revenues.[29]

[26] In the baseline specification, this is a dummy for the monitored diagnosis. Alternative specifications include dummies for diagnoses at the DXCCS and ICD-9 level. Results for the alternative and baseline specification are similar.

[27] In this case, $Z'_{idht} := [E_i,\ (E_i\cdot D_d),\ (E_i\cdot T_t), (A_{dh}\cdot D_d),\ (A_{dh}\cdot T_t),\ (A_{dh}\cdot E_i),\ (A_{dh}\cdot D_d\cdot E_i),\ (A_{dh}\cdot T_t\cdot E_i)]$

[28] See Gupta (2017) for an analysis which uses continuous variation in (expected) penalty levels to estimate the impacts of the policy.

[29] This is described further in Table 1.2.

Given Predictions 1 and 2, the two (alternative) hypotheses are:[30]

$$\textbf{(HA)}\colon \sum_{s=1}^{s=4} \beta_s < 0 \quad \text{and} \quad \textbf{(HB)}\colon \beta_1 + \beta_3 \neq 0 \quad \text{given} \sum_{s=1}^{s=4} \beta_s < 0$$

Hypothesis (HA) corresponds to Prediction 1 and tests whether the program led to a significant reduction in readmission rates among TM patients admitted for a monitored diagnosis at poor-performance hospitals. Hypothesis (HB) corresponds to Prediction 2 and tests whether the program had an effect on non-TM patients admitted for a monitored condition in poor-performance hospitals.

[30] The corresponding null hypotheses are:

$$\textbf{(HA0)}\colon \sum_{s=1}^{s=4} \beta_s = 0 \quad \text{and} \quad \textbf{(HB0)}\colon \beta_1 + \beta_3 = 0$$

5 | Data, Sample, and Descriptive Statistics

In this section, I discuss the data sources used in the empirical analysis and the sample selection criteria. Further information about the data can be found in Appendix A.3. I then provide some summary statistics for the main sample.

5.1 Data Sources

Patient data. The main data source is the Healthcare Cost and Utilisation Project's (HCUP) New York State inpatient database (NYSID) and emergency department database (NYSEDD) from 2008 to 2014. This dataset is suitable for this analysis for two reasons. First, it covers the universe of acute care hospitals that accept Medicare patients and are paid under the IPPS.[31] These are the types of hospitals that are affected by the HRRP. The hospitals included in the inpatient database correspond closely to the hospitals included in the American Hospital Association's Annual Survey of Hospitals. Second, the data encompasses all patients regardless of payer-type, allowing one to examine the effects of the HRRP on non-TM patients.

In my empirical analysis, the targeted patient group are defined as TM patients. The non-targeted patient group consists of Medicare Advantage (MA) patients and privately-insured patients. In my sample, roughly 72% of patients on Medicare have TM while the remaining 28% are on a Medicare health management organisation (HMO) plan, one of the main types of MA plans. Most patients with private health insurance acquire insurance via their employer. Private health insurance plans come in different forms, with Blue Cross/Blue Shield plans being the most common in my sample (35.39% among non-TM and non-MA patients). The omitted patient insurance groups are: self-insured patients, Medicare patients, patients with "other" insurance schemes (e.g. workers' compensation, CHAMPUS/VA coverage), and patients with "no charge" hospital stays.[32]

[31] Acute care hospitals are non-federal, short term hospitals. Such hospitals are usually contrasted with long-term care hospitals which specialise in providing treatment to patients with chronic conditions. Included among acute care community hospitals are academic medical centers and specialty hospitals such as obstetrics, gynecology, otolaryngology, short-term rehab, orthopedic, and pediatric hospitals. Non-community hospitals that are excluded from the sample include Federal hospitals (e.g. Veterans Affairs), long-term hospitals, psychiatric hospitals, alcohol/chemical dependency treatment facilities, and hospital units within institutions such as prisons.

[32] See Appendix A.3 for more details on the omitted insurance groups.

The data contains information on the diagnosis associated with the hospitalisation, the procedures performed (if any), total charges, expected payment source, patient co-morbidities, and patient demographics (including gender, race, age in months, and income at the zip-code level). Co-morbidities are underlying health conditions that may increase the risk of death. I use the Elixhauser Co-morbidity Index to summarise the set of existing co-morbidities.[33] Patient diagnoses are coded using the International Classification of Diseases, Clinical Modification (ICD 9-CM). I use this coding scheme to identify monitored and non-monitored conditions.[34] I turn the hospital charges into billable costs by applying the HCUP's cost-to-charge ratio data.[35] This ratio is hospital-specific and adjusts charges into costs by accounting for hospital-specific wage and capital costs. Thus, these cost figures reflect the actual expenses incurred in the production of hospital services such as wages, supplies, and utility costs. The data also contains detailed information on the expected payer for the hospital stay. This is the main variable I use to categorise patients into TM and non-TM groups.

Crucially, the data contains an identification number for the patient and for hospitals, allowing researchers to check whether the patient returned to (the same or a different) hospital within 30 days. To construct this readmission variable, I follow the CMS's definition of a hospital readmission as "an admission to an acute care hospital within 30 days of discharge from the same or another acute care hospital".[36] Same-day and consecutive hospital episodes are not considered as readmissions. The CMS uses an "all-cause" definition, meaning that readmissions can be for any diagnosis except for after-care (e.g. check-up after surgical procedures). Appendix A.3 contains more information about how index and readmission episodes are identified in the data.

[33] The Elixhauser Co-morbidity Index is a method of categorizing co-morbidities of patients based on the 9th Edition of the International Classification of Diseases (ICD) diagnosis codes found in administrative data, such as hospital abstracts data (Elixhauser et al., 1998). The index is based on 30 co-morbidities. Weights are assigned to these co-morbidities based on the relationship between a specific co-morbidity and readmission probabilities. The index takes integer values and higher values correspond to worse health status.

[34] See Appendix A.3 for diagnosis coding.

[35] See https://www.hcup-us.ahrq.gov/db/state/CCR_CD-SID_UserGuide_2001-2016.pdf [Accessed 13 December 2018.]

[36] See Horwitz et al. (2011) for more details on CMS calculation of 30-day readmission rates.

Hospital data. The patient-level data is matched to hospital data using the hospital ID.[37] Hospital information comes from two sources: (i) the CMS's *Hospital Compare* data which provides information on hospital performances regarding the HRRP[38] and (ii) the CMS's Provider of Services data which provides information on all inpatient hospitals in the US. Key variables of interest include: the excess hospital risk-adjusted readmission rate and hospital size measures (number of beds, physicians, nurses).

5.2 Sample Selection

Patient selection criteria. The unit of analysis is an "index admission", the first admission to hospital for a given diagnosis. An "index hospitalisation" is included in the main sample if it is an admission to hospital that satisfies the following criteria: (i) the patient's primary payer is Traditional Medicare, Medicare Advantage, or private insurance, (ii) the patient is between 40 and 90 years old when admitted, (iii) the patient is discharged alive and did not leave against medical advice, (iv) the patient is admitted for any of the monitored diagnoses or a "comparison" diagnosis, (v) the patient has a match using the matching procedure described in Section 4.

Hospital selection criteria. For hospital h to be included in the monitored diagnosis sample, the hospital must treat at least 25 patients diagnosed with the monitored condition both before and after the program is implemented.[39] The CMS uses the same threshold when deciding whether a hospital has sufficient eligible cases to calculate hospital performance.[40]

Since I estimate the effects of the HRRP separately for each of the three monitored conditions, the hospitals included in each diagnosis sample may differ. For example, if a hospital serves fewer

[37] The hospital ID in the HCUP data is the New York State Department of Health's permanent facility identifier (PFI). The hospital ID in the CMS data is the hospital Medicare's Provider Number (MPN). I crosswalk a hospital's PFI to MPN using the New York State Department of Health's hospital Health Profiles which contain information on a facility's PFI and MPN.

[38] See https://www.medicare.gov/hospitalcompare/search.html. [Accessed 6 April 2018.]

[39] Some hospitals may not treat patients diagnosed with any of the monitored conditions. For instance, some hospitals in the sample specialise in ear, nose, and throat conditions and therefore do not treat any patients diagnosed with heart attack, heart failure, or pneumonia.

[40] The CMS calculates performance variables for hospital h if the hospital has more than 25 cases over a three-year period.

than 25 AMI patients each year but more than 25 HF patients each year, this hospital would be included in the HF main sample but not in the AMI main sample. From the original sample of roughly 185 acute care hospitals in the dataset, this selection criteria produces 124, 143, and 142 hospitals in the AMI, HF, and PN samples respectively.

5.3 Descriptive Statistics

Patient characteristics. The summary statistics for the main sample are presented in Table 1.1. In panel A, I present descriptive statistics for traditional Medicare (TM) patients; in panel B, I present analogous statistics for non-TM patients (Medicare Advantage and privately-insured). Column (1) presents mean characteristics of AMI patients, column (2) presents mean characteristics of patients admitted for comparison diagnoses, and column (3) presents the p-value for the difference in means. The layout for the heart failure and pneumonia samples are the same.

Since I match on age bins, gender, and minority status, there is no significant difference in the distribution of these characteristics between the monitored and non-monitored patient groups across all diagnoses samples. Even though I do not match on the co-morbidity index, the mean co-morbidity index is quantitatively similar across patients admitted for monitored and non-monitored diagnoses. For example, in the heart attack TM sample, patients admitted for the monitored condition have a co-morbidity score of 14.64 while patients admitted for a comparison condition have a score of 14.70. For both TM and non-TM patient groups, health outcome and healthcare utilisation variables differ between patients admitted for monitored and non-monitored conditions.

Comparing the mean readmission rate of monitored and non-monitored patients, it is clear that the 30-day readmission rate for patients admitted for monitored conditions is on average higher than that among patients admitted for non-monitored conditions (apart from the PN sample).[41] This is unsurprising since CMS most likely decided to target performance in AMI/HF/PN diagnoses due to persistently high readmission rates in these diagnosis domains. Similarly, the probability of 30-day emergency department (ED) revisits is higher among patients admitted for monitored diagnoses

[41] For example, among TM patients admitted for AMI, the 30-day readmission rate is 21.12%; among TM patients admitted for comparison diagnoses, the 30-day readmission rate is 17.73%.

compared to those admitted for non-monitored diagnoses, apart from PN cases.[42] Lastly, total costs of the initial stay are generally higher for patients admitted for a monitored condition. The cost difference is particularly stark in the AMI sample: the average cost of AMI patients is $19,000 while the average cost for comparison-AMI patients is $9,000-$11,000.

The final row shows that this method produces a fairly high match rate. Across all three monitored diagnoses, over 86% of TM patients and over 80% of non-TM patients have an exact comparison patient. Table A.2 in Appendix A.4 compares matched and unmatched observations and shows that the unmatched observations tend to be sicker and have a higher proportion of men and non-white individuals. One concern is that the matched observations are driving the main results; however, robustness checks in Section 7 show that when the potential comparison group is enlarged (consisting of all non-monitored conditions excluding childbirths) so that almost all observations have matches, the baseline results are very similar. This suggests that the matching strategy is not driving the main results. Since the matches are one-to-one, the number of patients in the monitored and non-monitored groups are the same. Due to the prevalence of heart diseases, the heart failure sample for TM and non-TM patients is the largest with 307,226 ($153,613 \times 2$) in the TM group and 180,668 ($90,334 \times 2$) in the non-TM group.

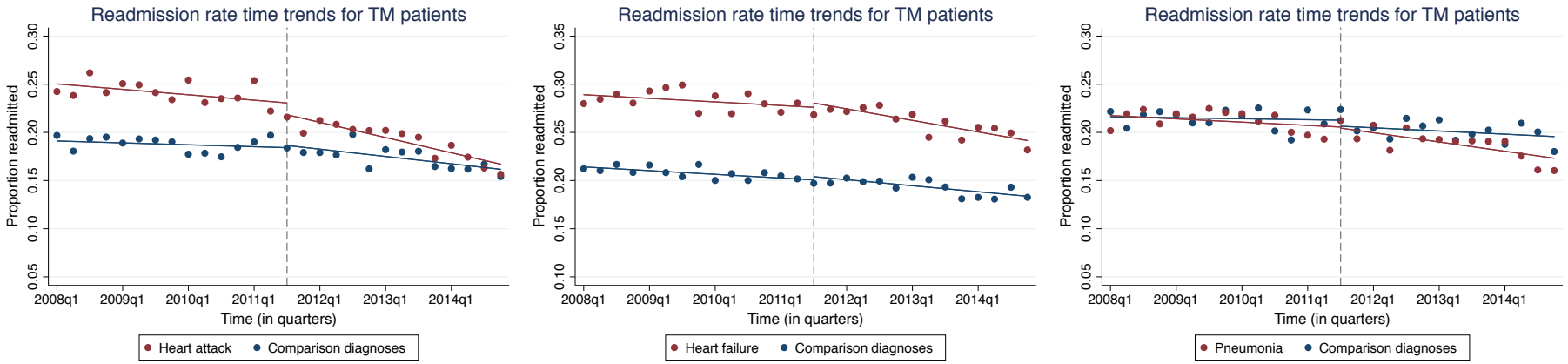
The key identifying assumptions for this empirical strategy are parallel readmission trends and SUTVA. To examine the evidence in support of these assumptions, Figure 1.2 plots the raw readmission trends separately by insurance group (TM or non-TM) and monitored condition (AMI/HF/PN) by quarter of first admission. For clarity, within each insurance-diagnosis group, the diagrams depict readmission trends for poor-performance hospitals. Appendix A.5 presents analogous figures for good-performance hospitals.[43] The vertical dashed line denotes the quarter after the announcement of the policy (2011 Q4). For example, Figure 1.2a presents readmission trends for TM patients admitted for AMI in red and readmission trends for TM patients admitted for comparison diagnoses in blue. After the announcement of the policy, the slope for the AMI group is steeper and there is a gradual convergence in readmission rates between the monitored and non-monitored diagnosis groups. This provides preliminary evidence of the impacts of the program.

[42] ED revisits are defined here as cases where the patient visits an ED within 30 days but is not readmitted.
[43] Since there are fewer good-performance hospitals in NYS, there is less precision in the depicted time trends.

Table 1.1: Mean characteristics and outcomes of patients admitted for monitored and non-monitored diagnoses

	Heart attack			Heart failure			Pneumonia		
	Monitored	Comparison	P-value	Monitored	Comparison	P-value	Monitored	Comparison	P-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. TM Patients									
Age	76.31	76.29	0.724	77.94	77.86	0.016	76.78	76.74	0.372
Female (%)	48.88	48.88	1.000	54.48	54.48	1.000	54.00	54.00	1.000
White (%)	76.87	76.87	1.000	71.19	71.19	1.000	77.70	77.70	1.000
Co-morbidity index	14.64	14.70	0.410	18.96	18.81	0.001	21.48	21.26	0.000
30-day readmission (%)	21.10	17.73	0.000	26.45	19.71	0.000	19.30	20.41	0.000
30-day ED revisit (%)	23.03	20.73	0.000	28.57	22.88	0.000	23.15	23.82	0.000
ED utilisation (%)	67.61	80.70	0.000	86.35	85.12	0.000	92.06	85.64	0.000
Total cost ($)	19,356	10,685	0.000	11,993	11,121	0.000	9,906	11,169	0.000
% Matched (# Patients)	93.41 (N=127,268)			87.26 (N=307,226)			86.36 (N=216,270)		
B. Non-TM Patients									
Age	64.72	64.67	0.417	72.13	72.06	0.150	67.58	67.53	0.516
Female (%)	34.88	34.88	1.000	47.21	47.21	1.000	51.72	51.72	1.000
White (%)	71.42	71.42	1.000	57.11	57.11	1.000	69.20	69.20	1.000
Co-morbidity index	9.01	9.15	0.016	16.75	16.61	0.019	16.94	16.73	0.004
30-day readmission (%)	14.03	12.32	0.000	22.85	16.85	0.000	13.56	16.11	0.000
30-day ED revisit (%)	15.98	14.50	0.000	24.59	19.77	0.000	16.67	18.83	0.000
ED utilisation (%)	65.58	76.13	0.000	84.37	84.07	0.076	91.86	83.43	0.000
Total cost ($)	19,418	9,414	0.000	13,034	10,719	0.000	8,699	10,280	0.000
% Matched (# Patients)	89.98 (N=140,492)			80.17 (N=180,668)			82.61 (N=133,896)		

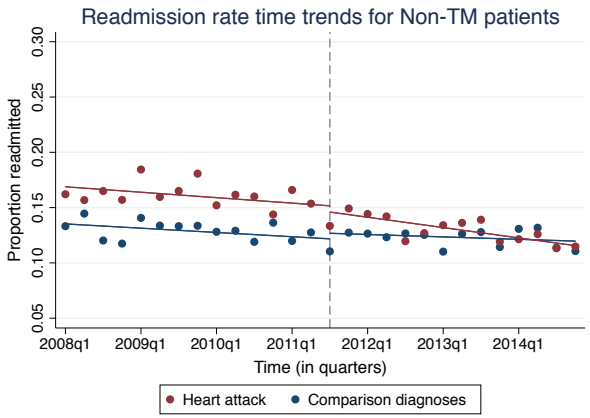
Notes. This table shows means of observable characteristics and outcomes among TM patients (Panel A) and non-TM patients (Panel B). Within each insurance group, the columns present statistics for patients admitted for monitored conditions (AMI/HF/PN), statistics for patients admitted for non-monitored conditions (gastrointestinal), and the p-value for the difference in means. The non-monitored comparison group consists of patients admitted for gastrointestinal conditions and patient matches are selected using the matching procedure described in Section 4. Unmatched observations are dropped. Since the matches are exact, the number of observations in the monitored and non-monitored groups are equal.



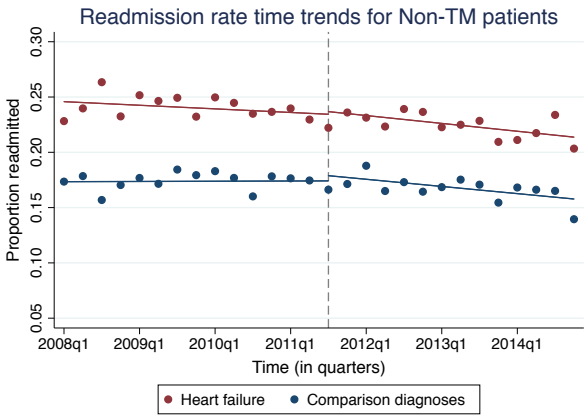
(a) Heart attack TM Patients

(b) Heart failure TM Patients

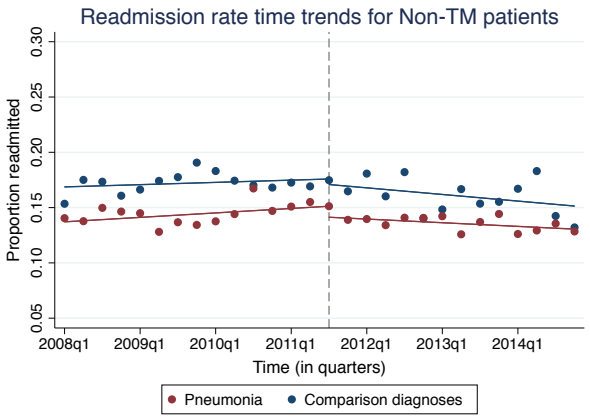
(c) Pneumonia TM Patients



(d) Heart attack non-TM Patients



(e) Heart failure non-TM Patients



(f) Pneumonia non-TM Patients

Figure 1.2: This figure presents readmission trends for each monitored diagnosis and its corresponding comparison group, by TM and non-TM patient groups, for hospitals that expect to be penalised by the program. The non-monitored comparison group consists of patients admitted for gastrointestinal conditions and patient matches are selected using the matching procedure described in Section 4.

The parallel trends assumption can be examined using the readmission trends before the policy announcement. Across all diagnoses samples, the readmission trends for both monitored and non-monitored groups appear parallel. The evidence for parallel readmission trends is particularly strong for AMI and HF samples. It is slightly weaker for pneumonia TM patients: the time trend for pneumonia appears to have a downward slope before the policy announcement.

The SUTVA assumption can be examined using the readmission trends for non-monitored conditions before and after the policy announcement. This assumption requires that there is no break in the trend for the comparison group before and after the program is introduced. These figures suggest that there is a downward trend for AMI and HF readmissions (as documented by Zuckerman et al. (2016)), but no break in this trend around the period of the HRRP implementation.

Hospital characteristics. Table 1.2 provides descriptive statistics for the hospitals in the sample. Performance statistics use the CMS' risk-adjusted models and therefore reflect actual penalties. There are 124, 143, and 142 hospitals in the AMI, HF, and PN samples respectively. Almost 80% of these hospitals have an (actual) excess readmission ratio that exceeds one (i.e. have a risk-adjusted readmission rate that exceeds the national average).[44] The mean excess readmission ratio is 1.06 across all diagnosis domains and this figure ranges from a low of 0.84 to a high of 1.36. The average penalty τ_h faced by hospitals in the first penalty year of the program (2012) is 0.55.

Over 85% of hospitals in the sample are private not-for-profit (NFP) hospitals while the remainder are public (state, federal, or local level) hospitals. An unusual feature of New York State is that there are no for-profit hospitals, which are common in the rest of the country (Hammond, 2018). Not-for-profit (NFP) hospitals are exempt from federal income tax and have charitable obligations to their communities. Although NFP hospital may make profits, any surplus funds must be reinvested in the hospital (rather than distributed to shareholders). The large proportion of NFP hospitals making up the sample motivates the earlier modelling decision to allow both quality and profit to enter the hospital payoff function.

[44] See Section 2 for more details on how NYS differs from other states.

Table 1.2: Descriptive statistics for hospitals in the sample

	Heart attack	**Heart failure**	**Pneumonia**
	(1)	(2)	(3)
Number of hospitals	124	143	142
% of hospitals w/ excess ratio>1	79.03	69.93	76.76
Mean excess ratio	1.06 [0.84,1.36]	1.05 [0.85,1.29]	1.06 [0.88,1.31]
Average penalty	0.55	0.55	0.55
% patients under TM	58.15	67.76	63.46
% Private NFP	89.52	86.71	87.32
% Public	10.48	13.29	12.68
Number of beds	376.21	348.54	344.92
Number of physicians	64	61	62
Number of nurses	516	466	462

Notes. This table presents descriptive statistics for the hospitals in the sample. Hospitals are included if they treat at least 25 patients with the specified diagnosis before and after the implementation of the HRRP. The excess ratio is the hospital's risk-adjusted predicted readmission rate divided by risk-adjusted expected readmission rate. A hospital is penalized if its excess ratio is greater than one. Figures in square brackets are minimum and maximum values. See Section 2 for more details on the program background.

The final three rows of the table provide information about the size of the hospitals, as measured by the total number of beds, physicians, and nurses. Hospitals in the AMI sample are on average the largest with 376 beds, 64 physicians, and 516 nurses.

6 | Results

6.1 Impact on Readmission Rates

Table 1.3 presents estimates of the direct and indirect effects of the policy on TM and non-TM patients based on equation (9) to test Predictions 1 and 2. In panel A, I present the estimated effect of the program on TM patients admitted for each of the three monitored conditions. In particular, the estimates in row (1) correspond to the $\beta_1 + \beta_2$ coefficients in equation (9). This coefficient captures the impact of the program on TM patients in hospitals that don't expect to be penalised. The estimates in row (2) correspond to the $\sum_{s=1}^{s=4} \beta_s$ coefficients in equation. This captures the impact of the program on TM patients in hospitals that expect to be penalised. In panel B, I present the estimated effects on non-TM patients. The estimates in row (3) correspond to the β_1 coefficient in equation (9). This is the spillover effect in hospitals that don't expect to be penalised. Lastly, the estimates in row (4) correspond to the $\beta_1 + \beta_3$ coefficients in equation (9). This is the spillover effect in hospitals that expect to be penalised. One-sided and two-sided tests are used respectively for hypotheses HA (panel A) and HB (panel B).

Columns (1) and (2) show the effect of the program on patients admitted for AMI. Panel A row (1) shows that among targeted TM patients treated at hospitals that do not expect to be penalised by the program, the program had no impact on readmission rates. The estimated impact is an economically small and statistically insignificant 0.37 percentage point (pp) reduction in readmission rates. In contrast, row (2) shows that among targeted patients treated at hospitals that do expect to be penalised by the program, readmission probabilities fell by 3.36 pp (p-value<0.001), relative to a baseline of 24.12%. This represents a 13.9% decrease in the number of readmissions among the TM group. This finding is consistent with Prediction 1.

Panel B shows that readmission rates also fell among among non-TM AMI patients treated at hospitals that expect to be penalised by the program (row 4). In particular, readmission rates among non-TM patients treated at these hospitals fell by 2.38 pp relative to a baseline readmission rate of 16.08%. This represents a 14.8% reduction in the number of patients being readmitted.

Table 1.3: Impact of program on 30-day readmissions for TM and non-TM patients

	Heart attack		Heart failure		Pneumonia	
	Pre-mean	Estimate	Pre-mean	Estimate	Pre-mean	Estimate
	(1)	(2)	(3)	(4)	(5)	(6)
A. Prediction 1: Impact on TM patients						
(1) Below average ($\beta_1 + \beta_2$)	19.56	-0.0037 (0.006)	23.09	0.0009 (0.007)	17.23	0.0037 (0.006)
(2) Above average ($\sum_{s=1}^{s=4} \beta_s$)	24.12	-0.0336*** (0.006)	28.28	-0.0073** (0.004)	21.21	-0.0094*** (0.004)
B. Prediction 2: Impact on non-TM patients						
(3) Below average (β_1)	12.91	-0.0072 (0.007)	21.58	-0.0056 (0.010)	12.08	0.0050 (0.006)
(4) Above average ($\beta_1 + \beta_3$)	16.08	-0.0238*** (0.005)	24.08	-0.0098** (0.004)	14.41	0.0036 (0.006)
# Patients	267,760 (47.53% TM)		487,894 (62.97% TM)		350,166 (61.76% TM)	
# Hospitals	121 (73.55% Above)		140 (66.43% Above)		139 (62.59% Above)	

Notes. This table presents the estimated impact of the program on TM and non-TM patients using equation (9). HA and HB denote hypotheses A and B respectively (see Section 4 for more details). "Above" denotes the effect among hospitals that expect to have an excess readmission ratio above one based on the observed readmission rates from Jul 2008 to Mar 2010. The estimate for "Above" corresponds to $\sum_{s=1}^{s=4} \beta_s$ for TM patients and $\beta_1 + \beta_3$ for non-TM patients. The estimate for "Below" corresponds to $\beta_1 + \beta_2$ for TM patients and β_1 for non-TM patients. Standard errors in parentheses, clustered at the hospital level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Among these hospitals, the estimated impact of the program on TM and non-TM patients is not statistically different.[45] These results indicate strong spillover effects in the case of AMI. Consistent with Prediction 2, given that good-performance hospitals did not reduce readmission rates for targeted patients (row 1), non-targeted patients treated at these hospitals also did not see a fall in readmissions (row 3). The estimated impact on non-TM patients treated at good-performance hospitals is a statistically insignificant -0.72 pp.

The estimates reported in columns (3) and (4) depict the effect of the program on patients admitted for heart failure. Among TM patients, panel A row (2) shows that readmission rates for heart failure TM patients treated at hospitals that expect to be penalised fell by 0.73 pp (p-value=0.037), relative to a baseline rate of 28.28%. Consistent with Prediction 1, there is no reduction in the readmission rate of heart failure TM patients treated at good-performance hospitals

[45] The $\beta_2 + \beta_4$ coefficients equal -0.009 with a p-value of 0.239.

(row 1).

Panel B suggests that the reductions among heart failure TM patients had spillover effects on non-TM patients treated at the same poor-performance hospitals. Row (4) shows that among poor-performance hospitals, readmission rates among non-targeted patients fell by -0.98 pp (p-value=0.028), relative to a baseline of 24.08%. Although the coefficients for the non-TM group are larger in magnitude than that for the TM group, the difference is not statistically significant.[46] Since good-performance hospitals did not respond to the program by reducing readmission rates for targeted patients, it is unsurprising that there are no spillover effects onto non-targeted patients treated at these hospitals (row 3).

Columns (5) and (6) present the effects of the policy on patients admitted for pneumonia. Panel A shows that the program reduced readmission rates among TM patients treated at hospitals with pre-policy readmission rates above the threshold. Readmission probabilities fell by 0.94 pp relative to a baseline of 21.21% (p-value=0.007), a 4.4% reduction. Among good-performance hospitals, the program had no impact on readmissions: the estimate is 0.004 (p-value=0.721). Panel B illustrates that the program had no effects on non-TM patients admitted for pneumonia. Among poor-performance hospitals, there is no spillover effect detected: the estimated impact is an increase of 0.036 pp (p-value=0.519). Since good-performance hospitals did not respond to the program by reducing readmission rates for TM patients, there is also no impact among TM patients treated at those hospitals.

The results point to differential policy impacts across diagnosis domains. Among TM patients, the effects are strongest for heart attack cases and moderate for heart failure and pneumonia cases. Existing studies also find that the impacts of the policy vary across diagnoses in a similar fashion.[47] One explanation is that responses depend on the biological nature of the diagnosis. For example, the acute nature of a heart attack may mean that patients are more responsive to treatments administered within hospital compared to more chronic conditions such as heart

[46] The $\beta_2 + \beta_4$ coefficients equal 0.0025 with a p-value of 0.683.

[47] For example, Gupta (2017) finds that AMI, HF, and PN readmission rates fell by 3.5pp, 1.0pp, and 2.8pp respectively. For Mellor et al. (2017), the corresponding estimates are 4.0pp, 2.6pp, and 1.6pp. Carey and Lin (2015) find that the program had effects on AMI and HF patients, but not among pneumonia patients in New York State.

failure.[48] Relative to the acute conditions, decreasing readmissions for heart failure patients may require more extensive coordination with community care providers as well as more intense patient follow-up. The variation in responses among non-TM patients, relative to the response among TM patients, also suggests variation in the degree of cost-complementarity across diagnoses. The spillover effects among AMI patients suggest that complementarities across insurance groups are strong for this diagnosis. There are also some spillover effects for HF. However, for PN cases, the estimates indicate negligible degrees of complementarity. This suggests that, from the perspective of the hospital, enhancing quality of care for TM and non-TM patients admitted for PN appear to be two distinct activities.

6.2 Impact on Treatment Intensity

The main results indicate that the program reduced readmission rates among patients admitted for the monitored diagnoses to varying degrees. One explanation for this is that, as policy-makers intend, hospitals responded by improving quality of care. This could be achieved by devoting more resources to patients admitted for monitored conditions or by improving hospital productivity in monitored diagnoses. Here, I examine the impacts of the program on several measures of treatment intensity: (i) the total (billable) costs of the initial stay, a function of the resources allocated to the patient, (ii) the number of procedures that the patient received during their initial hospitalisation, and (iii) the length of the initial hospitalisation. To do so, I estimate versions of equation (9) where the dependent variable is replaced by measures of treatment intensity.

The results are presented in Table 1.4. Panel A examines the impact of the program on the cost of the initial hospitalisation. The estimated effects for each of the four main subgroups, TM and non-TM patients treated at good- and poor-performance hospitals, are presented in each subsequent row. The results in panel A show that the program did not have noticeable effects on the cost of the initial hospitalisation for any of the three monitored diagnoses. For example, among TM

[48] Heart failure is a chronic condition that sometimes leads to hospitalisation due to exacerbations, but is mainly managed through office consultations with primary physicians or cardiologists and regular patient check-ups. In contrast, heart attack and pneumonia are highly acute conditions that are typically managed through hospitalisation with less frequent follow-up.

AMI patients treated at hospitals that expect to be penalised (the group where readmission rates reduced the most), the estimated effect of the policy on the total cost of the initial hospitalisation is a -0.12% change (p-value=0.499). This is consistent with existing studies that show that despite reductions in readmissions, there is little evidence of changes in the billable charges and costs of the initial hospital stay (Gupta, 2017; Mellor et al., 2017). Furthermore, it is consistent with the idea that hospitals may incur unbillable costs when improving quality due to the unassignable nature of the quality-enhancing activities. Among non-TM patients, the estimates suggest that costs fell for AMI patients treated at hospitals that expect to be penalised (column 2, row 4); there is also some evidence that costs fell for PN patients treated at good-performance hospitals (column 6, row 3).

Panel B examines the impact of the program on the number of treatment procedures received by the patient. Columns (1) and (2) indicate that the number of treatment procedures increased among AMI TM patients treated at poor-performance hospitals. Specifically, the number of treatment procedures increased by 0.27 from a baseline of 5 procedures before the program (a 5% increase). Among heart attack non-TM patients treated at poor-performance hospitals, the program did not have a significant impact on the number of procedures received (column 2, row 4). One possible reason for this is that before the program, non-TM patients received on average more procedures than TM patients (6.4 vs. 5) and thus additional procedures may not yield additional health benefits, akin to a "flat-end-of-the-curve" effect (Fuchs, 2004). There is also some evidence that non-TM AMI patients treated at good performance hospitals saw a reduction in the number of procedures received (column 2, row 3), significant at the 10% level. Columns (3) and (4) show the program increased the number of procedures received at the initial hospitalisation by 0.07 for TM heart failure patients treated at poor-performance hospitals, the group where a reduction in readmission rates is detected (see Table 1.3). Relative to a baseline of 1.64 procedures during the initial stay, this represents a 4% increase. For heart failure non-TM patients, there is no significant impact. Columns (5) and (6) show that the number of treatment procedures received by pneumonia patients do not appear to be affected by the program.

Lastly, panel C examines the impact of the program on length of stay. The results vary by diagnosis. Interestingly, columns (1) and (2) show that the length of hospital stay fell for TM AMI patients treated at both good and poor-performance hospitals. This suggests that readmission

Table 1.4: Impact of program on treatment intensity

	Heart attack		Heart failure		Pneumonia	
	Pre-mean	Estimate	Pre-mean	Estimate	Pre-mean	Estimate
	(1)	(2)	(3)	(4)	(5)	(6)
A. Log real total cost at initial hospitalization						
(1) TM Below	9.46	0.0326	8.80	0.0152	8.80	-0.0007
		(0.035)		(0.020)		(0.014)
(2) TM Above	9.52	-0.0124	9.03	-0.0147	8.93	-0.0176
		(0.018)		(0.012)		(0.012)
(3) Non-TM Below	9.56	-0.0040	8.79	0.0131	8.66	-0.0304**
		(0.039)		(0.031)		(0.014)
(4) Non-TM Above	9.56	-0.0409**	9.05	-0.0135	8.78	-0.0086
		(0.020)		(0.016)		(0.011)
B. Number of procedures						
(1) TM Below	5.00	0.0356	1.38	0.0424	1.04	0.0500
		(0.132)		(0.038)		(0.049)
(2) TM Above	4.99	0.2716***	1.64	0.0723**	1.47	-0.0046
		(0.079)		(0.030)		(0.034)
(3) Non-TM Below	6.48	-0.2380*	1.67	0.0311	0.92	-0.0085
		(0.125)		(0.048)		(0.053)
(4) Non-TM Above	6.40	-0.0892	1.86	-0.0294	1.42	-0.0175
		(0.076)		(0.030)		(0.032)
C. Length of stay						
(1) TM Below	7.43	-0.3618**	5.77	0.2667**	6.34	0.1278
		(0.139)		(0.121)		(0.132)
(2) TM Above	7.39	-0.2754***	6.66	0.2556***	6.87	-0.0646
		(0.100)		(0.072)		(0.092)
(3) Non-TM Below	5.77	-0.2808*	5.45	0.2674	5.52	-0.2341
		(0.149)		(0.174)		(0.148)
(4) Non-TM Above	5.57	-0.1001	6.34	0.3813***	5.75	-0.0437
		(0.114)		(0.105)		(0.066)

Notes. This table presents the estimated impact of the program on treatment intensity. The regression is based on equation 9 where the dependent variable reflects the outcome of interest. "Above" denotes the effect among hospitals that expect to have an excess readmission ratio above one based on the observed readmission rates from Jul 2008 to Mar 2010. "TM" denotes Traditional Medicare. Standard errors in parentheses, clustered at the hospital level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

outcomes may not be sensitive to length of stay or that other procedural or structural changes have led the required length of stay to fall. The existing literature (e.g. Batt et al., 2018; Gupta, 2017) also finds that reductions in readmissions are not accompanied by increases in length of stay. It is possible that the combination of more treatments but shorter lengths of stay roughly offset each other for TM AMI patients, leaving no significant changes in billable costs. It is also

possible that the increased treatments for TM AMI patients produced better health outcomes at similar billable costs. In contrast, columns (3) and (4) show that the program led to an increase in length of stay of heart failure TM patients by 0.26 days at both good and poor-performance hospitals.[49] Non-TM patients treated at poor-performance hospitals also see an increase in length of stay, which could partly explain the reductions in readmissions observed among this group in Table 1.3. Lastly, columns (5) and (6) show that the program had no effect on the length of stay of pneumonia TM or non-TM patients.

Overall, these results suggest that pinning down the treatment quality changes required to reduce readmissions is difficult. There is no pattern that holds across diagnoses and across hospital performance groups that is consistent with the pattern of readmission reductions observed in Table 1.3. Aside from observable changes in the number of procedures received for AMI and HF TM patients treated at poor-performance hospitals, other standard measures of treatment quality used in the literature do not seem to be driving the readmission reductions. This highlights the fact that even if hospitals incur higher costs from quality improvements, in the short run, these costs are not reflected in billable charges.

6.3 Impact on Gaming Behaviour

Although the goal of the program is to reduce readmission rates by improving treatment quality, it is possible that hospitals respond by changing their readmission strategies. Table 1.5 examines whether the baseline results are driven by hospitals "gaming" the system. First, I examine whether the program led hospitals to delay readmissions outside the 30-day window by estimating equation (9) with 31–45 day readmission rates as the dependent variable. Second, I examine whether the program affected the probability that patients return to an emergency department (ED) within 30 days after discharge (but are not readmitted). ED returns reflect a combination of patients seeking care and the hospital deciding to keep them in the outpatient/emergency department rather than formally admit them as an inpatient. Therefore, "missing" readmissions may show up as ED returns

[49] The fact that these changes in the length of stays for TM patients were common to both types of hospitals casts doubt on whether they were the result of the introduction of the HRRP.

instead. Third, I examine whether the program led hospitals to transfer patients during their initial stay to other hospitals. If a patient is transferred from Hospital A to Hospital B, any readmission within 30 days of being discharged from Hospital B is counted as a readmission against Hospital B, not Hospital A. Therefore, increased transfers may be one way to reduce readmissions costlessly.

Table 1.5: Impact of program on gaming behaviour

	Heart attack		**Heart failure**		**Pneumonia**	
	Pre-mean	Estimate	Pre-mean	Estimate	Pre-mean	Estimate
	(1)	(2)	(3)	(4)	(5)	(6)
A. 30-45 Day Readmissions						
(1) TM Below	4.54	0.0007	6.50	-0.0073*	4.43	0.0005
		(0.004)		(0.004)		(0.003)
(2) TM Above	5.44	-0.0084***	7.02	-0.0010	4.85	-0.0003
		(0.003)		(0.002)		(0.002)
(3) Non-TM Below	2.68	-0.0004	6.11	-0.0015	3.22	-0.0033
		(0.004)		(0.005)		(0.003)
(4) Non-TM Above	3.66	-0.0035	6.68	-0.0022	3.48	0.0042*
		(0.002)		(0.002)		(0.002)
B. ED returns (without readmissions)						
(1) TM Below	8.10	-0.0065	8.96	-0.0082**	7.44	-0.0023
		(0.004)		(0.004)		(0.004)
(2) TM Above	7.21	-0.0041	6.89	-0.0025	6.73	-0.0026
		(0.003)		(0.002)		(0.003)
(3) Non-TM Below	6.64	-0.0004	7.83	-0.0085*	5.87	-0.0001
		(0.005)		(0.005)		(0.005)
(4) Non-TM Above	5.96	0.0008	6.22	0.0002	5.34	-0.0021
		(0.004)		(0.003)		(0.003)
C. Transfer to other hospital						
(1) TM Below	3.04	0.0067	1.12	-0.0009	0.40	-0.0011
		(0.007)		(0.002)		(0.001)
(2) TM Above	4.16	-0.0008	0.82	0.0003	0.53	0.0000
		(0.003)		(0.001)		(0.001)
(3) Non-TM Below	3.02	-0.0018	1.49	0.0014	0.36	-0.0022
		(0.004)		(0.003)		(0.002)
(4) Non-TM Above	4.94	-0.0061	1.05	0.0002	0.42	0.0000
		(0.005)		(0.001)		(0.001)

Notes. This table presents the estimated impact of the program on treatment intensity. The regression is based on equation 9 where the dependent variable reflects the outcome of interest. "Above" denotes the effect among hospitals that expect to have an excess readmission ratio based on the observed readmission rates from Jul 2008 to Mar 2010. "TM" denotes Traditional Medicare. Standard errors in parentheses, clustered at the hospital level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results are presented in Table 1.5 and provide very little evidence that the readmission

reductions are driven by hospitals gaming the system. Panel A shows that 31–45 days readmission rates decreased (rather than increased) for heart attack TM patients treated at poor-performance hospitals. In particular, there is a 0.8 pp reduction in readmissions occurring 30–45 days after the first discharge among TM patients treated at hospitals that expect to be penalised; non-TM patients treated at these hospitals see no reduction. It is plausible that the treatment changes that are effective for reducing 30-day readmissions are also effective in reducing longer-term readmission outcomes. For heart failure cases, 31–45 days readmissions fell among TM patients treated at good-performance hospitals (significant at the 10% level). For pneumonia cases, there is some evidence that 31–45 days readmissions increased among non-TM patients treated at bad-performance hospitals.

Panel B examines the impact of the program on ED returns. Column (2) shows that the program did not have an impact on the probability of ED returns for AMI patients, providing little evidence that the fewer readmissions are due to returning patients being held in the emergency department. Column (4) shows that among HF patients, there is some evidence that ED returns fell among both TM and non-TM patients treated at good-performance hospitals by 0.8 pp. Among poor-performance hospitals where readmissions fell, ED return probabilities are not affected. Column (6) indicates that among PN patients, there is no change in ED return probabilities. Lastly, panel C shows that across all monitored diagnoses, there is also no increase in the probability of hospital transfers. This suggests that the reductions in readmissions are unlikely to be due to hospitals shifting riskier patients to other hospitals when they are first admitted.

In summary, the baseline results provide evidence that there are quality improvement for TM patients and spillovers across patient groups in the case of heart attack and heart failure. As suggested by the model in Section 3, these spillovers suggest that there are cost complementarities in quality improvements for these two diagnosis domains. There is little evidence that these readmission reductions are driven by hospitals finding ways to game the system.

7 | Robustness Checks and Confounding Factors

This section investigates the robustness of the baseline results and considers whether there are other factors confounding the interpretation of the baseline results.

7.1 Robustness Checks

To check that the timing of the reductions in readmissions coincides with the period when the HRRP is announced, I estimate dynamic versions of equation (9) by including leads and lags in the regression specification. In particular, I replace the binary pre/post-policy dummy P_{dt} (which equations 1 if diagnosis d is monitored in period t) with a set of dummies $P_{d,t+s}$ that equals 1 if diagnosis d is monitored in period $t+s$. Specifically, I include two pre-policy periods each spanning three quarters (Oct 2008–Jun 2009, July 2009–Mar 2010), two policy implementation periods (Apr 2010–Jul 2011 for the ACA period and August 2011–Jul 2012 for the post-HRRP period), and finally two post-policy periods (Aug 2012–Jul 2013 and Aug 2013–Dec 2014).This more general specification allows the estimated impacts to vary across periods. Coefficients on the ACA and HRRP periods should clarify when the effects of the program start to be observed. I estimate these dynamic models separately for TM and non-TM patients treated at poor-performance hospitals.

I present the time-varying estimated coefficients for both patient groups graphically in Figure A.1 of Appendix A.5. Subfigure A.1a presents the estimates for the AMI sample. It firstly shows that readmission rates started falling during the HRRP implementation period rather than in the year following the ACA. This is consistent with the interpretation that the HRRP is responsible for the fall in readmission rates in the post-policy period. Secondly, it shows that the impact of the HRRP is dynamic, consistent with the notion of learning over time as hospital staff make process adjustments.[50] Third, the figure illustrates the spillover effects of the HRRP, with the trend for non-TM patients tracking that of TM patients. Analogous figures for heart failure and pneumonia are presented in subfigures A.1b and A.1c respectively. Subfigure A.1b shows that HF readmission rates among TM patients began to decrease only in 2013. This figure also indicates

[50] These apparently dynamic effects could also be due to the increasing penalty caps imposed by the HRRP.

that readmission rates for non-TM patients decreased earlier, starting in the post-ACA period, suggesting that the evidence for spillover effects for HF cases is not as robust as that for AMI cases. Subfigure A.1c shows that the impact of the program for PN patients is only noticeable in 2014.

To allow for the possibility that readmission trends evolve differently across insurance groups, I include insurance-group specific linear time trends in Table A.3 of Appendix A.4.[51] This might be the case if there are other policies implemented by non-TM insurers/payers that might affect readmission trends. When these insurance-group specific time trends are included, the results are almost identical. This suggests that the spillover effects documented in Table 1.3 are unlikely to be due to different time trends across insurance groups.

A concern is that the baseline results may be sensitive to the definition of the comparison group. In Table A.4 of Appendix A.4, rather than restricting the potential comparison group to patients admitted for gastrointestinal conditions, I allow any patient admitted for a non-monitored condition, excluding childbirth, to be a potential match (in a similar way, Zuckerman et al. (2016) use patients admitted for any non-monitored condition as a comparison group). The matching algorithm used is the same as that described in Section 4 of this chapter. Given the larger sample size of the potential comparison group, the match rates are now much higher. The match rates are: 99.38% and 98.87% for AMI TM and non-TM patients respectively; 98.93% and 96.96% for HF TM and non-TM patients respectively; and 98.64% and 97.09% for PN TM and non-TM patients respectively. Using this alternative comparison group, the results are very similar for AMI patients. For example, the variation in responses across hospitals by past performance has the same pattern, with poor-performance hospitals responding but good-performance hospitals not responding. When using this broader comparison group, there are some differences for HF and PN patients compared to the baseline results. There are no spillovers to non-TM patients admitted for heart failure. The estimated coefficient for non-TM patients treated at poor-performance hospitals is -0.004. The estimated impact of the program on TM patients admitted for pneumonia is slightly weaker at poor-performance hospitals (estimate=-0.006, p-value=0.058) and there is a curious increase in

[51] This specification still imposes that within an insurance group, readmission trends are the same for monitored and comparison diagnoses.

readmission rates among non-TM patients treated at good-performance hospitals.

To test the robustness of the results to a different definition of the post-policy date, I present results with different post-policy dates in Table A.5 of Appendix A.4. The baseline post-policy date is defined to be admissions after August 2011, when the details of the HRRP were announced. An alternative post-policy date is March 2010 when the ACA was enacted and the HRRP was introduced (without precise details about the penalty structure). The baseline results are still very similar when this alternative post-policy date is used, with the variation across hospital performance-type and diagnoses being the same. The results suggest strong spillover effects in the case of AMI and HF. The main difference compared to the baseline results is that there appears to be an increase in readmission rates for non-TM patients admitted for pneumonia at poor-performance hospitals.

In Table A.6 of Appendix A.4, I use the ratio of risk-adjusted predicted readmission rates to risk-adjusted expected readmission rates for each monitored diagnosis to classify whether hospitals are likely to be affected by the introduction of the HRRP penalties (the CMS approach). In contrast, the baseline specification uses the ratio of observed readmission rate to the risk-adjusted national average rate to classify hospitals. The resulting differences from these two approaches can be seen in column (2). Hospitals that are not poor-performers according to the CMS approach also respond to the program. For example, among TM AMI patients treated at hospitals that do not have an excess risk-adjusted readmission rate according to the CMS approach, readmission probabilities fell by 2.11 pp (p-value=0.0024); among non-TM AMI patients treated at these hospitals, readmission probabilities fell by 2.22 pp (p-value=0.012). These differences suggest that hospitals are conservative in their expectations about future penalties for AMI cases.

Overall, these results suggest that the program had strong direct effects on TM patients admitted for the three monitored conditions. This result is robust to various alternative specifications and sample selection procedures. Regarding spillover effects, there is strong evidence of spillover effects for AMI cases. While our baseline results also provide some evidence of spillovers in the case of HF, these results are not robust to the definition of the control group and further investigation of the timing of when these effects occurred. Lastly, these results confirm that there are no spillover effects for pneumonia cases.

7.2 Other Confounding Factors

This section discusses other confounding factors that may undermine the current interpretation of the baseline results. First, a concern is that the estimates of hospitals' response may also be capturing the effects of other pay-for-performance policies that were implemented around the same time. If there are other contemporaneous policies targeting TM or non-TM patients, this might bias the estimated direct and indirect effects of the HRRP program respectively. An overlapping program targeted at TM patients is the Hospital Value Based Purchasing Program (HVBP) which was implemented in March 2010 under the ACA (Centers for Medicare & Medicaid Services, 2017b). The HVBP targets nearly 30 clinical- and process-quality measures in several domains, none of which explicitly target readmissions. Despite this timing overlap, it is unlikely to be the case that the HVBP is driving the main results. Werner and Dudley (2012) argue that the HVBP is expected only to have very small impacts on hospital payments while Gupta (2017) argues that, relative to the HRRP, the incentives created by the HVBP are low-powered and diffused since hospitals are graded on almost 30 metrics. These views are confirmed by Ryan et al. (2017) who find that HVBP was not associated with improvements in measures of clinical process or patient experience and was not associated with significant reductions in two of three mortality measures monitored by the HVBP.[52]

In addition, a few patterns in the existing results suggest that the incentives created by the HVBP are unlikely to be driving the main results. First, details of the HVBP were announced earlier in March 2010 compared to the HRRP. The extended dynamic models used (see Figure A.1) indicate that readmissions did not fall until 2011, the year when the details of the HRRP were announced. Second, the results indicate that the effects of the program are concentrated among hospitals that have high pre-policy readmission rates. Since the HVBP rewards are independent of this element of hospital performance, one would not expect hospital responses to the HVBP to depend on past readmission rates. Similarly, while it is not possible to control for all ongoing

[52] Another pay-for-performance scheme implemented recently is the Hospital Acquired Condition Reduction Program (HACRP) which reduces reimbursement for hospitals that rank among the lowest performing quartile in regard to hospital-acquired conditions (Centers for Medicare & Medicaid Services, 2017a). The details of this program were announced in July 2014 and thus is unlikely to be a confounding policy in this context (Sankaran et al., 2019).

non-TM policies that might affect readmission rates over this period, there are two reasons why this is unlikely to be driving the estimated spillover effects. First, when insurance-group specific time trends are included (see Table A.3), the estimates hardly change. Second, for such unobserved policies to be driving the impacts on non-TM patients, these policies would need to be announced at the same time as the HRRP, target the same diagnoses, and only be implemented in hospitals that expect to be penalised under the HRRP.

Second, a concern is that the indirect impact of the policy on non-TM patients is not due to cost complementarities, but due to physicians' lack of knowledge about the type of insurance a patient has. If physicians are unable to differentiate between patients with different insurance policies, they may exert equal effort into reducing readmissions for all patients admitted for a monitored condition. While this can't be ruled out with the existing data, qualitative evidence on the medical billing processes suggests that physicians generally have knowledge of the insurance policy that the patient has or has the ability to find this out. For example, most descriptions of the medical billing process state that patients must declare their insurance policy when they check in as an inpatient and therefore this information is available to medical staff during the patient's hospitalisation.[53]

[53] See https://www.medicalbillingandcodingonline.com/medical-coding-for-billers/ for one example.

8 | Conclusion

Many policies incentivise providers such as hospitals to improve outcomes for specific subgroups. Since providers often serve multiple groups simultaneously, such policies might have indirect effects on the non-targeted subgroups. Theory suggests that the impact of asymmetric incentives across groups depends on whether there are complementarities in production for the two groups. This chapter aims to understand the impacts of a policy that creates asymmetric incentives across subgroups in a healthcare setting and what such impacts imply about the quality provision process.

Using the Hospital Readmissions Reduction Program as a source of variation in the financial incentives that hospitals face, I provide evidence that the program (i) reduced readmission rates among the targeted patient group and that (ii) there are some spillovers to non-targeted patients for some diagnosis domains. In particular, among hospitals that could reasonably expect to be penalised by the program, heart attack readmission rates among the targeted patient group fell by over 3.3 pp while it fell by almost 2.4 pp among the non-targeted patient group. For heart failure cases, readmission rates fell by 0.7 pp and 0.9 pp for targeted and non-targeted patients respectively. For pneumonia cases, hospitals responded by reducing readmission rates for targeted patients; but the degree of spillover to non-targeted patients is not statistically significant. There is some evidence that treatment intensity, measured by the number of procedures received, increased for AMI and HF targeted patients treated at hospitals that expect to be penalised. These reductions in readmissions are not driven by hospitals gaming the system, suggesting they reflect improvements in process or structural quality.

The findings of this chapter are important for several reasons. First, in many healthcare systems, multiple actors (e.g. governments and private insurance companies) pay providers, leading providers to face a mix of incentives if these payment schemes are misaligned. Understanding how providers respond to mixed incentives is important for improving the design of payment systems. Second, given the interest in containing healthcare costs while improving healthcare quality, understanding any potential cost savings from complementarities is important.

Chapter II

The Impacts of Health Insurance on Readmissions: Evidence from Heart Attack Patients on Medicare

Chapter Abstract

Policy-makers often use post-discharge outcomes such as readmissions to measure hospital quality. The accuracy of these quality measures depends on the main drivers of such outcomes. Existing academic and policy debates focus on provider-side factors as the main driver. This chapter investigates the impacts of universal eligibility for Medicare insurance at age 65 on readmissions among heart attack patients, focusing on patient-side responses to changes in insurance eligibility. Using the age-65 eligibility threshold for Medicare as a source of variation in insurance coverage, I estimate discontinuities in 30-day readmission probabilities and treatment intensity at the age threshold. Among men, there is robust evidence that eligibility for Medicare insurance increases 30-day readmission probabilities by over 2.5 percentage points at age 65. Treatment intensity (measured by invasiveness and number of treatments, timely diagnosis, length of stay) is smooth across the age threshold, suggesting that the increase in readmissions is not due to changes in treatment intensity at the initial hospitalisation. Insofar as observable treatment procedures primarily capture supply-side responses, this finding suggests that the discontinuity primarily reflects demand-side responses to changes in insurance eligibility. Among women, there is no evidence that eligibility for Medicare insurance affects readmissions. Importantly, there are significant increases in the probability of female heart attack patients receiving more intensive treatments at the age threshold. The discontinuity in treatment intensity highlights gender disparities in treatments received and gender-specific impacts of insurance eligibility on treatment intensity.

Chapter Notation

- h = Health
- b = Positive health benefit
- $\mu \in [0, 1]$ = Adverse health shock
- $v(c, h)$ = Patient utility from consumption (c) and health (h)
- m = Medical expenditure
- ϕ = Co-payment rate
- I = Income
- q = Quality of care
- $\mathbf{C}$ = Insurance coverage characteristics
- rv = Running variable

1 | Introduction

The findings of the previous chapter suggest that supply-side factors matter for post-discharge outcomes: shifting providers' incentives helped decrease readmission rates for several diagnoses. These findings therefore lend support to the common assumption made in healthcare policy debates that providers are accountable for poor post-discharge outcomes (Benbassat and Taragin, 2000). Compared to the focus placed on supply-side factors, less attention has been devoted to understanding how demand-side determinants affect post-discharge outcomes. This contrasts with a long literature on moral hazard in healthcare utilisation, which places the role of patient incentives at the forefront.

This chapter examines the impacts of eligibility for Medicare insurance on readmissions, focusing on the role of demand-side responses to changes in insurance eligibility. The first way in which insurance may affect readmissions is via changes in patient-side factors. On the one hand, insurance lowers out-of-pocket costs and encourages the utilisation of healthcare facilities after a hospital discharge. This additional healthcare utilisation may uncover medical problems that require a readmission. In line with this, Chen et al. (2014) and Lasser et al. (2014) find that insurance expansion is associated with an increase in readmission rates in Massachusetts. On the other hand, insurance may increase access to affordable medications after discharge, which could lower the risk of readmissions (Choudhry et al., 2011). Insurance coverage may give previously uninsured patients the ability to seek more timely care in the first instance, lowering the likelihood of being readmitted. The second way in which insurance may affect readmissions is via changes in provider-side factors such as hospital incentives. As suggested by the previous chapter, different insurance policies may reimburse hospitals in different ways for both the initial stay and any subsequent readmission. Hospitals may therefore be incentivised to tailor treatment choices for patients during their initial stay based on their insurer, which may in turn affect their post-discharge outcomes.

Methodology. To examine the relationship between health insurance and readmissions, I begin by outlining a simple model of patient behaviour post discharge. Following a hospital discharge, a patient may experience a repeated negative health shock which leads them to seek additional

medical care. This decision to seek additional care depends on the out-of-pocket price they have to pay (and therefore on their insurance status) and the severity of the repeated health shock. This feature of the model implies that patients who have to pay less out-of-pocket are more likely to seek additional care at a given level of their illness severity. In the model, I allow hospitals to base their quality of care on the out-of-pocket costs paid by the patient. This helps account for the possibility that patients with different insurance policies receive different treatment, driving variation in health outcomes across insurance categories. This feature of the model implies that if hospitals base quality of care on the patient's out-of-pocket costs, then the patient's post-discharge behaviour reflects both patient incentives and quality of treatment received.

The empirical analysis focuses on patients admitted for heart attack at acute care hospitals in New York State. In this sample, readmission rates vary significantly by insurance category, suggesting a potential relationship between insurance and post-discharge outcomes. Among patients between 55 and 75 years old, those with Medicare insurance are 6 percentage points (pp) more likely than those with private or no insurance to be readmitted within 30 days of discharge. This variation of post-discharge health outcomes across insurance categories holds more generally across the U.S. (Barrett et al., 2015).

The main challenge in estimating the impact of insurance on post-discharge outcomes is that insurance coverage is likely to be endogenous to other factors that determine post-discharge outcomes. To deal with this, I use a regression discontinuity (RD) design and exploit the age-65 eligibility threshold for Medicare insurance. This age threshold is associated with various changes in insurance characteristics but not with discontinuous changes in other characteristics that might affect readmissions (Card et al., 2008, 2009). For example, there are sharp changes in the number of insurance policies that individuals have. Some patients gain some form of insurance coverage for the first time while others receive secondary insurance coverage in addition to pre-existing policies. Changes in the main source of insurance imply changes in the benefits received and restrictions faced under the insurance package. This age threshold therefore reflects various changes in insurance characteristics (e.g. changes in the number of insurance policies, the source of insurance coverage, and the benefits received) that may induce patient- or provider-side responses.

The main empirical specification in this analysis is a reduced-form regression discontinuity (RD) model that exploits the age-65 threshold to estimate the causal effect of universal eligibility for Medicare insurance at age 65 (henceforth, the effect of "eligibility for Medicare insurance") on readmission probabilities. The key identification assumption is that assignment to either side of the age threshold is as good as random. I provide evidence supporting this assumption; for example, the density of the number of heart attack patients is smooth around the age threshold and there are no discontinuities in other observable characteristics that might affect post-discharge outcomes.

Using this RD specification, I first estimate discontinuities in readmission probabilities at the age threshold separately for men and women. Estimating the RD models separately for men and women is useful because existing medical studies provide evidence that there are gender differences in the timing of when heart attacks occur, the symptoms experienced, and the treatments received (Maas and Appelman, 2010). In particular, cardiovascular diseases usually develop 7 to 10 years later in women than in men. The symptoms experienced by women may be different; for example, a relatively high proportion of women experience nausea rather than chest pain. Multiple studies have also shown that women are less likely than men to receive an intensive treatment when admitted for a heart attack (Blomkalns et al., 2005; Jneid et al., 2008; Maynard et al., 1992). The differences in time of onset, symptoms, and treatment of heart attacks suggest potential gender-specific impacts of health insurance on readmissions.

Second, to investigate whether the relationship between eligibility for Medicare insurance and readmissions could be driven by changes in treatment quality at the age threshold, I use the fact that heart attack treatments are broadly characterised by two competing management approaches. The first is an invasive approach marked by cardiac catheterisation shortly after admission followed by surgery (e.g. angioplasty or bypass surgery); the second is a conservative approach in which the patient is given drugs and then monitored (Molitor, 2018). It has been documented that patients who receive the intensive treatment early tend to have better health outcomes (Bhatt et al., 2004; Currie et al., 2016; Cutler et al., 1998; Winter et al., 2005). Drawing on this medical finding, I estimate discontinuities in various measures of treatment intensity at the age threshold to investigate whether insurance eligibility impacts treatments received.

Main Results. Among male heart attack patients, there is evidence that eligibility for Medicare insurance increases the probability of being readmitted within 30 days of discharge. There is a discontinuous increase of over 2.5 pp in the probability of 30-day readmissions at age 65, compared to a baseline rate of 11% among 63–64 year old male patients. This increase is driven by patients returning for non-major vascular events such as non-specific chest pain and coronary atherosclerosis. There is no significant increase in readmission rates for male patients returning with more critical conditions, such as a repeated heart attack. Since insurance is likely to have a stronger effect on patients returning with mild conditions compared to those with critical conditions (who are more likely to receive medical attention regardless of insurance status), the heterogeneous impacts across reasons for readmissions supports the interpretation that this increase captures the effect of eligibility for Medicare insurance.

Furthermore, the discontinuity is concentrated among non-white patients and those living in poorer zipcode areas, subgroups with the highest pre-65 rates of readmission and lowest pre-65 rates of insurance coverage. This heterogeneity in impacts across demographic subgroups is consistent with the interpretation that eligibility for Medicare insurance increases access to post-discharge medical services among patients who would not otherwise have sought or received care. In contrast, for female heart attack patients, there is no evidence that eligibility for Medicare insurance increases the probability of being readmitted within 30 days of discharge. The estimated discontinuity in the probability of readmission is negative but statistically insignificant. These results for men and women are robust to bandwidth choice, functional form specifications, the inclusion of various other controls, and alternative sample selection procedures.

Among male heart attack patients, there are no discontinuities in observed treatment intensity at the age threshold, where intensity is measured by invasiveness of treatment, timely diagnosis, length of stay, and number of procedures. This suggests that the discontinuity in readmissions for men is not driven by discontinuities in treatments received. Insofar as (observable) treatment procedures primarily capture supply-side responses, this finding suggests that demand-side responses to changes in insurance eligibility are likely to be the main drivers of the discontinuity in readmission probabilities at the age threshold for men.

In contrast, for women, there is a discontinuous increase of over 6.5 pp in the probability of receiving an intensive treatment at age 65, relative to a baseline of 80% among 63–64 year old female patients. This effect is concentrated among white female patients and those living in higher-income zipcode areas, groups where the switch from private to Medicare insurance is the largest at age 65.[54] This discontinuity in treatment intensity for women suggests that the estimated discontinuity in readmissions likely captures both the effect of changes in patient- and provider-side incentives associated with eligibility for Medicare insurance.

The discontinuity in treatment intensity among female patients has several important implications. First, it confirms the previous literature's documentation of gender differences in heart attack treatments. Among 63–64 year old patients, women are 3–4 pp less likely than men to receive an intensive treatment. This gap in treatment intensity may be one reason why pre-65 readmission rates are higher among women than men. Second, it illustrates the gender-specific effects of Medicare eligibility on intensity of treatment among heart attack patients. The asymmetric effects of eligibility for Medicare insurance between men and women reduce the gender disparity in treatment intensity at the age threshold.

Related Literature and Contribution. This chapter is related to several strands of literature. First, there is a handful of papers that provide evidence that access to health insurance increases readmission rates.[55] Two recent papers examine the impact of Massachusetts' health insurance expansion on 30-day readmissions. Lasser et al. (2014) use a difference-in-differences design to compare changes in readmissions in Massachusetts and readmissions among control hospitals in New York and New Jersey. They find a slight increase in the probability of all-cause readmissions in MA relative to the control states. Chen et al. (2014) examine the impact of Massachusetts' health insurance expansion on readmissions by using a hospital-level difference-in-differences design. Treated hospitals are those that initially treat many uninsured patients and therefore are most

[54] Under the Affordable Care Act, it is not a legal requirement for employers to offer coverage to spouses. However, the majority of employers do offer coverage to their spouses (Kaiser Family Foundation, 2018). It is increasing common for employers to restrict spousal coverage to spouses who don't have access to their own employer-sponsored plan, or to add a surcharge to their premiums if they have access to their own employer-sponsored plan but choose to be added to a spouse's plan instead.

[55] See http://insights.patientbond.com/blog/how-health-insurance-impacts-hospital-readmissions for a general discussion. [Accessed 11 May 2018]

affected by the expansion. The authors find that hospitals with the highest proportion of uninsured patients experienced a modest increase in their unadjusted readmission rate of 0.6 pp compared to other hospitals.

This chapter differs from these two preceding papers in the following ways. First, this chapter investigates potential supply-side responses to insurance eligibility (e.g. changes in medical treatments) that might generate a relationship between health insurance eligibility and readmissions. Second, this chapter examines a different setting. While the existing literature investigates the impact of expanding insurance to low-income individuals, this chapter focuses on the effect of gaining old age health insurance. This setting is of interest because Medicare is the largest government-funded health insurance program in the United States and increasing costs, including costs for post-discharge care, have been a concern for healthcare policy-makers.

Second, a large medical literature examines the key determinants of readmissions, focusing on observable patient and hospital characteristics (Jencks et al., 2009; Krumholz et al., 2017). The goal of this literature is to identify the patient groups that would benefit most from timely intervention. In a review of existing risk prediction models, Kansagara et al. (2011) write that most current risk prediction models that are designed for clinical purposes perform poorly. These prediction models do not examine the impact of insurance on readmissions.

Finally, this chapter is related to a more general health economics literature on the impact of insurance on healthcare utilisation and health outcomes (e.g. Anderson et al., 2014; Finkelstein et al., 2016). In terms of research design, this chapter is closest to Card et al. (2008) and Card et al. (2009) who use the Medicare eligibility threshold at age 65 to compare health-related outcomes and utilisation among people just before and after age 65. While the main focus of Card et al. (2009) is on utilisation and mortality,[56] the authors provide some evidence that readmission rates decrease at age 65 for patients treated at Californian hospitals. Whether this is driven by demand- or supply-side forces is ambiguous. While this literature usually interprets readmissions as a health outcome, this chapter suggests that it could also be interpreted as a form of healthcare utilisation

[56] Card et al. (2009) find that Medicare eligibility causes a sharp increase in the use of elective healthcare services, with a pattern of gains across groups that varies by the type of service. Furthermore, 7-day mortality rates drop at age 65.

among a group of individuals who are already utilising hospital services.

The rest of this chapter is organised as follows. In Section 2, I use a simple model to describe why there may be a relationship between health insurance and post-discharge outcomes. In Section 3, I provide some background information about the setting of the analysis, the data sources used, and how the variables are constructed. In Section 4, I present the empirical strategy used. In Section 5, I present the estimated impacts of eligibility for Medicare insurance on readmission rates. In Section 6, I discuss the impact of eligibility for Medicare insurance on observable treatment quality. In Section 7, I conclude and consider avenues for future research.

2| Conceptual Framework

This section describes why there may be a relationship between health insurance and readmissions through the lens of a simple model. Mathematical details underlying the predictions of this model can be found in Appendix B.1.

The first way through which insurance may affect readmissions is via changes in patient-side factors. Demand-side responses to changes in insurance eligibility may have conflicting effects on readmissions. On the one hand, insurance lowers out-of-pocket costs and encourages the utilisation of healthcare facilities after a hospital discharge.[57] This additional healthcare utilisation may uncover medical problems that require a readmission. For example, insurance may encourage patients to seek care at the emergency department for mild complications after an initial discharge which may then lead to a readmission (Brennan et al., 2015). On the other hand, insurance may also increase access to affordable medications after discharge, which could lower the risk of readmissions and revisits.[58] Insurance coverage may also give previously uninsured patients the ability to seek more timely care in the first instance at various medical institutions. This may help lower the probability of readmissions if the effectiveness of treatment for the initial condition is time-sensitive. The net effect of these conflicting effects may be different for different conditions.

The second way in which insurance may affect readmissions is via changes in provider-side factors such as hospital incentives. Since different insurance policies reimburse hospitals in different ways, hospitals may be incentivised to tailor treatment choices for patients based on their insurer. For example, insurance policies that reimburse hospitals for each day of the hospital stay may encourage longer initial hospitalisations; insurance policies that reimburse hospitals by procedures performed may encourage more costly treatment procedures. Furthermore, hospitals' inclination to readmit patients may depend on payments for readmissions, which may also depend on insurance.[59]

[57] A large literature documents moral hazard effects of insurance. For example, see Finkelstein et al. (2018) for a review.

[58] One of the potential effects of crossing the age 65 threshold is prescription drug coverage via Medicare Part D which came into effect on Jan 1, 2006. According to the Kaiser Family Foundation, in 2017, over 70% of Medicare beneficiaries are enrolled in Part D plans.

[59] Hospitals may also prefer to receive payment when the patient is discharged rather than wait for a reimbursement from the insurance company. Hospitals may therefore value patients who pay more out-of-pocket.

To put these elements into a model, consider a patient i who is discharged from hospital. The patient derives utility from consumption c_i and health h_i. The patient has the following quasi-linear utility function:

$$u(c_i, h_i) = v(c_i) + h_i \tag{10}$$

where $v'(c_i) > 0$ and $v''(c_i) < 0$. The assumption that utility is quasi-linear in consumption and health implies that (i) there are no income effects from consumption and that (ii) illness affects utility, but not the marginal utility of consumption. Without income effects, any change in the quantity of consumption is due to the relative price of the two goods. The property that illness doesn't affect the marginal utility of consumption means that utility from consumption doesn't depend on how healthy the patient is.[60]

Patient health. I assume that the patient's health level h_i is determined by the following equation

$$h_i = n_i - \mu_i + b_i$$

The term n_i captures the patients "natural" health endowment, which is assumed to be fixed prior to the initial admission. The term $\mu_i \in [0, 1]$ captures (adverse) health shocks that the patient may suffer from after being discharged. μ_i can be interpreted as either the probability that a patient suffers from a health shock (of the same severity across patients) or the severity of the health shock that is heterogeneous across patients. For example, if $\mu_i > 0$, then the patient suffers from a bad health shock such as a medical complication. If $\mu_i = 0$, then the patient does not suffer from any further illness after being discharged. If the patient experiences a repeated illness, she can seek additional care. This will produce a monetised health benefit $b_i := f(\mu_i)$, the last term in the health equation. I assume that f is increasing in μ_i so that health benefits from treatment are greater for those who are sicker (i.e. have a higher μ_i). If the patient does not suffer from a repeated adverse health shock, they are assumed not to seek care so that $f(0) = 0$. Lastly, patients are assumed to know μ_i and b_i.

[60] Viscusi and Evans (1990), Evans and Viscusi (1991), and Finkelstein et al. (2013) consider utility functions that allow the marginal utility of consumption to depend on health.

I assume that a proportion of patients who seek additional care are subsequently readmitted to hospital so that there is a positive monotonic relationship between the proportion seeking additional care and the proportion being readmitted. This model therefore just focuses on the decision to seek additional care.

Insurance. The patient's decision to seek additional care depends on their health insurance coverage and their experienced illness severity, where insurance coverage is exogenously determined. An insurance policy is defined as a contract between the insurer and insured specifying the insurance premium p and the co-payment schedule. In particular, let m_i be the medical expenditure of patient i. Then, the insurance contract specifies the co-payment rate $0 < \phi < 1$ and out-of-pocket costs of the patient $\phi m_i \leq m_i$. The premium p may depend on the patient's (exogenously determined) income I_i. More generous insurance packages are assumed to have lower co-payment rates (ϕ) and lower premia (p).

If the patient is insured by a contract that specifies a co-payment rate of ϕ and premium of p, income available for consumption when healthy is $I_i - p(I_i)$ and income available for consumption when sick is $I_i - p(I_i) - \phi m_i$. Therefore, the budget constraint of an insured patient can be written as:

$$c_i + \phi m_i \mathbb{1}\{\mu_i \geq \bar{\mu}_i^{INS}\} + p(I_i) \leq I_i \tag{11}$$

where $\bar{\mu}_i^{INS}$ is the threshold illness severity level above which insured individual i seeks additional care. On the other hand, if the patient is uninsured, the patient pays all of the medical payments and no premium. Their budget constraint is:

$$c_i + m_i \mathbb{1}\{\mu_i \geq \bar{\mu}_i^{UN}\} \leq I_i \tag{12}$$

where $\bar{\mu}_i^{UN}$ is the threshold illness severity level above which uninsured individual i seeks additional care

Patient's decision. The patient's utility from seeking or not seeking additional care can be expressed as:

$$u_i = \begin{cases} v(c_i^{\neg care}) - \mu_i & \text{if does not seek care} \\ v(c_i^{care}) + f(\mu_i) - \mu_i & \text{if seeks care} \end{cases} \tag{13}$$

where $c_i^{\neg care}$ is the consumption level if the patient does not seeks care and c_i^{care} is the consumption level if the patient seeks care. For an insured patient, $c_i^{\neg care} = I_i - p(I_i)$ and $c_i^{care} = I_i - \phi m_i - p(I_i)$. For an uninsured patient, $c_i^{\neg care} = I_i$ and $c_i^{care} = I_i - m_i$. Comparing the optimal decisions of the insured and uninsured patient, one arrives at the following prediction:

Prediction 1. *For a given level of illness severity μ_i, an insured patient is more likely to seek care than an uninsured patient.*

Prediction 1 suggests a positive relationship between insurance and readmission rates. A patient will seek care if the (marginal) health benefit of doing so is greater than the marginal cost of not doing so (the difference between not returning and returning). For an insured patient, the utility difference between seeking and not seeking care is smaller than that for the uninsured patient since medical expenditures are smoothed out for the insured. This inequality is therefore easier to satisfy for the insured patient.

Hospital quality provision. While the decision to seek additional care after discharge is primarily the choice of the patient, it is possible that hospitals can indirectly affect this choice by influencing the quality of care provided. In particular, a patient who received better quality of care at the initial hospitalisation may be less likely to suffer from repeated illness and/or suffer from less severe levels of illness. To allow for this possibility, I now consider a case where (i.) severity of the repeated health shock is a function of quality of care q provided initially by the hospital so that $\mu_i = \mu_i(q)$ and (ii.) providers choose the quality of care based on the patient's insurance status.

To model this, I assume that quality q depends on the proportion of costs paid out of pocket (ϕ) and investigate how small changes in ϕ affect the probability of returning for care.[61] In this

[61] For simplicity, this model assumes that out-of-pocket costs ϕm_i do not depend on quality.

setting, a change in the co-payment rate (e.g. an increase in $1-\phi$ so that the patient pays less out of pocket) has two effects: it changes the quality of care provided by the hospital and it changes the patient's incentive to seek additional care. The overall effect depends on whether and how hospitals respond to $1-\phi$ and, if so, the relative magnitude of the patient effect and the hospital effect. Define the function $E_i := f(\mu_i) - v(c_i^{\neg care}) + v(c_i^{care})$. This function determines whether the patient seeks additional care: if $E \geq 0$, then the patient will seek care. Since μ_i and c_i^{care} depend on the co-payment rate ϕ:

$$\frac{\partial E}{\partial(1-\phi)} = \underbrace{\frac{\partial f}{\partial \mu}\frac{\partial \mu}{\partial q}\frac{\partial q}{\partial(1-\phi)}}_{\substack{\text{effect via} \\ \Delta \text{ quality}}} + \underbrace{\frac{\partial v(c)}{\partial c}\frac{\partial c}{\partial(1-\phi)}}_{\substack{\text{effect via} \\ \Delta \text{ patient incentive}}} \tag{14}$$

The first term in equation (14) reflects the impact of a change in the co-payment rate on the monetised benefit of seeking care. This effect operates via the hospital's choice of quality of care in response to the co-payment rate. Under the assumptions that $f'(\mu_i) > 0$ so that health benefits are higher for sicker patients and $\mu_i'(q) < 0$ so that illness severity is decreasing in quality of care, the sign of this term is determined by whether hospitals provide better or worse quality of care to patients who pay less out of pocket (the sign of $\partial q/\partial(1-\phi)$).

The second term in equation (14) reflects the impact of a change in the co-payment rate on the value of consumption in the state where the patient seeks care. In particular, since $\partial c/\partial(1-\phi) = m$, this term is equal to the marginal utility of consumption ($\partial v/\partial c$) multiplied by the medical payments of the patient (m). The intuition behind this term is the same as that underlying prediction 1: when the patient has to pay less out of pocket, the cost of seeking care in terms of foregone consumption has decreased and therefore the patient is more likely to seek care.

The equation above shows that if hospitals change quality of care in response to the patient's out-of-pocket costs, then this estimated effect may not purely capture the impact of patient's incentives. This leads to the following prediction:

Prediction 2. *Assume that returns to care are increasing in illness severity ($f'(\mu_i) > 0$) and that illness severity is decreasing in quality of care received ($\mu_i'(q) < 0$). $\partial E/\partial(1-\phi)$ captures different objects depending on hospital behaviour. There are three possible cases:*

- $\frac{\partial q}{\partial(1-\phi)} = 0$*: In this case, out-of-pockets costs do not influence quality of care and* $\partial E/\partial(1-\phi)$ *only captures the patient incentive effect.*
- $\frac{\partial q}{\partial(1-\phi)} < 0$*: In this case, patients who pay more out of pocket receive better quality of care.* $\partial E/\partial(1-\phi)$ *will necessarily be positive and it will be a combination of the quality effect and patient incentive effect.*
- $\frac{\partial q}{\partial(1-\phi)} > 0$*: In this case, patients who pay less out of pocket receive better quality of care. The sign of* $\partial E/\partial(1-\phi)$ *is ambiguous and will depend on whether the patient incentive effect outweighs the quality effect.*

Prediction 2 suggests that if hospitals respond to the out-of-pocket costs paid by the patient, then the estimated effect of having insurance coverage on readmission rates may be an over- or under-estimate of the true patient incentive effect depending on the direction of the hospital's response. If the hospital provides worse care for those who pay less out of pocket, then a higher readmission rate among this group may not indicate patient-side responses to insurance. Rather, it may reflect worse care received leading to a more urgent need for medical attention. Conversely, if the hospital provides better care for those who pay less out of pocket and yet this group exhibits higher readmission rates, then this suggests that the demand-side effects are sufficiently large to offset any positive health benefits incurred from better quality of care.

3 | Background and Data Sources

To test the predictions of the model described in the previous section, I require variation in insurance coverage and a method to measure treatment quality. This section first provides details of the institutional and medical setting that enables these predictions to be tested. I then proceed to describe the data sources used and how the relevant variables are constructed.

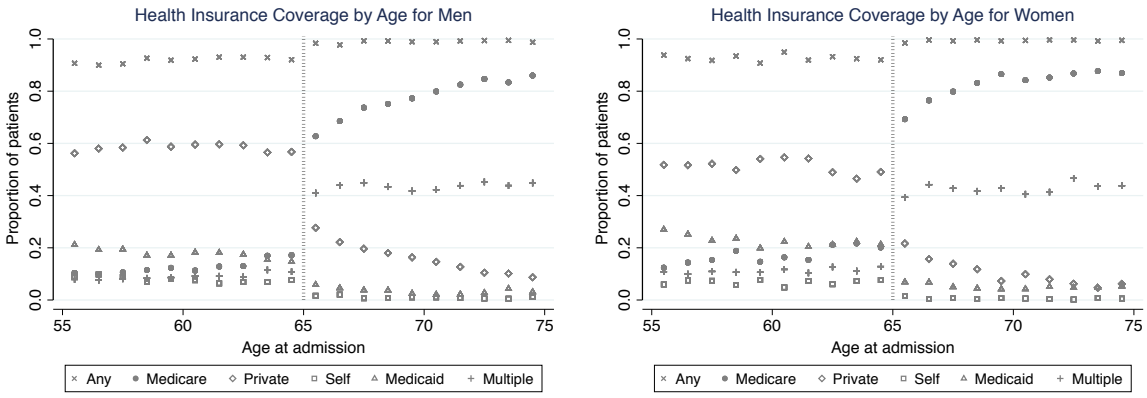
3.1 Background

Changes in Health Insurance Coverage

In the United States, health insurance characteristics shift at age 65 as individuals become eligible for public insurance under Medicare (Card et al., 2008, 2009). These changes are shown in Figure 2.1 which plots various insurance-related variables for the sample of heart attack patients in New York State by age at admission and gender. Although the age-65 threshold primarily symbolises eligibility for Medicare, it is associated with various changes in insurance status: whether an individual has any insurance coverage (i.e. an extensive margin change), the primary source of insurance coverage (e.g. private vs public), and the generosity of the insurance coverage (e.g. co-payment rates). The empirical analysis of this chapter uses this age threshold as a source of variation in Medicare insurance eligibility with the understanding that it reflects various changes in the insurance characteristics discussed.

As discussed in the thesis introduction, Medicare is the main public health insurance program. Nationally, around 75% of Medicare beneficiaries are enrolled in Traditional Medicare (TM) where they can freely choose their providers. The remainder are enrolled in Medicare Advantage (MA), which is an alternative Medicare plan run by private insurance companies but regulated by the government. TM comes in several parts. Part A is hospital insurance and covers inpatient hospital care and is available to people who are at least 65 and have worked 40 quarters or more in covered employment (or have a spouse that did).[62] Medicare coverage is also available to younger people

[62] It also covers other services such as skilled nursing facilities, hospice care, lab tests, and home health care.



(a) Men **(b)** Women

Figure 2.1: Discontinuities in various insurance characteristics by age at admission. Insurance coverage is identified using the primary expected payer on the medical billing record. "Multiple" denotes having more than one insurance coverage that is not the patient themselves.

with severe kidney disease and recipients of Social Security Disability Insurance. Most individuals don't pay a premium, but some have to pay a premium for this coverage (e.g. those who are under 65 and receive coverage due to chronic health conditions). Part B covers outpatient care, preventive services, and durable medical equipment. MA plans must offer coverage that is comparable to original Medicare parts A and B. Most also include prescription drug coverage, which is an optional add-on called Part D for those on TM.

Under Medicare Part A, for each benefit period between one and 60 days, Medicare beneficiaries pay a deductible and no coinsurance. A benefit period begins when a patient is admitted as an inpatient in a hospital and ends when the patient has not received any hospital care for 60 days in a row. For benefit periods above 60 days, individuals pay a coinsurance rate per day. This benefit period implies that most Medicare enrollees pay no coinsurance for 30-day readmissions.

Figure 2.1 shows that there is a large jump in the proportion of patients insured by Medicare at age 65. Among male patients (subfigure a), the proportion of patients insured by Medicare jumps from 17% at age 64 to 63% at age 65; among female patients (subfigure b), this proportion jumps from 20% to 69%. The proportion covered by Medicare increases with age, reaching 87% coverage for both genders at age 75.

At age 65, the proportion of patients with private insurance or no insurance drops as most transition from employer-based private insurance (or no insurance) to Medicare. Within the sample

of heart attack patients, over 58% of male patients and over 50% of female patients under 65 are privately insured; roughly 8% of male patients and 7% of female patients are uninsured.[63] For both genders, among the under-65 privately-insured, 42% are covered by a Blue Cross Blue Shield insurance plan while 41% are part of a health maintenance organisation (HMOs).[64] Among patients above age 65, less than 15% are on private insurance and less than 1% are uninsured.

Although Medicare and private insurance are the main insurance categories observed in the data, a non-negligible fraction of under-65 year old patients are insured by Medicare and "other" insurance schemes. Within my sample, 15% of male heart attack patients are insured by Medicaid at age 64 and this figure drops to 5% at age 65. For female patients, Medicaid coverage drops from 21% at age 64 to 7% at age 65. The "other" insurance category includes individuals with military health benefits offered to retired service members and individuals who purchase a state-sponsored health insurance plan (usually because they are medically uninsurable through private health insurance). Among male patients, roughly 3% of under-65 year olds and 2% over over-65 are insured by "other" insurance groups. Among female patients, the figures are 1.8% and 1%.

Figure 2.1 also shows that the proportion of patients with multiple health insurance policies and the proportion of patients with any health insurance coverage both rise at age 65. Overall, the age-65 threshold is associated with various changes in insurance characteristics: whether or not a patient has insurance and what type of insurance they have, which affects the benefit received.

Heart Attacks

The empirical analysis in this chapter focuses on patients admitted for heart attacks. Heart attacks, clinically referred to as acute myocardial infarction (AMI), occur when part of the heart's blood supply is blocked, starving the heart of oxygen and causing muscle cells to die. There are several reasons for focusing on this group of patients in the empirical analysis.

[63] Compared to the general inpatient population under 65, the proportion of uninsured patients is slightly higher in the heart attack sample ($\sim 8\%$ in the heart attack sample and $< 5\%$ in the general inpatient sample).

[64] The Blue Cross Blue Shield Association is a federation of 38 separate health insurance organisations and companies in the US. HMOs are a type of managed care organisation that provides a form of health care coverage that is fulfilled through hospitals, doctors, and other providers with which the HMO has a contract. Unlike traditional indemnity insurance, an HMO covers only care rendered by those doctors and other professionals who have agreed to treat patients in accordance with the HMO's guidelines and restrictions.

First, heart attack treatments are characterised by two competing treatment approaches. The first approach involves providing patients with an invasive treatment such as an angioplasty (balloon dilation of the blocked artery, with or without stenting) or open-heart bypass surgery (artery graft to "bypass" the blockage) soon after diagnosis. To determine whether a patient is a candidate for an invasive procedure, the physician uses a diagnostic technique called angiography. This procedure is usually included as part of a cardiac catheterisation in which a thin catheter is threaded into the coronary arteries. Cardiac catheterisation is often used as an indicator for invasive heart attack treatment (Chandra and Staiger, 2007; McClellan and Newhouse, 1997). The second treatment approach is a conservative approach in which patients are first given thrombolytic drugs and receive cardiac catheterisation only if symptoms persist. In this approach, thrombolytic drugs are used to dissolve blood clots blocking coronary arteries and are typically most effective when administered within three hours after the heart attack occurs. It has been documented that heart attack patients who receive the early invasive treatment tend to have better health outcomes (Bhatt et al., 2004; Currie et al., 2016; Winter et al., 2005). This classification allows one to check for discontinuities in treatment intensity at the age threshold which may in turn result in discontinuities in post-discharge outcomes.

Second, since heart attacks are an emergency condition and require immediate hospitalisation, it is unlikely that differential health outcomes are due to patients delaying care needed in their first admission. This means that selection into the sample of AMI-related index hospitalisations is not likely to be driven by insurance status. Using inpatient data from California, Card et al. (2009) provide evidence that selection is not a concern for emergency conditions. I provide support for this assumption in Section 4 by showing that the density of the number of heart attack inpatients is smooth across the age threshold. In contrast to emergency conditions like heart attacks, for elective procedures such as hip/knee replacements, individuals close to 65 may delay the procedure until they are eligible for Medicare (Card et al., 2009). Since the research design of this chapter uses the age-65 threshold for Medicare as a source of variation, it is important that there is no differential sample selection before and after the threshold.

Third, although there are many different types of heart attacks, focusing on one diagnosis group limits the extent to which outcomes are driven by the variation in diagnosis types. Heart attack

patients are likely to have similar underlying health conditions and require similar medical attention and treatment. By focusing on heart attack patients, this limits the extent to which variation in post-discharge patterns are driven by unobserved health heterogeneity of patients suffering from different conditions. Section 5 also shows that the baseline results hold when the sample is restricted to one type of heart attack.

Finally, many economic studies examining the impact of provider-side factors and outcomes (e.g. hospital efficiency) tend to focus on the outcomes of heart attack patients as a measure of performance (e.g. Chandra and Staiger, 2007; Skinner et al., 2005). One reason why heart attack is the condition most studied is that these patients are nearly always treated at the same hospital where they are taken to an emergency room. This makes it easy to assign patients to particular hospitals. In contrast, patients who suffer from other conditions might be treated at multiple hospitals.

3.2 Data

The following analysis uses the Healthcare Cost and Utilization Project's (HCUP) New York State (NYS) Inpatient Database from 2010 to 2013. The data contains the universe of inpatient discharge records in acute-care community hospitals in NYS. Since heart attacks are an emergency short-term condition, this database is a suitable source for analysis. A key benefit of this database is that it includes all patients regardless of payer-type, allowing one to examine how readmissions vary by insurance source.

The data contains information on the diagnoses associated with each visit, the procedures performed (if any), total charges, expected payment source, patient co-morbidities and patient demographics including gender, race, and age (in months). Patient diagnoses and procedures received are coded using the International Classification of Diseases, Clinical Modification (ICD 9-CM) codes. I use this coding scheme to identify patients admitted for a heart attack and the treatment/diagnostic procedures they received.[65] Both these databases contain patient and

[65] See Appendix B.2 for more details on medical codes.

hospital identifiers, allowing me to find patients who had an initial hospitalisation for a heart attack and then is readmitted.[66]

The data also contains detailed information about the expected payer for the hospital stay, which reflects the type of insurance used. The five main categories of insurance coverage identified in the data are: (i.) Medicare (the government insurance program for elderly individuals), (ii.) Medicaid (the government insurance program for low-income individuals), (iii.) private insurance, (iv.) the patient themselves (no insurance), or (v.) other insurance sources. The Medicare group includes patients on Traditional Medicare and Medicare Advantage. The Medicaid group includes patients on standard Medicaid and Medicaid HMO. The "private" category consists of Blue Cross, commercial insurance companies, and other forms of (non-Medicare) HMOs. The "other" insurance category includes worker's compensation, other government health insurance programs, and CHAMPVA/VA coverage (coverage for veterans and their spouses or children).

Although this dataset is rich, one limitation is that it only includes hospitals in NYS. This means that I cannot observe readmissions of patients who seek care outside of NYS within the 30 days after being discharged for a heart attack.[67] In general, given the severity of a heart attack, it is unlikely that patients recently discharged with a heart condition would move across states. A further shortcoming of the HCUP data is that it only contains information about the total charge for the hospital service. While there is information on the total charge for that hospitalisation, it does not record what proportion of the total charge the patient paid out of pocket.

3.3 Variable Definition and Sample Selection

The unit of analysis is a hospitalisation for heart attack. The key outcome variables are: an indicator for 30-day readmissions and measures of treatment intensity. This section discusses how these variables are constructed and how the sample is selected.

[66] This "panel" element of the data is important for constructing these variables. However, the econometric analysis below only uses the cross-sectional components of the data.

[67] For example, if a patient has a heart attack and is admitted to a hospital in NYS on day t and the patient leaves NYS between t and $t+30$ to seek medical care in another state's hospital, this patient would be recorded as having no readmission.

AMI hospitalisations. The sample consists of patients who are admitted to hospital with the primary diagnosis being a heart attack.[68] The exact sample used depends on the optimal bandwidth estimated for the RD specification; the bandwidths used in Section 5 mean that the sample roughly corresponds to patients between 55 and 75 years old. The hospitalisation episode must satisfy the following conditions: (i.) the patient is discharged alive and (ii.) the patient had a "routine" discharge (i.e. the patient is discharged home rather than transferred to another facility).[69] Criterion (ii) requires the patient's discharge procedure to be standard. Specifically, the patient must be discharged to their home, rather than transferred to some other medical facility such as a critical access hospital or other care facility such as a nursing home. The reason for this restriction is that if the patient has had a routine discharge, then it would be more credible to attribute any additional requirements for medical attention to the patient's behaviour at home. In contrast, if the patient is transferred to another facility, responsibility for further requirements of post-discharge care is less clear. Although criterion (i) - selecting patients who are discharged alive - is a prerequisite for the outcome variables of interest (readmissions) to make sense and criterion (ii) restricts the sample to standard discharges, one concern is that there may be differential death or transfer probabilities across the age eligibility threshold. This would affect the probability that a patient is selected in the sample. Section 4 shows that this is not a concern for the male or female samples. Index AMI admissions do not require a prior "clean period" with no hospitalisation: a hospital stay may be a readmission for a prior stay and the index admission for a subsequent readmission.[70]

30-day Readmissions. My main outcome variable of interest is an indicator for 30-day all-cause readmissions. I use the same method as that used in the previous chapter to construct this variable.

The all-cause definition of readmissions means that readmissions can be for reasons unrelated to the initial diagnosis. To gain further insights into the main factors driving readmissions, I break down readmissions by the primary diagnosis recorded on the billing record. The main diagnoses

[68] This is identified in the data using the following ICD9-CM codes: 41000–41091, excluding codes ending with 2.

[69] In the sample, roughly 4.4% of heart attack patients between 55 and 75 years old die in hospital. Section 4 explores whether these discharge outcomes (death or non-routine discharges) vary by insurance status.

[70] An alternative approach is to require index stays to be hospitalisations with no preceding hospitalisation in the previous 30 days. Results are robust to both definitions of index stays given the small number (<1%) of AMI hospitalisations that are themselves readmissions.

for readmissions that I focus on are: (i.) major vascular events (heart attacks, strokes, unstable angina, or heart failure), (ii.) non-major vascular events (non-specific chest pain, dysrhythmias, and coronary atherosclerosis), and (iii.) surgical complications. These categories are described in further detail in Section 5. Categorising readmissions by why they occur is helpful for learning about the determinants of post-discharge outcomes. For example, return visits for surgical complications are more likely to be driven by hospital-side quality factors than by post-discharge patient behaviour (Chen et al., 2014; Tsai et al., 2013).

Treatment intensity. The medical literature provides evidence on how observable treatment choices for heart attacks affect health outcomes. In light of these findings, I use the following variables as measures of treatment intensity: (i.) a binary variable indicating whether the patient received an angioplasty or cardiac catheterisation as the first procedure during their index admission, (ii.) a binary variable indicating whether the patient received catheterisation within 48 hours of admission, (iii.) the length of stay of the index admission, and (iv.) the number of procedures received at the index admission.

The first measure of intensity (definition i) concerns the invasiveness of the treatment while the second measure of intensity (definition ii) concerns the timely diagnosis of the heart attack (e.g. severity of attack), which usually precedes an invasive treatment. Given medical evidence that receiving invasive and earlier treatments is associated with better outcomes, more invasive and earlier treatments correspond to better quality in this context. Definition (i) is based on Currie et al. (2016) who define "aggressive" procedures as either angioplasties or catheterisation during the index visit, regardless of time of receipt. Definition (ii) follows Molitor (2018) and the existing medical literature (e.g. Bhatt et al., 2004) in defining early invasive treatment as receipt of catheterisation within 48 hours of admission. Technically, while catheterisation is an invasive diagnostic test that is usually followed by a surgical treatment, this need not be the case. In the sample, some patients who receive catheterisations do not proceed to have a surgery while some who have a surgical treatment do not receive a catheterisation. Thus, these two measures capture two aspects of treatment intensity: invasiveness of treatment and timely diagnosis.

The last two measures of treatment intensity capture the medical resources devoted to the

patient during their index admission. Under Medicare, hospitals are reimbursed according to the diagnostic group that the patient is admitted for, rather than the number of days the patient stays. Thus, the length of the index stay (definition iii), and number of procedures performed during the index stay (definition iv) all capture the input provided by the hospital to produce patient care. If insurance status does not alter the intensity of treatment provided by hospitals, these measurable inputs into post-discharge outcomes should stay constant across the age eligibility threshold.

3.4 Descriptive Statistics

Table 2.1 presents descriptive statistics for men and women between 55 and 75 separately by insurance category.[71] In panels A1 and B1, I examine healthcare utilisation and outcome variables by insurer type. This panel highlights the variation in readmission rates across insurance categories. Among men, Medicare patients have the highest readmission rate at 15.15%, followed by Medicaid patients (14.31%). Privately insured and uninsured patients have a 30-day all-cause readmission rate of roughly 7–9%. The readmission rate of patients insured by "other" insurance categories is similar to that among privately-insured patients. Among women, readmission rates are 0.8–2.0 pp higher than men across all insurance categories. The variation across insurance categories is similar, with Medicare and Medicaid patients having highest rates of 30-day readmissions ($\sim$16%). This pattern of variation in readmissions across insurance categories is not unique to NYS. This pattern holds more generally across the US with Medicare patients having higher readmissions than patients under other insurance categories (Barrett et al., 2015). The empirical strategy of this chapter aims to investigate whether there is a causal relationship between insurance and readmission outcomes.

Across all insurance categories and for both men and women, the most common reason for being readmitted is coronary atherosclerosis (CA) and the mean days to readmission is 12–14 days. Among patients who are readmitted, the distribution of days-to-readmission has a long right tail: over 60% of readmissions occur within two weeks after discharge. The probability of receiving an invasive treatment (catheterisation/angioplasty) varies across insurance categories and across

[71] This is approximately the age group that falls within the optimal bandwidth selected by the method of Calonico et al. (2014). See Section 5.

genders. For both men and women, privately insured and uninsured individuals are most likely to receive an invasive treatment (83–84% for men and ∼79–83% for women). This is partly due to these patients being younger and more suitable for invasive surgeries. Within insurance categories, women are less likely to receive an invasive treatment than men (except for patients under "other" insurance categories). If invasive treatments help reduce readmissions, lower rates of invasive treatments among women may be one reason for the higher readmission rates among women than men. The probability of receiving catheterisation within 48 hours of admission varies less across insurance categories, with Medicare and privately-insured individuals having similar rates of catheterisation within 48 hours of admission (18% for men and 25–27% for women). A higher proportion of Medicaid and uninsured patients receive catheterisation within two days of admission compared to other insurance categories.

Panels A2 and B2 summarise the average cost of the index admission and readmission. To construct the cost figures, I use the HCUP's cost-to-charge ratio data to adjust hospitalisation charges (which reflects the amount that the hospitals billed for services, not how much the services cost) into billable costs.[72] Thus, these cost figures reflect the actual expenses incurred in the production of hospital services such as wages, supplies, and utility costs. Among men, the average cost of an index admission for heart attacks ranges from $17,000 to $19,000 with Medicaid patients being the costliest. The average cost of a readmission ranges from $12,600 to $15,000. Among women, the average cost of an index admission ranges from $15,000 to $17,000, also with Medicaid patients being the costliest. The average cost of a readmission ranges from $12,000 to $14,000 and is highest for Medicare patients.

In panels A3 and B3, I present mean characteristics of patients by insurer, including their age, race, and co-morbidity index. Among men and women, Medicare patients are older than patients insured under other insurance schemes. The mean age of Medicare patients is 68 years while the mean ages of patients in other categories are roughly 60. There are fewest black patients in the private insurance category (6% for men, 12% for women) and most black patients in the Medicaid category (15% for men, 20% for women). The co-morbidity index indicates that Medicare patients

[72] See https://www.hcup-us.ahrq.gov/db/state/CCR_CD-SID_UserGuide_2001-2016.pdf for more details on the cost-to-charge adjustments. [Accessed 13 December 2018]

have more pre-existing medical conditions than all other insurance categories.[73] The descriptive statistics are consistent with previous studies documenting the socio-economic gradient in health outcomes and point to the likely endogeneity of insurance status. The fact that Medicare patients tend to be older and sicker means that their probability of being readmitted is higher.

There are over twice as many observations in the male sample than the female sample. Two factors could account for this. First, medical research has shown that men are twice as likely as women to suffer from a heart attack (Albrektsen et al., 2016).[74] Second, conditional on suffering from a heart attack, women tend to experience them later than men. This pushes the age distribution of female heart attack patients to the right. To ensure that the difference in results between genders is not driven by sample size, robustness checks in Section 5 include randomly dropping 50% of the male sample.

[73] Similar to Chapter 1, I use the Elixhauser co-morbidity Index.

[74] Women's naturally occurring hormone levels are thought to protect them against heart disease before menopause, when hormone levels drop.

Table 2.1: Descriptive statistics for main sample by insurance group and gender

		Medicare	Medicaid	Private	Uninsured	Other
		(1)	(2)	(3)	(4)	(5)
A.	**Men (55–75 years old)**					
1.	**Healthcare utilization and health outcome variables**					
	30-day readmission (%)	15.15	14.31	9.20	6.90	9.06
	Most common reason for readmission	CA	CA	CA	CA	CA
	Days to readmission	12.20	12.09	12.78	11.70	13.45
	Invasive treatment (%)	75.46	79.52	84.16	82.83	75.68
	Catheterization w/in 48 hours (%)	18.35	25.38	18.17	21.23	19.18
	Length of stay	4.35	4.83	3.59	4.10	4.13
	Number of procedures	6.39	6.89	7.21	6.94	7.04
2.	**Healthcare cost variables**					
	Total cost of index stay ($)	17,327	19,202	18,128	18,256	18,424
	Total cost of readmission ($)	13,605	13,554	12,660	15,369	12,560
3.	**Demographic variables**					
	Age	67.63	60.43	60.99	60.04	62.53
	Black (%)	8.94	15.11	5.86	12.43	10.73
	Co-morbidity Index	10.70	9.45	5.54	5.72	7.99
	Observations	9,194	2,754	9,553	1,159	662
B.	**Women (55–75 years old)**					
1.	**Healthcare utilization and health outcome variables**					
	30-day readmission (%)	15.94	16.04	11.36	8.76	11.96
	Most common reason for readmission	CA	CA	CA	CA	CA
	Days to readmission	12.30	11.91	12.25	13.94	13.41
	Invasive treatment (%)	75.21	75.05	82.88	79.12	81.52
	Catheterization w/in 48 hours (%)	24.64	28.38	26.77	24.23	29.89
	Length of stay	4.41	5.34	3.74	4.05	4.11
	Number of procedures	5.73	6.30	6.30	6.28	6.48
2.	**Healthcare cost variables**					
	Total cost of index stay ($)	15,116	17,230	15,585	16,551	15,285
	Total cost of readmission ($)	13,991	13,024	11,917	12,071	13,830
3.	**Demographic variables**					
	Age	68.06	61.49	61.26	60.66	63.34
	Black (%)	14.79	20.36	12.10	14.69	11.96
	Co-morbidity Index	11.65	11.65	7.06	6.10	8.35
	Observations	5,458	1,459	3,265	388	184

Notes. This table presents descriptive statistics for patients between 55 and 75 years old. "Co-morbidity index" refers to the Elixhauser co-morbidity index. "CA" refers to coronary atherosclerosis. Prices are in 2008 dollars.

4 | Empirical Strategy

The empirical analysis in this chapter uses a reduced-form regression-discontinuity (RD) model to estimate the impact of universal eligibility for Medicare insurance at age 65 on readmission probabilities. The patient's age at admission (in months) is the running variable and the discontinuity of interest occurs when the patient reaches 65. This section describes the reduced-form model and provides support for the identification assumption of this strategy.

4.1 Reduced-Form Model

To derive the reduced-form model, let post-discharge outcomes be determined by the following process:

$$y_i = f(a_i) + \mathbf{X}_i'\beta + \mathbf{C}_i'\gamma + \epsilon_i \tag{15}$$

where y_i equals one if patient i is readmitted within 30 days and zero otherwise, $f(a_i)$ is some smooth function of patient age a_i, $\mathbf{X}_i$ is a vector of patient characteristics (e.g. race, co-morbidities), and $\mathbf{C}_i$ is a $1 \times K$ vector of insurance coverage characteristics of patient i. For example, insurance characteristic $k = 1$ might indicate *any* coverage, insurance characteristic $k = 2$ might indicate a relatively generous insurance package (e.g. low co-payments and long benefit periods so that readmissions are covered), and insurance characteristic $k = 3$ might reflect how insurers reimburse hospitals for certain services, which may in turn affect the probability of receiving particular types of tests and treatments (e.g. catheterisation). Finally, ϵ_i is an unobservable error term which may include unobserved patient latent health (e.g. driven by factors such as lifestyle choice) and other aspects of treatment quality that do not depend on the patient's insurance status.

The main challenge in estimating the impact of insurance is that insurance coverage is likely to be endogenous to other factors that determine healthcare utilisation and post-discharge outcomes. To deal with insurance endogeneity, I use the age-65 eligibility threshold as a source of exogenous variation in insurance status. In particular, drawing on the evidence presented in Section 3 that

health insurance characteristics shift at age 65 as most people become eligible for Medicare, let insurance characteristic k be determined by a linear probability model of the form:

$$C_i^k = g^k(a_i) + \mathbf{X}_i' b^k + D_i \pi^k + v_i^k \tag{16}$$

where D_i denotes an indicator for being 65 or older at admission and the $g(a)$ functions are smooth functions of age.

Combining equations (15) and (16), the reduced form model for post-discharge outcomes is:

$$y_i = \alpha(a_i) + D_i \tau + \mathbf{X}_i' \eta + u_i \tag{17}$$

where $\alpha(a_i) = f(a_i) + \sum_k g^k(a_i)\gamma^k$ is the reduced-form age profile, $\tau = \sum_k \pi^k \gamma^k$ is the reduced-form impact of interest, $\eta = \beta + \sum_k b^k \gamma^k$ is the coefficient on the patient characteristics vector, and $u_i = \sum_k v_i^k \gamma^k + \epsilon_i$ is the error term.

The reduced-form effect τ can be interpreted as the effect of universal eligibility for Medicare insurance at age 65 on readmission probabilities (the effect of "eligibility for Medicare insurance"). Since each insurance characteristic k contributes additively to the reduced-form impact τ (e.g. the kth characteristic contributes $\pi^k \gamma^k$ to τ), a non-zero τ term could reflect multiple channels leading to an impact of Medicare eligibility on readmissions. For example, this effect could be attributable to patient-side factors (e.g. an increase in the overall fraction of the population with *any* insurance or an effect driven by people switching from other insurance schemes to Medicare) or hospital-side factors (e.g. changes in treatments due to different reimbursement methods under different insurance schemes).

I estimate the impact of Medicare eligibility on readmissions using equation (17). In the baseline regressions, I use a linear specification for age fully interacted with the post-65 dummy variable.[75] In the vector $\mathbf{X}_i$, I include indicators for race, the Elixhauser co-morbidity index, and year-month

[75] That is, $\alpha(a_i) = \beta_1 a_i + \beta_2 a_i D_i$

dummies. The Elixhauser co-morbidity index controls for co-existing medical conditions that may increase a patient's risk of being readmitted. The medical literature has shown that this index is helpful in predicting mortality and readmission outcomes (Moore et al., 2017). The year-month dummies control for variation in readmissions driven by common time trends. I follow Calonico et al. (2014) in selecting the optimal bandwidth (i.e. the optimal age window around the 65-threshold) which minimises the mean-square error (MSE) of the residual. Following Lee and Card (2008) and Card et al. (2008), I cluster standard errors at the level of the running variable (age in months). For inference, I report robust standard errors and p-values calculated using the method of Calonico et al. (2014). This method adjusts conventional confidence intervals by ensuring that they are robust to "large" bandwidths which may lead to an over-rejection of the null hypothesis. Appendix B.3 contains more information about this method.

In order to investigate whether Medicare eligibility affects treatments received, I also estimate versions of equation (17) where the dependent variable is various measures of treatment intensity (discussed in Section 3). Estimating discontinuities in treatment intensity may help shed light on whether a relationship between Medicare eligibility and readmissions is driven by changes in treatment received.

4.2 Identification Assumptions

For OLS to identify the effect of interest τ in equation (17), the key identification assumption is that assignment to either side of the discontinuity threshold is as good as random (Lee and Lemieux, 2010). More formally, let $rv_i = 65 - a_i$ be the running variable centred around the age-65 eligibility threshold, then this assumption can be written as:

$$\lim_{e \downarrow 0} \mathbb{E}[\sum_k v_i^k \gamma^k + \epsilon_i | rv_i + e, \mathbf{X}_i] - \lim_{e \uparrow 0} \mathbb{E}[\sum_k v_i^k \gamma^k | rv_i + e, \mathbf{X}_i] = 0$$

Under this assumption, the causal effect of Medicare eligibility on readmissions can be identified

by the discontinuity in post-discharge outcomes at the age 65 threshold. In particular, τ can be identified as:

$$\tau = \lim_{e \downarrow 0} \mathbb{E}[y_i | rv_i + e, \mathbf{X}_i] - \lim_{e \uparrow 0} \mathbb{E}[y_i | rv_i + e, \mathbf{X}_i] \tag{18}$$

In this setting of admitted hospital patients, this identification assumption may fail if insurance status affects (i.) the probability that a patient is admitted to hospital and/or (ii.) the probability that the patient satisfies the sample selection criterion (i.e. the patient is discharge alive and has a routine discharge). Regarding (i), one benefit of focusing on severe conditions which require immediate medical attention (such as heart attacks) is that admission to hospital is less likely to depend on patient insurance coverage than when focusing on selective procedures (such as hip replacements). To test that the probability that a patient is admitted to hospital is independent of whether they are under or over 65, I conduct a RD analysis of the (log) count of heart attack admissions by age measured in months. This procedure is similar to the test of manipulation proposed by McCrary (2008) except that I have a discrete running variable. Figure 2.2 provides a graphical representation of this test by showing the age profiles of the log of the total admission counts by age in months for patients aged between 55 and 75 years. The dark black line is the fitted line from a regression with a quadratic polynomial in age fully interacted with a dummy for patients older or equal to 65 years old at first admission. For both genders, the figure suggests that the density of age at admission is smooth across the age threshold. Formal regression discontinuity estimates confirm this. For men, the estimated discontinuity in log number of admissions is 0.104 with a robust p-value of 0.166. For women, the estimated discontinuity is -0.092 with a robust p-value of 0.173.

Regarding (ii), one concern is that even if patients above and below age 65 are equally likely to be admitted to hospital, patients may have different discharge outcomes based on age which affects the probability of admitted patients appearing in the sample. My sample selection criteria require patients to be discharged alive and to have a routine discharge (rather than to another medical facility). If there are differential death or transfer probabilities across the eligibility threshold, then this may affect the density around the cut-off age. To check this, I use the "complete" sample of patients admitted for a heart attack as their first diagnosis regardless of their discharge outcome

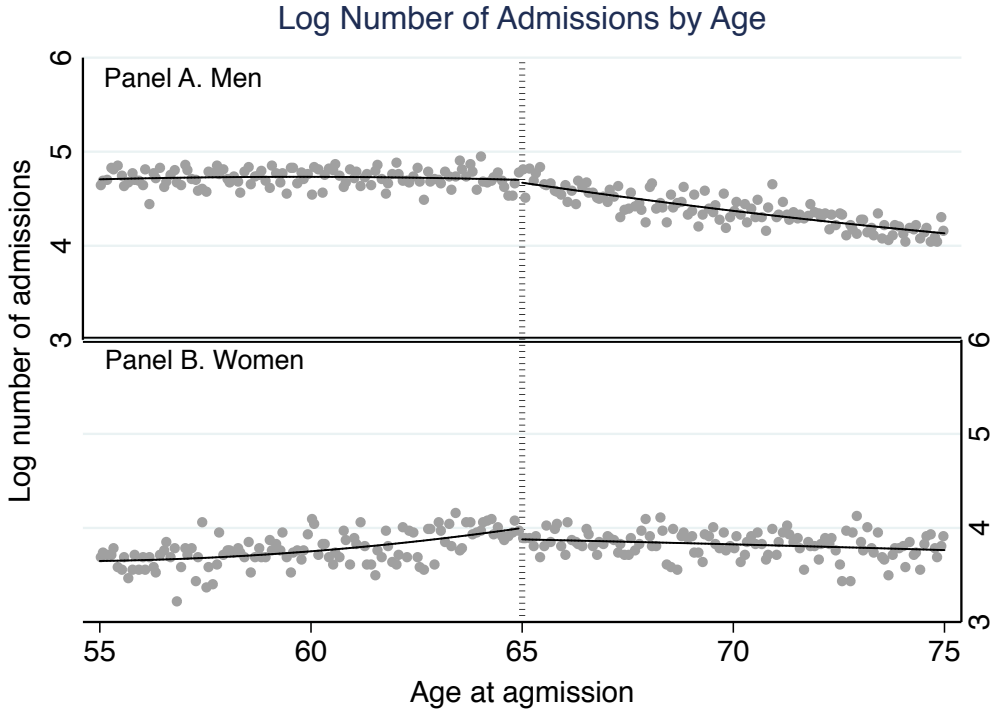


Figure 2.2: This figures plots the log number of admissions by age at admission in order to check the density around the age-65 threshold. Estimated discontinuity in log admissions for men: Estimated discontinuity= 0.104, robust p-value=0.166. Women: Estimated discontinuity=-0.092, robust p-value=0.173.

(i.e. for this exercise, patients who die in hospital or are discharged to another facility are included in the sample). I then estimate the discontinuity in the probability of having a discharge outcome that leads the observation to be omitted from the sample. These results are presented in Row A of Table B.4 in Appendix B.4. The estimated discontinuity is 0.211 for men (robust p-value: 0.836) and 2.17 for women (robust p-value: 0.124), suggesting that there is no discontinuity in the probability of having a discharge outcome that leads to the observation being omitted.

Even if there is no differential selection at age-65, discontinuities in other characteristics at the age threshold may invalidate the RD design. Rows B to D of Table B.4 present estimated discontinuities in several patient characteristics. For these regressions, the original sample selection procedure is used. Row B shows that there is no discontinuity in the proportion of black patients. Row C shows that there is no discontinuity in the proportion of patients living in zipcode areas where the median household income is in the bottom quartile of the national income distribution. Row D shows that there are no discontinuities in the illness severity of patients across the age threshold:

the estimated discontinuities is -0.42 for men (robust p-value: 0.44) and -0.37 for women (robust p-value: 0.63). This is consistent with the assumption that unobserved latent health variables are smooth at age 65. Note that the baseline regression controls for race and a co-morbidity index in the vector X_i.

A further concern is that there are changes in labour supply at age 65, a traditional age of retirement, which may affect health status. Although the patient-level discharge data does not contain information on employment, previous studies find no discontinuities in the likelihood of working at age 65 in the United States (e.g. Card et al., 2009). Using various nationally representative survey datasets, I also find no evidence of discontinuities in labour supply. For example, using the Medical Expenditures Survey, the estimated discontinuity in wage income and social security income are -0.047 and -0.001 respectively with corresponding p-values of 0.332 and 0.971.

Underlying the reduced-form empirical strategy is the assumption that the age-65 threshold has an impact on a patient's likelihood of being covered by Medicare. Evidence supporting this assumption was presented earlier in Figure 2.1 which plots the share of patients with various forms of primary insurance coverage. Table 2.2 presents formal regression discontinuity estimates of various insurance characteristics by gender and demographic group. The "non-white" category includes blacks, Hispanics, Asians, Native Americans, and other races. Although it would be desirable to distinguish between the various racial groups within the non-white category, stratifying the sample further produces noisy estimates.[76] In the bottom two rows of each panel, I stratify the sample by the estimated median household income of residents in the patient's zipcode. A patient is in the "highest quartile" sample if they live in zipcode areas where the median household annual income falls in the highest quartile of the income distribution ($\geq \$75,000$) and in the "lowest quartile" if they live in zipcode areas where the median household income falls in the lowest quartile of the income distribution ($< \$40,000$).[77]

Consistent with Figure 2.1, the formal RD estimates in the first row of columns (1) and (2)

[76] Among the non-white group, 27% are Black, 24% are Hispanic, 10% are Asian, less than 1% are Native American, and the remainder is of "other" race.

[77] Data on zipcode-level income is based on 2006 to 2010 of the American Community Survey and is provided by The University of Michigan's Population Studies Center (PSC).

show that the proportion of patients with any type of insurance coverage increases by roughly 5–7 pp at age 65 for men and women, from a baseline of over 90%.[78] Estimates in the first row of columns (3) and (4) show that the there is a 43 pp and 46 pp increase in the proportion of patients insured by Medicare for men and women respectively, compared to a baseline of 17% and 21%. For men, this effect is strongest among patients living in poorer zipcode areas where the baseline rate of coverage is the lowest. For women, this effect is concentrated among white patients and those living in higher income zipcode areas.

Columns (5) and (6) indicate a discontinuous drop in the proportion of patients covered by private insurance at age 65. The estimated discontinuity is -27 pp for men and -24 pp for women, from a baseline of 57% and 48% respectively. For both men and women, this effect is strongest among white patients and those living in high-income areas, where baseline rates of private insurance coverage are the highest. The estimates in columns (7) and (8) illustrate a discontinuous decrease in Medicaid coverage. This effect is larger for women than men (-15 pp vs -9 pp). For both men and women, this effect is concentrated among non-white patients and those living in low-income zipcode areas where rates of Medicaid enrollment for those under 65 are higher than other subgroups.

Finally, columns (9) and (10) show that there is a discontinuity in the proportion of patients with multiple insurance sources, defined as having more than one expected payer (that is not the patient themselves) recorded on the hospital billing record. The estimates suggest that the proportion of patients with more than one insurance source increases by 31 pp among men and 29 pp among women. In cases where patients have multiple insurance sources, the primary payer pays up to the limits of its coverage before the secondary payer steps in to pay the costs that the primary insurer didn't cover. Thus, having multiple insurance coverage sources is generally associated with having more generous medical coverage overall.

[78] In the AMI sample, the proportion of 63–64 year old patients with any insurance coverage is slightly higher than that found in nationally representative surveys. For example, using the Health and Retirement Study data over the same period, the proportion of 63–64 year old men with any insurance coverage is 90% and the figure for women is 91%. However, compared to patients admitted for other conditions, the proportion of patients with any insurance coverage in the AMI sample is slightly lower. See footnote 63.

Table 2.2: Estimated discontinuity in various insurance coverage characteristics

	Any		Medicare		Private		Medicaid		Multiple	
	Mean 63-64	Disc at 65	Mean 63-64	Disc at 65	Mean 63-64	Disc at 65	Mean 63-64	Disc at 65	Mean 63-64	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Men										
All	92.48	5.31*** (0.01)	17.08	42.68*** (0.02)	56.65	-26.83*** (0.02)	15.04	-8.59*** (0.01)	11.19	30.74*** (0.02)
	[w=12.06,N=27,429]		[w=4.16,N=10,213]		[w=4.77,N=11,774]		[w=5.62,N=13,686]		[w=7.25,N=17,298]	
White	93.18	5.38*** (0.01)	17.27	45.21*** (0.02)	64.23	-32.47*** (0.02)	7.21	-5.32*** (0.01)	11.51	34.22*** (0.02)
	[w=9.77,N=14,906]		[w=4.97,N=8,038]		[w=5.06,N=8,146]		[w=8.07,N=12,544]		[w=8.73,N=13,451]	
Non-White	91.10	5.46*** (0.01)	16.70	39.07*** (0.03)	41.76	-18.12*** (0.03)	30.44	-16.28*** (0.02)	10.55	24.50*** (0.02)
	[w=8.42,N=6,903]		[w=4.69,N=3,940]		[w=6.76,N=5,569]		[w=4.99,N=4,123]		[w=6.11,N=5,066]	
Highest qtrl income	94.55	3.57*** (0.01)	12.32	42.66*** (0.03)	67.70	-30.84*** (0.04)	11.28	-6.86*** (0.01)	10.64	34.45*** (0.04)
	[w=11.07,N=7,187]		[w=6.62,N=4,573]		[w=6.68,N=4,627]		[w=10.76,N=7,055]		[w=6.99,N=4,785]	
Lowest qtrl income	90.31	7.71*** (0.01)	22.17	47.11*** (0.03)	41.54	-22.68*** (0.03)	24.79	-16.47*** (0.02)	9.69	26.79*** (0.03)
	[w=8.10,N=4,270]		[w=8.95,N=4,669]		[w=8.64,N=4,521]		[w=8.31,N=4,351]		[w=8.27,N=4,351]	
B. Women										
All	92.23	6.65*** (0.01)	20.92	46.36*** (0.02)	47.78	-23.91*** (0.03)	21.79	-14.98*** (0.01)	11.97	28.74*** (0.02)
	[w=8.03,N=8,817]		[w=5.32,N=5,962]		[w=5.62,N=6,316]		[w=7.73,N=8,470]		[w=9.80,N=10,501]	
White	92.60	7.09*** (0.01)	20.92	51.58*** (0.03)	58.04	-32.78*** (0.03)	11.73	-9.60*** (0.01)	13.52	28.40*** (0.03)
	[w=8.17,N=5,475]		[w=6.27,N=4,314]		[w=5.61,N=3,844]		[w=9.64,N=6,286]		[w=12.14,N=7,763]	
Non-White	91.63	5.96*** (0.02)	20.92	40.28*** (0.04)	30.96	-10.13** (0.04)	38.28	-23.29*** (0.03)	9.41	29.10*** (0.03)
	[w=8.15,N=3,469]		[w=6.46,N=2,793]		[w=6.75,N=2,932]		[w=6.82,N=2,932]		[w=8.54,N=3,636]	
Highest qtrl income	93.49	3.75* (0.02)	17.47	55.10*** (0.03)	56.85	-39.29*** (0.03)	17.81	-10.83*** (0.02)	10.27	33.28*** (0.03)
	[w=8.65,N=2,264]		[w=10.55,N=2,665]		[w=9.81,N=2,509]		[w=8.78,N=2,288]		[w=11.43,N=2,874]	
Lowest qtrl income	92.35	6.42*** (0.02)	26.47	40.39*** (0.046)	33.82	-14.19** (0.053)	30.88	-19.95*** (0.032)	13.24	24.47*** (0.039)
	[w=9.32,N=2,736]		[w=5.96,N=1,794]		[w=6.03,N=1,815]		[w=6.75,N=2,023]		[w=9.23,N=2,719]	

Notes. This table presents the estimated discontinuity in the probability of having any insurance, various sources of insurance, and multiple sources of insurance by demographic groups. "Any" is an indicator for having any insurance coverage. "Multiple" is an indicator for having more than one insurance coverage. All models include a linear function for age at admission (in months) fully interacted with a dummy for age over 65, and also include race indicators, the Elixhauser co-morbidity index, and year-month dummies. Bandwidths ("w") are MSE-optimal and selected using the procedure of Calonico et al. (2014). "N" denotes the corresponding sample size. Standard errors (in parentheses) are robust and clustered at age in months. Kernel: Triangular. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 | Results: Impacts on 30-day readmissions

In this section, I present the RD estimates of the impact of eligibility for Medicare insurance on 30-day readmissions for men and women separately. In each table, I present the following statistics: the mean of the dependent variable for patients aged 63–64 years, the estimated discontinuity at age 65, the MSE-optimal bandwidth, the robust standard error, and corresponding significance levels denoted by asterisks. The bandwidth choice and robust inference methods follow Calonico et al. (2014). Unless stated otherwise, all regressions use a linear specification for age (in months) fully interacted with the post-65 dummy variable, and also include indicators for race, the Elixhauser co-morbidity index, and year-month dummies. Following Lee and Card (2008), I cluster standard errors at the level of the running variable (age in months). Robustness checks in Appendix B.4 show that the results are robust to functional form, bandwidth choices, control variables, and sample selection procedures.

5.1 Men: Discontinuities in readmissions

Table 2.3 presents results for male heart attack patients. Moving across the table, the subsequent columns break down the dependent variable by reasons for readmission. For example, the dependent variable in columns (1) and (2) is 30-day all-cause readmissions (readmissions for any primary diagnosis) while the dependent variable in columns (7) and (8) is 30-day readmissions for surgical complications only. Moving down the table, the subsequent rows break down the sample by demographic subgroups. For example, row B presents RD estimates for the white patient subsample and row E presents estimates for patients living in zipcode areas where the median household income is in the lowest quartile of the national income distribution. The definitions of the subgroups are the same as those in Table 2.2.

Row A, columns (1) and (2) indicate that among male AMI patients, the probability of being readmitted within 30-days increases by 2.68 pp (column 2) at age 65, relative to a baseline of 11.00% among patients aged 63–64 years (column 1). The optimal bandwidth selected is 7.18; this means patients between 58 and 72 are included in the sample, resulting in 17,298 observations

Table 2.3: Estimated discontinuity in 30-day readmissions at age 65 for male patients

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	11.00	2.681***	2.41	-0.465	4.04	2.321***	0.83	-0.021
		(1.082)		(0.473)		(0.643)		(0.283)
	[w=7.18,N=17,298]		[w=8.02,N=19,116]		[w=8.69,N=20,528]		[w=8.87,N=17,588]	
(B) **White**	10.17	1.474	2.12	-0.273	3.91	1.682**	0.91	-0.349
		(1.215)		(0.596)		(0.811)		(0.350)
	[w=10.11,N=15,345]		[w=8.32,N=12,888]		[w=9.11,N=13,992]		[w=7.25,N=9,850]	
(C) **Non-White**	12.64	4.412**	2.97	-0.719	4.29	3.432***	0.66	0.742
		(2.409)		(1.145)		(1.079)		(0.593)
	[w=7.53,N=6,174]		[w=8.57,N=6,955]		[w=8.57,N=6,955]		[w=7.06,N=4,735]	
(D) **Highest qtrl income**	10.64	2.637	2.08	0.284	3.89	1.604	0.60	0.329
		(1.904)		(0.975)		(1.284)		(0.574)
	[w=9.32,N=6,143]		[w=8.37,N=5,602]		[w=11.88,N=7,638]		[w=7.72,N=4,498]	
(E) **Lowest qtrl income**	12.32	2.863	3.12	-1.829*	3.45	2.349*	0.98	-0.110
		(1.919)		(1.294)		(1.352)		(0.819)
	[w=9.86,N=5,098]		[w=5.24,N=2,878]		[w=11.99,N=6,086]		[w=7.34,N=3,201]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns). In columns 1–2, the dependent variable is an indicator for all-cause 30-day readmissions. In columns 3–4, the dependent variable is an indicator for 30-day readmissions where the primary diagnosis is a major vascular condition (heart attack, stroke, unstable angina, or heart failure). In columns 5–6, the dependent variable is an indicator for 30-day readmissions where the primary diagnosis is a non-major vascular condition (non-specific chest pain, dysrhythmias, and coronary atherosclerosis/ other heart diseases). In columns 7–8, the dependent variable is an indicator for 30-day readmissions for surgical complications. For this outcome, each sample is limited to patients who received an angioplasty, catheterization, and/or coronary artery bypass graft as their first or second procedure during their initial stay. The dependent variable is an indicator for a readmission within 30 days. Other descriptions same as Table 2.2. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

for this specification.

Panel A of Figure 2.3 provides a graphical representation of this finding. This figure plots the proportion of male heart attack patients being readmitted within 30 days by age at admission. To create this figure, I first residualise the outcome variable and the running variable with respect to the baseline control variables. Second, I divide the residualised age variable into evenly-spaced bins (5 bins to the right and left). Third, I plot the means of the y-variable residuals within each bin against the mean x-value. Finally, I add back in the unconditional mean of the y-variable. The line is the linear line of best fit, fitted using the observations to the left or right of the discontinuity line. The slope of the line represents the relationship between the outcome variable and the running

variable while controlling for other regressors. This figure reveals several patterns. First, older male patients have a higher probability of being readmitted within 30 days. Second, for men, there is a discontinuous increase in the probability of being readmitted at the age eligibility threshold of 2.68 pp, consistent with the estimates in row A, columns (1) and (2) of Table 2.3.

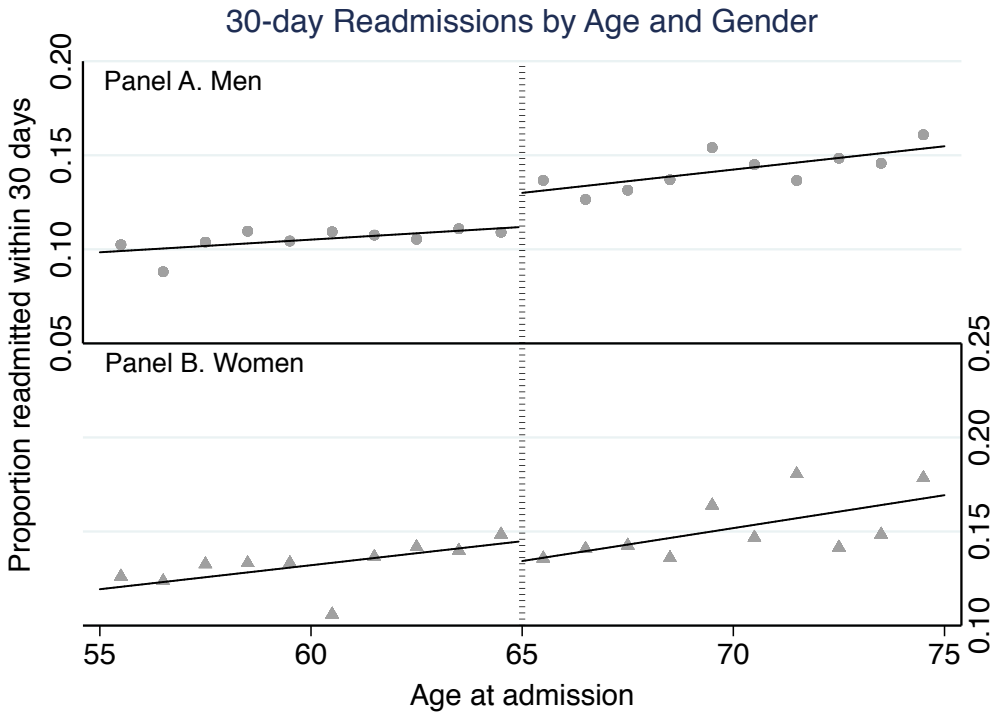


Figure 2.3: 30-day all-cause readmission rates by gender and age at admission.

Moving down the table, rows B to E of columns (1) and (2) indicate that the discontinuous increase in readmissions is driven by non-white patients. In particular, the estimated discontinuity is 4.41 pp for the non-white group relative to a baseline of 12.6%. This is the group with the highest pre-65 readmission rates. One possible reason for this is that Medicare eligibility increases healthcare access particularly for poorer and less healthy patients who require medical attention more than for other demographic groups, driving readmission rates upwards. Although the estimated coefficient among patients living in low-income zipcode areas (2.86 pp) is similar to that in row A, this estimate is statistically insignificant (robust p-value: 0.119).

Moving across the table, columns (3) to (8) of Table 2.3 break down the dependent variable by reasons for readmission. If different medical diagnoses during the readmission are caused by different factors between discharge and readmission, then distinguishing readmissions by the primary

diagnosis may help shed light on the channels via which eligibility for Medicare insurance affects post-discharge outcomes. For example, a common assumption is that readmissions for surgical complications are primarily driven by hospital quality rather than patient-side factors (Chin et al., 2016; Tsai et al., 2013). Under this assumption, if eligibility for Medicare insurance affects patient behavior rather than treatment quality, readmissions for surgical complications should not increase at age 65.

In columns (3) to (4), the dependent variable is an indicator for a readmission for a major vascular event such as a repeated heart attack, stroke, unstable angina, or heart failure. The dependent variable takes on a value of one if the patient is readmitted within 30 days for a major vascular event and takes on a value of zero either if the patient is not readmitted or if the patient is readmitted for anything not classified as a major vascular event. This classification follows the standard medical classification (Choudhry et al., 2011). In the full sample of men (row A), 2.41% of 63–64 year old patients are readmitted within 30 days for a major vascular condition and there is no significant discontinuity at age 65. The estimated discontinuity is -0.47 with a standard error of 0.47. This is consistent with the finding of Choudhry et al. (2011) that the elimination of co-payments for drugs prescribed after AMI among the treatment group did not reduce rates of major vascular events compared to the control group. Rows B to E show that there is also no significant discontinuity in readmissions for major vascular conditions among any of the subgroups except for patients living in low-income zipcode areas, among whom readmissions for major vascular conditions drop by 1.83 pp (robust p-value: 0.069).

In columns (5) to (6), the dependent variable is an indicator for a readmission for a non-major vascular event such as non-specific chest pain, dysrhythmias (abnormal heart beats), and coronary atherosclerosis/other heart diseases. Although there is no standard classification of non-major vascular conditions, there is consensus on common complications following a heart attack that do not fall under the major vascular events category. These common complications constitute my definition of non-major vascular events. In the sample, while over 80% of patients readmitted for major vascular events have an indication of ED use, only about 60% of patients readmitted for non-major vascular events have an indication of emergency department use. This suggests that patients being readmitted for major vascular events are in more urgent need of medical care than

those readmitted for non-major vascular events.

The estimates in columns (5) to (6) show that the increase in all-cause 30-day readmissions is driven by readmissions for non-major vascular events. In the full sample (row A), there is a 2.32 pp increase in the probability of being readmitted for a non-major vascular event from a baseline on 4.04%. Rows B to E indicate that this discontinuous increase is largest among non-white patients; white and low-income patients also contribute to the discontinuous increase. The estimated discontinuity is 3.43 pp (robust p-value: 0.000) for the non-white group relative to a baseline of 4.29% and 2.35 pp (robust p-value: 0.052) for those in poorer zipcode areas from a baseline of 3.45%. There is also evidence that part of this discontinuous increase is driven by white patients (row B), among whom there is a 1.68 pp increase in non-major vascular readmissions at age 65.

The heterogeneous results by patient demographics and reasons for readmissions support the interpretation that Medicare eligibility increases access to post-discharge medical services among patients with complications who would not have otherwise sought or received post-discharge medical attention. Patients with critical conditions after discharge are likely to seek care regardless of insurance status and hospitals are likely to treat these patients regardless of their ability to pay (Garthwaite et al., 2018). In contrast, among patients with mild conditions after discharge, the decision to seek additional care may depend on their insurance status. Therefore, one would expect insurance to have larger effects on readmissions for non-major issues than readmissions for major conditions. Increasing access can be beneficial for the patient since non-major vascular complications can lead to major heart issues if left unmonitored.[79]

In columns (7) to (8), the dependent variable is an indicator for readmissions related to surgical complications. For this specification, I limit the sample to patients who received an angioplasty, catheterisation, and/or coronary artery bypass graft as their first or second procedure during their initial stay (since patients who are not treated with a surgical procedure have no risk of having a related surgical complication). Among patients aged 63–64, the rate of readmissions for surgical complications is fairly low (0.83%) and there is no evidence of a discontinuity at age 65. The

[79] For example, coronary atherosclerosis can lead to heart failure or a repeated heart attack.

estimated discontinuity is -0.21 pp with a standard error of 0.28. Rows B to E show that among all demographic subgroups, there is no significant discontinuity in the probability of readmissions for surgical complications. Under the assumption that readmissions for surgical complications are likely to be driven by hospital and medical staff quality, the results in columns (7) and (8) suggest that quality of care is smooth across the age threshold for men.

Tables B.5 to B.7 (Appendix B.4) and Figure B.1 (Appendix B.5) test the sensitivity of the results to various changes in the baseline specification. First, following Kolesár and Rothe (2018), I use robust standard errors that are not clustered at the level of the running variable and examine whether the statistical significance of the results changes. Using these non-clustered standard errors results in slightly larger optimal bandwidths and therefore different point estimates; however, the point estimates are quantitatively similar to the baseline results. For example, the estimated discontinuity in all-cause readmissions for the full sample is 2.57 (relative to 2.68 in the baseline results), significant at the 1% level.

Second, I test whether the baseline results are sensitive to different functional forms and bandwidth choices. Two key challenges in using a RD design are selecting the correct bandwidth and selecting the most appropriate functional form for the relationship between the outcome and the running variable. On the one hand, using a larger bandwidth yields more precise estimates as more observations are available to estimate the regression, but choosing the correct functional form becomes more challenging. Misspecified functional forms can bias the estimate. On the other hand, using a smaller bandwidth means that the underlying functional form is more likely to be approximately linear, but the estimates may be less precise as fewer observations are available.

Table B.6 presents discontinuity estimates using a quadratic polynomial with interactions. In comparison to the linear interaction specification used in the baseline, the main difference is that the quadratic specification results in larger point estimates. For example, in the full sample (row A, column 2), the estimated discontinuity is 3.41 pp whereas it was 2.68 pp using the linear specification. The optimal bandwidth using the quadratic specification is also slightly larger (8.6 vs. 7.2).

Figure B.1 depicts the estimated discontinuity at age 65 for 30-day all-cause readmissions among

male patients and the corresponding confidence interval (shaded area) for different bandwidths using the linear interaction specification. Moving from left to right on the x-axis, I start with a bandwidth of 1 and increase it by 0.25 years until a bandwidth of 10 is reached. The dashed horizontal line denotes zero effect and the solid horizontal line denotes the baseline estimate. The figure firstly shows that for all bandwidths between 0 and 10, the estimated discontinuity is statistically significant and positive. Secondly, for bandwidths greater than 5.5, the estimated discontinuity is stable and similar to the baseline estimate of 2.68; for bandwidths below 5.5, the estimated discontinuity is larger but also less precisely estimated. Lastly, the baseline specification selects the optimal bandwidth by minimising the MSE of the underlying regression function. An alternative method of selecting the optimal bandwidth is to minimise the coverage error rate (CER) of the robust bias-corrected confidence interval. The CER-optimal bandwidth for the baseline specification is 5.102; the figure shows that the estimated discontinuity using the CER-optimal bandwidth is very similar to the baseline estimate (3.09 vs. 2.68).

Third, I test the sensitivity of the baseline results to the inclusion of other controls and other sample selection methods. Panel A of Table B.7 presents results for the full male sample when the list of control variables is altered, using the same optimal bandwidth selected in the baseline specification. The baseline regression includes indicators for race, a co-morbidity index, and month-year dummies. The list of control variables in the baseline specification is more extensive than the usual list used in this literature.[80] In row A, I follow Card et al. (2008) and additionally include an indicator for patients who are first admitted within one month of their 65th birthday. One assumption made in the empirical analysis is that the patient's insurance policy is constant over the 30-day window after a hospital discharge and a potential readmission. This assumption may be violated for patients who turn 65 over the 30-day window between discharge and readmission. The results indicate that controlling for this variable does not affect the estimated discontinuity; the estimated discontinuity in all-cause readmissions is 2.71 with the additional control. In row B, I additionally include indicators for the zipcode income quartile of the patient. For all outcomes of interest, the resulting estimates are very similar to the baseline estimates. For example, with controls for zipcode-level income quartiles, the estimated discontinuity for all-cause readmissions

[80] E.g. Card et al. (2009) include gender and race but no co-morbidity index.

is 2.66. In row C, rather than using year-month dummies to control for time trends, I use payer-specific year indicators to allow for different time trends by payer group. When payer-specific year indicators are included, the point estimate for all-cause readmissions is smaller (1.94 vs. 2.68) but quantitatively similar to the baseline estimate. The point estimate for non-major vascular readmissions is very similar (2.25 vs. 2.32), and there is some evidence of a discontinuous drop in major vascular readmissions (-0.74 pp, significant at the 10% level).

Panel B of Table B.7 presents results when the sample is altered. The baseline analysis includes all insurance categories. In row D, I drop Medicare patients and patients under "other" insurance schemes and focus on patients with Medicare insurance, private insurance, or no insurance. Focusing on these remaining categories is useful because most patients switch from private insurance to Medicare when they turn 65. The estimated discontinuity in readmissions is larger in this restricted sample. For example, the discontinuity in 30-day readmissions for men rises from 2.68 pp to 3.54 pp. Since most Medicaid patients are under 65 and post-discharge outcomes are poor among this group (as reflected in the high levels of readmissions relative to their age), including them in the original sample dilutes the difference in outcomes between over- and under-65 year old patients. Thus, the estimated impact is larger when Medicaid beneficiaries are omitted.

In row E, I restrict the sample to patients suffering from one type of AMI. Specifically, the baseline empirical specification treats all heart attacks as the same diagnosis. While all heart attacks occur when the heart does not receive sufficient blood supply, there are various forms of heart attacks that may require different treatments and have different symptoms.[81] To test whether the baseline results are driven by unobserved heterogeneity across patients suffering from different types of heart attacks, row E restricts the sample to patients diagnosed with non-ST elevation myocardial infarction (NSTEMI). This is the most common type of heart attack in the sample, with over 70% of heart attacks being of this type. The estimated discontinuities for this sample are larger in magnitude; for example, the discontinuity in all-cause readmissions is 4.1 pp.

In summary, the results suggest that eligibility for Medicare insurance increases the probability of 30-day readmissions by 2.68 pp among male AMI patients. This increase is primarily driven

[81] The three main types of heart attacks are ST segment elevation myocardial infarction, non-ST segment elevation myocardial infarction, and coronary spasm/unstable angina.

by patients returning for non-major vascular conditions and by non-white patients among whom the pre-65 rate of readmission is highest. These results are robust to other clustering methodologies, functional form specifications, bandwidth choices, control variables, and sample selection procedures.

5.2 Women: Discontinuities in readmissions

Table 2.4 presents results for female heart attack patients. The table is set up in the same way as that for men (Table 2.3). Comparing the pre-65 mean outcomes for the full sample of men and women in Tables 2.3 and 2.4 indicates that the mean readmission rates among 63–64 year old female patients is over 3 pp higher than the all-cause readmission rate among men (14.4% vs 11.00%). In contrast to the male sample, there is no significant discontinuity in the probability of readmission at age 65 for female patients. The estimated RD coefficient is negative (-1.48 pp) but not statistically significant (robust p-value: 0.395). Panel B of Figure 2.3 provides a graphical representation of this result. In contrast to the discontinuous increase in readmission rates at age 65 documented for men, there is no significant discontinuity for the female sample. Moving down the table, rows B to E of columns (1) to (2) indicate that there is no significant discontinuity among any of the demographic subgroups.

Each subsequent column of Table 2.4 breaks down readmissions by the primary reason for return. For women, 3.25% of 63–64 year old patients are readmitted within 30 days for a major vascular condition. Columns (3)–(4) show that there is also no discontinuity in these types of readmissions at age 65 for any of the subgroups. For example, in the full sample, the estimated discontinuity is 0.05 pp (robust p-value: 0.956). Columns (5)–(6) show that there is an estimated discontinuity of 0.98 pp in the probability of returning for a non-major vascular event, relative to a baseline of 3.57%. Although the estimate is positive, it is statistically insignificant (robust p-value: 0.311). Columns (7)–(8) show that for women the rate of readmissions for surgical complications is also low ($\sim$0.7%) and there is no evidence of a discontinuity at age 65. For the full female sample, the estimated discontinuity is 0.14 pp (robust p-value: 0.812).

Table 2.4: Estimated discontinuity in 30-day readmissions at age 65 for female patients

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	14.42	-1.483	3.25	0.050	3.57	0.978	0.70	0.141
		(1.959)		(1.057)		(0.969)		(0.444)
	[w=11.25,N=11,856]		[w=8.20,N=8,971]		[w=10.35,N=11,058]		[w=10.49,N=8,877]	
(B) **White**	14.54	-2.948	2.55	0.409	3.44	0.860	0.81	-0.471
		(2.316)		(1.198)		(1.127)		(0.498)
	[w=10.97,N=7,092]		[w=8.21,N=5,475]		[w=9.08,N=5,958]		[w=7.07,N=3,880]	
(C) **Non-White**	14.23	0.537	4.39	-0.572	3.77	1.508	0.52	0.841
		(3.052)		(1.554)		(1.434)		(1.030)
	[w=9.25,N=3,910]		[w=9.73,N=4,075]		[w=10.33,N=4,292]		[w=8.25,N=2,786]	
(D) **Highest qtrl zipcode income**	14.04	-1.740	2.05	1.580	4.11	-0.018	0.44	0.626
		(3.417)		(1.840)		(2.009)		(0.808)
	[w=10.24,N=2,592]		[w=6.57,N=1,750]		[w=8.08,N=2,113]		[w=6.89,N=1,479]	
(E) **Lowest qtrl zipcode income**	14.41	-2.931	3.53	0.198	4.12	-0.663	0.72	0.305
		(3.629)		(1.798)		(1.649)		(1.125)
	[w=8.38,N=2,501]		[w=9.24,N=2,719]		[w=7.70,N=2,319]		[w=8.99,N=2,060]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns) among women. Other descriptions same as Table 2.2. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tables B.8–B.10 (Appendix B.4) and Figure B.2 (Appendix B.5) perform the same set of robustness tests for women as those discussed previously for men. First, Table B.8 shows that when robust non-clustered standard errors are used, the optimal bandwidths are larger and the point estimates are larger in magnitude (e.g. -1.61 pp vs. -1.48 for all-cause readmissions in the full female sample). However, none of the estimated discontinuities are significantly different from zero.

Second, Table B.9 shows that when a quadratic polynomial is used, the magnitude of the estimated discontinuities are larger but are quantitatively very similar to the baseline estimates. The only difference is that with the quadratic specification, there is some evidence of a reduction in readmissions for surgical complications by 1.13 pp among white female patients (robust p-value: 0.037).

Third, Figure B.2 shows that the baseline result is robust to various bandwidths. Specifically, using a linear interaction specification with a bandwidth between 1 and 10, the estimated discontinuity for women is not statistically different from zero. The CER-optimal bandwidth for this specification is 8.006; Figure B.2 shows that using this bandwidth results in an estimate that is very similar in magnitude to the baseline estimate.

Fourth, Panel A of Table B.10 shows that including an indicator for admission one month before turning 65, indicators for zipcode income quartiles, and payer-specific time indicators does not affect the baseline results. Panel B shows that dropping Medicaid beneficiaries and restricting the sample to NSTEMI patients also does not affect these results.

Lastly, since the sample size for male patients is more than twice as large as the sample size for female patients, one concern is that the difference in results for men and women may be driven by sample size differences. To test whether this is the case, I randomly drop half of the male observations and estimate the same specifications on the 50% sample. Panel A of Table B.11 shows that even with this smaller sample, the estimated discontinuity is quantitatively similar. For example, columns (1) and (2) show the estimated discontinuity in all-cause readmissions is 3.06 pp based on the smaller sample (robust p-value: 0.043), larger but similar in magnitude to the baseline estimate of 2.68 pp. Columns (3) to (4) show that in this smaller sample, there is some evidence of a reduction in readmissions for major vascular conditions (estimate=-1.24 pp, robust p-value: 0.057). Columns (5) to (6) confirm the baseline finding that the increase in all-cause readmissions is driven by readmissions for non-major vascular conditions. Columns (7) to (8) confirm the earlier finding that there are no significant discontinuities in readmissions for surgical complications.

In summary, the results suggest that eligibility for Medicare insurance at age 65 has no effect on the probability of 30-day readmissions among female AMI patients. These results are robust to various alterations to the baseline specification.

6 | Mechanisms: Impacts on treatment intensity

A relationship between eligibility for Medicare insurance and readmissions could capture both patient-side and provider-side factors. In particular, on the provider side, hospitals may alter treatment intensity at the initial stay based on insurance status. To examine whether the relationship between eligibility for health insurance and readmissions is driven by changes in treatment intensity, I estimate discontinuities in treatment intensity around the age eligibility threshold. In particular, I estimate the same specification as equation (17) and replace the dependent variable with the classifications of heart attack treatment procedures discussed in Section 3 (treatment invasiveness, timeliness, length of stay, and number of procedures).

6.1 Men: Discontinuities in treatment intensity

In Table 2.5, I present estimated discontinuities in treatment intensity at age 65, where treatment intensity is measured in four different ways. In columns (1) and (2), treatment intensity is indicated by whether or not the patient received an invasive procedure such as catheterisation or angioplasty during the index stay. This definition is motivated by the finding that patients who receive invasive treatments tend to have better health outcomes (Bhatt et al., 2004; Currie et al., 2016; Cutler et al., 1998; Winter et al., 2005). In columns (3) and (4), treatment intensity is indicated by whether or not the patient received catheterisation within 48 hours of admission. This definition follows Molitor (2018) and can be seen as both a precondition for further invasive treatment and an indication of timely diagnosis.[82] In columns (5) to (6), the dependent variable is the length of stay of the first admission. In columns (7) to (8), the dependent variable is the log of the number of procedures recorded on the bill of the index stay.

Among male AMI patients, there is limited evidence that the discontinuous increase in readmissions could be driven by patients over 65 receiving less intensive treatments. Using the invasiveness definition (columns 1–2), the results show that almost 80% of male patients receive a catheterisa-

[82] As discussed in Section 3, catheterisation is an invasive diagnostic procedure that often precedes further invasive treatments such as bypass surgery.

Table 2.5: Estimated discontinuity in treatment intensity at age 65 for male patients

	Invasive treatment		Cathsn w/in 48 hrs		Length of stay		Num. procedures	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	79.92	0.200	19.30	-1.167	4.09	-1.096	6.92	-2.604
		(1.422)		(1.284)		(2.179)		(1.848)
	[w=9.36,N=21,909]		[w=6.94,N=16,724]		[w=8.90,N=20,858]		[w=7.90,N=18,733]	
(B) **White**	80.27	-0.784	18.73	-3.790***	3.95	-2.130	6.98	-3.479**
		(1.744)		(1.363)		(2.737)		(1.955)
	[w=8.96,N=13,789]		[w=9.76,N=14,906]		[w=7.61,N=11,977]		[w=8.69,N=13,451]	
(C) **Non-White**	79.23	2.849	20.44	2.858	4.37	2.736	6.80	0.020
		(2.184)		(2.599)		(3.203)		(3.585)
	[w=6.23,N=5,144]		[w=6.23,N=5,144]		[w=6.65,N=5,451]		[w=7.20,N=5,898]	
(D) **Highest qtrl income**	82.62	-2.548	17.77	-0.489	3.91	3.001	7.15	-2.379
		(2.766)		(2.791)		(3.011)		(3.005)
	[w=7.51,N=5,161]		[w=8.91,N=5,893]		[w=8.87,N=5,893]		[w=7.86,N=5,331]	
(E) **Lowest qtrl income**	79.15	-1.227	19.38	3.299	4.38	-5.647	6.75	-3.652
		(3.039)		(3.366)		(4.093)		(4.428)
	[w=8.75,N=4,595]		[w=6.24,N=3,342]		[w=12.55,N=6,367]		[w=7.07,N=3,714]	

Notes. This table presents the estimated discontinuity in various measures of quality of care at the patient's initial admission for a heart attack among men. In columns 1–2, the dependent variable is an indicator for whether the patient received cardiac catheterisation or an angioplasty during their initial hospitalization. In columns 3–4, the dependent variable is an indicator for whether the patient received cardiac catheterisation within 48 hours of admission. In columns 5–6, the dependent variable is the log length of stay. In columns 7–8, the dependent variable is the log number of procedures received. Other descriptions same as Table 2.2. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

tion or an angioplasty during their index stay and that there is no significant discontinuity in this proportion at age 65 (estimate: 0.20 pp, robust p-value: 0.934). Rows B to E show that across all demographic subsamples, there is no discontinuity in the probability of receiving an invasive treatment.

Using the "timely diagnosis" definition, columns (3) and (4) show that there is a statistically insignificant fall of -1.17 pp in the proportion of patients receiving catheterisation within 48 hours of admission, relative to a baseline of 19.30%. Row B column (4) indicates that among white patients, the probability of receiving catheterisation within 48 hours decreases discontinuously by 3.79 pp. However, robustness checks below show that this discontinuity is not statistically significant with a quadratic functional form. Row C column (4) shows that among the group which saw readmissions increase discontinuously (non-whites), the timely-diagnosis measure is smooth across the threshold.

Together, these results do not suggest that the discontinuity documented earlier is due to patients at the age threshold receiving worse quality of care at their initial hospitalisation.

Figure 2.4 provides a graphical representation of these RD estimates for the full sample. The method used to construct this figure is as described before in Section 5 of this chapter. In panel A of this figure, I plot the age profiles of invasive treatment; in panel B, I plot the age profiles of timely diagnosis. The age profiles in panel A is downward sloping as older patients become less suitable for invasive treatment procedures. The age profiles in panel B is fairly flat from ages 55 to 75. For men (shaded circles/triangles), the age profiles for both treatment intensity measures appear smooth across the threshold.

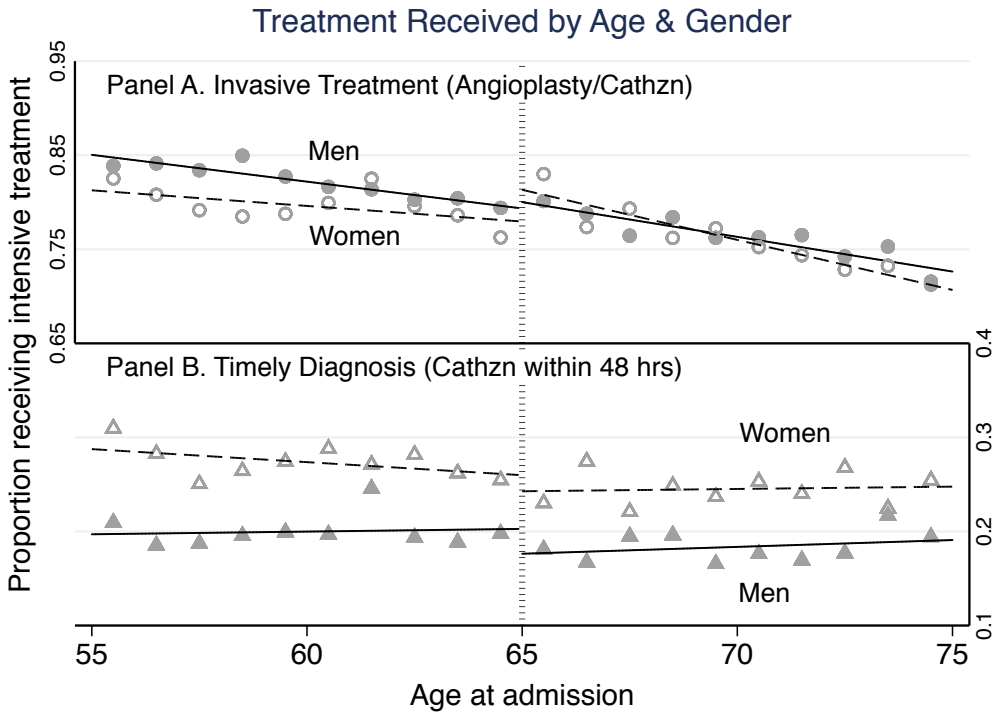


Figure 2.4: Treatment intensity received at initial hospitalisation by gender and age at admission. "Cathzn" stands for cardiac catheterisation.

Columns (5) to (6) of Table 2.5 indicate that there are no discontinuities in length of stay across all subgroups. Among the full sample, the average number of days in hospital is 4 and the estimated discontinuity at age 65 is -1.10. Columns (7) to (8) indicate some heterogeneity in the impact of eligibility for Medicare insurance on the number of procedures received during the patient's initial admission. Among white patients, there is a significant fall in the number of

procedures received. In particular, eligibility for Medicare insurance leads to a 3.5% drop in the number of procedures for this group.

Table B.12 in Appendix B.4 estimates the treatment intensity specifications using a quadratic polynomial in age. Most of the results are similar. In particular, there are no significant discontinuities in any measure of treatment intensity for most demographic subgroups, apart from a reduction in the number of procedures for white AMI male patients. Figure B.3 shows that using a linear specification with varying bandwidths produces very similar results. For all possible bandwidths between 1 and 10 years, the estimated discontinuity in the invasive treatment indicator is not statistically different from zero. For bandwidths between 3 and 10, the estimated discontinuity is stable and very close to the baseline estimate of -0.20. Panel B of Table B.11 drops half the male sample to test whether these results are affected by sample size. The results indicate that for this smaller sample, the estimated discontinuities are statistically insignificant for all measures of treatment intensity.

In summary, the results do not suggest that there is a relationship between treatment intensity during the initial hospital stay and the discontinuous increase in readmissions documented in Section 5. In particular, among groups that saw a discontinuous increase in readmissions (non-whites), treatment intensity is smooth across the age threshold. Under the assumption that treatment quality primarily captures provider-side responses, the smoothness of treatment intensity and the discontinuous increase in readmissions suggests that the increase is likely to be driven by patient-side responses to insurance eligibility.

6.2 Women: Discontinuities in treatment intensity

Table 2.6 examines the impact of eligibility for Medicare insurance on treatment intensity for women and reveals several interesting patterns. First, comparing the pre-mean variables for men in Table 2.5 with the pre-mean variables for women in Table 2.6 shows that women are less likely to be treated with an invasive procedure during their initial stay. This is despite the fact that women are more likely to receive catheterisation which is usually a precursor for invasive treatments. In

particular, women just under 65 are 3 pp less likely than men to receive an invasive treatment and 5 pp more likely than men to receive catheterisation. This holds even conditional on characteristics that might influence suitability for such procedures.[83] This "gender gap" in heart attack treatments has been documented by Swannell (2018), Mosca et al. (2011) and others.

Table 2.6: Estimated discontinuity in treatment intensity at age 65 for female patients

	Invasive treatment		**Cathsn w/in 48 hrs**		**Length of stay**		**Num. procedures**	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	77.42	6.941*** (1.838)	25.83	-1.882 (2.117)	4.40	-0.536 (2.233)	6.10	6.759** (3.622)
	[w=4.35,N=4,971]		[w=10.72,N=11,387]		[w=8.46,N=9,236]		[w=5.58,N=6,227]	
(B) **White**	76.40	11.154*** (2.604)	25.77	-2.122 (2.536)	4.35	-3.078 (2.691)	6.08	6.587* (4.295)
	[w=4.30,N=2,994]		[w=10.84,N=7,048]		[w=9.75,N=6,405]		[w=5.71,N=3,907]	
(C) **Non-White**	79.08	1.108 (2.442)	25.94	-1.361 (3.240)	4.47	3.964 (4.381)	6.12	5.072 (5.247)
	[w=7.59,N=3,289]		[w=8.19,N=3,496]		[w=6.62,N=2,873]		[w=8.06,N=3,438]	
(D) **Highest qtrl income**	74.66	7.694* (3.934)	27.05	-7.659** (4.254)	4.73	-2.195 (5.107)	5.98	5.146 (6.607)
	[w=7.74,N=2,026]		[w=8.76,N=2,288]		[w=11.65,N=2,912]		[w=6.64,N=1,767]	
(E) **Lowest qtrl income**	78.53	1.062 (3.511)	25.59	-3.094 (3.365)	4.55	1.401 (5.335)	6.05	6.886 (5.335)
	[w=6.69,N=1,999]		[w=9.14,N=2,694]		[w=6.36,N=1,902]		[w=6.43,N=1,926]	

Notes. This table presents the estimated discontinuity in various measures of quality of care at the patient's initial admission for a heart attack among women. Other descriptions same as Table 2.2. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Second, in contrast to the smoothness of treatment intensity for men across the age threshold, there is a discontinuous increase of 6.94 pp in the proportion of women receiving an invasive treatment (row A, column 2) even though there is no discontinuous increase in the proportion of women receiving a catheterisation (row A, column 4). This can be seen graphically in Panels A and B of Figure 2.4. Panel A shows that before age 65, women's age profile for invasive treatment (hollow circles) lies below that of men (solid circles); at age 65, there is an increase of roughly 7 pp in the probability of receiving an invasive treatment at the age-threshold which reduces the gap between the male and female age profiles. Furthermore, the increase in treatment intensity

[83] For example, the coefficient on the female dummy is significantly negative in a linear probability model which regresses the indicator for receiving an invasive treatment on gender, race, age and co-morbidities.

appears to be concentrated among women admitted at age 65, rather than diffused across women admitted at older ages. Panel B shows that the age profile of receiving catheterisation within 48 hours of admission is 5 pp higher for women (hollow triangles) than men across all ages. The age profile is fairly flat and there is no discontinuity at the age threshold.

Rows B to E of column (2) provide evidence that the increase in treatment intensity is concentrated among white women and those living in high-income zipcode areas. Among white females, the estimated discontinuity in the probability of an invasive treatment is 11.15 pp (p-value=0.000). Among females living in zipcodes where the median household income falls in the top quartile, the estimated discontinuity is 7.69 pp (robust p-value=0.0528). One possible reason for the heterogeneous effects on treatment intensity is that white and high-income groups exhibit larger increases in Medicare enrolment at 65 than other racial and income groups (see Table 2.2).

Columns (5) to (6) of Table 2.6 indicate that on average women have a longer length of stay than men and that there is no discontinuity in the length of stay at the initial admission. For example, the estimated discontinuity in the full sample is -0.54 pp (robust p-value: 0.953). Rows B to E show that there is no discontinuity in length of stay for any of the demographic subsamples. Finally, columns (7) to (8) suggest that there is an increase in the number of procedures received by female AMI patients at age 65. In the full sample, there is a 6.76% increase in the number of procedures received relative to a baseline of 6.1 procedures during the index stay. This is almost fully accounted for by the increase in the proportion of women at age 65 receiving an angioplasty which increases by over 6 pp at age 65. Rows B to E indicate that this increase is driven by white female patients, consistent with the findings in columns (1) and (2).

There are several implications and interpretations of the gender gap in treatment invasiveness and the estimated discontinuity at age 65 for women. First, the estimated discontinuity in treatment invasiveness suggests that for women the age-65 cutoff captures both changes in patient-side and provider-side behaviour. It is possible that these changes to the treatments received by women at age 65 offset any effects on readmission rates of patient-side responses to changing insurance eligibility. Second, the discontinuous increase in the proportion of women receiving an angioplasty or bypass surgery suggests that Medicare has the effect of reducing gender differences in treatment

intensity. Before age 65, the raw gap between the proportion of men and women receiving invasive treatments is roughly 3 pp; at age 65, this gap reduces significantly.

Lastly, these results raise the question of why there are supply-side responses for women but not for men. Card et al. (2008) suggest that increases in procedures at age 65 are consistent with there being fewer restrictions in Medicare than in many private insurance plans; however, since there are unlikely to be gender differences in restrictions under Medicare, it is unclear why such effects would exist for women but not for men. One potential reason is that the gender gap in treatment intensity among patients under 65 reflects women receiving suboptimal levels of invasive treatment; Medicare decreases this gap by establishing a set of guidelines or standards for how patients should be treated at their initial hospitalisation.

To check the robustness of these results, Table B.13 shows that these results are robust to using a quadratic polynomial. In particular, using the quadratic specification, the probability of receiving an invasive treatment increases by 7.14 pp and the number of procedures received increases by 6.48% at age 65. Figure B.4 examines the sensitivity of the estimated discontinuity in the proportion of women receiving invasive treatments to varying bandwidths. The figure shows that for all bandwidths from 1.75 to 10, the estimated discontinuity in the probability of receiving an invasive treatment is positive and statistically significant. For bandwidths between 2 to 5, the estimate is similar to the baseline estimate of 6.94. The point estimate then decreases from 6.94 at a bandwidth of 5 to 4.13 at a bandwidth of 10. This decrease is likely due to the fact that the discontinuity is concentrated around the age 65 threshold, rather than dispersed through the age distribution (see Figure 2.4).

7 | Discussion and Conclusion

This chapter asks whether changes in insurance eligibility influences post-discharge outcomes such as readmission rates, focusing on demand-side responses to insurance. I answer this question by examining how the age-65 eligibility threshold for Medicare insurance affects readmission probabilities and treatment intensity among a group of heart attack patients.

The results differ by gender. Among men, there is evidence that eligibility for Medicare insurance affects post-discharge outcomes. In particular, the estimated discontinuity in 30-day readmissions is an increase of 2.68 pp at age 65, relative to a baseline of 11% among patients aged 63–64. The heterogeneous results by reason for readmission and by demographic subgroups also support the interpretation that this discontinuity captures the effect of insurance. There are no discontinuities in treatment intensity received at the initial hospitalisation. This suggests that the increase in readmission rates is not driven by patients above 65 years old receiving worse quality of care at their initial hospitalisation. Insofar as observable treatment procedures primarily capture supply-side responses, these results provide some evidence that patient-side responses to insurance eligibility may also affect post-discharge outcomes.

Among women, there is no evidence that eligibility for Medicare insurance affects post-discharge outcomes. Interestingly, there are significant increases in the probability of female heart attack patients receiving more invasive treatments at the age threshold. One possibility is that these changes to the treatments received by female heart attack patients at age 65 offset any effects on readmission rates of patient-side responses to changing insurance eligibility. These findings also raise questions about gender differences in the level and change in quality of treatment. In particular, below 65, women are less likely than men to receive an invasive treatment at their initial hospitalisation; at age 65, this gender difference in treatment procedures narrows. Part of this reduction in the gender treatment gap may be due to the impact of Medicare eligibility increasing invasive treatment for women but not for men.

Future work can proceed in two directions. The first is to study the implications of the increase in readmissions documented among men. This chapter does not shed light on whether the increase

in readmission rates due to Medicare is good or bad for the patient. It would be interesting to investigate whether increased access to post-discharge healthcare for male patients above 65 is beneficial for longer-term health outcomes such as mortality. The second direction for future research is to investigate why eligibility for Medicare increases the probability of receiving an invasive treatment among female patients but not for male patients.

Chapter III

Obstacles on the Road to School:

The Impact of Mobility Restrictions on Educational Performance

Chapter Abstract

This chapter turns to the subject of educational performance. It examines the impact of conflict-related mobility restrictions on educational performance in the West Bank between 2000 to 2006 during the Israeli-Palestinian (IP) conflict. A prominent feature of this conflict is the system of mobility restrictions enforced through various physical barriers such as checkpoints, infrastructures manned by security personnel who conduct checks on commuters. Using data on the home and school location of a student and data on the location of barriers over time, we find strong evidence that mobility restrictions in the form of checkpoints have adverse impacts on educational performance. Crossing one or more checkpoints on the road to school reduces the probability of passing the final high school exam by over 3 percentage points (pp) and the overall score by 0.07 standard deviations. The impacts of mobility restrictions operate through a distinct channel to conflict-related violence which has been the focus of the previous literature. We find evidence of three mechanisms at play. First, each additional minute of travel time due to delays at checkpoints reduces the probability of passing by 0.11 pp. Second, mobility restrictions significantly increase the probability that students have issues concentrating. Third, mobility restrictions negatively impact the physical resources at school, potentially worsening the school learning environment.

This chapter is co-authored with Sami Miaari. We thank the Palestinian Ministry of Education and Higher Education (MOEHE) for providing data on exam results. We are also grateful to the Palestinian Central Bureau of Statistics for the Palestinian Labor Force Survey data.

1 | Introduction

Conflicts have adverse impacts on individual and aggregate human capital accumulation, impeding the long-term development of countries under strife (Justino, 2012; UNESCO, 2011). A growing literature illustrates the negative impact of conflict-related violence on educational attainment, an important aspect of human capital accumulation that often serves as a stepping stone for good future labour market outcomes (Brück et al., 2019; Leon, 2012; Monteiro and Rocha, 2017). This literature has not explored the impacts of another prominent feature of contemporary conflicts: conflict infrastructures such as walls, barriers, and buffer zones (Pullan, 2013).[84] These conflict infrastructures often result in restricted mobility and affect the daily routine of those living in close proximity; for example, by increasing travel costs required to arrive at school. Despite the major role played by conflict infrastructures in contested areas, little is known about their impact on the educational attainment of students.

This chapter investigates the impact of conflict infrastructure and resulting mobility restrictions on the educational performance of high school students. We study this question in the context of the Israeli-Palestinian (IP) conflict between 2000 and 2006, a period of intensified conflict known as the Second Intifada. We focus on performance in the final high school exam (the *Tawjihi General Examination* or TGE) among students studying in state schools in the West Bank. This context offers a unique setting to study this question for two main reasons. First, a prominent feature of the IP conflict is the system of mobility restrictions enforced in the West Bank through various physical barriers. The most prominent of these barriers are checkpoints which Palestinians often cross to commute between Palestinian villages and between Israel and the West Bank. These checkpoints are usually manned by Israeli security personnel who conduct checks and searches on Palestinian commuters. Second, exam outcomes in this context have important implications for young Palestinians. Achieving a high TGE score is a necessary prerequisite for admission to university and to apply for well-paid public sector jobs. To the best of our knowledge, this is the first study that examines the impact of conflict-related mobility restrictions on educational performance.

[84] Recent examples of conflict infrastructure include the buffer zone in Nicosia during the Greek-Turkish crisis, the Berlin wall between 1961 and 1989, peace walls in Northern Ireland, the green line between Sunni and Shia communities in Beirut, the main Boulevard in Mostar which divides East and West, and contested lands in Kirkuk (Pullan et al., 2012).

Methodology. To answer this question, we use several unique and novel data sources on conflict exposure and educational performance. Our data has at least three advantages that allow us to answer our question of interest. First, it allows us to measure the multiple dimensions of the IP conflict. These dimensions include exposure to conflict infrastructure such as checkpoints, our primary measure of mobility restrictions, and exposure to conflict-related violence such as fatalities. This allows us to distinguish between the impacts of conflict-related violence, which have been the focus of the existing literature, and those of conflict-related mobility restrictions which are the focus of this study. Second, it allows for a more granular measure of exposure to conflict. We measure exposure to conflict-related violence at the level of Palestinian localities (or villages) with an average size of 8.5 km^2 and at a frequent time interval. Using information on the student's home and school locality as well as data on the location of barriers, we construct measures capturing whether a particular student encounters a barrier on the shortest route to school in the year they sit the TGE. Third, the data allows us to explore the mechanisms underlying a potential relationship between conflict and educational attainment.

With this data, we examine the impact of exposure to conflict on performance in the TGE for a given student living in home locality h and attending a school in locality l. Our main empirical strategy is a within-home locality and within-school locality specification that exploits variation across time in the probability that a student faces a barrier along the shortest route to school.[85] The key identifying assumption is that the probability that student i encounters a barrier on the road to school in academic year t is independent of their potential outcomes, conditional on time-invariant home-locality characteristics, time-invariant school-locality characteristics, distance travelled to school, time-specific characteristics, and other observable student and school characteristics that we can control for.

Main Results. Our baseline results provide strong evidence that mobility restrictions in the form of checkpoints have adverse impacts on the educational performance of students. Crossing one or more checkpoints reduces the probability of passing the final exam by 3.05 percentage points (pp) and the overall test score by 0.071 standard deviations. The magnitude of these effects are larger

[85] More specifically, the main specification includes both home locality and school locality fixed effects, but doesn't include fixed effects for each home locality-school locality pair.

than our estimates of the impact of an additional fatality in the school locality. Conditional on exposure to barriers, an additional fatality in the school locality reduces the probability of passing by 0.07 pp and the overall test score by 0.001 standard deviations. The estimated impacts of mobility restrictions are also comparable in magnitude to the impacts of other inputs into the education production function examined in the existing literature. For example, the estimated impact of crossing at least one checkpoint is equivalent to one-half to two-thirds of the decrease in test scores associated with a one standard deviation decrease in teacher quality in the US (Rivkin et al., 2005; Rockoff, 2004) or a two standard deviation reduction in school management quality in Israel (Lavy and Boiko, 2017).

The effects of mobility restrictions on educational performance appear to be heterogeneous along several dimensions. First, the effects vary across exam subjects, with Maths exam scores being particularly adversely impacted. In particular, our baseline estimates suggest that crossing at least one checkpoint on the road to school reduces Maths scores by over 0.073 standard deviations and English scores by 0.053 standard deviations. Second, the adverse impacts of exposure to mobility restrictions and violence are concentrated among male students. Third, different types of barriers have different effects on educational performance. While encountering checkpoints on the road to school has a strong adverse effect on TGE performance, encountering other types of barriers such as roadblocks, earthmounds, and gates do not appear to have an adverse impact.

We vary our main empirical specification in multiple ways to examine the robustness of our baseline results. When we interact home and school locality fixed effects with academic year fixed effects to allow for time-varying home and school locality unobserved characteristics, the estimated impact of crossing at least one checkpoint on the road to school remains very stable. When we include fixed effects for each home locality-school locality pair so that identification comes from comparing differences in outcomes of students who travel on routes that experience a change in the number of checkpoints with differences in outcomes for those who travel on routes that do not experience a change, the estimates are not statistically different from the baseline results. We also provide evidence that our baseline results are not driven by other potential confounding factors such as students moving endogenously within the West Bank in response to conflict or sample selection of the group of students sitting the final exam.

Motivated by the finding that mobility restrictions have an adverse impact on educational performance, we proceed to investigate four potential mechanisms that might give rise to a negative relationship between mobility restrictions and educational performance. The first mechanism concerns time loss due to delays at checkpoints. Students who cross checkpoints on the road to school lose time in the morning and early evening, reducing their study time. To investigate whether this mechanism is driving our results, we use newly-collected data on delay factors incurred at specific checkpoints across the West Bank. We examine how the delays faced by student i living and studying in localities h and l in year t affect their exam performance, conditional on the number of checkpoints they cross, the distance between h and l, time-invariant confounders at the home and school locality level, and other observable student and school characteristics. The variation in delay times across specific checkpoints allows us to separately identify the impact of delays from the impact of crossing an additional checkpoint. We find that each additional minute delay reduces the probability of passing the exam by 0.11 pp and the overall score by 0.002 standard deviations. Existing literature indicates that delay times of at least 15 minutes were the norm during the Second Intifada. Therefore, these estimates suggest the delays incurred by crossing a checkpoint reduce the probability of passing the exam by 1.65 pp and the overall score by 0.03 standard deviations. The adverse impact of the delays can account for over half of the estimated impact of crossing at least one checkpoint on the road to school.

The second mechanism suggests that mobility restrictions have adverse impacts on the psychological wellbeing of students which may leave them in a worse state to learn and study. Since the administrative exam data does not contain information on psychological wellbeing, we supplement our analysis with additional survey data which measures the cognitive, emotional, and behavioural problems experienced by students in the West Bank. We link the survey data to measures of conflict at the home locality level. We find evidence that among students who study outside their home locality, the presence of one or more checkpoints within 10 km of their home locality significantly increases the probability that a student reports suffering from a lack of concentration. Exposure to checkpoints around the home locality also increases tendencies towards violent behaviour. These findings hold even when we condition on a rich set of student and household characteristics, suggesting that these results are not driven by particular types of students being more likely to be

exposed to mobility restrictions.

The third mechanism we consider is that mobility restrictions have negative effects on the physical school learning environment. We estimate school-level within-school models that exploit variation across academic years in the number of physical barriers surrounding a given school and the number of fatalities occurring in the school locality in the months leading up to the exam. We find that the introduction of checkpoints within 10 km of the school locality centre reduces the probability of schools having a library, science lab, or computer lab. One potential reason for this is that over this period goods crossing checkpoints often have to be off-loaded and inspected by security personnel. These additional transport costs may deter schools from ordering books, lab equipment and computers which are required to keep these facilities running.

The fourth and final mechanism we consider is the impact of conflict on teachers. For each school, we calculate the proportion of teachers who cross a checkpoint on the way to school using information on the home and school localities of teachers. We then regress school-level average educational outcomes on teacher exposure measures, the number of checkpoints surrounding the school locality, the number of fatalities, allowing for school fixed effects and academic year fixed effects. We find that the coefficients on the teacher exposure variables are statistically insignificant and economically small, suggesting that measures of teacher exposure to conflict do not help explain within-school variation in student outcomes in our context. We are unable to examine other ways in which teachers may impact exam performance, such as via teacher absenteeism, which could be important in this context.

Related Literature and Contribution. This chapter is related to several strands of literature. First, it is most closely related to a small literature that examines the relationship between exposure to conflict and educational performance. Most similar is Brück et al. (2019) who study the effects of the IP conflict on the probability of Palestinian students passing their final high school exam. They find that exposure to conflict, measured by Palestinian fatalities, reduces the probability of passing the final exam and being admitted to university. Shany (2016) examines the effects of violence in Israel during the Second Intifada and provides evidence that the occurrence of fatal attacks shortly before an exam has significant adverse effects on the exam performance of Israeli

high school students.

A handful of papers examine this relationship in other contexts. Kibris (2015) combines data on university entrance exam scores with data on exposure to fatalities resulting from the Turkish-Kurdish conflict and finds a significant negative association between exposure to conflict and exam performance. Monteiro and Rocha (2017) investigate the impact of being exposed to drug-related conflicts in Brazilian favelas. They find that gang violence has a negative impact on Maths test scores among fifth-grade students. Furthermore, they provide evidence that this relationship is primarily driven by supply-side factors: local violence by drug gangs has a large negative impact on the supply of school resources.

This chapter contributes to this literature by studying the impact of a different aspect of conflict, mobility restrictions, on educational performance while accounting for other dimensions of conflict such as fatalities which might be correlated with mobility restrictions and affect educational attainment. Our use of multiple novel data sources also allows us to explore several mechanisms that might explain such impacts.

Second, this chapter is related to a growing literature examining the impacts of conflict on years of schooling. Di Maio and Nandi (2013) study the impact of the IP conflict on school attendance and child labour of Palestinian children in the West Bank during the Second Intifada. They find that an increase in the number of days when the mobility of Palestinians is severely restricted (via borders, curfews, and sieges) increases child labour while it weakly reduces school attendance. Consistent with this, Al Kafri (2002) finds that between 2000 and 2001, child labour increased while school attendance decreased for male children.

Other papers examine the impact of conflicts in other countries on years of completed schooling. Several papers document heterogeneous effects by gender and by type of schooling (primary or secondary). Leon (2012) finds that the average person exposed to the Peruvian civil conflict between 1980 and 1993 accumulated 0.31 fewer years of education as an adult, with the effect being stronger for girls. Shemyakina (2011) finds that girls who were of school age during the 1992–1998 Tajikistan conflict were less likely to complete their mandatory schooling which spans ages 7 to 15. Swee (2015) finds that cohorts more exposed to the 1992–1995 Bosnian war were less

likely to complete secondary schooling but not primary schooling and that this effect was stronger for males than females. Justino et al. (2013) find that the 1999 wave of violence during the Timor Leste conflict had strong negative impacts on the primary school completion among boys.

Third, this chapter is related to the literature on the impacts of commuting time on labour market outcomes and education. In the context of the IP conflict, it is well documented that physical barriers in the West Bank increase commuting time. In this vein, this chapter is most similar to Calì and Miaari (2018) who provide evidence that these physical barriers have a significant negative effect on labour market outcomes of Palestinian workers in the West Bank. In particular, checkpoints have a significant negative effect on employment, wages, and days worked. Aranki (2006) and Miaari and Sauer (2011) both estimate the effect of mobility restrictions on the Palestinian labour market using data from the Palestinian Labor Force Survey. Their results suggest that mobility restrictions increase the probability of being unemployed and decrease the monthly earnings of Palestinian workers regardless of whether they work in Israel or in Palestinian territories. In the context of schooling, increased commuting times due to physical barriers may hinder the ability of students and teachers to reach their schools, reducing the number of school days attended and the supply of school resources. Consistent with this hypothesis, Tigre et al. (2017) use Brazilian data and find strong evidence that longer commuting times to school (in non-conflict situations) have a negative effect on academic achievement.

To the best of our knowledge, this analysis is the first to explore how conflict-related physical barriers increase commuting time and affect academic performance. Our rich data sources allow us to construct delay factors due to checkpoints experienced by students on the road to school.

Fourth, this chapter is related to a literature that documents the impact of conflict-related violence on students' cognitive, emotional, and behavioural traits. A large psychiatric and psychological literature provides evidence that students' psychological well-being is an important determinant of academic achievement (Roeser et al., 1998) and that exposure to violence is associated with increased risk of Post-Traumatic Stress Disorder (Hoven et al., 2003; Pfefferbaum et al., 1999, 2000; Schwarzwald et al., 1993). With regard to the IP conflict, high levels of emotional and behavioural problems and PTSD are well documented among children and adolescents in the West Bank, with

a greater effect being found among those who live in more intense conflict zones (Al-Krenawi et al., 2009; Dimitry, 2012; Mataria et al., 2009; Schiff et al., 2007; Thabet et al., 2002).

While the existing literature focuses on the impact of conflict-related violence on cognitive and psychological wellbeing, our analysis draws on detailed survey data to investigate whether conflict infrastructure and resulting mobility restrictions also adversely impact psychological wellbeing.

The rest of this chapter is structured as follows. In Section 2, we provide details about the setting of the IP conflict, the educational system in the West Bank, and potential reasons for a relationship between mobility restrictions and academic performance. In Section 3, we discuss the data sources used. In Section 4, we outline our empirical specification and the identification assumptions underlying the specification. In Section 5, we present our main results. In Section 6, we explore potential mechanisms that may explain our main findings. Finally, Section 7 concludes.

2 | Background

This section provides information on the key features of the Israeli-Palestinian conflict and the educational system in the West Bank that will inform our empirical analysis.

2.1 The Israeli-Palestinian conflict and Mobility Restrictions

The Israeli-Palestinian (IP) conflict is one of the longest-lasting conflicts in the world. As a result of the Six Days War in 1967, the West Bank and Gaza Strip (previously part of Jordan and Egypt, respectively) fell under Israeli control. In December 1987, a Palestinian uprising against Israeli control (the "First Intifada") broke out, culminating in the signing of the Oslo Accords in 1993. The Accords established a set of agreements between Israel and Palestine and granted Palestinians control over civilian matters (e.g. education, health, and taxation) in the West Bank and Gaza Strip under the direction of the Palestinian National Authority. After a relatively peaceful period following the Oslo Accords, the "Second Intifada" broke out in September 2000, characterised by violent clashes between Palestinians and the Israeli Defence Force (IDF). While there is no official end date to the Second Intifada, violence decreased substantially after 2006. The data used in this analysis focuses on the West Bank and spans the period of the Second Intifada (2000 to 2006).

The primary focus of this chapter is the impact of conflict infrastructure on the educational performance of high school students. The IP conflict offers a unique setting to study this question because a main feature of this conflict is the multiple types of physical barriers (such as checkpoints, road blocks, separation walls, and gates) placed across the West Bank by the IDF for security reasons. These physical barriers result in a system of mobility restrictions. In the 1990s, this system was loosely enforced and did not represent a serious obstacle to mobility within the West Bank. However, following the outbreak of the Second Intifada, this system of mobility restrictions was strengthened, restricting the movement of goods and labour across Palestinian-Israeli borders and within the West Bank.

These restrictions are enforced through a system of manned and unmanned physical barriers

placed on roads and at the entrance to villages, towns, and cities. The most prominent of these barriers are permanent checkpoints (henceforth, *checkpoints*), which were first imposed in 1995 (Calì and Miaari, 2018).[86] Checkpoints are infrastructures, usually manned by Israeli security personnel, preventing the movement of vehicles and pedestrians without documents deemed adequate by the personnel manning the barriers. Checkpoints have the stated aim of enhancing the security of Israel and Israeli settlements. The majority of permanent checkpoints are located well within the West Bank (internal checkpoints) and increase the commuting cost of travelling within the West Bank. The remaining (external checkpoints) are the last inspection points before entering Israel. Over this period, the movement of Palestinians in and out of Israel was controlled by a permit regime.

These physical barriers to mobility affect the daily routine of Palestinian students, teachers, and workers by increasing the travelling time and costs required to arrive at their school or work place. Who can cross checkpoints and when depends on the type and location of the checkpoint (external or internal) and how frequently it is staffed.[87] Whether commuters can cross a checkpoint by foot or car also depends on various factors.[88] While there is no standardised procedure to cross checkpoints, this process usually involves several features (Braverman, 2011, 2012). At internal checkpoints that are manned by security staff, commuters are usually required to form a single-file line and are processed one at a time by showing their identification card to the security personnel (B'Tselem, 2007b). At external checkpoints, the checks involved are more stringent. Commuters must first present their crossing permit and pass through a revolving metal gate ("carousels") equipped with a metal detector (OCHA, 2006; Rijke and Minca, 2019). This activity is monitored by soldiers and security guards working at the checkpoint. At checkpoints that allow vehicles, drivers must also exit their vehicle so that it can be searched. Delays incurred at checkpoints vary depending on various factors such as the traffic volumes at the checkpoint, the number of personnel at the checkpoint, and even the mood of those guarding the checkpoint. Over this

[86] In contrast to fixed checkpoints, "flying" checkpoints are temporary and may be erected along roads in the West Bank at short notice.

[87] See https://www.btselem.org/freedom_of_movement/checkpoints_and_forbidden_roads for examples of rules for crossing current checkpoints in the West Bank and Gaza Strip.

[88] During the Second Intifada, Palestinian vehicles were restricted from crossing external checkpoints (Braverman, 2011). Some internal checkpoints allow both cars and pedestrians. Most internal checkpoints within the district of Hebron can only be crossed by foot.

period, typical delay times range from 15 minutes to 3 hours (B'Tselem, 2007b; Eklund, 2010; Eklund and Martensson, 2012)

In addition to permanent checkpoints, other types of physical barriers (henceforth, *other barriers*) have also been erected across the West Bank to prevent vehicular movement. These barriers include roadblocks (concrete blocks stretched across roads), earth mounds (mounds of rubble), road gates, agricultural and barrier gates (metal gates which provide the only access through the West Bank wall to the so-called "seam zone"), trenches and earth walls. These additional barriers began to be imposed from 2001 onwards. These other barriers are likely to affect commuters who travel by vehicle rather than foot. While commuters who travel by foot can circumvent these barriers by foot and then find an alternative mode of transport on the other side if they so wish, vehicles must stop and find an alternative route when such barriers are encountered.

Figure 3.1 depicts two maps of the West Bank at the start and end of our sample period. The maps indicate the location of checkpoints in large red circles and the location of other physical barriers in smaller yellow circles. The maps firstly indicate temporal variation in the number of physical barriers scattered across the West Bank. The number of checkpoints grew by twofold from 2000 to 2006 and there were no other types of barriers in 2000. The maps also indicate geographic variation in the physical barriers faced over time. The map depicts the 11 districts (or governorates) in the West Bank,[89] which contain smaller geographic units called localities.[90] In the district of Hebron which contains 12 localities, the average number of checkpoints within a 10 km radius of the center of each locality grew from 0.25 in 2000 (max=2) to 5.16 in 2006 (max=9). In Tubas, the growth in the number of checkpoints was less significant, rising from 0.42 (max=4) in 2000 to 1.83 in 2006 (max=9).[91]

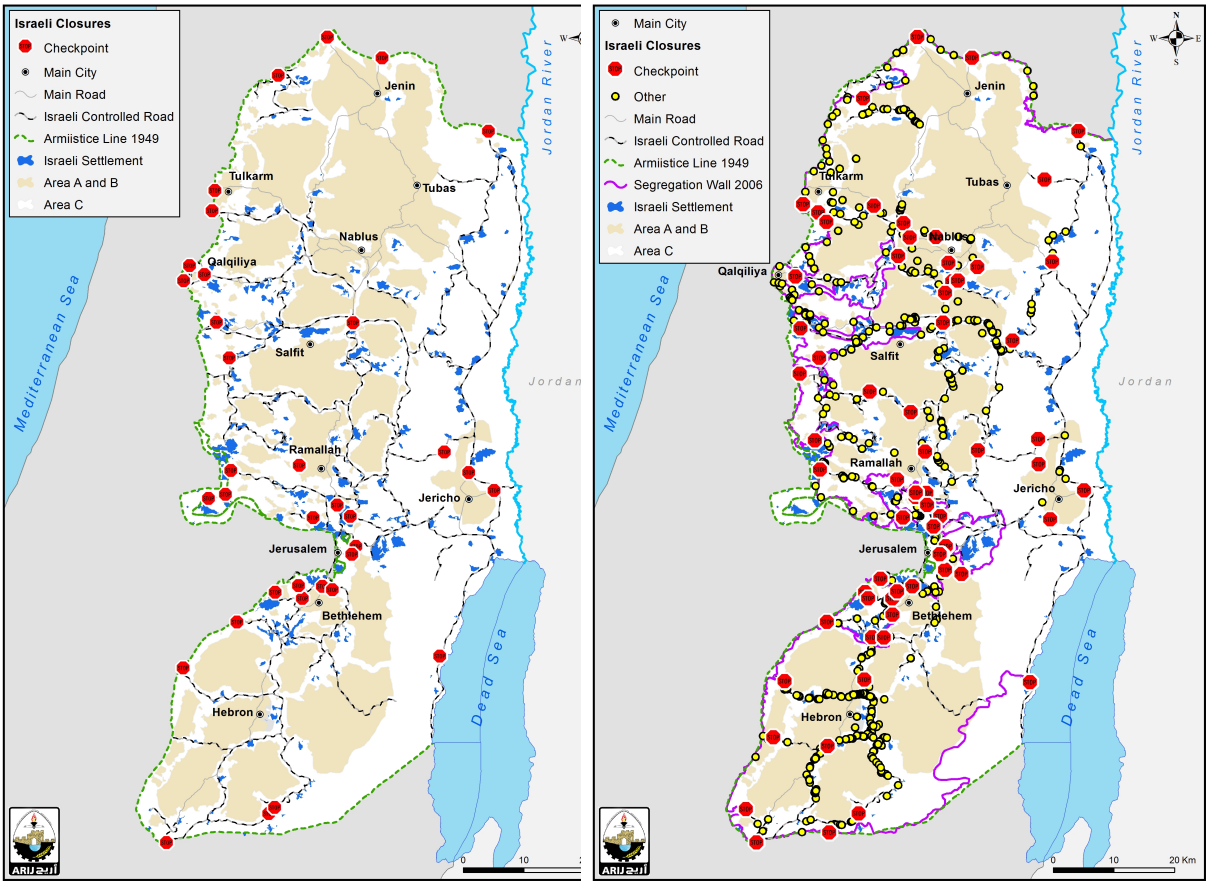
Figure 3.2 provides further evidence on how the number of physical barriers changed over time. Over our sample period, the number of checkpoints rose from 35 in 2000 to 69 in 2006.[92] The number of other barriers fluctuates from year to year but averages 400 over the sample period.

[89] There are 5 additional districts in the Gaza Strip.

[90] As discussed in Section 3, we use localities as our main geographic unit in the analysis below.

[91] Source: Calculated using authors' data sources discussed below.

[92] The decrease in the number of checkpoints in 2004 is due to the numerous internal checkpoints being supplanted by fewer but more permanent checkpoint crossings.(Braverman, 2012).



(a) West Bank, 2000 **(b)** West Bank, 2006

Figure 3.1: These two maps depict the temporal and spatial variation in physical barriers in the West Bank in 2000 (subfigure a) and 2006 (subfigure b). Checkpoints, the main physical barrier this chapter focuses on, are denoted by large red circles. Other barriers (roadblocks, earth mounds, gates, and walls) are depicted by smaller yellow circles. *Source:* Applied Research Institute of Jerusalem (ARIJ).

The number of other physical barriers rose from 120 in 2001 to a peak of almost 600 in 2006. It is estimated that in 2007, more than 40% of the West Bank area was subject to some form of access restrictions for Palestinians (UN, 2007).

Our identification strategy exploits spatial and temporal variation in conflict intensity, measured by conflict infrastructure and conflict-related violence. Therefore, the placement of physical barriers is an important consideration. According to the IDF, checkpoints and other barriers are erected for security reasons (Israel Ministry of Defence, 2007). They tend to be located near Israeli settlements and on Israeli-controlled roads that serve as access points to settlements (United Nations Office for the Coordination of Humanitarian Affairs, 2018). Figure 3.1 shows that localities situated near

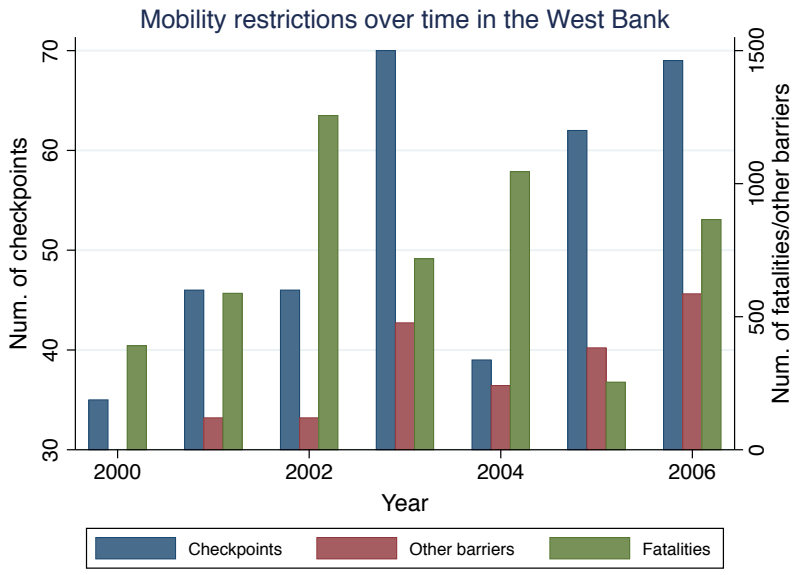


Figure 3.2: This figure shows the number of checkpoints, other physical barriers, and fatalities over time. Other physical barriers include roadblocks, earth mounds, road gates, agricultural and barrier gates, trenches and earth walls. *Source:* Applied Research Institute of Jerusalem (ARIJ) for barriers data; The Israeli Information Center for Human Rights in the Occupied Territories (B'Tselem) for fatalities data.

East Jerusalem and Bethlehem where there is a large Israeli settler population have numerous checkpoints within a 10 km radius. In comparison, in Jericho where few Israeli settlers reside, the average locality has 1.6 checkpoints within a 10 km radius. This suggests that the location of checkpoints and other barriers is primarily determined by the location of the historic Israeli settler population, rather than determined by contemporaneous socio-economic factors which may influence educational performance.[93]

In addition to the physical barriers discussed, a separation wall (the "West Bank wall") was constructed in 2002 mostly along the Green Line, the internationally recognised border between the West Bank and Israel established during the 1949 Jordanian-Israeli armistice. In Figure 3.1 , the Green Line is depicted by the dashed green line while the separation wall is depicted by the purple solid line. This wall is constructed mainly inside West Bank territory (85%) and has a total length of 708 km. The wall restricts the movement of Palestinians from the West Bank into Israel. It also makes it difficult for landowners and their workers to access land in the "seam zone", the

[93]We show this more formally in Table C.9. This historic settlement policy began in 1967 under the leadership of Israeli Deputy Prime Minister Yigal Allon and was largely motivated by national and religious sentiments (Segev, 2006; Zertal and Eldar, 2009).

area between the Green Line and the West Bank wall.

On top of the system of mobility restrictions, another important feature of the IP conflict is the numerous violent actions perpetrated by both Palestinians and the IDF. The killings of civilian and Palestinian militants, Palestinian suicide attacks in Israel, assassinations of Palestinian leaders and demolitions of Palestinian houses by the IDF were common occurrences during the Second Intifada. Between 2000 and 2006, there were 703 Israeli civilian deaths, 316 Israeli military deaths, and over 4,000 Palestinian deaths (B'Tselem, 2007a). The evolution of the number of Palestinian fatalities is depicted by the green bars in Figure 3.2.

2.2 The Education System in the West Bank

The Palestinian Ministry of Education and Higher Education (MoEHE), the body in charge of the education system in the West Bank and Gaza Strip since the Oslo Accords in 1993, places high priority on education. In the West Bank, schooling is compulsory for children aged between 6 and 16 (Grades 1 to 10). Enrolment rates in these grades has been consistently high since the signing of the Accords and the establishment of the national curriculum (UNESCO, 2007). Entry into secondary school is not compulsory and is based on the successful completion and results of Grade 10 education. Between 2000 and 2006, conditional on completing Grade 10, the secondary school enrolment rate is above 80% among students aged 17 to 19, the typical age of secondary school students.[94]

The secondary school education of Palestinian children consists of two years, Grades 11 and 12, during which they specialise in a stream of study within the academic or vocational track.[95] Within the academic track, the two streams include Arts (*Abadi*) and Science (*Elmi*). Within the vocational track, the streams include Commerce (*Tejar'i*), Agriculture (*Zera'i*), and Manufacturing (*Sena'i*).[96] While Science stream students can study all disciplines in university, Arts stream

[94] Author's calculations using the Palestinian Labour Force Survey.

[95] Until 2006, schools in the West Bank followed the Jordanian curriculum; from 2007 onwards, schools followed an independent Palestinian curriculum. Secondary school education in the West Bank is provided by government schools and private schools.

[96] The list of vocational subjects was extended in 2007 after the change in curriculum to include hotel and management (*fonduqi*), economics and finance (*iqtesad*), and Islamic studies (*shar'i*).

students cannot major in Science stream degrees (e.g. engineering or medicine). For this reason, it is generally accepted that academically capable students choose to study the scientific stream rather than any other streams. The majority of secondary school students (roughly 75%) are enrolled in schools provided by the Palestinian Authority.

At the end of Grade 12, students take a final exam called the *Tawjihi General Examination* (TGE). Performance in the TGE is the main outcome of interest in our study and is an important exam for young Palestinians. Achieving a high exam score in the TGE is a necessary prerequisite for admission to university in the West Bank and abroad. Although the quality of the secondary school may influence student performance, entrance to university depends solely on the student's performance in the TGE and is independent of the ranking of the secondary school the student attends. The MoEHE sets a minimum score required to enter university (usually 65%) and specific universities may set further requirements such as good performance in additional tests. A good TGE score also enables students to apply for well-paid public sector jobs. An external commission nominated by the MoEHE grades the exam and exam results are public. The TGE period starts at the end of June and lasts for roughly two and a half weeks. The exam for a given subject (e.g. Maths) takes place on the same day in all schools in the West Bank under the same conditions.

The subject of the exam depends on which stream the student is studying. Correspondingly, the weighting system used to calculate the total exam score differs by stream. For instance, Arts stream students take exams in Islamic education (maximum grade: 100), Arabic (300), English (280), History (120), Mathematics (100), Geography (100), and Scientific culture (100). The total score for Arts students is calculated by summing the grades for the following subjects: Arabic, English, History, and the highest two grades of the remaining subjects. The maximum exam score is 1000. Science stream students take exams in Islamic education (100), Arabic (200), English (200), Mathematics (240), Physics (160), Chemistry (100), and Biology (100). The total score for Science students is calculated by summing the grades for the following subjects: Arabic, English, Mathematics, Physics, and the highest scores of any other two subjects. Although only two of the optional subjects contribute to the total exam score calculation, the student must pass each individual exam in order to pass overall. Appendix C.1 contains more information about the score calculation for each study stream. Across study streams, some subjects overlap. For instance, both

Arts and Science students must take exams in Islamic education and Arabic. The content of the Islamic education for both Arts and Science stream students is the same, as reflected by the same maximum grade achievable in their respective exams (100). However, the content of the Arabic syllabus is more intensive for Arts students, as reflected by the higher maximum score achievable (300 vs. 200).

2.3 How do mobility restrictions affect educational performance?

Educational performance may be hindered by exposure to conflict infrastructure or conflict-related violence for a myriad of reasons. We focus on the following potential mechanisms that are likely to be important in the context of the IP conflict.

Loss of time for studying. The first mechanism through which conflict infrastructure and the resulting mobility restrictions may negatively impact academic performance is by increasing commuting costs and reducing students' available time to study and learn. The existing literature documents the detrimental effects of high transport costs and poor infrastructure (Gibbons and Machin, 2006), checkpoints (Barka, 2012), and blockades (Calì and Miaari, 2018) on labour market participation and wages.

In this context, loss of time due to checkpoints and other physical barriers within the West Bank is likely to affect students' and teachers' ability to move efficiently between their home and school localities. Two channels are likely to be particularly important. First, physical barriers increase the time it takes for students to arrive at school and return home, reducing the time to study at home in the evenings and possibly making it more difficult for students to concentrate while studying at school. The unpredictability of delays could also be costly for students as they must leave home sufficiently early to allow for potentially long delays. As noted previously, the lack of a formal procedure for crossing checkpoints means that delays incurred at a specific checkpoint can be unpredictable. Interviews with checkpoint commuters during the Second Intifada indicate that it was common to wait for over 3 hours at a checkpoint and that the shortest delay was on average 15 minutes (Eklund, 2010; Eklund and Martensson, 2012).

Second, increased commuting costs may increase student absenteeism as students find it physically harder to attend school (Di Maio and Nandi, 2013; Giacaman et al., 2007; World Bank, 2007a). Similarly, conflict-related violence may result in school closures and thus decrease the quantity of school days attended. Andecdotal evidence suggests that the Palestinian Authority closed schools and interrupted normal school activities during the Second Intifada for security reasons: roughly 1,135 school days were lost and 580 schools faced temporary closure during the Second Intifada (Save the Children, 2003).

Psychological well-being. The second way through which conflict may affect educational outcomes is via its negative effects on students' psychological and emotional wellbeing and cognitive capabilities, which in turn reduce academic performance. Numerous studies document the negative relationship between exposure to violence and cognitive performance among youths (Aizer, 2009; Wilson and Rosenthal, 2003). Potential mechanisms for this relationship include exposure to violence causing reduced sleep, increased anxiety, and difficulties in concentrating, resulting in suboptimal academic performance (Gershenson and Tekin, 2018; McGill et al., 2014; Pynoos et al., 1987). In view of this, exposure to fatalities or daily interaction with security personnel manning the checkpoints may impact students' emotional and behavioural mindset and hinder their cognitive ability to perform well at school (Schiff et al., 2007; Shany, 2016).

Furthermore, conflict may also have a negative effect on the psychological well-being of parents. In turn, this may negatively impact children's academic performance as they receive less emotional and academic support at home. Several psychological studies provide evidence that adults who experience conflict-induced stress modify their behaviour towards their own children (Srour, 2005; Thabet et al., 2002).

School learning environment. The third way through which conflict may affect educational outcomes is via its impact on the supply of resources to schools, which may have negative impacts on the learning environment. The physical infrastructure of the schools may be affected by violence. Mobility restrictions on goods may impede on the ability of schools to upgrade or replace damaged infrastructure. Consistent with this hypothesis, Brück et al. (2019) find evidence that the IP conflict

increased classroom crowding in the West Bank, possibly indicating that the conflict restricted the ability of schools to expand and accommodate the growing population of secondary-school age students.

Teachers. Another potential mechanism through which conflict may reduce students' academic performance is via its effect on teachers (Hanushek and Woessmann, 2011). Physical barriers and conflict-related violence may increase teacher absenteeism, reducing instructional intensity and creating disruptions to teaching routines.[97] This may adversely impact student achievement if consistent interaction between teachers and students is important for exam performance (Miller et al., 2008). Teachers' psychological well-being may also be impacted due to exposure to violence or mobility restrictions on the road to school (Nicolai, 2007). In turn, this may impact teaching quality even if teacher absenteeism is not affected.

The aforementioned mechanisms are not mutually exclusive or exhaustive. For example, the time spent waiting at checkpoints may reinforce the psychological stress experienced by students. Additional channels that may reinforce the relationship between exposure to mobility restrictions and academic performance include perceptions of lower returns to investments in education in labour markets disrupted by conflict and higher discount rates linked to higher mortality risk.

[97] See B'Tselem (2011) for a Palestinian teacher's account of crossing a checkpoint.

3 | Data

This analysis uses several unique and novel data sources to identify the causal effects of mobility restrictions and conflict-related violence on the academic achievement of Palestinian children. The main unit of analysis is the exam performance of each student. The two main geographic units at which exposure to conflict is measured at are: (a) a locality or (b) a home locality-school locality pair. Localities represent the smallest spatial unit for which economic data is available in the West Bank. There are 660 localities as defined by the Palestinian Central Bureau of Statistics (PCBS) in the West Bank with an average size of 8.5 km^2.[98]

3.1 Education Data

All data on educational outcomes, school characteristics, and teacher characteristics are provided by the Palestinian Ministry of Education and Higher Education (MoEHE). We have data for 2000–2006, a period of intense conflict covering the Second Intifada.

Student exam scores and characteristics. Our main data source is administrative data on student exam scores for the population of Palestinian students enrolled in their final year of state high schools in the West Bank. The data contains information on (a) whether the student passed the TGE, (b) their overall score (out of 100), and (c) their exam score for each of the main exams they take for their stream. The data also contains information on student demographics such as their age, gender, religion, locality of residence while in secondary school, stream of study, and an identification number for the school that they attend. Henceforth, we refer to this dataset as the *exam scores data*.

Our main sample focuses on state school students between 17 and 19 years old who take the TGE in the period 2000–2006. We restrict our sample to state-school students for two reasons.

[98] The PCBS defines a locality as "A permanently inhabited place, which has an independent municipal administration or a permanently inhabited, separated place not included within the formal boundaries of another locality" (Palestinian Central Bureau of Statistics, 2009). The West Bank is approximately 5,700 km^2; therefore each of the 660 localities is on average 8.5 km^2 in area.

First, the majority of West Bank students (roughly 75% in our sample) attend state schools and therefore they are the most representative student group. Second, many students attending private schools in fact enrol to sit for the TGE via "private-study" centres. Details of these private-study centres are not consistently collected by the MoEHE. We restrict our sample to students between 17 and 19 years old to limit the likelihood of having repeated test takers in our sample (who cannot be identified in the data).

School characteristics. The *school data* provides information on the characteristics of the universe of state high schools in the West Bank. The school data includes information on the locality of the school and information on the learning environment which may affect student performance. These variables include the number of classrooms, the number of teachers, the percentage of dropouts at that school, and the availability of facilities such as a computer lab. We merge the school data with the exam scores data using the identification number of the school which the student attends.

Teacher characteristics. The *teacher data* provides demographic information on the whole population of high school teachers in the West Bank. The data contains information on their locality of residence and the school they work at in the year they appear in the data. We merge the teacher data to the school data using the school identification number, allowing us to observe a teacher's locality of residence and work.[99] The teacher data also contains various characteristics of teachers including their age, gender, tenure, date of appointment, qualifications (BA, MA, PhD), status (permanent or not), and subject specialisation.

3.2 Conflict Data

We use various data sources to measure exposure to (a) conflict infrastructure (e.g. physical barriers) that result in mobility restrictions and (b) conflict-related violence (e.g. fatalities). All data on conflict measures is provided by the Applied Research Institute of Jerusalem (ARIJ). Since

[99] We are unable to match the student to the specific teacher that taught them.

we match conflict data to the education data sources, we focus on a similar time-frame and use conflict data spanning the period 1999 to 2006.[100]

Mobility restrictions. We have Geographic Information System (GIS) data on the location of various physical barriers for each year in the West Bank. The physical barriers contained in this data include checkpoints (permanent and partial), roadblocks, earth mounds, and gates (road, agricultural, and barrier gates). We use this GIS data to construct measures of exposure to mobility restrictions at the individual (home-school locality pair) level and at the school-locality level. These measures are discussed further in Section 4. We match (calendar) year t measures of mobility restrictions to (academic) year t education data.[101]

Delay factors. A contribution of this chapter is the collection of new data on delay factors at checkpoints. In order to calculate the delay factor incurred at each checkpoint, we conducted a survey in collaboration with ARIJ between January and June 2018. The purpose of this survey was to calculate the average delay time incurred when crossing internal and external checkpoints. For internal checkpoints, 70 drivers covered the major transport routes obstructed by these checkpoints. Data on the car location, time, and speed was collected every 5 seconds during the survey, allowing us to construct the time between when the car entered the checkpoint and exited the checkpoint.[102] For external checkpoints that serve as entry points to Israel, a random sample of labourers and students crossing the checkpoint at the end of the work day was interviewed.[103] Information was gathered about the average delay time at a checkpoint. More details about the collection of the delay factor data can be found in Appendix C.2.

A benefit of this delay factor data is that it captures variation in delays between individual internal checkpoints and variation between individual external checkpoints, reflecting the reality of a lack of uniformity in checkpoint-crossing procedures. A shortcoming is that since this data

[100] Since some of our regressions used lagged conflict variables, data from 1999 is included.

[101] The academic year starts in September and ends in July.

[102] For each internal checkpoint, we calculate the average time it takes cars to travel from 750 meters before the checkpoint to 750 meters after the checkpoint (a total distance of 1500 meters). We take the average delay for each month and then the average over the 6 months. Due to the inaccessibility of checkpoints in Hebron by car, internal checkpoints in Hebron were not covered by our sample.

[103] The GPS method is not available for external checkpoints since Palestinian vehicles are prohibited from crossing these checkpoints.

was collected in 2018, the delay factors collected are likely to underestimate the true waiting times incurred during the Second Intifada given that the intensity of conflict has decreased since then. For this reason, rather than making use of the absolute magnitude of delay times measured in 2018, we exploit the relative delay times across individual checkpoints and assume that the variation across checkpoints measured in 2018 is similar to the variation across checkpoints over our sample period. This is a reasonable assumption given that most checkpoints present in 2006 are still present in 2018 and that the main checkpoints during the Second Intifada still remain the major checkpoints today. For checkpoints that are present over our period but not in 2018, we evaluated the delay factor by the closest similar checkpoint covered by the survey. We emphasise that in our empirical analysis, we exploit the relative delay factor lengths across checkpoints to estimate the impact of delay time on educational performance. However, we do not make use of measures of the absolute delay times in 2018. We follow existing studies and assume an average delay of 15 minutes across all checkpoints over the period 2000–2006 (Eklund, 2010; Eklund and Martensson, 2012).

Conflict-related violence. Our main measure of conflict-related violence is fatalities. The *fatalities data* include information on all Palestinian fatalities due to politically-motivated violence in each month since the outbreak of the first Intifada in 1987 for each locality. This data is collected by the Israeli Information Center for Human Rights in the Occupied Territories, B'Tselem, and is considered to be accurate and reliable by both Israelis and Palestinians (Mansour and Rees, 2012). We use locality-year level measures of fatalities in the 12 months prior to the exam (July $t-1$ to June t).

Other policy measures. We have data on other policy variables that may affect exposure to mobility restrictions and fatalities and be correlated with educational outcomes. Given that the placement of barriers is likely to be determined by the presence of Israeli settlements, in our baseline analysis we make use of data on the location and Israeli settler population in the West Bank. In our robustness checks, we also make use of data on the number of jailed Palestinian prisoners in each locality, house demolitions in each locality, and the length of the separation wall between the West Bank and Israel running through the locality. Additional details on these measures can be found in Appendix C.2.

3.3 Other Data Sources

In addition to the main data sources on conflict and educational outcomes, we also use several other data sources. First, we use the 2001 wave of the Survey on the Impact of the Israeli Measures on the Well-being of the Palestinian Children, Women and the Palestinian Household (henceforth, *wellbeing survey*) to examine whether mobility restrictions might affect education attainment via psychological factors. This survey is discussed further in Section 6.

Second, we use the Palestinian Labour Force Survey (PLFS) collected by the PCBS. The PLFS is used to provide support for the identification assumptions of the baseline empirical strategy. We also use the PLFS data to compute economic variables such as the hourly wage and the unemployment rate at the locality level. For some specifications, we merge this locality-level data with the test scores data to control for time-varying locality characteristics.

3.4 Descriptive Statistics

Table 3.1 provides summary statistics for every other year in our main sample. Panel A presents statistics for various conflict variables where the unit of observation is a school. The statistics depict average exposure to conflict across all schools in a particular year. The first row documents the average number of checkpoints within a 10 km radius from the school locality's centroid. In 2000, on average each school has 1 checkpoint within a 10 km radius (with some schools having up to 6 checkpoints within this vicinity). In each subsequent year, the number of checkpoints within the 10 km radius increased. In 2006, the average school faces 2.74 checkpoints within a 10 km radius, with some schools facing up to 12 checkpoints in this area. The second row documents the average number of other barriers (e.g. roadblocks and earth mounds) within a 10 km radius. In 2000, checkpoints were the only type of barriers in the West Bank. Between 2002 to 2006, the average number of other barriers within a 10 km radius of the school locality increased from 12 to 16. The third row describes the evolution of fatalities in the locality where the school is located. The number of fatalities is relatively low in 2000 (0.29), peaks in 2002 (6.67), and then decreases

in the subsequent years (0.94 in 2006).

Panel B presents summary statistics for students in the sample where the unit of observation is a student exam taker. The average age of students is 18 years across all years in the data. Slightly over half the exam takers are female. Almost all students are Muslim (over 99%). In terms of study streams, each year, the majority of students (roughly 66%) sit exams for the Arts stream. Approximately a quarter of students sit exams in the Science stream. The proportion of students in the Science stream appears to be decreasing overtime; there were 28% of Science stream students in 2000 and 24% in 2006. At the same time, the proportion of students studying a vocational stream (commerce, agriculture, or manufacturing) appears to be increasing. 4.4% of students studied a vocational subject in 2000 and this figure rose to almost 8% in 2006.

The percentage of students who pass the TGE fluctuates across the sample period. In 2000, 60.7% of students passed the TGE. This figure rises in 2002 (75.9%) and stays above 70% for the rest of the sample period. Similarly, the overall score varies across the sample period. For example, in 2000, the average grade was 60.3% with a standard deviation of 21.1; in 2006, the average grade was 65.3% with a standard deviation of 20.9.

The final three rows of panel B report the commuting patterns of students. Over 30% of students attend a school that is in a locality that differs from their home locality ("traversing" students). This figure varies across years: in 2000, 36% of students study and live in different localities; in 2006, this figure is 30%. Assuming that students who attend schools in different localities travel by the shortest route between the centroids of the two localities, the average distance travelled is 14–15 km across the sample period. Since our main specification exploits variation in the barriers faced by traversing students on the road to school, it is worth examining whether traversing and non-traversing students differ on observable dimensions. To this end, Table C.2 in Appendix C.3 presents variable means for these two groups of students. The summary statistics suggest that traversing and non-traversing students are similar in terms of baseline characteristics and educational outcomes. For example, in both groups, the average age is 18.2 years and the pass rate is 73%.

Among students who attend school in a different locality, the proportion of students who cross a

Table 3.1: Descriptive statistics for sample

	2000	2002	2004	2006
	(1)	(2)	(3)	(4)
A. Conflict variables				
# CP within 10km	0.927 [0,6]	2.480 [0,10]	2.255 [0,11]	2.744 [0,12]
# Other barriers within 10km	0.000 [0,0]	11.927 [0,47]	9.480 [0,40]	16.002 [0,88]
# Fatalities in school locality	0.285 [0,4]	6.666 [0,64]	2.227 [0,36]	0.937 [0,14]
B. Student variables				
Background and Stream of Study				
Age	18.247	18.250	18.192	18.137
Female	0.511	0.520	0.525	0.527
Muslim	0.991	0.992	0.993	0.994
Art	0.681	0.683	0.673	0.683
Science	0.276	0.252	0.258	0.240
Vocational	0.044	0.065	0.069	0.077
Educational performance				
Pass	0.607 (0.488)	0.759 (0.428)	0.730 (0.444)	0.764 (0.425)
Grade	60.263 (21.153)	65.953 (19.826)	63.069 (21.449)	65.313 (20.856)
Commuting patterns				
School in diff locality	0.359 (0.480)	0.331 (0.471)	0.294 (0.456)	0.300 (0.458)
Distance travelled (KM) if school in diff locality	15.136 (17.207)	15.333 (18.992)	13.681 (17.651)	13.641 (18.153)
Cross checkpoint if school in diff locality	0.042 (0.200)	0.175 (0.380)	0.141 (0.348)	0.136 (0.343)
C. School variables				
Number of classrooms	13.961 [2,29]	14.108 [3,34]	13.847 [2,28]	13.689 [2,30]
Class size	37.810 [20,56]	38.997 [11,79]	39.335 [16,68]	40.173 [20,61]
Student-teacher ratio	23.113 [8,36]	21.674 [8,35]	21.939 [5,117]	20.967 [7,72]
Teacher experience	8.642 [0,20]	8.197 [2,19]	8.528 [1,25]	8.637 [1,23]
Number of schools	330	371	444	511
Number of students	14,362	18,542	22,920	29,437
Number of localities	187	210	240	271

Notes. This table presents summary statistics for the main sample. In panels A and C, the averages are taken across schools. In panel B, the averages are taken across students. Standard deviation in parentheses. Min and max in square brackets.

checkpoint on the shortest route varies across years as the volume and location of these checkpoints evolve across the sample period. In 2000, fewer than 5% of traversing students crossed a checkpoint on the shortest route. This figure increases to 17.5% in 2002 and falls again in subsequent years to 13% in 2006.

Lastly, panel C presents summary statistics for schools in the sample. The unit of observation is a school. The average number of classrooms in each school is 14 and this figure is fairly constant across all years. There is variation across schools with some schools having as few as two classrooms and some schools having as many as 34 classrooms. The average class size is 37–40 students. There is also considerable variation across schools in class size, ranging from 11 to 79 students per class. The average student-to-teacher ratio is 21–23 across all years. This finding is consistent with statistics reported by the World Bank (2004).[104] Lastly, teachers on average have a tenure of 8.5 years (at a given school) and this figure is stable across the sample period.

The final two rows of the table report the number of schools and students per year. Both the number of students and schools increased across the sample period. The upward trend in student numbers reflects two factors: an increase in secondary school enrolment rates among the school age population[105] and population growth in the decade before the Second Intifada (Nicolai, 2007; UNICEF, 2010; World Bank, 2007b).[106]

The increase in the number of schools partly reflects an attempt to accommodate the larger student population and could be one reason for the decreasing proportion of students who commute to a different locality to attend school across the years. One concern is that the number of schools and the location of schools responds endogenously to conflict variables. To test this, Table C.3 in Appendix C.3 estimates locality-level specifications to examine the drivers of the number of schools. In particular, we regress the number of secondary schools in locality l on the following explanatory variables (lagged by one year): secondary-school age population, the presence of checkpoints within 10 km, economic variables (hourly wage, unemployment rate), and the Israeli settler population.

[104] This report states that the average student-teacher ratio has stayed constant over the Second Intifada period except for secondary schools in Gaza, where the ratio has increased somewhat.

[105] Using the PLFS data, we find that secondary school enrolment rates increased from 76% in 2000 to 83% in 2006.

[106] World Bank (2002) reports that the average population growth rate before the Second Intifada was about 4% per year, one of the highest in the world, with about half of the population being under 15 years old.

The estimates indicate that the main determinant of the number of schools is the secondary-school age population: conditional on locality and year fixed effects, an increase in the population aged 17–19 by 1000 is associated with an increase in the number of secondary schools of 0.26. The presence of checkpoints within 10km and other economic variables do not appear to impact the number of schools in the locality.

Our final sample consists of 542 schools, 146,942 student-year observations, and 276 localities. On average, each of these localities has 2 schools that teach TGE students.[107]

[107] Conditional on having a school in the locality, the distribution of schools per locality is as follows. 33.33% of localities have one school, 57.97% have two schools, and the remaining 8.7% have three or more schools.

4 | Empirical Strategy

4.1 Estimating Equations

Our main empirical specification, the *barrier matrix specification*, uses information on student i's home and school locality to construct a measure of the student's exposure to physical barriers on his or her journey to school. For the following discussion, we use the following notation. i denotes the student, $h(i)$ denotes the home locality of student i, $s(i)$ denotes the school that student i attends, $l(s)$ is the locality of school s, and t is the academic year when the exam is taken. For notational convenience, we denote the school, locality of the school, and locality of home by s, l, and h respectively.

When constructing a measure of student i's exposure to physical barriers, we face the challenge that although we observe the student's home and school locality, we do not know the exact route the student took to school and therefore do not know the barriers encountered along this route. To address this problem, we use the following steps. First, we use network analysis (Ballas et al., 2017; Kharel et al., 2018) to calculate the distance between all West Bank localities. When calculating the distance, we assume that (a) the starting and ending points of the route are the centres of the respective localities and (b) the shortest route on the existing road system is taken.[108] Second, using the GIS data on the location of barriers, we calculate the number of barriers crossed on the shortest route for a given year. The second step of this method provides us with a "barrier matrix" that contains information about the number of barriers a commuter faces when travelling on the shortest route from one West Bank locality to another.

We use the barrier matrix to construct an indicator variable $cross_{hlt}$ to measure whether student i living and studying in localities h and l crosses one or more barriers on the road to school in the year that they sit the exam. We use a binary indicator rather than a count variable because the majority of students (58%) who cross at least one checkpoint cross a single checkpoint rather

[108] We use ArcGIS 10.5 to complete this step.

than multiple checkpoints.[109] It is worth nothing that although the shortest distance between two localities h and l (constructed in step 1) does not change over time, whether a student encounters a barrier along the shortest route from h to l (constructed in step 2) changes across years as the conflict evolves.

Using the constructed measure, we estimate the following regression:

$$\begin{aligned} y_{ihlt} = \alpha + \beta cross_{hlt} + D'_{hl}\delta + \rho_h + \sigma_l + \tau_t \\ + F_{l,t-1}\gamma_l + F_{h,t-1}\gamma_h + P'_{l,t-1}\eta_l + P'_{h,t-1}\eta_h + X'_{it}\zeta_1 + W'_{st}\zeta_2 + \epsilon_{ihlt} \end{aligned} \tag{19}$$

In equation (19), the dependent variable is a measure of educational attainment of student i living and studying in localities h and l respectively taking the TGE in academic year t. Using the exam scores data, we use the following measures for y: (a) whether the student passes the final exam (the student achieves over 50% on all subjects examined), (b) the overall score for the TGE, (c) Maths test score, and (d) English test score. Although students taking different study streams sit exams in different subjects, all students take exams in English and most students (apart from students in the Agricultural stream) take exams in Maths. All scores are normalised to be out of 100% and expressed in standard deviations from the mean where the mean and standard deviation are calculated using the pooled data for all years.[110]

The main variable of interest is $cross_{hlt}$, an indicator for whether a student who lives in locality h and studies in locality l in year t crosses one or more checkpoints on the shortest route to school. We also construct analogous $cross$ variables for other barrier types. Although the $cross_{hlt}$ provides a granular measure of conflict exposure, without observing the exact path that the student took to school, it is possible that there is misclassification error in this constructed variable. If the student lives and studies in the same locality, the method we use to construct this variable sets the distance

[109] Among those who cross a checkpoint on the road to school, the distribution of the number of checkpoints crossed is as follows. 58% cross one checkpoint, 29% cross 2 checkpoints, 11% cross 3 checkpoints, and the remaining cross between 4–6 checkpoints.

[110] Since different study streams place different weights on each of these subjects when computing the overall exam score, we normalise the score in each subject by the maximum score achievable in that stream so that the (normalised) maximum score is 100% for each subject in each stream. See Appendix C.1 for information on how the total score is calculated for each discipline.

travelled to zero and therefore the number of barriers faced to zero. Students who study and live in the same locality may still have to cross checkpoints on the road to school. If the student lives and studies in different localities, it is possible that students take a longer route in order to avoid having to cross a checkpoint or a shorter route if the precise location of their home/school is towards the edge of the locality rather than centre. Assuming that the other explanatory variables in equation (19) are measured accurately, this potential measurement error in the binary *cross* variable would bias the β coefficient towards zero (Aigner, 1973).

Since distance travelled is positively correlated with the probability of crossing a barrier, we control flexibly for the distance between locality h and l using a vector of distance bin dummies D'_{hl}. Our distance bins are divided into 0 km (home and school locality are the same), 5 km intervals from 0 km up to 50 km, and then 50 km or above. This gives us 11 distance bin dummies in total, excluding the reference group.[111] Furthermore, since students who live further away from their schools might differ from those who live closer to school, these distance bin dummies help control for student heterogeneity along this dimension.

Home and school locality fixed effects ρ_h and σ_l control for home and school locality time-invariant characteristics that might impact performance. These fixed effects control for student selection into residential areas and student selection into school areas. Academic year fixed effects τ_t help account for differences in the content and difficulty of the final exam in different years. They also control for the influence of time-varying macro-economic conditions (such as labour market conditions and national education policies) on student exam results.

Our main coefficient of interest is β, which can be interpreted as the impact of crossing one or more checkpoint on educational outcomes. Inclusion of home- and school-locality fixed effects means that identification of the β coefficient comes from two main sources: (a) within home-locality variation in exposure to barriers across time and across students who attend schools in different localities and (b) within school-locality variation in exposure to barriers across time and across students who live in different localities.

The variables $F_{l,t-1}$ and $F_{h,t-1}$ consist of the number of Palestinian fatalities that occurred

[111] We also tried classifying distance travelled into deciles. Both approaches produce very similar results.

within 12 months before the exam in the school and home locality respectively. The γ_l and γ_h coefficients are identified using variation in fatality rates within school-locality and within home-locality across time. Given the existing literature's focus on the impact of fatalities on educational outcomes, we also report estimates of γ_l and γ_h in our main tables.

The vectors $P'_{l,t-1}$ and $P'_{h,t-1}$ consist of other policy variables that might influence student achievements and may be correlated with exposure to conflict. In our baseline analysis, we include in this vector measures of the Israeli settlement population size (in 1000s) within a 10 km radius of the school and home locality centres. In our robustness checks, we extend this vector to include the number of jailed Palestinian prisoners (in 1000s) in localities l and h, an indicator for one or more house demolitions in localities l and h, and the length of the separation wall (in km) going through localities l and h.

The vector X_{it} consists of student-level characteristics. We include indicators for gender, religion, (calendar) year of birth, and the students' study stream.[112] Year-of-birth fixed effects help control for any shocks common to all students born in the same year. This is important given the large literature documenting the importance of the in utero environment on later life outcomes (Almond and Currie, 2011). Study stream fixed effects help account for differences across branches in the difficulty and content of the TGE exam. The vector W_{st} denotes time-varying school-level characteristics. We include the following variables: the number of classrooms, the total number of students, the total number of teachers, and a dummy indicating whether the school is mixed-gender or male/female-only. We cluster standard errors at home-school locality pair level. In our robustness checks, we also allow for multi-way clustering by clustering at both the home and school locality levels.

In addition to our main specification, we also estimate a *within-school specification* that exploits variation in student-level outcomes across different academic years among students attending the same school. The within-school specification is a natural complementary estimation strategy because such models exploiting variation in conflict levels in a geographic area close to schools are

[112] That is, we include dummies for Arts, Science, Commerce, Agriculture, and Manufacturing streams.

common in the education literature (Brück et al., 2019; Monteiro and Rocha, 2017).[113] Furthermore, the within-school specification forms the basis of our strategy to examine the mechanisms behind the relationship between mobility restrictions and educational performance and to check for violations of our identification assumptions.

The main difference between the within-school and barrier-matrix specifications is that the former uses school-locality variation in conflict measures to identify the impacts of mobility restrictions on educational performance whereas the latter uses home-school locality pair variation in conflict intensity. Our within-school specification can be expressed as follows:

$$y_{islt} = \alpha + \beta_1 B_{lt} + \beta_2 B_{lt} \cdot T_{it} + \kappa T_{it} + \lambda_s + \tau_t + F_{l,t-1}\gamma + P'_{l,t-1}\eta + X'_{it}\zeta_1 + W'_{st}\zeta_2 + \epsilon_{islt} \quad (20)$$

In equation (20), the main variable of interest is B_{lt}. This is an indicator that equals one if there is one or more checkpoints within 10 km of the center of school locality l in year t and zero otherwise. We focus on checkpoints within this distance because localities are roughly 8.5 km^2 in area and thus students are likely to be exposed to these checkpoints regardless of where within the locality the school is situated. We interact this mobility restriction measure with a dummy T_{it} which denotes whether student i taking exam in academic year t attends a school that is located in a different locality to his/her home locality (i.e. $T_{it} = 1$ if $h \neq l$). Since students who traverse localities in order to attend school are more likely to have direct contact with a physical barrier, mobility restrictions may have a larger impact on these students. School fixed effects λ_s control for unobservable time-invariant differences across school, such as school quality, that may influence student performance. The other variables in equation (20) are as defined before. For this specification, we cluster standard errors at the school locality level.

The inclusion of school fixed effects implies that β and γ in equation (20) are identified using within-school across-time variation in the intensity of mobility restrictions surrounding the school

[113] To examine the causal impact of drug battles on student achievement in Brazil, Monteiro and Rocha (2017) exploit variation over time in the number of drug battles occurring within 250 meters of a given school during the academic year. To examine the impact of conflict-related casualties, Brück et al. (2019) exploit variation in the number of fatalities over time within Palestinian localities where the school is located.

vicinity and conflict violence occurring within the school locality. This strategy removes cross-sectional variation in time-invariant locality socio-economic characteristics that may be correlated with conflict intensity. This helps deal with potential spurious negative relationships between conflict intensity and educational performance if low-quality schools are located in areas with more conflict or if poor-performing students are likely to attend schools in these areas. The estimated β coefficient is likely to be a lower bound on the true effect of mobility restrictions on academic performance because students are not necessarily exposed to barriers surrounding their school.[114]

4.2 Identification Assumptions

This section presents the main assumptions required for the β and γ coefficients in equations (19) and (20) to identify the causal impact of measured exposure to conflict. We discuss several potential threats to these assumptions. In Section 5.2, we present evidence suggesting that our baseline results are not due to violations of these assumptions.

Identification of the β coefficient in the barrier matrix specification (equation 19) comes from within home-locality variation in exposure to barriers across time and across students who attend schools in different localities and from within school-locality variation in exposure to barriers across time and across students who live in different localities. The key identifying assumption is that the probability that student i encounters (at least) one barrier on the road to school is independent of the potential outcomes of that student, conditional on time-invariant home-locality confounders, time-invariant school-locality confounders, distance travelled to school, time-specific confounders, and observable student and school characteristics.

Identification of the β coefficients in the within-school model (equation 20) comes from variation within school-locality across time. The key identification assumption is that exposure to conflict is orthogonal to unobservable time-varying school and school-locality characteristics and unobservable time-varying or time-invariant student characteristics.

In both the barrier matrix and within-school models, the coefficients on the fatality variables

[114] The within-school specification captures intent-to-treat effects of exposure to mobility restrictions.

(γ) are identified using within-locality across-time variation in fatality rates. The key identification assumption is that there are no time-varying locality characteristics that affect exposure to conflict-related violence and influence academic performance of students living or studying in that locality.

The first potential threat to our identification strategy is *time-varying omitted variables* at the locality level. A concern is that there are unobservable changes at the home- or school-locality level (e.g. political attitudes or perceptions about threats to security) that we cannot control for and that may be correlated with both measured exposure to conflict as well as academic performance.

Second, even if there are no time-varying school-locality and home-locality characteristics that affect exposure and academic outcomes contemporaneously, these locality effects may not be *additively separable* as equation (19) stipulates. For example, the home-locality fixed effect may differ across students who live in the same home locality but attend schools in different localities.

Third, student characteristics may change in unobservable ways across academic years within a given locality or school. One concern is *endogenous mobility*: higher-ability students may move to localities with fewer conflict infrastructures in response to increased conflict intensity or move to localities nearer their school in order to avoid high commuting costs. This would bias the coefficient β downwards as the negative relationship between exposure to mobility restrictions and exam performance would be driven by unobservable student ability that is also correlated with exposure to conflict.

Another concern is that there is *sample selection* of students sitting the exam if conflict affects probability of students staying on to take the TGE at the end of the academic year. In particular, if exposure to fatalities or mobility restrictions causes higher-ability students to drop out during the academic year, then a negative relationship between conflict and educational performance could be the consequence of lower-ability students taking the exam in more conflict-intense periods.

Fourth, a potential concern is that there is *reverse causality*. For example, if security forces react to an expectation that exam performance is going to worsen in a given locality in year t by placing more checkpoints in that locality, then a negative relationship between mobility restrictions and academic performance would be found.

5 | Results

5.1 Main Results

Panel A of Table 3.2 presents the estimated impact of crossing at least one checkpoint (β) from equation (19) while panel B presents the estimated impact of conflict-related violence in the school and home locality (γ_l and γ_h respectively). We examine the impact of these aspects of conflict on four educational performance measures: whether the student passes the TGE (column 1), the overall TGE score (column 2), and the scores for Maths and English (columns 3 and 4). The scores are normalised to have a maximum value of 100% and are measured in standard deviations from the overall sample mean.

Column (1) indicates that crossing at least one checkpoint on the road to school reduces the probability of passing the exam by 3.05 pp, relative to a baseline of 73% (a 4.2% reduction). Panel B of this column shows that an additional fatality in the school locality in the twelve months before the exam is taken reduces the probability of passing by 0.07 pp. Over this period, the average number of fatalities in the school locality was 2.6 fatalities per year. Therefore, this estimate suggests that on average exposure to fatalities reduced the probability of passing by 0.18 pp (0.07×2.6). In contrast, the fatality rate in the student's home locality does not appear to have an adverse impact on academic performance. One possible reason for this is that clashes with the IDF are more likely further away from home. These findings firstly indicate that multiple aspects of conflict – mobility restrictions and violence – have detrimental impacts on educational performance. Furthermore, comparing the magnitudes of the estimated β and γ_l coefficients suggests that the impact of crossing one or more checkpoints is larger in magnitude than the effect of one additional fatality in the school locality.

Column (2) shows that crossing at least one checkpoint on the road to school reduces the overall TGE score by 0.071 standard deviations. This is equivalent to a 1.48 pp (0.071×20.79) decrease in the overall score, relative to a mean of score of 64.19% (a 2.31% decrease). Panel B shows that an additional fatality in the school locality reduces overall TGE scores by 0.0012

Table 3.2: Impact of crossing one or more checkpoints on educational performance

	Pass	Overall score	Maths	English
	(1)	(2)	(3)	(4)
Mean	0.73	64.19	63.37	57.79
SD	0.44	20.79	25.88	22.55
A. Mobility restrictions				
Cross checkpoint	-0.0305***	-0.0714***	-0.0727***	-0.0529***
	(0.0091)	(0.0188)	(0.0199)	(0.0176)
B. Other conflict variables				
Fatalities in school locality	-0.0007***	-0.0012**	-0.0017*	-0.0010**
	(0.0002)	(0.0005)	(0.0009)	(0.0004)
Fatalities in home locality	0.0001	0.0005	0.0010	0.0005
	(0.0002)	(0.0004)	(0.0007)	(0.0004)
# Student-Years	146,942	146,942	146,268	146,942
Adjusted R^2	0.107	0.256	0.165	0.327

Notes. This table presents estimated coefficients from equation (19). Scores expressed in standard deviations. Mean and standard deviation of untransformed dependent variables (e.g. exam scores) presented in the first two rows. All regressions include the following fixed effects: school locality fixed effects, home locality fixed effects, distance bin fixed effects, and academic year fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of home and school locality, student indicators (gender, religion, year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, gender of school). Standard errors in parentheses, clustered at the home-school locality pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

standard deviations whereas fatalities occurring in the student home locality have no impact on school performance. Comparing the estimates in panel A with those in panel B shows that the estimated impact of crossing at least one checkpoint is larger in magnitude than the impact of an additional fatality in the school locality.

Panel A of columns (3) and (4) shows that although crossing at least one checkpoint on the road to school adversely impacts performance in both Maths and English exams, Maths exam scores are particularly negatively impacted. Crossing at least one checkpoint on the road to school reduces Maths exam scores by 0.073 standard deviations and English exam scores by 0.053 standard deviations. Panel B shows that an additional fatality in the school locality also has a detrimental impact on Maths scores, reducing it by 0.002 standard deviations. English scores are also negatively affected by fatalities in the school locality, falling by 0.001 standard deviations with each additional fatality. Fatalities occurring in the home locality have no impact on test scores for Maths or English

scores.

These results provide some evidence that Maths scores are particularly sensitive to exposure to conflict. One possible reason for these heterogeneous impacts by exam subjects is that the inputs required to achieve good scores in these subjects differ. For example, if achieving good scores in some subject requires constant interaction between teachers and students, then if mobility restrictions impact student (or teacher) absenteeism, these subjects will be more adversely affected compared to subjects that require less interaction between teachers and students. The existing literature finds evidence that Maths scores are particularly adversely affected by conflict. For example, Monteiro and Rocha (2017) find that Maths test scores are negatively affected by conflict whereas English scores are not affected.

The magnitude of the coefficients on the cross checkpoint variable suggests that mobility restrictions are quantitatively important for educational performance. Having to cross a checkpoint reduces the probability of passing the TGE by 3.05 pp and overall test scores by 0.071 standard deviations. These effects are comparable to the impacts of conflict-related violence on educational performance found in the existing literature. For example, Monteiro and Rocha (2017) find that exposure to violence triggered by drug gangs in Brazil reduces Maths test scores by 0.054 standard deviations. Shany (2016) finds that an increase of one Israeli fatality in the student's area within five days before the exam is associated with a 0.006 standard deviation decline in exam scores.

Furthermore, the estimated impacts of mobility restrictions are also comparable in magnitude to the impacts of other inputs into the education production function examined in the existing literature. For example, the effect of exposure to barriers is equivalent to one-half to two-thirds of the drop in test scores associated with a one standard deviation decrease in teacher quality in the US (Rivkin et al., 2005; Rockoff, 2004).[115] The effect of exposure to barriers is almost equivalent to a two standard deviation reduction in school management quality (as measured by superintendent value added) in Israel (Lavy and Boiko, 2017).[116] The impact of removing checkpoints in this

[115] Rivkin et al. (2005) find that a one standard deviation increase in average teacher quality for a grade raises average student achievement in the grade by at least 0.11 standard deviations of the total test score distribution in Maths and 0.095 standard deviations in Reading. Rockoff (2004) finds that a one standard deviation increase in teacher quality raises both reading and math test scores by about 0.1 standard deviations on the national scale.

[116] The authors find that a one standard deviation improvement in superintendent value added increases test scores by about 0.04 standard deviations in the test score distribution.

context is comparable to the impact of mothers in India attending a training program that teaches literacy, numeracy, and engagement in their children's education at home on their children's Maths test scores (Banerji et al., 2017).[117]

Table 3.3 investigates whether the impact of mobility restrictions on educational performance differs across genders by adding interactions between the cross checkpoint and fatality variables and a *Male* dummy, allowing the coefficients on these conflict measures to take different values for male and female students. There are several differences between male and female students that should be accounted for when comparing the impacts of mobility restrictions across genders. First, students in the West Bank often attend single-sex schools and therefore the schools attended by male and female students differ.[118] Second, males and females sort into different study streams. Among females, 74.1% and 22.2% study the Arts and Science streams respectively; among males, these figures are 62.7% and 27.4% respectively. On average, the pass rate for the Arts stream is lower than that for the Science stream (66.1% vs 87.9%). The stream fixed effects in our regressions help control for this sorting into study streams. Despite these differences, the commuting patterns of female and male students are similar. 32.90% and 30.90% of male and female students respectively attend a school outside their home locality. The proportion of traversing students who cross checkpoints are also similar for males and females (10.67% and 12.55% respectively).

The first row of Table 3.3 is suggestive of heterogeneity across genders, showing that female students perform better on average than male students. The pass rate is 79% for females and just above 67% for males; the average overall scores are 67.7% for females and 60.3% for males. The following rows of the table indicate that the impacts of conflict differ by gender. First, the impact of crossing at least one checkpoint on educational performance is concentrated among male students. For example, column (1) shows that for female students, the estimated impact of crossing at least one checkpoint on the probability of passing is not statistically different from zero (-0.06 pp). For male students, the estimated impact is the sum of the first two rows in the table: a 6.17 pp ($-0.06 - 6.11$) decrease in the probability of passing the exam. Columns (2) to (4) show that although the estimated impacts of crossing at least one checkpoint on the overall TGE score,

[117] In this experiment mothers receiving this training increased their child's Maths score by 0.056 standard deviations.

[118] Roughly 9% of our sample attends a mixed-gender school.

Maths score, and English scores of female students are mostly negative, they are not statistically significant. In contrast, for male students, the impacts are economically and statistically significant. For example, among male students, the estimated impact of crossing at least one checkpoint is a -0.142 standard deviation decrease in the overall TGE score.

Table 3.3: Impact of crossing one or more checkpoints on educational performance by gender

	Pass	**Overall score**	**Maths**	**English**
	(1)	(2)	(3)	(4)
Female: Mean (SD)	0.79 (0.41)	67.74 (19.35)	67.60 (24.19)	61.08 (22.14)
Male: Mean (SD)	0.67 (0.47)	60.32 (21.60)	58.74 (26.86)	54.21 (22.45)
A. Mobility restrictions				
Cross checkpoint	-0.0006	-0.0005	-0.0197	0.0057
	(0.0113)	(0.0247)	(0.0246)	(0.0235)
Cross checkpoint × Male	-0.0611***	-0.1423***	-0.1055***	-0.1158***
	(0.0153)	(0.0372)	(0.0370)	(0.0342)
B. Other conflict variables				
Fatalities in school locality	0.0003	0.0015*	0.0007	0.0015
	(0.0003)	(0.0009)	(0.0008)	(0.0010)
Fatalities in school locality × Male	-0.0019***	-0.0057***	-0.0051***	-0.0053***
	(0.0006)	(0.0016)	(0.0014)	(0.0016)
Fatalities in home locality	0.0000	0.0003	0.0008	0.0003
	(0.0002)	(0.0005)	(0.0006)	(0.0006)
Fatalities in home locality × Male	-0.0002	-0.0005	-0.0002	-0.0003
	(0.0004)	(0.0011)	(0.0012)	(0.0011)
# Student-Years	146,942	146,942	146,268	146,942
Adjusted R^2	0.107	0.255	0.165	0.326

Notes. Scores expressed in standard deviations. All regressions include the following fixed effects: school locality fixed effects, home locality fixed effects, distance bin fixed effects, and academic year fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of home and school locality, student indicators (religion, year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, gender of school). Standard errors in parentheses, clustered at the home-school locality pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Second, panel B of this table shows that the negative impact of fatalities on educational performance is also concentrated among male students. For example, column (1) shows that an additional fatality in the school locality does not have a statistically significant impact on the probability that a typical female student passes the TGE. However, among male students, an additional fatality in the school locality reduces the probability of passing by 0.16 pp ($0.03 - 0.19$). Columns (2) to (4) confirm that the adverse impact of fatalities on exam scores is concentrated

among male students, with each additional fatality in the school locality reducing overall, Maths, and English scores by 0.005–0.006 standard deviations. One explanation for the larger impact of fatalities on males than females is that males are more likely to clash with the IDF and the fatalities variables capture elements of this exposure to violence.

Table 3.4 investigates the impact of crossing other barrier types (e.g. roadblock, earthmounds, gates) on the shortest route to school by replacing the cross checkpoint indicator in equation (19) with a cross other barrier indicator.[119] Panel A shows that these other barrier types do not appear to have a detrimental impact on educational performance in the way that checkpoints do. For example, the estimated impact of encountering some other barrier type on the shortest route to school is a 0.36 pp reduction in the probability of passing the exam and a 0.009 standard deviation reduction in the overall exam score, both statistically insignificant at the 5% significance level. When we allow the coefficients on the cross other barrier dummy to take different values for male and female students, we find that there are no significant effects for either male or female students. The different coefficients on the cross checkpoint indicator (Table 3.2) and the cross other barriers indicator (Table 3.4) suggests that variation over time in exposure to checkpoints and to other barriers are only weakly correlated: the other barriers indicator does not proxy for the checkpoints indicator when the latter is omitted.

Panel B shows that the estimated impact of fatalities occurring before the exam is similar in magnitude to the impact discussed earlier when the main barrier measure used was checkpoints. In particular, fatalities occurring a year prior to the exam have a negative impact on the probability of passing (-0.07 pp in Tables 3.2 and 3.4) and on students' overall grade (-0.001 standard deviations in both tables). The similarity of the estimated impacts of fatalities in Table 3.2 (when mobility restrictions is measured using checkpoints) and Table 3.4 (when mobility restrictions is measured using other barrier types) indicates that variation over time in the checkpoints and fatalities measures of conflict intensity are only weakly correlated with each other: fatalities do not proxy well for checkpoints when checkpoints are omitted.

There are several reasons why other barriers do not have a negative impact on educational

[119]The cross other barrier indicator takes on a value of one if the student passes one or more non-checkpoint barriers when travelling between localities h and l.

performance in the way that checkpoints do. First, these additional barriers are unlikely to add to the travel time of the student if he or she is commuting by foot since pedestrians can cross over earth mounds and walk by roadblocks. These other barriers are likely to be a more serious impediment to vehicles since vehicles must U-turn and take another route when they encounter such barriers. Second, these other types of barriers are less likely to be manned by security forces. Compared to checkpoints, students are therefore less likely to encounter personnel when crossing these other barriers. This might cushion the potential impact such barriers have on the psychological or emotional wellbeing of students.

Table 3.4: Impact of crossing one or more other barrier types on educational performance

	Pass	**Overall score**	**Maths**	**English**
	(1)	(2)	(3)	(4)
Mean	0.73	64.19	63.37	57.79
SD	0.44	20.79	25.88	22.55
A. Mobility restrictions				
Cross other barriers	-0.0036	-0.0091	-0.0044	-0.0144
	(0.0060)	(0.0131)	(0.0148)	(0.0120)
B. Other conflict variables				
Fatalities in school locality	-0.0007***	-0.0012**	-0.0017*	-0.0011**
	(0.0002)	(0.0005)	(0.0009)	(0.0004)
Fatalities in home locality	0.0001	0.0005	0.0010	0.0005
	(0.0002)	(0.0004)	(0.0007)	(0.0004)
# Student-Years	146,942	146,942	146,268	146,942
Adjusted R^2	0.107	0.256	0.165	0.327

Notes. This table presents estimated coefficients from equation (19) where *cross* is defined as crossing one or more other barrier types on the road to school. Scores expressed in standard deviations. Mean and standard deviation of untransformed dependent variables (e.g. exam scores) presented in the first two rows. All regressions include the following fixed effects: school locality fixed effects, home locality fixed effects, academic year fixed effects, calender year of birth fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of home and school locality, student controls (gender, religion, study branch) and school controls (number of classrooms, total number of students, and total number of teachers). Standard errors in parentheses, clustered at the home-school locality pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5 presents estimates using our supplementary *within-school specification*. Panel A of this table examines the impact of one or more checkpoints within 10 km of the school locality centre on traversing and non-traversing students. Column (1) shows that the presence of at least one checkpoint within 10 km of the school locality centre decreases the probability that traversing students pass the exam by 1.93 pp but does not have a statistically significant impact

on the performance of non-traversing students. Column (2) indicates that the TGE scores of both traversing and non-traversing students are adversely affected by mobility restrictions. In particular, their test scores fall by 0.035–0.042 standard deviations. Columns (3) and (4) examine outcomes for specific subjects and indicate that both traversing and non-traversing students are negatively affected by the presence of checkpoints in the school vicinity. This effect is larger in magnitude for traversing students than non-traversing students but the estimated impacts are not statistically different. For example, the presence of one or more checkpoints reduces Maths scores by 0.060 pp and 0.072 standard deviations for traversing and non-traversing students respectively. The analogous figures for English scores are 0.029 and 0.050 standard deviations respectively.

Table 3.5: Impact of checkpoints (CPs) near school on educational performance

	Pass	**Overall score**	**Maths**	**English**
	(1)	(2)	(3)	(4)
A. Mobility restrictions				
(i) $\geq$ 1 CP within 10 km	-0.0085	-0.0351**	-0.0604***	-0.0285*
	(0.0055)	(0.0136)	(0.0138)	(0.0163)
(ii) $\geq$ 1 CP within 10 km	-0.0108	-0.0061	-0.0117	-0.0220
$\times$ commute	(0.0079)	(0.0178)	(0.0167)	(0.0159)
Row (i) + (ii)	-0.0193**	-0.0412**	-0.0720***	-0.0504***
	(0.0082)	(0.0181)	(0.0197)	(0.0156)
B. Other conflict variables				
Fatalities in school locality	-0.0007***	-0.0013***	-0.0015*	-0.0011*
	(0.0002)	(0.0004)	(0.0008)	(0.0006)
# Student-Years	146,942	146,942	146,268	146,942
# Schools	543	543	542	543
Adjusted R^2	0.120	0.276	0.187	0.344

Notes. This table presents estimated coefficients from equation (20). Scores expressed in standard deviations. All regressions include the following fixed effects: school fixed effects and academic year fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of school locality, student indicators (gender, religion, year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, gender of school). Standard errors in parentheses, clustered at the school locality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

There are two patterns in panel A that are worth highlighting. First, columns (2) to (4) indicate that the test scores of both traversing and non-traversing students are adversely impacted by mobility restrictions. Thus, although the barrier-matrix specification (equation 19) uses variation among traversing students to identify the effects of mobility restrictions on educational performance,

the within-school estimates indicate that the effects of physical barriers also affects students who live and study in the same locality.

Second, although the estimated impact of checkpoints on traversing students from the within-school specification are quantitatively similar to the estimated impact of checkpoints from the barrier-matrix specification (Table 3.2), the former estimates are slightly smaller for several outcomes (probability of passing and the overall score). For example, the estimated impact of crossing at least one checkpoint on the probability of passing is -3.05 pp in Table 3.2 while the estimated impact of at least one checkpoint within 10 km of the school locality centre is -1.93 pp in Table 3.5. For subject-specific scores, the estimated impact of mobility restrictions for traversing students using the within-school specification are very similar to the estimated impact using the barrier-matrix specification. For example, the estimated impact of crossing at least one checkpoint on Maths scores is -0.073 standard deviations using the barrier-matrix specification and the estimated of a checkpoint within 10 km of the school locality on traversing students is -0.072 standard deviations using the within-school specification.

Panel B of Table 3.5 examines the impact of conflict-related violence on educational performance. Estimates of the γ coefficient presented in this panel depict the detrimental impact of exposure to conflict-related violence on academic performance, consistent with the estimated coefficients of the barrier matrix specification. Columns (1) and (2) indicate that an additional fatality occurring 12 months before the exam reduces a typical student's probability of passing the exam by 0.07 pp and their overall score by 0.001 standard deviations. Columns (3) and (4) show that fatalities have a negative impact on performance in Maths and English: each additional fatality reduces scores by 0.002 and 0.001 standard deviations respectively. Compared to the estimates in Table 3.2 using the barrier-matrix specification, the estimated impacts of fatalities (γ) using the within-school specification are similar in magnitude.

Overall, our results provide evidence that conflict infrastructure, particularly in the form of checkpoints, can be detrimental to academic performance. The impact of mobility restrictions operate through a distinct channel to other conflict measures such as fatalities, which have been the focus of the previous literature. The estimated impacts of crossing at least one checkpoint

are statistically significant and economically important. In comparison to findings in the existing literature on the determinants of educational performance, our estimates indicate that the impact of encountering a checkpoint on the way to school is at least as important as being exposed to an additional fatality in the school locality and comparable to the importance of other determinants of educational performance such as teacher and school management quality.

5.2 Robustness Checks and Threats to Identification

In Section 4.2, we discussed several identification assumptions that are important for our empirical specification. This section examines whether violations of these assumptions might be driving our main results by presenting several robustness checks.

Time varying omitted variables. An important condition for the OLS estimates of the β, γ_l and γ_h coefficients in equation (19) to identify the causal impacts of exposure to conflict is that the school or home locality variables that affect exam performance and are correlated with exposure to conflict are time-invariant.

To check whether omitted time-varying locality characteristics are driving our main results, we conduct two tests. Our first test augments the list of observable school and home locality variables that might be correlated with the intensity of conflict over time in the vectors P_l' and P_h'. In our baseline analysis, we control for the Israeli settler population size (in 1000s) within a 10 km distance from the home or school locality centre. In this test, we also include the following controls at the home and school locality level: the number of political prisoners (in 1000s), an indicator for one or more house demolitions, and the length of the separation wall (in km) going through that locality. The results, presented in Table C.4, firstly show that the coefficient on the *cross* variable is similar but smaller in absolute value than the estimates when these additional controls are not included. For example, the estimated impact of crossing at least one checkpoint on the probability of passing the TGE is -2.63 pp when these additional controls are included at both the home and school locality level compared to -3.05 pp when these controls are omitted. Second, conditioning on all other variables included in the baseline regressions, most of these additional policy variables

do not have a statistically significant impact on educational performance. There is some evidence that the presence of prisoners in the school locality and the separation wall in the home locality reduces overall and Maths exam scores. Overall, our baseline results do not appear to be driven by these additional observable time-varying characteristics.

Although the previous test controls for an extensive list of policy variables that might be correlated with exposure to conflict and educational performance, unobservable time-varying locality characteristics may still be driving our main results. Our second test allows the home and school locality fixed effects of equation (19) to take on different values for each academic year.[120] The advantage of this strategy is that it helps control non-parametrically for unobservable time-varying characteristics at both the school locality and home locality. Such changes may affect the placement of barriers around those localities and therefore affect the probability that a student encounters a barrier when travelling from h to l. Examples of such time-varying variables include the political atmosphere and the perceived security risks in these localities. The disadvantage of this strategy is that we cannot estimate the impact of other time-varying observable locality variables such as the number of fatalities, since these variables are now collinear with the time-varying locality effects. Therefore, in specifications that include academic-year specific locality fixed effects, we omit all other time-varying locality characteristics from the regression (e.g. $F_{l,t-1}$ and $F_{h,t-1}$) and examine whether the coefficient on our main variable of interest ($cross$) is affected. When these year-specific home and school locality fixed effects are included, the coefficents are identified by variation within home (school) locality across students attending schools (living) in different localities. Results for these models are presented in the odd columns of Table C.5. In the even columns of this table, we include school fixed effects in addition to time-varying home and locality fixed effects.

The estimates in the table are very similar to those in the baseline specification. For example, column (1) shows that the estimated impact of crossing at least one checkpoint on the probability of passing when time-varying home and school locality fixed effects are included is -2.79 pp (vs. -3.05 pp in the baseline specification). When school fixed effects are included in addition to time-varying home and school locality fixed effects, the estimates are slightly smaller in magnitude but not statistically different. For example, column (2) shows that the estimated impact of crossing

[120] In particular, we interact each locality fixed effect with a set of academic year fixed effects.

at least one checkpoint when both school fixed effects and academic year-specific locality effects are included is -2.49 pp. The remaining columns of Table C.5 perform the same exercise using test scores as the outcome variable. The estimates are very similar to those in the baseline specification. Overall, these results suggest that time-varying unobservables at the locality level are unlikely to be a major confounding factor.

Additively separable home and locality fixed effects. The baseline specification uses additive home and school locality fixed effects, potentially restricting the way in which unobservable time-invariant home and school characteristics can affect educational outcomes. To test whether this is a restrictive assumption, the odd columns of Table C.6 present estimates from versions of equation (19) that include home-school locality pair fixed effects rather than additive home and school locality fixed effects. These interactive fixed effects allow the effect of the school locality to differ across students who live in different localities. For example, under this specification, the equation for a student living in h and studying in l includes the fixed effect σ_{hl} while the equation for a student living in h' and studying in l includes the fixed effect $\sigma_{h'l}$.[121] When home-school locality pair fixed effects are included, this absorbs the distance bin dummies and identification comes from comparing changes in outcomes of students commuting on routes that saw changes in the number of checkpoints over time with changes in outcomes of students commuting on routes that did not see a change in the number of checkpoints. In the even columns of the table, the estimates are from models that include both home-school locality pair fixed effects and school fixed effects.

When these alternative fixed effects are used, the estimated coefficients are almost identical to those in the baseline specification with additive home and school locality fixed effects. For example, when home-school locality pair fixed effects are included, the estimated coefficient on the *cross* variable is -3.08 pp (vs. -3.05 when additive fixed effects are used) for the probability of passing and -0.0764 standard deviations (vs. -0.0723 when additive fixed effects are used) for the overall test score. When school fixed effects are included in addition to home-school locality pair fixed effects, the estimates are slightly smaller in magnitude but still remain very stable. For example, for the outcome of passing the exam, the estimated coefficient is 3.08 pp with home-school locality

[121] In comparison, the additive fixed effects model would include $\rho_h + \sigma_l$ for the former student and $\rho_{h'} + \sigma_l$ for the latter student.

pair fixed effects and 3.05 with school fixed effects in addition to home-school locality pair fixed effects.

Multi-way clustering. The standard errors of our baseline estimates are clustered at the home-school locality pair level. Although this allows for errors to be correlated by home-school locality pairs, it does not allow for possible two-way error correlation across both the home and school locality in the pair. Table C.7 estimates equation (19) and clusters standard errors at both the home locality and school locality level to allow for this two-way correlation. The standard errors presented in this table are very similar to those in the baseline specification and the statistical significance of our estimated impacts are unchanged. For example, the standard error in panel A of column (1) is 0.0090 with multi-way clustering and 0.0091 with clustering at the home-school locality pair.

Sample selection of students sitting exams. Conflict may affect the sample of students who who enrol in secondary school or affect the subset of students who actually sit the exam. To examine whether sample selection due to conflict intensity is likely, we follow Brück et al. (2019) and perform two tests. First, we use the PLFS to regress school attendance among 17 to 19 year olds in locality l on conflict measures B_{lt} and $F_{l,t-1}$ of equation (20) in that locality, conditioning on locality fixed effects. These results are presented in column (1) of Table C.8 in Appendix C.3. The results indicate that a high proportion of students attend school (92.7%) and provide little evidence that increased exposure to checkpoints within and around the locality affects levels of school attendance. For example, the estimated impact of the introduction of at least one checkpoint within 10 km of the school locality centroid is negative (-0.011) but statistically insignificant (p-value=0.378). Additional fatalities in the locality and increases in the Israeli settler population also do not affect school attendance rates.

Second, we use the schools data to examine determinants of school-level drop-out rates. We regress the school drop-out rate, reported in the schools data, on conflict measures, locality controls, school, and academic year fixed effects. These results are shown in column (2) of Table C.8. The first row above the column numbers shows that the school drop-out rate is low (3.1%). Estimates

in panels A and B also indicate that increases in checkpoints and fatalities in the school locality do not appear to affect the school dropout rate.

Endogenous mobility. Another concern is that higher-ability students move to localities with fewer conflict infrastructures in response to increased conflict intensity or move to localities nearer their school in order to avoid high commuting costs. While it is difficult to check the mobility patterns of students as a function of their latent ability, anecdotal evidence suggests that internal and external mobility was severely restricted in the West Bank over our sample period (Central Intelligence Agency, 2008; World Bank, 2007a). To examine this further, we first use the 1997 and 2007 Palestinian census to document patterns of mobility. Across these two census years, roughly 25% of individuals report ever having lived in a locality that is different to their current residential locality. Of those who have moved, the majority report having moved for marriage (18.56% of ever-movers, most of whom are females) or to accompany a family member (36.69%). Fewer than 2.3% of ever-movers report having moved for study-related reasons. Furthermore, among this 2.3% who moved for study-related reasons, the median (mean) age is 22 (23), suggesting that they are likely to be university rather than high school students.

Second, using the rotating quarterly panel nature of the PLFS, we investigate whether year-on-year migration across localities can be predicted by exposure to conflict. For individual i living in locality l and interviewed in year t, we examine whether exposure to conflict in year $t-1$ leads the individual to move localities in the next survey wave in year $t+1$.[122] In these models, we include home locality fixed effects to control for selection into residential location and individual characteristics such as age, gender, and educational attainment. These results, presented in column (3) of Table C.8, firstly indicate that year-on-year mobility is low. Only 0.2% of the sample are recorded as having moved locality over the time frame they are surveyed. Second, there is no evidence that increased exposure to mobility restrictions or fatalities leads to a higher probability of moving localities between periods t and $t+1$.

[122] In regressions where we use conflict in year t as our measure of conflict exposure, the estimated coefficients are quantitatively and statistically very similar.

Reverse causality. To check whether reverse causality could be a threat to our identification strategy, we regress the number of checkpoints within a 10 km radius of the locality centroid on lagged or contemporaneous measures of either the average pass rate and the overall score (measured in standard deviations). We control for the number of fatalities in the locality, the population size (in 1000s) of Israeli settlers within a 10 km radius of the school locality, and locality and year fixed effects.

The estimates, presented in panel A of Table C.9, provide no evidence that the number of checkpoints is determined by either contemporaneous (columns 1 and 2) or lagged (columns 3 and 4) educational outcomes. For example, the estimated coefficients on the contemporaneous and lagged pass rates are 0.416 and 0.452 respectively, with corresponding p-values of 0.70 and 0.75. The coefficients on contemporaneous and lagged overall TGE scores are also statistically insignificant, suggesting that reverse causality is unlikely to be a serious concern. Panel B shows that an important factor for determining the number of checkpoints is the population size of Israeli settlers within 10 km of the locality centre, as discussed in Section 2 and depicted in Figure 3.1. Since we include locality fixed effects, the estimates suggest that increases in the number of Israeli settlers are followed by increases in the number of checkpoints. It is worth emphasizing that all of our regressions control for the settler population size via the vector P.

6 | Mechanisms

The negative coefficients on the *cross* variable in the barrier matrix specification suggests that mechanisms that affect the individual student may be important for explaining the negative relationship between mobility restrictions and educational performance. Furthermore, the negative coefficients on the number of checkpoints within 10 km of the school locality in the within-school specification suggests that factors that impact the school surroundings could be important. This section explores four mechanisms that are consistent with these findings. The first two impact students directly (time loss and psychological effects) and the remaining two impact teachers and the physical resources at the school.

6.1 Impact via Time Loss

Time loss due to commuting is the most direct channel behind the relationship between mobility restrictions and academic achievement. Students who cross checkpoints on the road to school lose time in the morning and early evening, reducing their available time for home study. The unpredictability of delays could also be costly as students either have to leave home early enough to allow for long delays or risk missing part of the school day. To investigate the role of delays, we modify our barrier matrix specification and estimate the following regression:

$$\begin{aligned} y_{ihlt} = \alpha + \beta_1 Delay_{hlt} + \beta_2 \#Checkpoints_{hlt} + D'_{hl}\delta + \rho_h + \sigma_l + \tau_t \\ + F_{l,t-1}\gamma_l + F_{h,t-1}\gamma_h + P_{l,t-1}\eta_l + P_{h,t-1}\eta_h + X'_{it}\zeta_1 + W'_{st}\zeta_2 + \epsilon_{ihlt} \end{aligned} \tag{21}$$

In equation (21), $Delay_{hlt}$ captures the delay factor incurred due to checkpoints encountered along the shortest route from h to l in year t. Since those who cross more checkpoints are likely to face a longer total delay time, we also control for the total number of checkpoints encountered on the shortest route from h to l in year t through the variable $\#Checkpoints_{hlt}$. Since there is variation across checkpoints in the delay time incurred, the effects of delay time and an additional

checkpoint encountered can be separately identified. The remaining variables in equation (21) are as described before.

Equation (21) modifies baseline equation (19) in two ways. First, it replaces the binary *cross* variable by the count $\#Checkpoints$ variable since total delay time is influenced by the total number of checkpoints crossed, rather than a binary indicator of having to cross a checkpoint. Second, it augments the baseline equation with the $Delay$ variable. The odd columns of Table 3.6 first establish that the variable $\#Checkpoints$ affects exam performance by estimating equation (21) without the $Delay$ variable; the even columns presents results for regressions that also include the $Delay$ variable.

The odd columns of panel A provide evidence that each additional checkpoint encountered along the route to school has a negative impact on the probability of passing (column 1), the overall TGE score (column 3), Maths exam scores (column 5), and English exam scores (column 7). Comparing columns (5) and (7) highlights the particularly adverse effects of additional checkpoints on Maths scores compared to English scores.

Table 3.6: Mechanisms – Time loss due to mobility restrictions

	Pass		Overall score		Maths		English	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Mobility restrictions								
Total number of CP	-0.0120**	-0.0108**	-0.0311***	-0.0285***	-0.0345***	-0.0326***	-0.0167*	-0.0136
	(0.0052)	(0.0052)	(0.0109)	(0.0109)	(0.0115)	(0.0115)	(0.0099)	(0.0100)
Delay factor (in minutes)		-0.0011**		-0.0022*		-0.0016		-0.0027**
		(0.0005)		(0.0011)		(0.0012)		(0.0011)
B. Other conflict variables								
Fatalities in school locality	-0.0007***	-0.0007***	-0.0012**	-0.0012**	-0.0017*	-0.0017*	-0.0010**	-0.0010**
	(0.0002)	(0.0002)	(0.0005)	(0.0005)	(0.0009)	(0.0009)	(0.0004)	(0.0004)
Fatalities in home locality	0.0001	0.0001	0.0004	0.0004	0.0009	0.0009	0.0005	0.0005
	(0.0002)	(0.0002)	(0.0004)	(0.0004)	(0.0007)	(0.0007)	(0.0004)	(0.0004)
# Student-Years	146,942	146,942	146,942	146,942	146,268	146,268	146,942	146,942
Adjusted R^2	0.107	0.107	0.255	0.255	0.165	0.165	0.327	0.327

Notes. This table presents estimates from Table (21). Scores expressed in standard deviations. All regressions include the following fixed effects: school locality fixed effects, home locality fixed effects, distance bin fixed effects, and academic year fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of home and school locality, student indicators (gender, religion, year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, gender of school). Standard errors in parentheses, clustered at the home-school locality pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The even columns of panel A provide evidence that delays at checkpoints contribute to the negative impact of mobility restrictions on educational performance. In column (2), the β_1 estimate suggests that conditional on the number of checkpoints encountered and on the distance travelled from home to school, each additional minute delayed at checkpoints reduces the probability of passing the TGE by 0.11 pp. Existing literature suggests that over the Second Intifada delay times at checkpoints were typically at least 15 minutes (Eklund, 2010; Eklund and Martensson, 2012). Therefore, this estimate suggests that being delayed at one or more checkpoints reduces the probability of passing by at least 1.65 pp (0.11×15). This is a non-trivial component of the overall effect of crossing at least one checkpoint (a reduction of 3.05 pp in the probability of passing). The β_2 estimate indicates that each additional checkpoint reduces the probability of passing the TGE by 1.08 pp, similar to the estimate in column (1). This suggests the checkpoint count measure and the constructed delay times are not strongly correlated with each other, after partialing out all the other explanatory variables and fixed effects included in equation (21). Panel B presents the γ_l and γ_h coefficients and shows that an additional fatality at the school locality reduces the probability of passing the TGE by 0.07 pp whereas fatalities at the home locality do not have an impact on educational performance.

The estimates in column (4) also indicate that delays at checkpoints may be an important reason why mobility restrictions have a negative impact on overall TGE scores. Panel A shows that each additional minute of delay at checkpoints reduces the overall TGE score by 0.0022 standard deviations. Evaluated at an average delay of 15 minutes per checkpoint, this suggests that the time delay mechanism can account for a 0.0333 (0.0022×15) standard deviation decrease in overall TGE scores. This calculation is roughly half the estimated impact of crossing at least one checkpoint on overall TGE scores in Table 3.2 (-0.071 standard deviations). Panel B of this column indicates that additional fatalities at the school locality have a strong negative impact on the overall score, conditioning on the delay time and the total number of checkpoints encountered.

Columns (6) and (8) explore the impact of delays on Maths and English scores. Column (6) shows that although an additional checkpoint encountered has a strong negative impact on Maths scores (-0.033 standard deviations), the coefficient on the *Delay* variable is negative but statistically insignificant. Column (8) suggests that each additional minute delayed at a checkpoint

reduces English scores by 0.0027 standard deviations. Evaluated at an average delay time of 15 minutes per checkpoint, this would suggest that being delayed at a checkpoint reduces English scores by 0.041 standard deviations (0.0027×15), almost 80% of the baseline effect. When the *Delay* variable is included, the coefficient on the total number of checkpoints encountered becomes statistically insignificant. This suggests that the main avenue through which mobility restrictions affect English exam scores is via time loss. Panel B of these two columns shows that even when conditioning on the delay time and total number of checkpoints encountered, the estimated impact of an additional fatality in the home or school locality remains very similar to the estimates in the baseline specification.

Overall, the estimated coefficient on the *Delay* variable in equation (21) establishes that delays at checkpoints contains information over and above the measure of the number of checkpoints crossed and the distance travelled from home to school. These results therefore support the hypothesis that one reason why physical barriers in the form of checkpoints worsen academic performance is that students lose time when travelling to and from school, reducing the time available to study.

6.2 Impact via Psychological Well-being

Since the test scores data do not contain information on the psychological wellbeing of students, to examine whether exposure to conflict has a negative impact on the wellbeing of students, we supplement our analysis with the *wellbeing survey* conducted by the PCBS. The survey sample consists of 2,301 representative households in the West Bank. The survey primarily focuses on pre-secondary school age children; thus, the student sample covered by the well-being survey are younger than those in the test scores data. We focus on students between 8 and 17 years old who are enrolled in school.[123]

The survey questions that we use can be broadly categorised into three domains. First, the survey provides information about the locality of residence of the household. This allows us to

[123] Over 99.2% of children within this age group report being enrolled in school.

match measures of mobility restrictions and conflict-related violence to the student home locality. Although the survey does not include information on the school locality, it asks whether students attend school in a different locality to their home locality. We use this information to construct an indicator to capture whether the student traverses localities to attend school. Among the sample of students enrolled in school, 27.57% attend a school in a different locality than their home locality. Second, the survey asks an extensive list of questions that allow us to control for household characteristics that might be correlated with psychological well-being and affect exposure to mobility restrictions and violence.

Third, and most importantly, the survey contains several questions that aim to measure well-being of students. We focus on three questions asked to children or their caregiver, respectively capturing cognitive, emotional, and behavioural problems suffered by children. The cognitive question asks whether the child suffers from a "lack of ability to concentrate" in the half year before the survey (yes or no).[124] The psychological question asks whether the student suffers from "feelings of hopelessness and frustration" or "anger and nervous breakdown" (yes or no).[125] The behavioural question asks whether the child has tendencies of "violence", "screaming" or to "beat/imprecate others" (yes or no).[126] We use these questions to construct binary outcomes of cognitive, emotional, and behavioural problems.

Using this data, we estimate the following cross-sectional regression for student i living in home locality h:

$$y_{ih} = \alpha + \beta_1 B_h + \beta_2 B_h \cdot T_{ih} + \kappa T_{ih} + F_h\gamma + P_h'\eta + X_{ih}'\zeta_1 + \epsilon_{ih} \tag{22}$$

The dependent variable y_{ih} consists of binary indicators for cognitive, emotional, and behavioural problems. In columns (1) and (2) of Table 3.7, the dependent variable takes on a value of 1 if the

[124]The survey question is: "Does X suffer from a lack of concentration?"

[125]The survey question is: "Which of the following psychological/emotional problems suffered by X? (a) Feelings of hopelessness and frustration, (b) Anger and nervous breakdown, (c) None of the above."

[126]Which of the following behavioural problems suffered by X? (a) Tendency to violence, (b) Tendency to screaming, (c) Tendency to beat/imprecate others, (d) None of the above.

student reports suffering from issues of concentration and zero otherwise. In columns (3) and (4), the dependent variable takes on a value of 1 if the student reports having feelings of hopelessness, frustration, anger, or having experienced a nervous breakdown and zero otherwise. In columns (5) and (6), the dependent variable takes on a value of 1 if the student reports a tendency of violent behaviour and zero otherwise. In this sample, the baseline rates of reported concentration, emotional, and behavioural problems are high. Almost 41% of students report experiencing a "lack of ability to concentrate". 53.1% experience feelings of hopelessness, anger or report having experienced a nervous breakdown. 26.4% of students report a tendency to be violent.

Similar to the within-school specification, B_h is an indicator variable that equals one if there are one or more checkpoints within 10 km of the home locality centre in the (calender) year before the survey. We interact this mobility restriction measure with a dummy T_{ih} which denotes whether student i living in home locality h attends a school that is located in a different locality to his/her home locality. The variable F_h measures the number of fatalities in home locality h in the 12 months before the survey.[127]

Since the well-being survey is only available for one year, we are unable to use a fixed-effects strategy to deal with time-invariant observables that might influence exposure to conflict as well as psychological wellbeing. Our empirical strategy therefore involves conditioning on a rich set of controls in the vectors P'_h and X'_{ih}. The vector P'_h consists of the extended list of lagged policy variables (population size of Israeli settlers within 10 km of home locality (in 1000s), number of Palestinian prisoners in locality h (in 1000s), an indicator for one or more house demolitions, the length of the separation wall going through that locality) and economic variables (average hourly wage and unemployment rate) in locality h. The vector X_{ih} includes a set of household characteristics. In the odd columns of the table, we present estimates from specifications where X_{ih} consists of a restricted set of student controls (age and gender) similar to those used for the main specification (equation 19). In the even columns of the table, we include a more extensive list of controls: refugee status (registered refugee, non-registered refugee, non-refugee), dwelling type (e.g. villa, house, apartment), availability of health insurance, household income (in Israeli

[127] Since the survey took place in April to May 2001, this variable consists of fatalities from April 2000 to March 2001.

Shekels), and the educational attainment of the household head.

Table 3.7 presents the β_1, β_2, and γ coefficients from equation (22). Panel A examines whether exposure to one or more checkpoints within 10 km of the home locality affects the probability of students experiencing any of these problems, conditioning on other conflict measures that might influence well-being. Row (i) of column (1) shows that among students that live and study in the same locality, mobility restrictions do not appear to have an impact on their ability to concentrate. However, among students that attend school in a different locality, the presence of one or more checkpoints within 10 km of the home locality centre increases the probability of suffering from a lack of concentration by 15.24 pp. When additional controls are included in column (2), this estimate is very stable at 15.84 pp, suggesting that the estimated impact is unlikely to be driven by selection based on observable characteristics.

Columns (3) and (4) indicate that the presence of one or more checkpoints within 10 km of the home locality centre does not have an impact on the probability of experiencing feelings of hopelessness or frustration for either traversing or non-traversing students.

Columns (5) and (6) show that exposure to checkpoints in the home locality increases the likelihood that both traversing and non-traversing students exhibit violent behaviours. The estimates in row (i) suggest that the presence of at least one checkpoint within 10 km of the home locality increases the probability of violent behaviour among non-traversing students by 5 pp. This effect is even more pronounced among traversing students, increasing their probability of behaving in violent ways by over 9 pp. These estimates are stable when additional individual and household-level controls are included.

Panel B examines the impact of exposure to fatalities in the months prior to the survey on the three measures of well-being. Column (1) indicate that each additional fatality occurring in the 12 month period prior to the survey increases the probability of students suffering from a lack of concentration by 0.6 pp, significant at the 5% significance level. When additional controls are included, the estimated impact of an additional fatality is identical. Columns (3) and (4) suggest that fatalities occurring in the 12 month period prior do not have a detectable impact on the probability of students reporting feelings of hopelessness or anger. Lastly, columns (5) and

Table 3.7: Mechanisms – Impact of mobility restrictions on wellbeing

		Cognitive (Lack of concentration)		**Emotional** (Hopeless/Anger)		**Behaviour** (Violent acts)	
		(1)	(2)	(3)	(4)	(5)	(6)
	Dep. Var. Mean	0.409		0.531		0.264	
A.	**Mobility restrictions**						
(i)	$\geq$ 1 CP within 10 km	-0.0104	-0.0039	0.0128	0.0091	0.0477**	0.0499**
		(0.0219)	(0.0223)	(0.0224)	(0.0229)	(0.0197)	(0.0200)
(ii)	$\geq$ 1 CP within 10 km	0.1628***	0.1623***	0.0305	0.0234	0.0435	0.0434
	$\times$ commute	(0.0624)	(0.0626)	(0.0634)	(0.0635)	(0.0567)	(0.0577)
	Row (i) + (ii)	0.1524**	0.1584***	0.0433	0.0325	0.0913*	0.0934*
		(0.0608)	(0.0611)	(0.0615)	(0.0617)	(0.0554)	(0.0565)
B.	**Other conflict variables**						
	Fatalities in home locality	0.0059***	0.0059**	0.0035	0.0031	0.0095***	0.0084***
		(0.0023)	(0.0023)	(0.0023)	(0.0024)	(0.0022)	(0.0023)
	Basic student controls	Yes	Yes	Yes	Yes	Yes	Yes
	Additional student controls		Yes		Yes		Yes
	Other policy variables	Yes	Yes	Yes	Yes	Yes	Yes
	Locality controls	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	2,706	2,698	2,706	2,698	2,706	2,698
	Adjusted R^2	0.057	0.061	0.044	0.043	0.041	0.047

Notes. This table presents estimated coefficients of equation (22). Sample includes students aged 8–17 in the Palestinian wellbeing survey. We focus on students who attend school in the West Bank to be consistent with the main test score data. Basic student controls include gender and age. Additional student controls include: refugee status, dwelling type, household income, educational attainment of household head, whether the individual has insurance. Other policy variables number of political prisoners in the home locality (in 1000s), indicator for at least one house demolition in the home locality, and population size (in 1000s) of Israeli settlements within 10 km of home locality. Locality controls include average hourly wage and average unemployment rate in the previous year. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data source: wellbeing survey.

(6) provide evidence that fatalities occurring in the 12 month period prior to the survey month increase tendency towards violent behaviour. Each additional fatality in the home locality increases the probability of violent behaviour by just under 1 pp. It is interesting to note that the effect of conflict-related mobility restrictions has a larger impact on students' ability to concentrate compared to the impact on violent behaviour (panel A) whereas the effect of conflict-related violence has a larger impact on violent behaviour relative to its impact on ability to concentrate (panel B).

Overall, these results lend support to the hypothesis that one potential reason for the negative impact of mobility restrictions on educational performance is the detrimental impact mobility restrictions have on the psychological well-being of these students, which may in turn influence their

ability to perform well in school. The cross-sectional nature of the well-being survey data leads us to be more cautious about ascribing a causal interpretation to these findings. However, the fact that the estimated effects are fairly stable when an extensive list of student, household, and locality characteristics is included somewhat alleviates the concern that these results are driven by students who are more prone to psychological problems being more likely to be exposed to conflict.

6.3 Impact via School-learning Environment

It has been documented that mobility restrictions impede the movement of goods across the West Bank (Calì and Miaari, 2018; World Bank, 2004). This might impact the availability of physical resources such as computers at school which may be important inputs into educational performance. To examine this mechanism, we estimate school-level versions of the within-school specification (equation 20), replacing the dependent variables with measures that capture elements of the school learning environment. We also replace the vector X_{it} with school-level averages of percentages in the Arts and Science stream since this is likely to affect the importance of some facilities (e.g. science labs). Our main outcomes of interest are binary variables for the availability of a library, science lab, computer lab, and electricity.

These results are presented in Table 3.8. Column (1) studies the impact of conflict on the availability of a library in school. The results indicate that the introduction of at least one checkpoint within 10 km of the school locality centre is followed by a reduction in the probability that schools in that locality have a functioning library by 4.75 pp, relative to 83% of schools having a library. Column (2) shows that the presence of at least one checkpoint within 10 km of the school locality centre significantly reduces the probability of a school having a science lab by 8.35 pp, relative to 71% of schools having a science lab. Column (3) suggests that checkpoints within 10 km of the school locality centre reduce the probability of a school having a computer lab by 7.5 pp. Lastly, column (4) indicates that the availability of electricity in the school is not negatively affected by the presence of checkpoints around the school. These results hold when we restrict the sample to schools that are present throughout the entire sample (57% of the overall sample of schools), suggesting that these findings are not driven by the changes in resources of new schools entering

Table 3.8: Mechanisms – Impact of mobility restrictions on school learning environment

	Library	Science lab	Computer lab	Electricity
	(1)	(2)	(3)	(4)
Dep. Var. Mean	0.832	0.710	0.730	0.986
A. Mobility restrictions				
$\geq$ 1 CP within 10 km	-0.0475**	-0.0835**	-0.0745**	0.0032
	(0.0237)	(0.0384)	(0.0306)	(0.0057)
B. Other conflict variables				
Fatalities in school locality	-0.0155	-0.0495	0.0069	-0.0031
	(0.0261)	(0.0342)	(0.0343)	(0.0054)
# School-Years	2,761	2,419	2,761	2,748
# Schools	539	539	539	539
Adjusted R^2	0.205	0.247	0.247	0.201

Notes. Scores expressed in standard deviations. All regressions include the following fixed effects: school fixed effects and academic year fixed effect. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of school locality and school controls (number of classrooms, total number of students, total number of teachers). Regressions are weighted by the number of TGE takers in the school. Standard errors in parentheses, clustered at the school locality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

during this period.

These estimates raise the question of whether mobility restrictions surrounding the school lead to a permanent loss of these facilities. Descriptive statistics from the schools data suggest that these effects are temporary. For example, among schools that report losing a library in a particular year, 70.83% of them report having a library in the following academic year. Similar patterns are found for other facilities: among schools that report losing a science lab or computer lab in a given year, 58.47% and 61.77% of them report re-opening these facilities in the following academic year respectively.

Panel B shows that fatalities in the school locality do not have a negative impact on any of the four measures of the physical environment of the school. This is consistent with the idea that the measures of checkpoints surrounding the school capture the impact of restricted mobility of goods, rather than other unobservable time-varying school characteristics such as financial constraints that may also affect the supply of physical resources to the school. If the conflict variables were merely capturing unobservable time-varying school characteristics such as financial constraints, it is likely that this would also be reflected in the fatalities measures.

One potential reason for the negative impact of mobility restrictions on the presence of these physical resources is that these restrictions reduce the supply of books, science lab equipment, and computer stock required to keep these facilities running. It has been documented that goods crossing checkpoints have to be off-loaded, inspected by security personnel, and then reloaded onto local trucks for delivery (World Bank, 2004). These increased transport costs may deter schools from ordering additional supplies for these facilities in years when the movement of goods is highly restricted. Another potential explanation is that mobility restrictions affect the supply of the personnel such as librarians, lab demonstrators, or I.T. technicians who help run these facilities. Overall, the results in Table 3.8 provide some evidence that mobility restrictions have a negative impact on the physical resources in schools.

6.4 Impact via Teachers

To examine whether teachers could be an important channel through which mobility restrictions affect student performance, we make use of information on teachers' home and school locality which is present in the teacher data. For each school, we construct measures of the percentage of teachers who cross a checkpoint on the road to school and the percentage of teachers who live in a locality that differs from their home locality. We then estimate the following regression:

$$y_{slt} = \alpha + \beta_1 TeacherCross_{st} + \beta_2 TeacherTraverse_{st} + \beta_3 B_{lt} + F_{l,t-1}\gamma + W'_{st}\zeta_2 + \lambda_s + \tau_t + \epsilon_{st} \quad (23)$$

In equation (23), the dependent variable y_{slt} are school-level average educational outcomes. The variable $TeacherCross_{st}$ measures the percentage of teachers in school s and academic year t who cross a checkpoint on their shortest route to school. Since the proportion of teachers who cross a checkpoint is positively correlated with the proportion of teachers who live in a locality that differs to their school locality, we control for the proportion of teachers who traverse localities to work via the variable $TeacherTraverse_{st}$. The final two terms of equation (23) are equivalent to the main conflict measures used in the within-school specification. The remaining terms are

defined as before. We weight these regressions by the number of TGE takers in the school and cluster standard errors at the school-locality level.

The results for this regression are shown in Table 3.9. Panel A shows that the teacher exposure measures have no impact on any school-average performance measures. For example, a percentage point increase in the proportion of teachers who cross a checkpoint has a statistically and economically insignificant impact on the probability of passing the exam (0.009 pp) and on the overall TGE score (-0.007 standard deviations). The standard errors on the teacher exposure variables are large, suggesting that there is limited variation in teacher exposure to conflict over time within schools. Panel B of columns (1) and (2) show the estimated impact of introducing one or more checkpoints within 10 km of the school locality centre is a 1.39 pp reduction in the probability of passing the TGE and a -0.049 standard deviation in the overall TGE score, similar to the baseline estimates of the within-school specification. Panel C shows that additional fatalities in the school locality have a negative impact on average school-level outcomes. For example, an additional fatality in the school locality reduces the probability of passing by 0.08 pp, similar to the within-school specification estimates.

Overall, the results in this table firstly suggest that the additional information contained in the teacher exposure variables are not informative for explaining variation within schools in academic outcomes. Secondly, including the teacher exposure variables (panel A) does not alter the magnitude of the coefficients on the other conflict variables (panels B and C). There are some shortcomings of our analysis on the teacher mechanism which may explain why we do not find an effect. An important caveat is that although we know which school each teacher works at in each year, we do not have information on the classes or students they taught. Thus, we cannot construct average pass rates of average exam scores for the students who were taught by individual teachers. Even if the quality of some teachers is negatively impacted by exposure to conflict, since their instructional quality may not matter for all students in the school, such effects may not be detected by our school-level regressions. Another caveat is that we do not observe individual-level or school-level teacher absenteeism rates which have been shown to be important in the existing literature (Monteiro and Rocha, 2017).

Table 3.9: Mechanisms - Impact of teachers' exposure to mobility restrictions

	Pass	Overall score	Maths	English
	(1)	(2)	(3)	(4)
A. Teacher Mobility Restrictions				
% Teachers cross CP	0.0009	-0.0071	0.0041	-0.0019
	(0.0273)	(0.0627)	(0.0721)	(0.0635)
% Teachers commuting	-0.0071	-0.0309	-0.0496	-0.0367
	(0.0203)	(0.0542)	(0.0583)	(0.0545)
B. School/Student Mobility restrictions				
$\geq$ 1 CP within 10 km	-0.0139**	-0.0491***	-0.0649***	-0.0476***
	(0.0054)	(0.0138)	(0.0137)	(0.0137)
C. Other conflict variables				
Fatalities in school locality	-0.0008***	-0.0019***	-0.0019**	-0.0018***
	(0.0002)	(0.0004)	(0.0008)	(0.0006)
# School-Years	2,820	2,820	2,813	2,820
# Schools	478	478	477	478
Adjusted R^2	0.666	0.768	0.723	0.754

Notes. Scores expressed in standard deviations. All regressions include the following fixed effects: school fixed effects and academic year fixed effect. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of school locality and school controls (number of classrooms, total number of students, total number of teachers). Regressions are weighted by the number of TGE-takers in the school. Standard errors in parentheses, clustered at the school locality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7 | Conclusion

A prominent feature of contested regions are conflict infrastructures such as walls and checkpoints (Pullan et al., 2012). These infrastructures restrict mobility and affect the daily routine of those living nearby. This chapter studies the impact of such physical barriers and resulting mobility restrictions on the educational performance of high school students. We study this question in the context of the West Bank between 2000 to 2006 during the IP conflict. This is a suitable setting to study this question because a prominent feature of this conflict is the system of mobility restrictions enforced through various physical barriers such as checkpoints, infrastructures manned by security personnel. We combine unique and novel data on the final high school exam performance of the entire population of high school students in the West Bank with geographic information data on the location of physical barriers and various other measures of conflict such as fatalities.

We find strong evidence that mobility restrictions in the form of checkpoints have adverse impacts on educational performance. Having to cross one or more checkpoints reduces the probability of passing the final exam by 3.05 pp and overall test scores by 0.071 standard deviations. The impacts of mobility restrictions operate through a distinct channel to conflict-related violence which has been the focus of the previous literature. Our estimates indicate that, conditional on exposure to barriers, an additional fatality in the school locality reduces the probability of passing by 0.07 pp and the overall test score by 0.001 standard deviations. The effects of mobility restrictions are particularly detrimental to Maths scores and among male students.

We find evidence of three mechanisms at play. First, increased travel time due to delay at checkpoints lead to less time for studying. Our results suggest that each additional minute travelled reduces the probability of passing the final exam by 0.11 pp. Second, mobility restrictions significantly increase the probability of students suffering from a lack of concentration and increase the tendency of violent behaviour. These effects are concentrated among students who study outside their home locality and are therefore more likely to be exposed to checkpoints. Third, checkpoints surrounding a school negatively impacts the probability that a school has a library, science lab, or computer lab, potentially worsening the school learning environment.

To the best of our knowledge, this is the first study of the impact of mobility restrictions on educational performance. Given the importance of educational performance for future labour market prospects and for aggregate human capital accumulation, understanding the relationship between the various facets of conflict and educational outcomes is valuable. Furthermore, understanding the mechanisms through which conflict affects educational outcomes and the potential role of mobility restrictions in determining future academic success will help in the planning of educational policies.

Conclusion to Thesis

This thesis has explored the determinants of healthcare quality, health outcomes, and educational performance. In the first chapter, I studied the impact of asymmetric financial incentives on hospital quality provision. I demonstrated that hospitals are responsive towards these financial incentives by reducing readmission rates among targeted patients. Furthermore, readmission rates among non-targeted patients also decreased. This suggests that there are cost complementarities in the provision of quality within hospitals.

In the second chapter, I investigated whether demand-side responses to health insurance affects readmission rates. Existing literature and policy debates tend to focus on provider-side factors as the main driver of post-discharge outcomes such as readmissions. The results of this chapter suggest that patient-side factors may also be important for understanding the drivers of readmissions. In particular, I provide evidence that among male heart attack patients, eligibility for Medicare insurance increases the probability of 30-day readmissions but does not affect the quality of treatment received.

In the third chapter, Sami Miaari and I examine how exposure to mobility restrictions such as checkpoints affect the performance of high-school students in the West Bank on an important exam. We show that having to cross a checkpoint on the road to school has detrimental impacts on educational performance, even when exposure to other aspects of conflict such as conflict-related fatalities are controlled for. Existing literature overlooks the role of mobility restrictions which is an important feature in many contemporary conflicts.

Bibliography

Aigner, Dennis J (1973) "Regression with a binary independent variable subject to errors of observation," *Journal of Econometrics*, Vol. 1, No. 1, pp. 49–59.

Aizer, Anna (2009) "Neighborhood violence and urban youth," *The Problems of Disadvantaged Youth: An Economic Perspective*.

Al Kafri, Saleh (2002) "Impact of Israeli measures on Palestinian child labor and schooling," *Economic Research Forum Working Paper Series 215*.

Al-Krenawi, Alean, John R Graham, and Yaniv Kanat-Maymon (2009) "Analysis of trauma exposure, symptomatology and functioning in Jewish Israeli and Palestinian adolescents," *The British Journal of Psychiatry*, Vol. 195, No. 5, pp. 427–432.

Albrektsen, Grethe, Ivar Heuch, Maja-Lisa Løchen, Dag Steinar Thelle, Tom Wilsgaard, Inger Njølstad, and Kaare Harald Bønaa (2016) "Lifelong gender gap in risk of incident myocardial infarction: the Tromsø Study," *JAMA Internal Medicine*, Vol. 176, No. 11, pp. 1673–1679.

Almond, Douglas and Janet Currie (2011) "Killing me softly: The fetal origins hypothesis," *Journal of Economic Perspectives*, Vol. 25, No. 3, pp. 153–72.

Anderson, Michael L, Carlos Dobkin, and Tal Gross (2014) "The effect of health insurance on emergency department visits: Evidence from an age-based eligibility threshold," *Review of Economics and Statistics*, Vol. 96, No. 1, pp. 189–195.

Aranki, Ted N (2006) "The effect of Israeli closure policy on wage earnings in the West Bank and Gaza Strip," *Working paper*.

Ballas, Dimitris, Graham Clarke, Rachel S Franklin, and Andy Newing (2017) *GIS and the social sciences: Theory and applications*: Routledge.

Banerji, Rukmini, James Berry, and Marc Shotland (2017) "The impact of maternal literacy and participation programs: Evidence from a randomized evaluation in India," *American Economic Journal: Applied Economics*, Vol. 9, No. 4, pp. 303–37.

Barka, Habiba Ben (2012) "Border posts, checkpoints, and Intra-African trade: Challenges and solutions," *OPEV Newsletter*, pp. 1–18.

Barrett, ML, LM Wier, J Jiang, and CA Steiner (2015) "All-cause readmissions by payer and age, 2009–2013," *Agency for Healthcare Research and Quality, Rockville, MD*.

Batt, Robert J., Hessam Bavafa, and Mohamad Soltani (2018) "Quality Improvement Spillovers: Evidence from the Hospital Readmissions Reduction Program," *Working Paper*.

Benbassat, Jochanan and Mark Taragin (2000) "Hospital readmissions as a measure of quality of health care: advantages and limitations," *Archives of Internal Medicine*, Vol. 160, No. 8, pp. 1074–1081.

Berchick, Edward R, Jessica C Barnett, and Rachel D Upton (2019) "Health Insurance Coverage in the United States: 2018," Available at https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-267.pdf.

Bhatt, Deepak L, Matthew T Roe, Eric D Peterson, Yun Li, Anita Y Chen, Robert A Harrington, Adam B Greenbaum, Peter B Berger, Christopher P Cannon, David J Cohen et al. (2004) "Utilization of early invasive management strategies for high-risk patients with non–ST-segment elevation acute coronary syndromes: results from the CRUSADE Quality Improvement Initiative," *JAMA*, Vol. 292, No. 17, pp. 2096–2104.

Bleakley, Hoyt (2010) "Health, human capital, and development," *Annual Review of Economics*, Vol. 2, No. 1, pp. 283–310.

Blomkalns, Andra L, Anita Y Chen, Judith S Hochman, Eric D Peterson, Kelly Trynosky, Deborah B Diercks, Gerard X Brogan, William E Boden, Matthew T Roe, E Magnus Ohman et al. (2005) "Gender disparities in the diagnosis and treatment of non–ST-segment elevation acute coronary syndromes: Large-scale observations from the CRUSADE (Can Rapid Risk Stratification of Unstable Angina Patients Suppress Adverse Outcomes With Early Implementation of the American College of Cardiology/American Heart Association Guidelines) national quality improvement initiative," *Journal of the American College of Cardiology*, Vol. 45, No. 6, pp. 832–837.

Boccuti, Cristina and Giselle Casillas (2015) "Aiming for fewer hospital U-turns: The Medicare hospital readmission reduction program," *Policy Brief*.

Braverman, Irus (2011) "Civilized borders: A study of Israel's new crossing administration," *Antipode*, Vol. 43, No. 2, pp. 264–295.

——— (2012) "Checkpoint Watch: Bureaucracy and resistance at the Israeli/Palestinian border," *Social & Legal Studies*, Vol. 21, No. 3, pp. 297–320.

Brennan, Jesse J, Theodore C Chan, James P Killeen, and Edward M Castillo (2015) "Inpatient readmissions and emergency deparment visits within 30 days of a hospital admission," *Western Journal of Emergency Medicine*, Vol. 16, No. 7, pp. 1025–1029.

Brück, Tilman, Michele Di Maio, and Sami H Miaari (2019) "Learning the hard way: The effect of violent conflict on student academic achievement," *Journal of the European Economic Association*, Vol. 17, No. 5, pp. 1502–1537.

B'Tselem (2007a) "Fatality Statistics," Available at https://web.archive.org/web/20100105120054/http://www.btselem.org/english/Statistics/Casualties.asp.

——— (2007b) "Ground to a Halt - Denial of Palestinians' Freedom of Movement in the West Bank."

——— (2011) "Testimony: Teacher from Jordan Valley tells of daily delays at checkpoints onhis way to and from school in Northern West Bank," Available at https://www.btselem.org/testimonies/testimony-teacher-jordan-valley-tells-daily-delays-checkpoint-his-way-and-school-northern-.

Butler, JR (1995) *Hospital cost analysis*, Vol. 3: Springer Science & Business Media.

Calì, Massimiliano and Sami H Miaari (2017) "Do Israeli settlements radicalize Palestinians?" *Working paper*.

——— (2018) "The labor market impact of mobility restrictions: Evidence from the West Bank," *Labour Economics*, Vol. 51, pp. 136–151.

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell, and Rocıo Titiunik (Forthcoming) "Regression discontinuity designs using covariates," *Review of Economics and Statistics*.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014) "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, Vol. 82, No. 6, pp. 2295–2326.

——— (2015) "Optimal data-driven regression discontinuity plots," *Journal of the American Statistical Association*, Vol. 110, No. 512, pp. 1753–1769.

Card, David, Carlos Dobkin, and Nicole Maestas (2008) "The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare," *American Economic Review*, Vol. 98, No. 5, pp. 2242–58.

——— (2009) "Does Medicare save lives?" *The Quarterly Journal of Economics*, Vol. 124, No. 2, pp. 597–636.

Carey, Kathleen and Meng-Yun Lin (2015) "Readmissions to New York hospitals fell for three target conditions from 2008 to 2012, consistent with Medicare goals," *Health Affairs*, Vol. 34, No. 6, pp. 978–985.

Centers for Medicare & Medicaid Services (2017a) "Hospital-Acquired Condition Reduction Program," Available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/HAC-Reduction-Program.html.

——— (2017b) "Hospital Value-Based Purchasing," Available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/HAC-Reduction-Program.html.

——— (2017c) "National Health Expenditure Data: Health Expenditures by State of Residence," Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsStateHealthAccountsResidence.html.

Central Intelligence Agency (2008) "The World Factbook.," Available at https://www.cia.gov/library/publications/the-world-factbook/geos/we.html.

Chalkley, Martin and James M Malcomson (1998) "Contracting for health services when patient demand does not reflect quality," *Journal of Health Economics*, Vol. 17, No. 1, pp. 1–19.

Chandra, Amitabh and Douglas O Staiger (2007) "Productivity spillovers in health care: Evidence from the treatment of heart attacks," *Journal of Political Economy*, Vol. 115, No. 1, pp. 103–140.

Chang, Tom and Mireille Jacobson (2011) "What do nonprofit hospitals maximize? Evidence from California's seismic retrofit mandate," *NBER Working Paper*, Vol. 5.

Chen, Christopher, Gabriel Scheffler, and Amitabh Chandra (2014) "Readmission penalties and health insurance expansions: a dispatch from Massachusetts," *Journal of Hospital Medicine*, Vol. 9, No. 11, pp. 681–687.

Chen, Min Michelle and David C Grabowski (2017) "Hospital readmissions reduction program: Intended and unintended consequences," *Medical Care Research and Review*.

Chernew, Michael, Katherine Baicker, and Carina Martin (2010) "Spillovers in Health Care Markets: Implications for Current Law Projections," Available at https://www.cms.gov/Reports-TrustFunds/downloads/spillovereffects.pdf.

Chin, David L, Heejung Bang, Raj N Manickam, and Patrick S Romano (2016) "Rethinking thirty-day hospital readmissions: Shorter intervals might be better indicators of quality of care," *Health Affairs*, Vol. 35, No. 10, pp. 1867–1875.

Choudhry, Niteesh K, Jerry Avorn, Robert J Glynn, Elliott M Antman, Sebastian Schneeweiss, Michele Toscano, Lonny Reisman, Joaquim Fernandes, Claire Spettell, Joy L Lee et al. (2011) "Full coverage for preventive medications after myocardial infarction," *New England Journal of Medicine*, Vol. 365, No. 22, pp. 2088–2097.

Clark, Jonathan R and Robert S Huckman (2012) "Broadening focus: Spillovers, complementarities, and specialization in the hospital industry," *Management Science*, Vol. 58, No. 4, pp. 708–722.

Commonwealth Fund (2014) "Mirror, Mirror on the Wall, 2014 Update: How the US Healthcare System Compares Internationally," Available at www.commonwealthfund.org/publications/fund-reports/2014/jun/mirror-mirror.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora (2013) "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment," *The Quarterly Journal of Economics*, Vol. 128, No. 2, pp. 531–580.

Currie, Janet, W Bentley MacLeod, and Jessica Van Parys (2016) "Provider practice style and patient health outcomes: the case of heart attacks," *Journal of Health Economics*, Vol. 47, pp. 64–80.

Cutler, David, Mark McClellan, and Joseph Newhouse (1998) "The costs and benefits of intensive treatment for cardiovascular disease," *NBER Working Paper*.

Darden, Michael, Ian McCarthy, and Eric Barrette (2018) "Hospital Pricing and Public Payments," *NBER Working Paper*.

Deneffe, Daniel and Robert T. Masson (2002) "What do nonprofit hospitals maximize?" *International Journal of Industrial Organization*, Vol. 20, No. 4, pp. 403–428.

Di Maio, Michele and Tushar K Nandi (2013) "The effect of the Israeli–Palestinian conflict on child labor and school attendance in the West Bank," *Journal of Development Economics*, Vol. 100, No. 1, pp. 107–116.

Dimitry, Lydia (2012) "A systematic review on the mental health of children and adolescents in areas of armed conflict in the Middle East," *Child: Care, health and development*, Vol. 38, No. 2, pp. 153–161.

Donabedian, Avedis (1988) "The quality of care: how can it be assessed?" *JAMA*, Vol. 260, No. 12, pp. 1743–1748.

Duffin, Erin (2019) "Percentage added to the Gross Domestic Product (GDP) of the United States of America in 2018, by industry (as a percentage of GDP)," Available at https://www.statista.com/statistics/248004/percentage-added-to-the-us-gdp-by-industry/.

Dumont, Etienne, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer (2008) "Physicians' multitasking and incentives: Empirical evidence from a natural experiment," *Journal of Health Economics*, Vol. 27, No. 6, pp. 1436–1450.

Eklund, Lina (2010) "Accessibility to health services in the West Bank, Occupied Palestinian Territory," *Working Paper*.

Eklund, Lina and U Martensson (2012) "Using geographical information systems to analyse accessibility to health services in the West Bank, occupied Palestinian territory," *Eastern Mediterranean Health Journal*, Vol. 18, No. 8, pp. 796–802.

Elixhauser, Anne, Claudia Steiner, D Robert Harris, and Rosanna M Coffey (1998) "Comorbidity measures for use with administrative data," *Medical Care*, Vol. 36, No. 1, pp. 8–27.

Evans, William N and W Kip Viscusi (1991) "Estimation of state-dependent utility functions using survey data," *Review of Economics and Statistics*, pp. 94–104.

Feng Lu, Susan (2012) "Multitasking, information disclosure, and product quality: Evidence from nursing homes," *Journal of Economics & Management Strategy*, Vol. 21, No. 3, pp. 673–705.

Finkelstein, Amy, Erzo FP Luttmer, and Matthew J Notowidigdo (2013) "What good is wealth without health? The effect of health on the marginal utility of consumption," *Journal of the European Economic Association*, Vol. 11, No. suppl_1, pp. 221–258.

Finkelstein, Amy, Neale Mahoney, and Matthew J Notowidigdo (2018) "What does (formal) health insurance do, and for whom?" *Annual Review of Economics*, Vol. 10, pp. 261–286.

Finkelstein, Amy N, Sarah L Taubman, Heidi L Allen, Bill J Wright, and Katherine Baicker (2016) "Effect of Medicaid coverage on ED use—further evidence from Oregon's experiment," *New England Journal of Medicine*, Vol. 375, No. 16, pp. 1505–1507.

Fuchs, Victor R (2004) "More Variation In Use Of Care, More Flat-Of-The-Curve Medicine: Why does it occur? What should be done about it?" *Health Affairs*, Vol. 23, No. Supplement 2.

Garthwaite, Craig, Tal Gross, and Matthew J Notowidigdo (2018) "Hospitals as insurers of last resort," *American Economic Journal: Applied Economics*, Vol. 10, No. 1, pp. 1–39.

Gaynor, Martin, Kate Ho, and Robert J Town (2015) "The industrial organization of health-care markets," *Journal of Economic Literature*, Vol. 53, No. 2, pp. 235–84.

Gershenson, Seth and Erdal Tekin (2018) "The effect of community traumatic events on student achievement: Evidence from the beltway sniper attacks," *Education Finance and Policy*, Vol. 13, No. 4, pp. 513–544.

Giacaman, Rita, Harry S Shannon, Hana Saab, Neil Arya, and Will Boyce (2007) "Individual and collective exposure to political violence: Palestinian adolescents coping with conflict," *The European Journal of Public Health*, Vol. 17, No. 4, pp. 361–368.

Gibbons, Stephen and Stephen Machin (2006) "Transport and labour market linkages: Empirical evidence, implications for policy and scope for further UK research," *Department of Transport*.

Gupta, Atul (2017) "Impacts of performance pay for hospitals: The Readmissions Reduction Program," *Working Paper*.

Hammond, Bill (2018) "Revisiting New York's Restrictive Hospital Ownership Laws," Available at https://www.empirecenter.org/wp-content/uploads/2018/05/Hospital-Ownership_May-2018_FINAL.pdf.

Hanushek, Eric A and Ludger Woessmann (2011) "The economics of international differences in educational achievement," in *Handbook of the Economics of Education*, Vol. 3: Elsevier, pp. 89–200.

Hines, Anika L, Marguerite L Barrett, H Joanna Jiang, and Claudia A Steiner (2014) "Conditions with the largest number of adult hospital readmissions by payer, 2011: statistical brief number 172."

Holmstrom, Bengt and Paul Milgrom (1991) "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," *Journal of Law, Economics, & Organization*, Vol. 7, p. 24.

Horwitz, Jill R and Austin Nichols (2009) "Hospital ownership and medical services: Market mix, spillover effects, and nonprofit objectives," *Journal of Health Economics*, Vol. 28, No. 5, pp. 924–937.

Horwitz, Leora, Chohreh Partovian, Zhenqiu Lin, Jeph Herrin, Jacqueline Grady, Mitchell Conover, Julia Montague, Chloe Dillaway, Kathleen Bartczak, Joseph Ross et al. (2011) "Hospital-wide (all-condition) 30-day risk-standardized readmission measure," *Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. Retrieved September*, Vol. 10, p. 2012.

Hoven, Christina W, Cristiane S Duarte, and Donald J Mandell (2003) "Children's mental health after disasters: The impact of the World Trade Center attack," *Current Psychiatry Reports*, Vol. 5, No. 2, pp. 101–107.

Israel Ministry of Defence (2007) "Israel's Security Fence," Available at www.securityfence.mod.gov.il/pages/eng/purpose.htm.

Jencks, Stephen F, Mark V Williams, and Eric A Coleman (2009) "Rehospitalizations among patients in the Medicare fee-for-service program," *New England Journal of Medicine*, Vol. 360, No. 14, pp. 1418–1428.

Jneid, Hani, Gregg C Fonarow, Christopher P Cannon, Adrian F Hernandez, Igor F Palacios, Andrew O Maree, Quinn Wells, Biykem Bozkurt, Kenneth A LaBresh, Li Liang et al. (2008) "Sex differences in medical care and early death after acute myocardial infarction," *Circulation*, Vol. 118, No. 25, pp. 2803–2810.

Joynt, Karen E and Ashish K Jha (2012) "Thirty-day readmissions—truth and consequences," *New England Journal of Medicine*, Vol. 366, No. 15, pp. 1366–1369.

Justino, Patricia (2012) "Violent conflict and human capital accumulation," in Brown, Graham and Arnim Langer eds. *Elgar Companion to Civil War and Fragile States*, Chap. 13.

Justino, Patricia, Marinella Leone, and Paola Salardi (2013) "Short-and long-term impact of violence on education: The case of Timor Leste," *The World Bank Economic Review*, Vol. 28, No. 2, pp. 320–353.

Kaiser Family Foundation (2018) "Employer Health Benefits Annual Survey."

Kaiser Health News (2014) "364 Hospitals Have High Rates Of Overall Readmissions, New Medicare Data Show," Available at https://khn.org/news/364-hospitals-have-high-rates-of-overall-readmissions-new-medicare-data-show/.

Kansagara, Devan, Honora Englander, Amanda Salanitro, David Kagen, Cecelia Theobald, Michele Freeman, and Sunil Kripalani (2011) "Risk prediction models for hospital readmission: a systematic review," *Journal of the American Medical Association*, Vol. 306, No. 15, pp. 1688–1698.

Kharel, Subham, P. Shivananda, K. S. Ramesh, Shri K. Naga Jothi, and K. Ganesha Raj (2018) "Transportation network model for route And closest facility analysis in central Bengaluru," *International Journal of Application or Innovation in Engineering and Management*, Vol. 7, No. 4, pp. 58–62.

Kibris, Arzu (2015) "The conflict trap revisited: Civil conflict and educational achievement," *Journal of Conflict Resolution*, Vol. 59, No. 4, pp. 645–670.

Kolesár, Michal and Christoph Rothe (2018) "Inference in regression discontinuity designs with a discrete running variable," *American Economic Review*, Vol. 108, No. 8, pp. 2277–2304.

Krumholz, Harlan M, Kun Wang, Zhenqiu Lin, Kumar Dharmarajan, Leora I Horwitz, Joseph S Ross, Elizabeth E Drye, Susannah M Bernheim, and Sharon-Lise T Normand (2017) "Hospital-Readmission Risk—Isolating Hospital Effects from Patient Effects," *New England Journal of Medicine*, Vol. 377, No. 11, pp. 1055–1064.

Lasser, Karen E, Amresh D Hanchate, Danny McCormick, Meredith G Manze, Chieh Chu, and Nancy R Kressin (2014) "The effect of Massachusetts health reform on 30 day hospital readmissions: Retrospective analysis of hospital episode statistics," *British Medical Journal*, Vol. 348, p. 2329.

Lavy, Victor and Adi Boiko (2017) "Management quality in public education: Superintendent value-added, student outcomes and mechanisms," *NBER Working Paper*.

Lee, David S and David Card (2008) "Regression discontinuity inference with specification error," *Journal of Econometrics*, Vol. 142, No. 2, pp. 655–674.

Lee, David S and Thomas Lemieux (2010) "Regression discontinuity designs in economics," *Journal of economic literature*, Vol. 48, No. 2, pp. 281–355.

Leon, Gianmarco (2012) "Civil conflict and human capital accumulation the long-term effects of political violence in perú," *Journal of Human Resources*, Vol. 47, No. 4, pp. 991–1022.

Maas, AHEM and YEA Appelman (2010) "Gender differences in coronary heart disease," *Netherlands Heart Journal*, Vol. 18, No. 12, pp. 598–603.

Mansour, Hani and Daniel I Rees (2012) "Armed conflict and birth weight: Evidence from the al-Aqsa Intifada," *Journal of Development Economics*, Vol. 99, No. 1, pp. 190–199.

Mataria, Awad, Rita Giacaman, Angelo Stefanini, Nirmala Naidoo, Paul Kowal, and Somnath Chatterji (2009) "The quality of life of Palestinians living in chronic conflict: Assessment and determinants," *The European Journal of Health Economics*, Vol. 10, No. 1, pp. 93–101.

Maynard, Charles, Paul E Litwin, Jenny S Martin, and W Douglas Weaver (1992) "Gender differences in the treatment and outcome of acute myocardial infarction: Results from the Myocardial Infarction Triage and Intervention Registry," *Archives of Internal Medicine*, Vol. 152, No. 5, pp. 972–976.

McClellan, Mark and Joseph P Newhouse (1997) "The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach," *Journal of Econometrics*, Vol. 77, No. 1, pp. 39–64.

McCrary, Justin (2008) "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, Vol. 142, No. 2, pp. 698–714.

McGarry, Brian E, Albert A Blankley, and Yue Li (2016) "The impact of the Medicare hospital readmission reduction program in New York state," *Medical Care*, Vol. 54, No. 2, pp. 162–171.

McGill, Tia, Shannon R Self-Brown, Betty S Lai, Melissa Cowart, Ashwini Tiwari, Monique LeBlanc, Mary Lou Kelley et al. (2014) "Effects of exposure to community violence and family violence on school functioning problems among urban youth: The potential mediating role of posttraumatic stress symptoms," *Frontiers in Public Health*, Vol. 2, p. 8.

McGuire, Alistair (1985) "The theory of the hospital: A review of the models," *Social Science & Medicine*, Vol. 20, No. 11, pp. 1177–1184.

Mellor, Jennifer, Michael Daly, and Molly Smith (2017) "Does it pay to penalize hospitals for excess readmissions? Intended and unintended consequences of Medicare's hospital readmissions reductions program," *Health Economics*, Vol. 26, No. 8, pp. 1037–1051.

Miaari, Sami H and Robert M Sauer (2011) "The labor market costs of conflict: Closures, foreign workers, and Palestinian employment and earnings," *Review of Economics of the Household*, Vol. 9, No. 1, pp. 129–148.

Miller, Raegen T, Richard J Murnane, and John B Willett (2008) "Do teacher absences impact student achievement? Longitudinal evidence from one urban school district," *Educational Evaluation and Policy Analysis*, Vol. 30, No. 2, pp. 181–200.

Molitor, David (2018) "The evolution of physician practice styles: Evidence from cardiologist migration," *American Economic Journal: Economic Policy*, Vol. 10, No. 1, pp. 326–56.

Monteiro, Joana and Rudi Rocha (2017) "Drug battles and school achievement: Evidence from Rio de Janeiro's favelas," *Review of Economics and Statistics*, Vol. 99, No. 2, pp. 213–228.

Moore, Brian J, Susan White, Raynard Washington, Natalia Coenen, and Anne Elixhauser (2017) "Identifying increased risk of readmission and in-hospital mortality using hospital administrative data," *Medical care*, Vol. 55, No. 7, pp. 698–705.

Mosca, Lori, Elizabeth Barrett-Connor, and Nanette Kass Wenger (2011) "Sex/gender differences in cardiovascular disease prevention: What a difference a decade makes," *Circulation*, Vol. 124, No. 19, pp. 2145–2154.

Neidell, Matthew and Jane Waldfogel (2010) "Cognitive and noncognitive peer effects in early education," *The Review of Economics and Statistics*, Vol. 92, No. 3, pp. 562–576.

Newhouse, Joseph P (1970) "Toward a theory of nonprofit institutions: An economic model of a hospital," *American Economic Review*, Vol. 60, No. 1, pp. 64–74.

Nicolai, Susan (2007) *Fragmented Foundations: Education and Chronic Crisis in the Occupied Palestinian Territory.*: UNESCO International Institute for Educational Planning (IIEP).

OCHA (2006) "Preliminary Analysis of the Humanitarian Implications of the April 2006 Barrier Projections - Update number 5."

Palestinian Central Bureau of Statistics (2009) "Statistical Atlas of Palestine," Available at http://www.pcbs.gov.ps/Downloads/book1624/book1624_09.pdf.

Pfefferbaum, B, VL Moore, NB McDonald, BT Maynard, Robin H Gurwitch, and Sara J Nixon (1999) "The role of exposure in posttraumatic stress in youths following the 1995 bombing.," *The Journal of the Oklahoma State Medical Association*, Vol. 92, No. 4, pp. 164–167.

Pfefferbaum, Betty, Thomas W Seale, Nicholas B McDonald, Edward N Brandt Jr, Scott M Rainwater, Brian T Maynard, Barbara Meierhoefer, and Peteryne D Miller (2000) "Posttraumatic stress two years after the Oklahoma City bombing in youths geographically distant from the explosion," *Psychiatry*, Vol. 63, No. 4, pp. 358–370.

Psacharopoulos, George (1972) "Rates of Return to Investment in Education around the World," *Comparative Education Review*, Vol. 16, No. 1, pp. 54–67.

Psacharopoulos, George and Harry Anthony Patrinos (2018) "Returns to investment in education," *Education Economis*, Vol. 26, No. 5, pp. 445–458.

Pullan, Wendy (2013) "Spatial Discontinuities: Conflict infrastructures in contested cities," in *Locating Urban Conflicts*: Springer, pp. 17–36.

Pullan, Weny, James ANderson, mick Dumper, and Liam O'Dowd (2012) "Rethinking Conflict Infrastructure: How the built environment sustains divisions in contested cities," Available at http://www.urbanconflicts.arct.cam.ac.uk/downloads/briefing-paper-2.

Pynoos, Robert S, Calvin Frederick, Kathi Nader, William Arroyo, Alan Steinberg, Spencer Eth, Francisco Nunez, and Lynn Fairbanks (1987) "Life threat and posttraumatic stress in school-age children," *Archives of General Psychiatry*, Vol. 44, No. 12, pp. 1057–1063.

Rau, Jordan (2015) "Half of nation's hospitals fail again to escape Medicare's readmission penalties," *Kaiser Health News*.

Rijke, Alexandra and Claudio Minca (2019) "Inside Checkpoint 300: Checkpoint regimes as spatial political technologies in the Occupied Palestinian Territories," *Antipode*, Vol. 51, No. 3, pp. 968–988.

Rivkin, Steven G, Eric A Hanushek, and John F Kain (2005) "Teachers, schools, and academic achievement," *Econometrica*, Vol. 73, No. 2, pp. 417–458.

Rockoff, Jonah E (2004) "The impact of individual teachers on student achievement: Evidence from panel data," *American Economic Review*, Vol. 94, No. 2, pp. 247–252.

Rodríguez-Lesmes, Paul and Marcos Vera-Hernández (2018) "Using two-part contracts to identify cross price elasticities: An application to family doctors," *Working paper*.

Roeser, Robert W, Jacquelynne S Eccles, and Karen R Strobel (1998) "Linking the study of schooling and mental health: Selected issues and empirical illustrations at the level of the individual," *Educational Psychologist*, Vol. 33, No. 4, pp. 153–176.

Roland, Martin and Frede Olesen (2016) "Can pay for performance improve the quality of primary care?" *British Medical Journal*, Vol. 354, p. i4058.

Rosenthal, Elisabeth (2017) *An American Sickness: How Healthcare Became Big Business and How You Can Take It Back*: Penguin Press.

Ryan, Andrew M, Sam Krinsky, Kristin A Maurer, and Justin B Dimick (2017) "Changes in hospital quality associated with hospital value-based purchasing," *New England Journal of Medicine*, Vol. 376, No. 24, pp. 2358–2366.

Sankaran, Roshun, Devraj Sukul, Ushapoorna Nuliyalu, Baris Gulseren, Tedi A Engler, Emily Arntson, Hanna Zlotnick, Justin B Dimick, and Andrew M Ryan (2019) "Changes in hospital safety following penalties in the US Hospital Acquired Condition Reduction Program: Retrospective cohort study," *British Medical Journal*, Vol. 366, p. l4109.

Save the Children (2003) "Growing Up during Curfew. Safeguarding the Basic Rights of Palestinian Children," *Save the Children UK*.

Sawyer, Bradley and Cynthia Cox (2018) "How does health spending in the U.S. compare to other countries?," Available at https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries/#item-start.

Sawyer, Bradley and Daniel McDermott (2019) "How does the quality of the U.S. healthcare system compare to other countries?," Available at https://www.healthsystemtracker.

org / chart-collection / quality-u-s-healthcare-system-compare-countries / #item-overall-years-life-lost-1990-2017.

Schiff, Miriam, Hillah Haim Zweig, Rami Benbenishty, and Deborah S Hasin (2007) "Exposure to terrorism and Israeli youths' cigarette, alcohol, and cannabis use," *American Journal of Public Health*, Vol. 97, No. 10, pp. 1852–1858.

Schwarzwald, Joseph, Matisyohu Weisenberg, Mark Waysman, Zahava Solomon, and Avigdor Klingman (1993) "Stress reaction of school-age children to the bombardment by SCUD missiles.," *Journal of Abnormal Psychology*, Vol. 102, No. 3, p. 404.

Segev, Tom (2006) "A bitter prize: Israel and the occupied territories," *Foreign Affairs*, Vol. 85, pp. 145–150.

Shany, Adi (2016) "Too Scared for School? The effects of terrorism on Israeli student achievement," *Working Paper*.

Shemyakina, Olga (2011) "The effect of armed conflict on accumulation of schooling: Results from Tajikistan," *Journal of Development Economics*, Vol. 95, No. 2, pp. 186–200.

Sisko, Andrea M, Sean P Keehan, John A Poisal, Gigi A Cuckler, Sheila D Smith, Andrew J Madison, Kathryn E Rennie, and James C Hardesty (2019) "National Health Expenditure Projections, 2018–27: Economic And Demographic Trends Drive Spending And Enrollment Growth," *Health Affairs*, Vol. 38, No. 3, pp. 491–501.

Skinner, Jonathan, Amitabh Chandra, Douglas Staiger, Julie Lee, and Mark McClellan (2005) "Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients," *Circulation*, Vol. 112, No. 17, pp. 2634–2641.

Sloan, Frank A (2000) "Not-for-profit ownership and hospital behavior," *Handbook of Health Economics*, Vol. 1, pp. 1141–1174.

Srour, Roney W. (2005) "Children living under a multi-traumatic environment: The Palestinian case," *Israel Journal of Psychiatry and Related Sciences*, Vol. 42, No. 5, p. 88.

Swannell, Cate (2018) "Gender gap in heart attack treatment 'perplexing'," *The Medical Journal of Australia*, p. 1.

Swee, Eik Leong (2015) "On war intensity and schooling attainment: The case of Bosnia and Herzegovina," *European Journal of Political Economy*, Vol. 40, pp. 158–172.

Thabet, Abdel Aziz Mousa, Yehia Abed, and Panos Vostanis (2002) "Emotional problems in Palestinian children living in a war zone: A cross-sectional study," *The Lancet*, Vol. 359, No. 9320, pp. 1801–1804.

Tigre, Robson, Breno Sampaio, and Tatiane Menezes (2017) "The Impact of Commuting Time on Youth's school performance," *Journal of Regional Science*, Vol. 57, No. 1, pp. 28–47.

Tsai, Thomas C, Karen E Joynt, E John Orav, Atul A Gawande, and Ashish K Jha (2013) "Variation in surgical-readmission rates and quality of hospital care," *New England Journal of Medicine*, Vol. 369, No. 12, pp. 1134–1142.

UN (2007) "The humanitarian impact on Palestinians of Israeli Settlements and other infrastructure in the West Bank," Available at https://www.ochaopt.org/sites/default/files/ocharpt_update30july2007.pdf.

UNESCO (2007) "Global Education Digest," Vol. Institute for Statistics. Data Centre. UNESCO, Paris.

——— (2011) "The Hidden Crisis: armed conflict and education," *EFA global monitoring report*.

UNICEF (2010) "The Situation of Palestinian Children in the Occupied Palestinian Territory, Jordan, Syria and Lebanon," *UNICEF, Jordan*.

United Nations Office for the Coordination of Humanitarian Affairs (2018) "Humanitarian Bulletin – occupied Palestinian territory," Available at https://www.ochaopt.org/sites/default/files/hummonitor_september_2018.pdf.

Viscusi, W Kip and William N Evans (1990) "Utility functions that depend on health status: Estimates and economic implications," *American Economic Review*, pp. 353–374.

Werner, Rachel M and R Adams Dudley (2012) "Medicare's new hospital value-based purchasing program is likely to have only a small impact on hospital payments," *Health Affairs*, Vol. 31, No. 9, pp. 1932–1940.

Wilson, W Cody and Beth Spenciner Rosenthal (2003) "The relationship between exposure to community violence and psychological distress among adolescents: A meta-analysis," *Violence and Victims*, Vol. 18, No. 3, p. 335.

de Winter, Robbert J, Fons Windhausen, Jan Hein Cornel, Peter HJM Dunselman, Charles L Janus, Peter EF Bendermacher, H Rolf Michels, Gerard T Sanders, Jan GP Tijssen, and Freek WA Verheugt (2005) "Early invasive versus selectively invasive management for acute coronary syndromes," *New England Journal of Medicine*, Vol. 353, No. 11, pp. 1095–1104.

World Bank (2002) "West Bank and Gaza: An Evaluation of Bank Assistance," Available at http://ieg.worldbankgroup.org/sites/default/files/Data/reports/west_bank_and_gaza.pdf.

——— (2004) "Four Years – Intifada, Closures and Palestinian Economic Crisis."

——— (2007a) "Movement and access restrictions in the West Bank: uncertainty and inefficiency in the Palestinian economy," Available at http://documents.worldbank.org/curated/en/964061468339036297/pdf/404450GZ0West0Bank0restrictions01PUBLIC1.pdf.

——— (2007b) "West Bank and Gaza - Public expenditure review: From crisis to greater fiscal independence."

——— (2018) "The Human Capital Project.," Available at https://openknowledge.worldbank.org/handle/10986/30498.

Zertal, Idith and Akiva Eldar (2009) *Lords of the land: The war over Israel's settlements in the occupied territories, 1967-2007*: Hachette UK.

Zhang, Dennis J, Itai Gurvich, Jan A Van Mieghem, Eric Park, Robert S Young, and Mark V Williams (2016) "Hospital readmissions reduction program: An economic and operational analysis," *Management Science*, Vol. 62, No. 11, pp. 3351–3371.

Ziedan, Engy (2018) "The Intended and Unintended Consequences of theHospital Readmission Reduction Program," *Working paper*.

Zuckerman, Rachael B, Steven H Sheingold, E John Orav, Joel Ruhter, and Arnold M Epstein (2016) "Readmissions, observation, and the hospital readmissions reduction program," *New England Journal of Medicine*, Vol. 374, No. 16, pp. 1543–1551.

Appendix

A | Appendix to Chapter 1

A.1 Background Appendix

CMS Risk-adjustment Methodology

When calculating HRRP penalties, CMS adjusts the raw readmission rate for differences in hospital case mix. The goal of the risk-adjustment is to account for variation in readmission rates due to observable patient risk factors. More formally, let y_{ih} be a binary outcome variable of patient i treated at hospital h and x_{ih} a vector of patient specific characteristics. The risk adjustment model is:

$$logit(\mathbb{E}[y_{ih}|\alpha_h, \beta, x_i]) = logit(y_{ih}) = \alpha_h + x'_{ih}\beta \tag{24}$$

where α_h is a hospital-specific intercept and $[\alpha_h|\mu, \sigma^2] \sim N(\mu, \sigma^2)$. α_h is estimated for each hospital h using empirical Bayesian methods.

σ^2 is the between-hospital variation in outcomes after accounting for patient risk factors at admission. If there are no between-hospital effects, then $\sigma^2 = 0$ and $\alpha_h = \mu$ for all hospitals h. In this case, any observed differences in the unadjusted hospital outcomes would be due to differences in case-mix and hospital-specific factors do not account for these differences at all.

The estimated predicted readmission rate for hospital h, diagnosis d and insurance group j is:

$$\hat{r}_{dhj} = \frac{1}{N_{dhj}} \sum_{i=1}^{N_{dhj}} logit(\hat{\alpha}_h + x'_{ih}\hat{\beta}) \tag{25}$$

where N_{dhj} is the total number of patients in hospital h diagnosed with d and insured by j.

The estimated expected readmission rate is what CMS expects the readmission rate to be if the hospital's specific intercept is equal to the national mean (i.e. $\alpha_h = 0$):

$$\tilde{r}_{dhj} = \frac{1}{N_{dhj}} \sum_{i=1}^{N_{dhj}} logit(x'_{ih}\hat{\beta}) \tag{26}$$

The excess risk-adjusted readmission rate is then $\hat{r}_{dhj}/\tilde{r}_{dhj}$.

HRRP Penalty Formula

Penalties for all monitored diagnosis. To arrive at the total penalty rate, the CMS aggregates payments for excess readmissions across all monitored conditions and divides this by the Medicare revenue from all diseases. The readmission penalty ratio is capped at the maximum penalty level.

Let d denote a specific diagnosis and let $\mathcal{D}$ denote the set of all diagnosis groups and $\mathcal{D}^M \subset \mathcal{D}$ be the set of monitored conditions under the HRRP..[128] For example, in 2012, $\mathcal{D}^M = \{\text{AMI, HF, PN}\}$. Let $\Pi_{all} := \sum_{d \in \mathcal{D}} \Pi_{dj}$ denote total Medicare revenue of the hospital across all diagnoses domains. Thus, the total penalty rate for a hospital can be written as:

$$\tau = \min \left\{ \frac{\sum_{d \in \mathcal{D}^M} \max \left\{ \frac{r_{dj}}{\tilde{r}_{dj}} - 1, 0 \right\} \Pi_{dj}}{\Pi_{all}}, P^{cap} \right\} \tag{27}$$

where $j = \text{TM}$ and Π_{dj} is revenue from TM patients admitted for diagnosis d.

To see how equation (27) can be used to calculate the penalties for a hospital, suppose that a hospital receives \$2.4 million from Medicare overall. Of the total revenue, \$20,000 is from AMI (with an excess readmissions ratio of 1.1), \$50,000 is from HF (with an excess readmissions ratio of 1.2), \$30,000 is from PN (with an excess readmissions ratio of 0.7) and the remaining is from other (non-monitored) conditions. In this case, the hospital's penalty rate is $(0.1 \times 20,000 + 0.2 \times 50,000)/2.4\text{m} = 0.005$. The total penalty is applied to the entire Medicare revenue, which evaluates to $0.005 \times 2.4\text{m} = \$12,000$.

Penalties for single diagnosis. To move from penalties at the "global" level to penalties at the "local" diagnosis level, note that each monitored diagnoses' excess readmissions induces penalties on all diagnoses. Thus, the penalty rate induced by excess readmissions in diagnosis d is:

$$\tau_d = \min \left\{ \max \left\{ \frac{r_{dj}}{\tilde{r}_{dj}} - 1, 0 \right\} \frac{\Pi_d}{\Pi_{all}}, P_d^{cap} \right\}$$

where P_d^{cap} is the penalty cap assigned to diagnosis d by the hospital.

I follow Zhang et al. (2016) and assume that the hospital allocates the penalty proportionally to the

[128] Subscripts dh were omitted from the main text for simplicity with the understand that the unit of analysis is a hospital department specialising in a particular diagnosis.

revenue contribution of diagnosis d relative to the revenue from all the monitored conditions:

$$P_d^{cap} = P^{cap} \frac{\Pi_d}{\sum_{d \in \mathcal{D}^M} \Pi_d}$$

Then, the penalty rate can be written as:

$$\tau_d = \min \left\{ \max \left\{ \frac{r_{dj}}{\tilde{r}_{dj}} - 1, 0 \right\}, \frac{P^{cap}}{\frac{\sum_{d \in \mathcal{D}^M} \Pi_d}{\Pi_{all}}} \right\}$$

To arrive at the dollar penalty amount, the penalty rate is multiplied by the total revenue from that diagnosis: $\Pi_d \times \tau_d$

A.2 Model Appendix

Prediction 1 (Direct impact of penalties on TM patients). *Hospital departments will respond to the penalties by increasing quality for TM patients admitted for monitored diagnoses if the hospital expects that its pre-policy readmission rate is greater than the national average risk-adjusted readmission rate $\tilde{r}_m$. There will be no change otherwise.*

Proof. Let subscripts $\{0, 1\}$ denote pre- and post-periods respectively.

(1) Suppose that $r_{m0} < \tilde{r}$, then the hospital doesn't expect to face any penalties and the maximand is the same as in the pre-policy period. Assuming that q_{m0} was set optimally, there will be no change in quality.

(2) Suppose that $r_{m0} \in [\tilde{r}, r^{cap}]$ where r^{cap} is the readmission rate such that the maximum penalty rate is reached, then $\tau_m = \frac{r_m}{\tilde{r}} - 1$. The first-order condition for q_m in the post-policy period evaluated at $q_m = q_{m0}$ becomes:

$$\underbrace{\frac{\partial V}{\partial q_m} - (p_m^r - c_m^r)x_m - \frac{1}{\kappa}\frac{\partial \phi}{\partial q_m}}_{=0 \text{ evaluated at } q_m = q_{m0}} + \underbrace{(p_m^r - c_m^r)\left(\frac{r_m}{\tilde{r}} - 1\right)x_m + \frac{1}{\tilde{r}}\Big[(p_m - c_m)x_m + (p_m^r - c_m^r)r_m x_m\Big]}_{>0 \text{ since at } q_m = q_{m0} \text{ implies } r_m > \tilde{r}} > 0$$

Therefore, the hospital will increase quality.

(3) Suppose that $r_{m0} > r^{cap}$, then $\tau_m = \frac{P^{cap}}{\psi_m}$ where $\psi_m := \frac{\sum_{d \in \mathcal{D}^M} \pi_d}{\pi_{all}}$. Thus, evaluated at $q_m = q_{m0}$:

$$\underbrace{\frac{\partial V}{\partial q_m} - (p_m^r - c_m^r)x_m - \frac{1}{\kappa}\frac{\partial \phi}{\partial q_m}}_{=0 \text{ evaluated at } q_m = q_{m0}} + \underbrace{(p_m^r - c_m^r)\left(\frac{P^{cap}}{\psi_m}\right)x_m}_{>0}$$

Intuitively, although increasing quality to $r_{m1} > r^{cap}$ does not reduce the penalties, the marginal loss in revenues from fewer readmissions is lower due to the total payments being reduced by the penalty program. □

Prediction 2 (Indirect impact of penalties on non-TM patients). *(A.) If the penalties had no effect on TM patients, then the hospital department will not alter quality for non-TM patients. (B.) If the penalties reduced readmissions for TM patients, then the impact of the penalties on readmission rates among non-*

TM patients admitted for monitored diagnoses depends on the sign of ϕ''_{mn}. In particular, if q_m and q_n are complements ($\phi''_{mn} < 0$), then $\frac{\partial q_n}{\partial \tau_m} > 0$. If q_m and q_n are substitutes ($\phi''_{mn} > 0$), then $\frac{\partial q_n}{\partial \tau_m} < 0$. If there are no dependencies between q_m and q_n ($\phi''_{mn} = 0$), then $\frac{\partial q_n}{\partial \tau_m} = 0$.

Proof. **(1)** Suppose that the policy had no impact on TM patients. This means that in the post-policy period, among TM patients, the marginal cost of decreasing readmissions outweighs the marginal benefit of doing so. Since the policy is imposed on TM patients only, this means the cost of decreasing readmissions among non-TM patients is weakly greater than the net costs of among TM patients. Hence, if the cost of decreasing readmissions among TM patients outweighs the benefits of doing so, this must also be true for non-TM patients.

(2) Suppose that the policy had an impact on TM patients. From equation (8):

$$\frac{\partial q_n}{\partial \tau_m} = -\underbrace{\frac{\partial^2 \phi}{\partial q_m \partial q_{dhn}}}_{=:\phi''_{mn}} - \underbrace{\left[\frac{\partial^2 \phi}{\partial q_n^2} - \frac{\partial^2 V}{\partial q_n^2}\right]^{-1}}_{>0} \times \frac{\partial q_m}{\partial \tau_m}$$

Assuming that the second order condition for q_m is satisfied, the term in square brackets is positive. Hence, $\partial q_n/\partial \tau_m > 0$ if and only if $\phi''_{mm} < 0$ and $\partial q_m/\partial \tau_m < 0$, $\partial q_n/\partial \tau_m > 0$ if and only if $\phi''_{mn} > 0$ and $\partial q_m/\tau_m > 0$, and $\partial q_n/\partial \tau_m = 0$ if and only if $\phi''_{mn} = 0$ and/or $\partial q_m/\tau_m = 0$. □

A.3 Data Appendix

Variable Definitions

Index Admissions Index admissions do not require a prior "clean period" with no hospitalisations: a hospital stay may be a readmission for a prior stay and the index admission for a subsequent readmissions. As such, I focus on index admissions from January 2008 to November 2014 (since I cannot observe readmissions in January 2015). An alternative approach is to require index stays to be hospitalisations with no preceding hospitalisation in the previous 30 days. This latter approach would require focusing on index admissions from February 2008 (since I cannot observe hospitalisations in December 2007) to November 2014. My empirical results are robust to both these methods of defining index admissions. Indeed, since multiple readmissions within 30 days do not incur additional penalties for the hospital, the definition of index admissions used in this chapter produces more conservative estimates of treatment effects.[129]

30-day Readmissions. The key outcome variable is a binary variable indicating whether the patient returns to a hospital within 30 days. I identify readmissions within 30 days after discharge using the CMS's hospital-wide readmission measure. Readmissions are repeated hospitalisations occurring within 30 days of the last discharge for any diagnosis ("all-cause" readmissions), where the main purpose of the hospitalisation is not rehabilitation or after-care.[130] Readmissions may or may not be for the same diagnosis as the index admission.

One challenge with the discharge claim data is that some hospital stays overlap or are adjacent to each other. This can complicate the construction of readmission rates because the latter stay may be mistaken as a readmission when in fact it is part of the same hospitalisation. In my construction of readmission rates, same-day and consecutive hospital stays are not considered as readmissions. Adjacent discharge claims can arise because of transfers within hospital units or across hospitals. If a patient was transferred to a different hospital on the same day or was transferred within the same hospital, I combine the two events into a single stay and do not count the second event as a readmission. Overlapping discharge claims can arise if the hospital at which the patient had their first stay submits the billing form after the patient has moved to another hospital. In these instances, I treat this as one stay. The discharge date of the combined stay is set

[129] Under this definition, some index admissions are readmissions. Since hospitals have little incentive to reduce second-readmissions to these types of index stays, including readmissions as index stays will provide a lower bound of the estimated treatment effect.

[130] These are (re)admissions where the main diagnosis code is 254 using HCUP's Clincal Classification Software.

to the discharge date of the latter overlapping stay while the admission date is set to the admission date of the former overlapping stay. Total charges for the combined stay is the sum of charges for both stays. If the patient was treated at two different hospitals across these two overlapping stays, then the hospital ID for the combined stay is set to the that of the hospital at which the patient had the longer stay.

Post-policy date. Since details of the policy was announced in August 2011, I set the post-policy date to September 2011. This follows the convention of most studies analyzing the impact of the HRRP with the exception of Zuckerman et al. (2016) who use April 2010 as the post-program date. The estimated impacts of the program are robust to both definitions of the post-policy date.

Sample selection

The data contains detailed information on the expected payer for the hospital stay. This is the main variable I use to categorise patients into TM and non-TM groups. As discussed in Section 2, the two non-TM insurance groups I focus on are Medicare HMO/Medicare beneficiaries and privately-insured patients. The main remaining payer groups excluded from the non-TM group are self-insured patients and Medicaid beneficiaries.[131] I exclude Medicaid beneficiaries because this group of patients tend to be of a different socio-economic background than privately-insured patients and have worse health outcomes due to the negative health-income gradient. Robustness checks in Section 7 show that the main results are not sensitive to the exclusion of this group.[132] There are two smaller omitted categories of insurance types: hospital stays with no charge and hospitals stays covered by some "other" insurance policy. The "no charge" hospital stays include hospitalisations for medically indigent patients and hospitalisations covered by charity or research purposes[133] The "other" insurance category includes workers' compensation, other government health insurance programs, and CHAMPUS/VA coverage (coverage for Veterans, their spouses or children). In my sample, roughly 0.07% of stays incur no charge while 0.16% of patients have CHAMPUS/VA coverage.

Summary of sample construction methodology

The procedure I use to construct the main sample is:

[131] Medicaid is a joint federal-state program that provides health coverage or nursing home coverage to certain categories of low-asset people, including children, pregnant women, parents of eligible children, people with disabilities and elderly needing nursing home care.

[132] Indirect effects are larger when the Medicaid sample is included in the group of non-TM patients.

[133] I exclude these hospitalisations because it is unclear how the structure for payment of medical services would affect the hospital's incentives.

1. For each monitored diagnoses and comparison diagnosis, create a dataset consisting of patients admitted for those diagnoses.

2. Restrict the sample to patients (i) aged between 40 and 90 years old when admitted and (ii) insured by Traditional Medicare, Medicare Advantage, or private insurance, (iii) who did not leave against medical advice or die in hospital during their first admission, (iv) has a valid patient ID and hospital ID.

3. Further restrict sample to patients treated at hospitals affected by the HRRP and that serve at least 25 patients diagnosed with the monitored diagnosis before and after the program is implemented.

4. Match the patients in each monitored diagnoses dataset to a patient in the non-monitored diagnosis dataset.

5. Restrict the sample to matched pairs.

ICD9 Codes for monitored and non-monitored conditions

Table A.1: ICD9-CM/DXCCS codes for monitored and non-monitored diagnoses

Diagnosis group	ICD-9 Codes/DXCCS Codes
Heart attack (Monitored)	ICD-9: 41000 – 41091, excluding codes ending with "2" (for after care)
Heart failure (Monitored)	ICD-9: 40201, 40211, 40291, 40401, 40403, 40411, 40413, 40491, 40493, 4280 – 4289
Pneumonia (Monitored)	ICD-9: 4800 – 4870 and 48811
Gastrointestinal (Non-monitored)	DXCCS: 135 – 155

Notes. This table consists of the ICD-9 codes used to identify the monitored and non-monitored diagnoses. The CMS uses the same codes to identify the monitored conditions.

Heart attacks, referred to clinically as acute myocardial infarction (AMI), occur when part of the heart's blood supply is blocked, starving the heart of oxygen and causing muscle cells to die. Heart attacks are an emergency condition and require immediate hospitalisation. Heart failure (HF) occurs when the heart muscle cannot pump blood around the body properly. Heart failure usually develops gradually and is a chronic condition. Heart attacks can lead to heart failure by weakening the heart's ability to pump blood around the body. Lastly, pneumonia (PN) is an inflammatory

condition of the lung primarily affecting the small air-sacs known as alveoli. It is usually caused by viral- or bacterial-infections and less commonly by other micro-organisms.

A.4 Additional Tables

Table A.2: Mean characteristics and outcomes of matched and unmatched patients

	Heart attack		Heart failure		Pneumonia	
	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched
	(1)	(2)	(3)	(4)	(5)	(6)
A. TM Patients						
Age	76.30	72.31	77.90	75.19	76.76	73.13
Female (%)	48.88	35.43	54.48	40.05	54.00	43.11
White (%)	76.87	47.19	71.19	51.11	77.70	67.08
CM Index	14.67	18.47	18.89	26.59	21.37	34.46
30-day readmission (%)	19.42	21.83	23.08	30.52	19.86	24.12
Observations	127,268	4,490	307,226	22,419	216,270	17,078
B. Non-TM Patients						
Age	64.69	68.67	72.10	73.82	67.55	71.54
Female (%)	34.88	30.01	47.21	37.84	51.72	43.50
White (%)	71.42	53.27	57.11	48.91	69.20	64.16
CM Index	9.08	14.38	16.68	24.94	16.84	30.62
30-day readmission (%)	13.17	17.45	19.85	27.38	14.84	20.78
Observations	140,492	7,822	180,668	22,342	133,896	14,089

Notes. This table shows means of observable characteristics and outcomes for matched and unmatched TM patients (Panel A) and non-TM patients (Panel B) admitted for the three monitored conditions.

Table A.3: Baseline regression with insurance-group specific linear time trends

	Heart attack		Heart failure		Pneumonia	
	Pre-mean	Estimate	Pre-mean	Estimate	Pre-mean	Estimate
	(1)	(2)	(3)	(4)	(5)	(6)
A. Prediction 1: Impact on TM patients						
(1) Below average	19.56	-0.0037	23.09	0.0010	17.23	0.0038
		(0.006)		(0.007)		(0.006)
(2) Above average	24.12	-0.0336***	28.28	-0.0072**	21.21	-0.0094***
		(0.006)		(0.004)		(0.004)
B. Prediction 2: Impact on non-TM patients						
(3) Below average	12.91	-0.0073	21.58	-0.0058	12.08	0.0049
		(0.007)		(0.010)		(0.006)
(4) Above average	16.08	-0.0238***	24.08	-0.0098**	14.41	0.0036
		(0.005)		(0.004)		(0.006)
# Patients	267,760 (47.53% TM)		487,894 (62.97% TM)		350,166 (61.76% TM)	
# Hospitals	121 (73.55% Above)		140 (66.43% Above)		139 (62.59% Above)	

Notes. Same as Table 1.3 except that insurance-group specific linear time trends are included. There are three timetrends (one for TM, Medicare advantage, and private insurance). Standard errors in parentheses, clustered at the hospital level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Impact of program on 30-day readmissions using alternative comparison group

	Heart attack		Heart failure		Pneumonia	
	Pre-mean	Estimate	Pre-mean	Estimate	Pre-mean	Estimate
	(1)	(2)	(3)	(4)	(5)	(6)
A. Prediction 1: Impact on TM patients						
(1) Below average	19.45	-0.0069 (0.007)	23.66	0.0093 (0.005)	17.68	0.0127 (0.005)
(2) Above average	24.10	-0.0301*** (0.006)	28.72	-0.0088** (0.004)	21.76	-0.0060* (0.004)
B. Prediction 2: Impact on non-TM patients						
(3) Below average	13.33	-0.0057 (0.006)	22.42	-0.0054 (0.006)	12.94	0.0212*** (0.006)
(4) Above average	16.26	-0.0184*** (0.005)	24.94	-0.0044 (0.004)	15.48	0.0001 (0.006)
# Patients	289,770 (46.73% TM)		566,820 (61.45% TM)		404,382 (61.09% TM)	
# Hospitals	122 (72.95% TM)		141 (66.67% TM)		141 (62.41% TM)	

Notes. Same as Table 1.3 except that the comparison group from which matches are selected consists of all non-monitored diagnosis excluding admissions for childbirth. The match rates are: 99.38% and 98.87% for AMI TM and non-TM patients respectively; 98.93% and 96.96% for HF TM and non-TM patients respectively; and 98.64% and 97.09% for PN TM and non-TM patients respectively. See Section 4 for details on matching procedure. Standard errors in parentheses, clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Impact of program on 30-day readmissions using alternative post-policy date

	Heart attack		Heart failure		Pneumonia	
	Pre-mean	Estimate	Pre-mean	Estimate	Pre-mean	Estimate
	(1)	(2)	(3)	(4)	(5)	(6)
A. Prediction 1: Impact on TM patients						
(1) Below average	19.67	-0.0034	23.08	-0.0013	16.82	0.0046
		(0.010)		(0.007)		(0.006)
(2) Above average	24.62	-0.0281***	28.64	-0.0075**	21.67	-0.0106***
		(0.005)		(0.003)		(0.004)
B. Prediction 2: Impact on non-TM patients						
(3) Below average	13.42	-0.0121	21.54	-0.0017	11.62	0.0055
		(0.011)		(0.009)		(0.007)
(4) Above average	16.53	-0.0184***	24.37	-0.0107***	13.97	0.0104**
		(0.005)		(0.003)		(0.005)
# Patients	267,760 (47.53% TM)		487,894 (62.97% TM)		350,166 (61.76% TM)	
# Hospitals	121 (73.55% Above)		140 (66.43% Above)		139 (62.59% Above)	

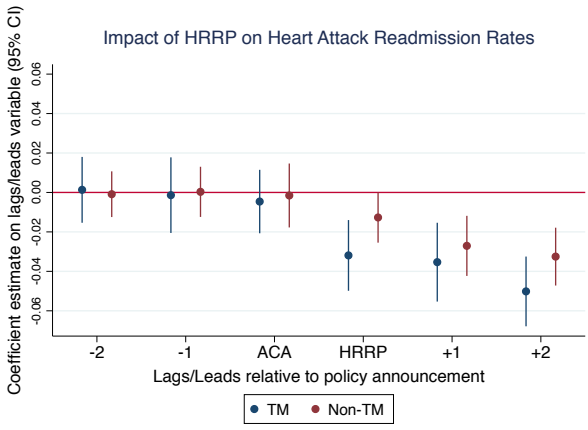
Notes. Same as Table 1.3 except that the post-policy date is now March 2010 (when the ACA is implemented) rather than August 2011 (when the final HRRP rules are announced). Standard errors in parentheses, clustered at the hospital level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Impact of program on 30-day readmissions using CMS approach to classify hospital into "above average" and "below average" categories

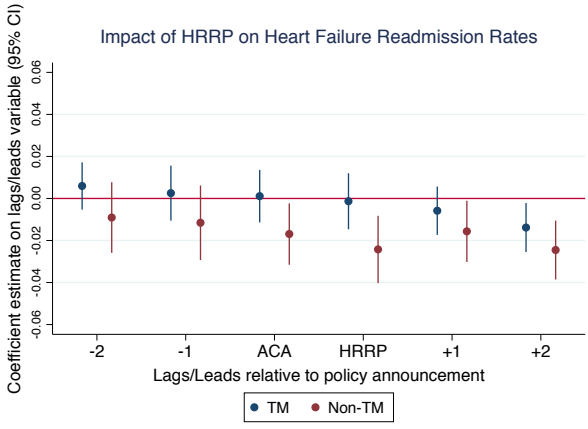
	Heart attack		Heart failure		Pneumonia	
	Pre-mean	Estimate	Pre-mean	Estimate	Pre-mean	Estimate
	(1)	(2)	(3)	(4)	(5)	(6)
A. Prediction 1: Impact on TM patients						
(1) Below average	19.42	-0.0211***	23.06	0.0028	16.71	0.0057
		(0.007)		(0.006)		(0.007)
(2) Above average	23.58	-0.0264***	28.25	-0.0075**	20.64	-0.0076**
		(0.006)		(0.004)		(0.004)
B. Prediction 2: Impact on non-TM patients						
(3) Below average	13.26	-0.0222**	22.14	-0.0156	12.19	0.0058
		(0.009)		(0.010)		(0.008)
(4) Above average	15.67	-0.0185***	23.99	-0.0072*	14.00	0.0037
		(0.005)		(0.004)		(0.005)
# Patients	267,760 (47.53% TM)		487,894 (62.97% TM)		350,166 (61.76% TM)	
# Hospitals	121 (73.55% Above)		140 (66.43% Above)		139 (62.59% Above)	

Notes. Same as Table 1.3 except that rather than using the observed readmission rate between July 2008 and March 2010 to determine whether the hospital will be affected by the HRRP, the ratio between risk-adjusted predicted readmission rates and risk-adjusted expected readmission rates (the CMS approach) is used. Standard errors in parentheses, clustered at the hospital level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

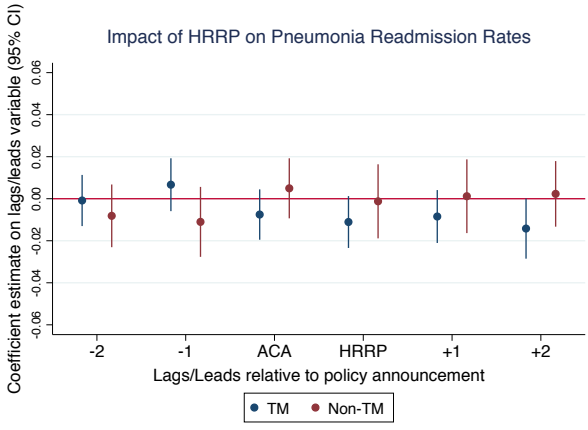
A.5 Additional Figures



(a) Heart attack

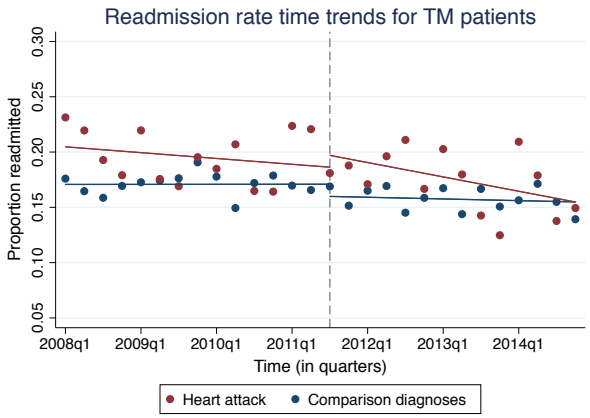


(b) Heart failure

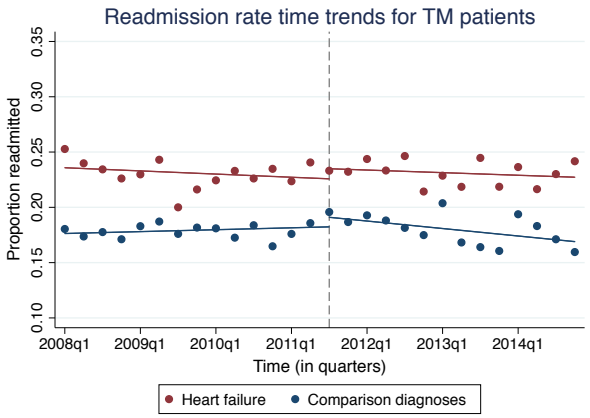


(c) Pneumonia

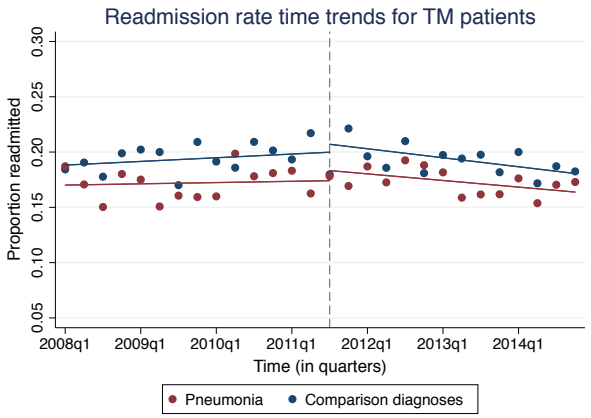
Figure A.1: This figure presents coefficients on the year dummies in a modified version of equation (9).



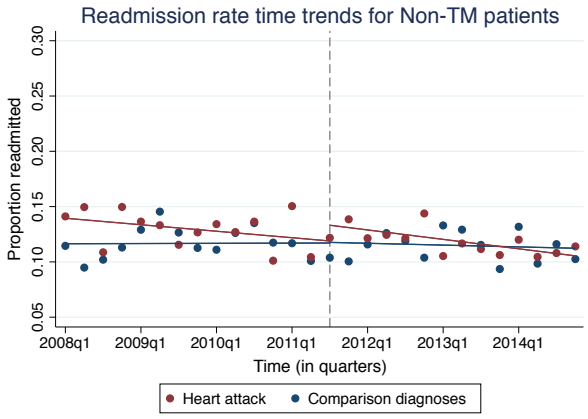
(a) Heart attack TM Patients



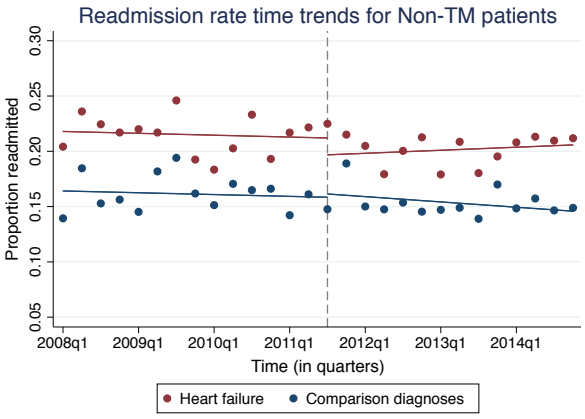
(b) Heart failure TM Patients



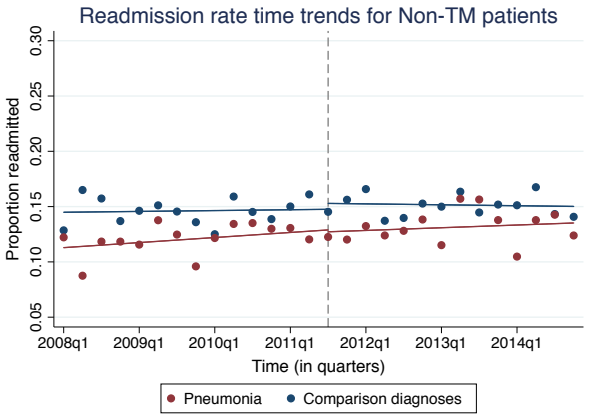
(c) Pneumonia TM Patients



(d) Heart attack non-TM Patients



(e) Heart failure non-TM Patients



(f) Pneumonia non-TM Patients

Figure A.2: This figure presents readmission trends for each monitored diagnosis and its corresponding comparison group, by TM and non-TM patient groups, for hospitals below the national average. The non-monitored comparison group consists of patients admitted for gastrointestinal conditions and patient matches are selected using the matching procedure described in Section 4.

B | Appendix to Chapter 2

B.1 Model Appendix

Prediction 1. *For a given level of illness severity μ_i, an insured patient is more likely to return to the ED than an uninsured patient.*

The patient returns if $v(c_i^{care}) + f(\mu_i) \geq v(c_i^{\neg care})$, which is equivalent to $f(\mu_i) \geq v(c_i^{\neg care}) - v(c_i^{care})$. For an insured patient, the RHS equals $v(I_i - p(I_i)) - v(I_i - \beta m_i - p(I_i))$. For an uninsured patient, the RHS equals $v(I_i) - v(I_i - m_i)$. Suppose $f(\mu_i) \geq v(I_i) - v(I_i - m_i)$ so that the uninsured returns to the ED. Since $v'(\cdot) > 0$ and $0 < \beta < 1$, this implies $f(\mu_i) \geq v(I_i) - v(I_i - m_i) > v(I_i - p(I_i)) - v(I_i - \beta m_i - p(I_i))$ and so the insured will also return. However, the converse is not true for $0 < \beta < 1$. This holds for any sickness level μ_i since the impact of experienced illness (and natural health endowments) on the decision to return to the ED is "differenced" out in the utility comparisons. Therefore, an insured patient is more likely to return to the ED regardless of their illness severity.

Prediction 2. *Assume that returns to care are increasing in illness severity ($f'(\mu_i) > 0$) and that illness severity is decreasing in quality of care received ($\mu_i'(q) < 0$). $\partial E/\partial(1-\phi)$ captures different objects depending on hospital behaviour. There are three possible cases:*

- $\frac{\partial q}{\partial(1-\phi)} = 0$*: In this case, out-of-pockets costs do not influence quality of care and $\partial E/\partial(1-\phi)$ only captures the patient incentive effect.*
- $\frac{\partial q}{\partial(1-\phi)} < 0$*: In this case, patients who pay more out of pocket receive better quality of care. $\partial E/\partial(1-\phi)$ will necessarily be positive and it will be a combination of the quality effect and patient incentive effect.*
- $\frac{\partial q}{\partial(1-\phi)} > 0$*: In this case, patients who pay less out of pocket receive better quality of care. The sign of $\partial E/\partial(1-\phi)$ is ambiguous and will depend on whether the patient incentive effect outweighs the quality effect.*

If $\frac{\partial q}{\partial(1-\phi)} = 0$, then from equation (14), $\frac{\partial E}{\partial(1-\phi)} = 0 + \underbrace{\frac{\partial v(c)}{\partial c}\frac{\partial c}{\partial(1-\phi)}}_{+\text{ve}}$.

If $\frac{\partial q}{\partial(1-\phi)} < 0$, then from equation (14), $\frac{\partial E}{\partial(1-\phi)} = \underbrace{\frac{\partial f}{\partial \mu}\frac{\partial \mu}{\partial q}\frac{\partial q}{\partial(1-\phi)}}_{+\text{ve}} + \underbrace{\frac{\partial v(c)}{\partial c}\frac{\partial c}{\partial(1-\phi)}}_{+\text{ve}}$.

If $\frac{\partial q}{\partial(1-\phi)} > 0$, then from equation (14), $\frac{\partial E}{\partial(1-\phi)} = \underbrace{\frac{\partial f}{\partial \mu}\frac{\partial \mu}{\partial q}\frac{\partial q}{\partial(1-\phi)}}_{\text{-ve}} + \underbrace{\frac{\partial v(c)}{\partial c}\frac{\partial c}{\partial(1-\phi)}}_{+\text{ve}}$.

B.2 Background Appendix

Medical diagnosis codes

Table B.1: Classification of heart attack codes

Diseases	ICD-9 Codes
Heart attack	41000 – 41091, excluding codes ending with "2" (for after care)

Notes. This table contains the ICD-9 codes to identify heart attack patients. These codes are based on the Center for Medicare and Medicaid Services methodology to identify heart attack admissions.

Treatment and diagnosis procedure codes

Table B.2: Classification of procedure codes

Procedure	ICD-9 Codes
Cardiac Catheterization	3722 – 3723
Angioplasty	0066 or 3601 – 3609
Coronary artery bypass surgery	Codes beginning with "361"

Notes. This table contains the ICD-9 codes to identify heart attack patients. These codes are based on the Center for Medicare and Medicaid Services methodology to identify heart attack admissions.

The main heart attack treatment and diagnostic procedures discussed in this chapter are cardiac catheterisation, angioplasty, and coronary bypass. Here, I provide some further information on these procedures.

- **Cardiac catheterisation**: Catheterisation is a medical procedure used to diagnose and treat some heart conditions. A long, thin, flexible tube called a catheter is put into a blood vessel in the arm, upper thigh, or neck and threaded into the heart to perform diagnostic tests. This procedure is performed while the patient is awake. Patients are expected to stay at the hospital overnight after undergoing catheterisation and to limit movement to avoid bleeding from the site where the catheter was inserted.

- **Angioplasty**: An angioplasty is the most common heart attack surgical treatment. In the sample, over 50% of heart attack patients receive an angioplasty.[134] During balloon angioplasty, a balloon-tipped catheter pushes plaque back against the arterial wall to allow for improved blood flow in the artery. Angioplasties usually take 1 to 2 hours to complete and is done with local anaesthesia on patients who are mildly sedated. Most patients will stay overnight in the hospital for observation, will be discharged the following morning, and can resume normal activities within a week.

- **Coronary bypass**: A coronary artery bypass graft (CABG) is a surgical procedure used to treat coronary heart disease. The aim of coronary bypass surgery is to bypass the narrowed sections of your coronary arteries. This is achieved by grafting a blood vessel between the aorta (the main blood vessel leaving the heart) and a point along the coronary artery, past the narrowed area. During this procedure, surgeons usually make a cut down the middle of the breastbone. Patients usually stay in hospital for a week after surgery.

Categorisation of reason for readmission

Table B.3: Classification and coding of reasons for readmissions

Reason for readmission	Diagnosis at readmission	Medical codes
(1)	(2)	(3)
Major vascular events	AMI, Stroke, unstable angina, heart failure	CCS 102, CCS 106, CCS 101 excluding those w/ ICD-9 code 4111
Non-major vascular events	Non-specific chest pain, dysrhythmias, atherosclerosis	CCS 100, CCS 108, ICD-9 43301 to 4350, excluding those ending with 0, ICD-9 4111
Treatment complications	Device complications/surgical complications	CCS 237, 238

Notes. This table presents the categorization of reasons for readmissions used in the analysis Column (1) describes the category used in the analysis. Column (2) describes the types of diagnoses or procedures that fall under those categories. Column (3) lists the medical codes used to identify those diagnoses or procedures. "CCS" refers to the Clinical Classifications Software method of classifying ICD-9-CM codes.

[134] Nearly one million balloon angioplasty procedures are performed each year in the United States alone. See: http://circ.ahajournals.org/content/106/23/e187.

B.3 Empirical Strategy Appendix

This section discusses the estimation procedure used in this chapter. Further technical details can be found in Calonico et al. (2014), Calonico et al. (2015), and Calonico et al. (Forthcoming).

Estimand and Estimator. For each patient i, let the scalar random variable Y_i denote the outcome of interest (e.g. 30-day readmission probability) and the scalar regressor Z_i denote their age at admission relative to the 65 years cut-off. Let the superscripts 0 and 1 denote potential outcomes. In the reduced-form model, the estimand of interest is:

$$\tau = \mathbb{E}[Y_i^1|Z=0] - \mathbb{E}[Y_i^0|Z=0] \tag{28}$$

The corresponding estimator is:

$$\hat{\tau}(w_n) = \hat{\lambda}_{Y+}(w_n) - \hat{\lambda}_{Y-}(w_n) \tag{29}$$

where w_n is some positive bandwidth and where $\hat{\lambda}_{Y+}$ and $\hat{\lambda}_{Y-}$ are the intercepts at the threshold of a weighted (local-linear) regression for dependent variable Y from above and below the threshold respectively. The weights are based on the kernel choice; this chapter uses the triangular kernel. More generally, the regression function can be a pth-order polynomial.

Let $\mathbf{Z}_n = (Z_1, \ldots, Z_n)'$ and p the of the polynomial regression function. Under appropriate regularity conditions, the mean squared-error of the regression can be expanded as follows:

$$MSE(w_n) = \mathbb{E}\Big[\{\hat{\tau}(w_n) - \tau\}^2|\mathbf{Z}_n\Big] \approx w_n^{2(p+1)} B_{n,p}^2 + \frac{1}{nw_n} V_{n,p} \tag{30}$$

where $B_{n,p}$ is the leading asymptotic bias of $\hat{\tau}(w_n)$ and $V_{n,p}$ is the asymptotic variance of $\hat{\tau}(w_n)$. The reduced-form estimator will be consistent if $w_n \to 0$ and $nw_n \to \infty$.

Robust inference. The conventional approach to constructing asymptotically valid confidence intervals (CI) based on $\hat{\tau}(w_n)$ assumes away the bias of the estimator and constructs $100(1-\alpha)$ percent CIs of the form:

$$CI_{1-\alpha,n} = \left\{\hat{\tau}(w_n) \pm \mathbf{\Phi}^{-1}_{1-\frac{\alpha}{2}}\sqrt{\hat{v}_n}\right\} \tag{31}$$

where α is the significance level, $\mathbf{\Phi}^{-1}_{\alpha}$ is the appropriate quantile of the normal distribution, and $\hat{v}_n$ is an appropriate choice of the variance estimator.

The robust p-values in this chapter are the counterparts of CIs based on bias-corrected local polynomials. There are two steps to constructing these CIs: (i.) bias-correct the RD estimator and (ii.) rescale the bias-correct t-statistic using a new variance formula. The bias-corrected estimator uses higher-order local polynomials to estimate the unknown derivatives in the leading bias term. Let q be the order of the polynomial used to estimate the bias such that $q \geq p+1$. The resulting bias-corrected estimator is:

$$\hat{\tau}^{bc}_{p,q} = \hat{\tau}_p(w_n) - w_n^{p+1}\hat{B}_{n,p,q} \tag{32}$$

where the formula for $\hat{B}_{n,p,q}$ can be found in Calonico et al. (2014). The bias-estimate is introduced in order to remove the potentially large effects of the unknown leading bias of the RD estimator.[135]

The new variance formula uses an alternative asymptotic approximation to the bias-corrected RD estimator (Calonico et al., 2014). This formula accounts for the additional variability introduced by estimating the bias. Combining the bias-corrected estimate and the new variance formula $\hat{v}^r_n$, these robust CIs take the form:

$$CI^r_{1-\alpha,n} = \left[\left\{\hat{\tau}_p(w_n) - w_n^{p+1}\hat{B}_{n,p,q}\right\} \pm \mathbf{\Phi}^{-1}_{1-\frac{\alpha}{2}}\sqrt{\hat{v}^r_n}\right\} \tag{33}$$

[135] In order to implement this bias-correction approach, one needs a pilot bandwidth b_n to estimate the bias. Calonico et al. (2014) propose an MSE-optimal choice of the pilot bandwidth:

$$b = \left\{\frac{(2p+3)V_{p+1,q}}{2(q-p)B^2_{p+1,q}}\right\}^{\frac{1}{(2q+3)}} n^{\frac{-1}{2q+3}}$$

Bandwidth selection. The point estimator $\hat{\tau}(w_n)$ will be optimal in an asymptotic MSE sense if the bandwidth w_n is chosen so that:

$$w_{MSE,n,p} = \left\{ \frac{V_p}{2(p+1)B_p^2} \right\}^{\frac{1}{2p+3}} w^{\frac{-1}{(2p+3)}} \tag{34}$$

The baseline results of this chapter uses the Calonico et al. (2014) method to estimate the MSE-optimal bandwidth of equation (34):

$$\hat{w}_{MSE,n,p} = \left\{ \frac{\hat{V}_p}{2(p+1)\hat{B}_p^2 + \hat{R}_p} \right\}^{\frac{1}{2p+3}} w^{\frac{-1}{(2p+3)}} \tag{35}$$

where $\hat{R}$ is a regularization term to avoid small denominators in moderate-size samples and where $\hat{V}_p$ and $\hat{B}_p$ are consistent estimators of their population counterparts.

I also test the sensitivity of the baseline results to different bandwidth selection methods. In particular, further results implement the coverage error rate (CER) optimal bandwidth choice. This bandwidth choice minimizes the CER of the biased-corrected confident interval and takes the form:

$$\hat{w}_{CER,n,p} = n^{-\frac{p}{(3+p)(3+2p)}} \cdot \hat{w}_{MSE,n,p} \tag{36}$$

B.4 Additional Tables

Table B.4: Estimated discontinuity in discharge outcomes at age 65

		Men		Women	
		Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
		(1)	(2)	(3)	(4)
(A)	**Non-routine outcome**	29.68	0.211	34.69	2.165
			(1.022)		(1.408)
		[w=9.56,N=32,329]		[w=8.48,N=15,343]	
(B)	**Black**	8.08	-1.064	12.84	1.978
			(0.984)		(1.691)
		[w=7.03,N=16,910]		[w=9.39,N=10,083]	
(C)	**Poor**	39.61	2.376	44.69	1.598
			(1.680)		(2.575)
		[w=12.53,N=28,374]		[w=8.85,N=9,620]	
(D)	**CM Index**	8.14	-0.419	9.73	0.366
			(0.437)		(0.632)
		[w=6.78,N=16,340]		[w=9.61,N=10,324]	

Notes. This table presents the estimated discontinuity in the probability of having a non-routine outcome that doesn't satisfy the sample selection criteria (dying in hospital, transfer to another hospital etc.), probability of being black, probability of being in the bottom quartile of the income distribution, and the Elixhauser co-morbidity index. In row (A), all heart attack patients admitted to New York State hospitals between 2010 and 2013 are included, regardless of discharge outcome. In rows (B) to (D), the sample is constructed in the manner as the main sample. All models include a linear function for age at admission (in months) fully interacted with a dummy for age over 65. Row (A) also includes race indicators, the Elixhauser co-morbidity index, and year-month dummies. Coefficients and standard errors in rows (A) to (C) have been multiplied by 100. Standard errors in parentheses. "w" denotes the optimal bandwidth and "N" denotes the corresponding sample size. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: Baseline results for men without clustered standard errors

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	11.00	2.571**	2.41	-0.394	4.04	2.219***	0.83	-0.045
		(0.977)		(0.444)		(0.621)		(0.295)
	[w=8.33,N=19,648]		[w=8.85,N=20,858]		[w=9.64,N=22,444]		[w=9.88,N=19,356]	
(B) **White**	10.17	1.423	2.12	-0.260	3.91	1.502**	0.91	-0.353
		(1.035)		(0.529)		(0.713)		(0.361)
	[w=10.62,N=16,042]		[w=8.72,N=13,451]		[w=10.18,N=15,486]		[w=7.50,N=10,150]	
(C) **Non-White**	12.64	4.477**	2.97	-0.864	4.29	3.441***	0.66	0.721
		(1.879)		(0.863)		(1.152)		(0.651)
	[w=7.31,N=5,968]		[w=7.60,N=6,243]		[w=9.46,N=7,645]		[w=7.50,N=5,068]	
(D) **Highest qtrl zipcode income**	10.64	2.617	2.08	0.300	3.89	1.614	0.60	0.297
		(1.718)		(0.826)		(1.060)		(0.612)
	[w=9.55,N=6,290]		[w=8.56,N=5,701]		[w=11.51,N=7,452]		[w=8.18,N=4,732]	
(E) **Lowest qtrl zipcode income**	12.32	2.853	3.12	-1.822*	3.45	2.344*	0.98	-0.101
		(1.902)		(1.205)		(1.203)		(0.768)
	[w=10.94,N=5,643]		[w=5.15,N=2,841]		[w=12.07,N=6,133]		[w=7.88,N=3,410]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns). Same description as Table 2.3 except standard errors are robust to heteroskedasticity but not clustered at the level of the running variable. Bandwidths ("w") are MSE-optimal and selected using the procedure of Calonico et al. (2014). "N" denotes the corresponding sample size. Kernel: Triangular. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Baseline results for men using quadratic polynomial

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	11.00	3.405***	2.41	-0.748	4.04	2.562***	0.83	0.095
		(1.284)		(0.557)		(0.753)		(0.340)
	[w=8.59,N=20,358]		[w=9.95,N=23,154]		[w=12.62,N=28,540]		[w=9.82,N=19,222]	
(B) **White**	10.17	1.988	2.12	-0.382	3.91	2.070**	0.91	-0.292
		(1.681)		(0.724)		(0.946)		(0.398)
	[w=9.67,N=14,785]		[w=9.58,N=14,556]		[w=11.77,N=17,606]		[w=9.20,N=12,048]	
(C) **Non-White**	12.64	4.814*	2.97	-1.482	4.29	3.658***	0.66	0.830
		(2.697)		(1.487)		(1.228)		(0.648)
	[w=11.35,N=9,052]		[w=9.82,N=7,899]		[w=12.80,N=10,030]		[w=11.48,N=7,485]	
(D) **Highest qtrl zipcode income**	10.64	2.899	2.08	0.309	3.89	1.848	0.60	0.874
		(2.114)		(1.096)		(1.651)		(0.673)
	[w=14.60,N=9,014]		[w=11.81,N=7,596]		[w=14.03,N=8,731]		[w=9.17,N=5,211]	
(E) **Lowest qtrl zipcode income**	12.32	6.334**	3.12	-2.440**	3.45	5.067***	0.98	-0.261
		(2.767)		(1.396)		(2.131)		(1.031)
	[w=8.22,N=4,304]		[w=8.28,N=4,351]		[w=9.03,N=4,718]		[w=8.62,N=3,721]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns). Same description as Table 2.3 models include a quadratic polynomial for age at admission (in months) fully interacted with a dummy for age over 65. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.7: Estimated discontinuity in 30-day readmissions at age 65 for male patients using different controls and samples

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covariate changes								
(A) **Month before 65** [N = 17298]	11.00	2.709*** (0.968)	2.41	-0.515 (0.418)	4.04	2.354*** (0.570)	0.83	-0.005 (0.245)
(B) **Zip-code income qtrl** [N = 16686]	11.00	2.664*** (1.007)	2.41	-0.410 (0.433)	4.04	2.332*** (0.604)	0.83	-0.166 (0.249)
(C) **Payer x Year dummies** [N = 17298]	11.00	1.943** (0.976)	2.41	-0.744* (0.423)	4.04	2.254*** (0.577)	0.83	-0.122 (0.245)
Sample changes								
(D) **Exclude Medicaid**	10.02	3.540*** (1.077)	2.29	-0.241 (0.444)	3.48	3.441*** (0.710)	0.91	-0.132 (0.309)
	[w=7.97,N=16,238]		[w=9.66,N=19,194]		[w=6.78,N=14,040]		[w=10.15,N=16,894]	
(E) **One type of AMI**	11.66	4.055*** (1.448)	2.88	-0.402 (0.680)	3.67	3.802*** (0.984)	1.03	0.023 (0.389)
	[w=6.49,N=9,811]		[w=7.38,N=11,134]		[w=6.45,N=9,811]		[w=9.96,N=11,803]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns). In columns 1–2, the dependent variable is an indicator for all-cause 30-day readmissions. In columns 3–4, the dependent variable is an indicator for 30-day readmissions where the primary diagnosis is a major vascular condition (heart attack, stroke, unstable angina, or heart failure). In columns 5–6, the dependent variable is an indicator for 30-day readmissions where the primary diagnosis is a non-major vascular condition (non-specific chest pain, dysrhythmias (abnormal heart beats), and coronary atherosclerosis/ other heart diseases). In columns 7–8, the dependent variable is an indicator for 30-day readmissions for surgical complications. For this outcome, each sample is limited to patients who received an angioplasty, catheterisation, and/or coronary artery bypass graft as their first or second procedure during their initial stay. The dependent variable is an indicator for a readmission within 30 days. Mean statistics in the even columns are calculated for patients who are 63 or 64 in that subgroup. All models include a linear function for age at admission (in months) fully interacted with a dummy for age over 65, race indicators, the Elixhauser co-morbidity index, and year-month dummies. Description of bandwidth, standard errors, kernel same as baseline specification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: Baseline results for women without clustered standard errors

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	14.42	-1.606	3.25	0.044	3.57	1.002	0.70	0.162
		(1.650)		(0.973)		(0.957)		(0.456)
	[w=9.94,N=10,665]		[w=8.14,N=8,894]		[w=9.71,N=10,418]		[w=10.94,N=9,264]	
(B) **White**	14.54	-2.974	2.55	0.421	3.44	0.861	0.81	-0.316
		(2.047)		(1.118)		(1.215)		(0.554)
	[w=10.20,N=6,653]		[w=8.34,N=5,573]		[w=9.06,N=5,958]		[w=8.00,N=4,310]	
(C) **Non-White**	14.23	0.432	4.39	-0.575	3.77	1.515	0.52	0.847
		(2.831)		(1.567)		(1.556)		(0.999)
	[w=8.94,N=3,786]		[w=9.82,N=4,096]		[w=10.44,N=4,344]		[w=8.15,N=2,740]	
(D) **Highest qtrl income**	14.04	-1.725	2.05	1.620	4.11	0.364	0.44	0.630
		(3.267)		(1.859)		(2.031)		(0.866)
	[w=10.35,N=2,630]		[w=6.87,N=1,834]		[w=8.90,N=2,307]		[w=6.92,N=1,495]	
(E) **Lowest qtrl income**	14.41	-3.237	3.53	0.242	4.12	-0.613	0.72	0.247
		(3.320)		(1.742)		(1.864)		(1.110)
	[w=8.01,N=2,408]		[w=8.99,N=2,657]		[w=7.83,N=2,333]		[w=8.66,N=1,987]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns). Same description as Table 2.4 except standard errors are robust to heteroskedasticity but not clustered at the level of the running variable. Bandwidths ("w") are MSE-optimal and selected using the procedure of Calonico et al. (2014). "N" denotes the corresponding sample size. Kernel: Triangular. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.9: Baseline results for women using quadratic polynomial

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	14.42	-1.867	3.25	-0.034	3.57	1.023	0.70	-0.249
		(2.359)		(1.166)		(1.200)		(0.598)
	[w=15.64,N=15,727]		[w=13.39,N=13,844]		[w=12.61,N=13,184]		[w=9.24,N=7,948]	
(B) **White**	14.54	-4.195	2.55	0.425	3.44	0.984	0.81	-1.128**
		(3.086)		(1.371)		(1.376)		(0.628)
	[w=11.11,N=7,172]		[w=12.87,N=8,190]		[w=11.05,N=7,135]		[w=7.52,N=4,103]	
(C) **Non-White**	14.23	0.306	4.39	-0.779	3.77	1.540	0.52	0.992
		(3.711)		(1.834)		(1.698)		(1.180)
	[w=12.85,N=5,215]		[w=14.38,N=5,708]		[w=13.27,N=5,343]		[w=11.49,N=3,661]	
(D) **Highest qtrl income**	14.04	-1.882	2.05	1.547	4.11	-0.796	0.44	0.337
		(4.078)		(1.966)		(2.716)		(0.978)
	[w=14.47,N=3,564]		[w=11.08,N=2,779]		[w=9.82,N=2,509]		[w=7.55,N=1,590]	
(E) **Lowest qtrl income**	14.41	-5.595	3.53	0.945	4.12	-1.811	0.72	-0.144
		(4.801)		(2.180)		(1.980)		(1.485)
	[w=10.21,N=2,994]		[w=11.97,N=3,426]		[w=8.87,N=2,633]		[w=9.24,N=2,110]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns). Same description as Table 2.4 except models include a quadratic polynomial for age at admission (in months) fully interacted with a dummy for age over 65. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.10: Estimated discontinuity in 30-day readmissions at age 65 for female AMI patients using different controls and samples

	All-cause		Major vascular		Non-major vascular		Surgical complication	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covariate changes								
(A) **Month before 65** [N = 11856]	14.42	-1.683 (1.596)	3.25	-0.033 (0.866)	3.57	0.864 (0.823)	0.70	0.114 (0.375)
(B) **Zip-code income qtrl** [N = 11314]	14.42	-1.653 (1.633)	3.25	-0.305 (0.841)	3.57	1.057 (0.884)	0.70	0.081 (0.385)
(C) **Payer x Year dummies** [N = 11856]	14.42	-2.748* (1.616)	3.25	-0.396 (0.884)	3.57	0.613 (0.837)	0.70	0.242 (0.377)
Sample changes								
(D) **Exclude Medicaid**	14.57	-2.516 (1.887)	2.50	0.756 (0.944)	3.95	0.676 (1.055)	0.65	0.098 (0.450)
	[w=11.70,N=10,442]		[w=7.23,N=6,798]		[w=10.08,N=9,111]		[w=10.15,N=7,343]	
(E) **One type of AMI**	14.10	-1.459 (2.108)	3.18	-1.203 (1.196)	3.50	1.519 (0.967)	0.68	0.129 (0.490)
	[w=10.81,N=8,395]		[w=6.99,N=5,601]		[w=11.86,N=9,123]		[w=9.89,N=5,947]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for various subgroups (rows) and for various types of readmissions (columns). In columns 1–2, the dependent variable is an indicator for all-cause 30-day readmissions. In columns 3–4, the dependent variable is an indicator for 30-day readmissions where the primary diagnosis is a major vascular condition (heart attack, stroke, unstable angina, or heart failure). In columns 5–6, the dependent variable is an indicator for 30-day readmissions where the primary diagnosis is a non-major vascular condition (non-specific chest pain, dysrhythmias (abnormal heart beats), and coronary atherosclerosis/ other heart diseases). In columns 7–8, the dependent variable is an indicator for 30-day readmissions for surgical complications. For this outcome, each sample is limited to patients who received an angioplasty, catheterisation, and/or coronary artery bypass graft as their first or second procedure during their initial stay. The dependent variable is an indicator for a readmission within 30 days. Mean statistics in the even columns are calculated for patients who are 63 or 64 in that subgroup. Description of bandwidth, standard errors, kernel same as baseline specification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.11: Results for random 50% of male sample

A. Impact on readmissions for random 50% of male sample

All-cause		Major vascular		Non-major vascular		Surgical complication	
Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
11.13	3.059**	3.03	-1.244*	0.04	1.879**	0.01	-0.159
	(1.772)		(0.760)		(0.891)		(0.424)
[w=5.38,N=6,519]		[w=7.64,N=9,100]		[w=8.64,N=10,178]		[w=6.40,N=6,484]	

B. Impact on quality for random 50% of male sample

Intensive treatment		Cathzn w/in 48 hrs		Length of stay		Num. procedures	
Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
78.88	1.032	19.23	-2.312	4.21	-2.263	7.01	0.637
	(1.931)		(1.681)		(2.181)		(2.276)
[w=9.61,N=11,239]		[w=9.08,N=10,590]		[w=9.81,N=11,392]		[w=7.73,N=9,187]	

Notes. This table presents the estimated discontinuity in 30-day readmission probabilities for a random 50% of the male sample. Same description as Table 2.3 except standard errors are robust to heteroskedasticity but not clustered at the level of the running variable. Bandwidths ("w") are MSE-optimal and selected using the procedure of Calonico et al. (2014). "N" denotes the corresponding sample size. Standard errors (in parentheses) are robust (following Calonico et al. (2014)) and clustered at age in months. Kernel: Triangular. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.12: Discontinuity in treatment intensity for men using quadratic specification

	Invasive treatment		Cathsn w/in 48 hrs		Length of stay		Num. procedures	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	79.92	0.233	19.30	-0.442	4.09	-1.204	6.92	-2.862
		(1.717)		(1.498)		(2.565)		(2.005)
	[w=12.45,N=28,207]		[w=10.01,N=23,322]		[w=12.14,N=27,589]		[w=12.65,N=28,540]	
(B) **White**	80.27	-1.294	18.73	-2.588	3.95	-3.261	6.98	-4.519*
		(2.197)		(1.939)		(3.477)		(2.390)
	[w=12.38,N=18,313]		[w=9.72,N=14,785]		[w=10.31,N=15,576]		[w=10.33,N=15,576]	
(C) **Non-White**	79.23	3.262	20.44	3.553	4.37	4.225	6.80	1.468
		(2.362)		(2.844)		(3.734)		(4.395)
	[w=9.17,N=7,452]		[w=10.12,N=8,161]		[w=8.63,N=7,020]		[w=8.73,N=7,077]	
(D) **Highest qtrl income**	82.62	-3.060	17.77	-0.248	3.91	0.109	7.15	-4.189
		(3.291)		(3.689)		(3.666)		(3.550)
	[w=10.95,N=7,140]		[w=11.05,N=7,187]		[w=8.57,N=5,701]		[w=11.16,N=7,246]	
(E) **Lowest qtrl income**	79.15	-1.657	19.38	4.107	4.38	-5.958	6.75	-2.824
		(3.410)		(3.620)		(4.990)		(4.879)
	[w=13.45,N=6,738]		[w=10.35,N=5,368]		[w=16.95,N=7,939]		[w=10.28,N=5,328]	

Notes. This table presents the estimated discontinuity in various measures of quality of care at the patient's initial admission for a heart attack. Other descriptions same as Table 2.5 except models include a quadratic polynomial for age at admission (in months) fully interacted with a dummy for age over 65. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.13: Discontinuity in treatment intensity for women using quadratic specification

	Invasive treatment		Cathsn w/in 48 hrs		Length of stay		Num. procedures	
	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65	Mean 63-64 (%)	Disc at 65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) **All**	77.42	7.140***	25.83	-1.855	4.40	-0.590	6.10	6.484*
		(1.960)		(2.535)		(2.800)		(3.773)
	[w=8.43,N=9,236]		[w=14.95,N=15,207]		[w=10.92,N=11,629]		[w=10.27,N=10,989]	
(B) **White**	76.40	9.303***	25.77	-2.404	4.35	-4.075	6.08	5.386
		(2.595)		(3.540)		(3.271)		(4.298)
	[w=10.11,N=6,611]		[w=11.76,N=7,557]		[w=12.53,N=8,006]		[w=10.67,N=6,895]	
(C) **Non-White**	79.08	1.107	25.94	0.444	4.47	5.655	6.12	5.214
		(3.115)		(3.729)		(4.672)		(6.761)
	[w=9.57,N=4,003]		[w=10.62,N=4,407]		[w=11.51,N=4,750]		[w=10.84,N=4,508]	
(D) **Highest qtrl zipcode income**	74.66	7.177	27.05	-9.122**	4.73	-1.917	5.98	4.609
		(5.342)		(4.564)		(6.130)		(7.691)
	[w=9.58,N=2,437]		[w=16.11,N=3,886]		[w=15.59,N=3,806]		[w=9.80,N=2,509]	
(E) **Lowest qtrl zipcode income**	78.53	-0.830	25.59	-3.025	4.55	3.927	6.05	6.163
		(4.022)		(4.172)		(6.439)		(5.670)
	[w=9.18,N=2,719]		[w=11.93,N=3,426]		[w=8.97,N=2,657]		[w=10.64,N=3,102]	

Notes. This table presents the estimated discontinuity in various measures of quality of care at the patient's initial admission for a heart attack. Same description as Table 2.6 except models include a quadratic polynomial for age at admission (in months) fully interacted with a dummy for age over 65. All coefficients and their standard errors have been multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.5 Additional Figures

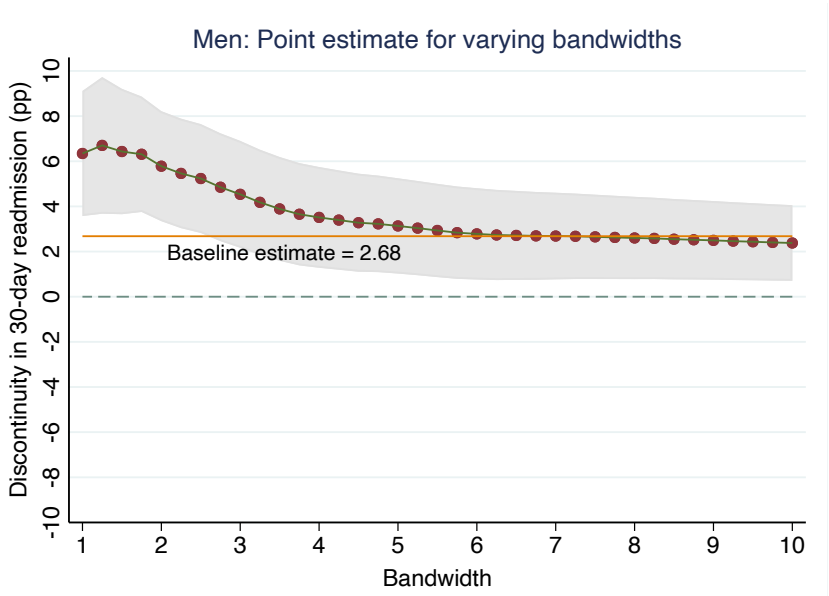


Figure B.1: Men: Estimated discontinuity in 30-day readmissions using linear specification and varying bandwidths.

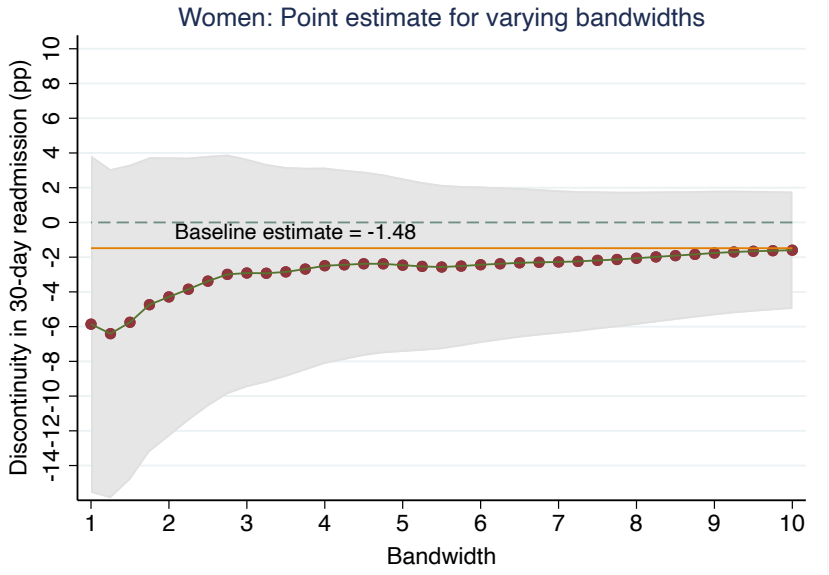


Figure B.2: Women: Estimated discontinuity in 30-day readmissions using linear specification and varying bandwidths.

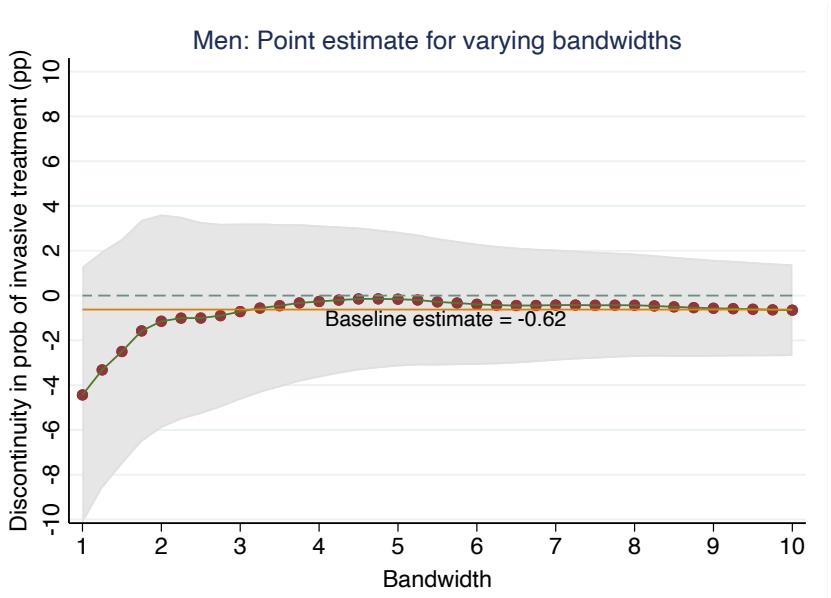


Figure B.3: Men: Estimated discontinuity in probability of receiving invasive treatment using linear specification and varying bandwidths.

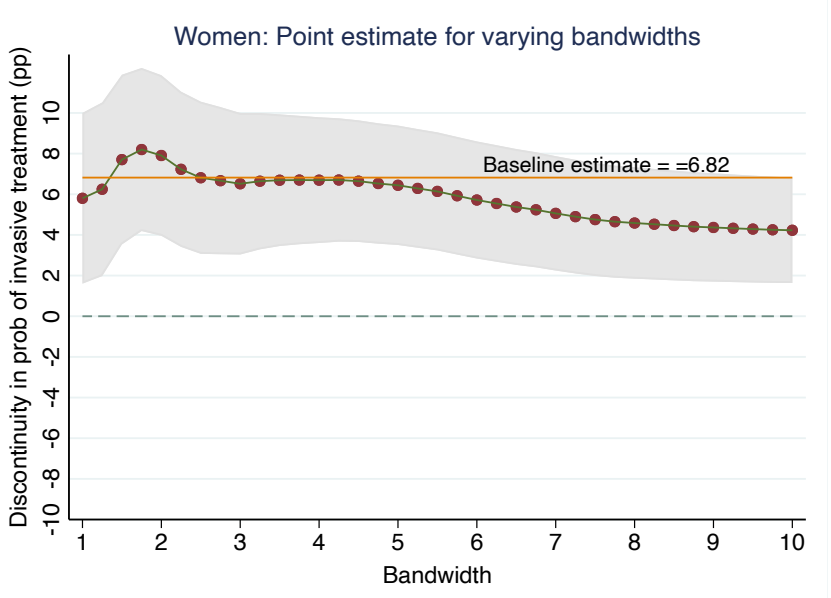


Figure B.4: Women: Estimated discontinuity in probability of receiving invasive treatment using linear specification and varying bandwidths.

C | Appendix to Chapter 3

C.1 Background Appendix

Study Streams and Exam Scores

As discussed in Section 2.2, the subjects and exam scores in the TGE varies across study streams. Table C.1 illustrates how the exam subjects differ by study stream. For example, the exam subjects of Arts students are: Islamic education (maximum grade 100), Arabic (300), English (280), History (120), Mathematics (100), Geography (100), and Scientific culture (100). The total score is calculated by summing the grades for the following subjects: Arabic, English, History, Mathematics and the highest two grades between two of the remaining subjects (Geography, Islamic education and Scientific culture). The total maximum exam score is 1000. For Science students, the subjects and the maximum grade are as follows: Islamic education (100), Arabic (200), English (200), Mathematics (240), Physics (160), Chemistry (100) and Biology (100). The total score is calculated by summing the grades for the following subjects: Arabic, English, Mathematics, Physics, and the highest score of any other two subjects (Islamic education, Chemistry, Biology). Although only two of the remaining optional subjects contribute to the final score, the student must pass each individual subject in order to pass the TGE.

Table C.1: Calculation of total score by stream of study

Stream of study	**Subjects examined**	**Max score**	**Part of final score**
(1)	(2)	(3)	(4)
Art	Islamic education	100	
(*Adabi*)	Arabic	300	Yes
	English	280	Yes
	History	120	Yes
	Mathematics	100	Yes
	Geography	100	
	Scientific culture	100	
Science	Islamic education	100	
(*Elmi*)	Arabic	200	Yes
	English	200	Yes
	Mathematics	240	Yes
	Physics	160	Yes
	Chemistry	100	
	Biology	100	
Commerce	Islamic education	100	
(*Tejari*)	Arabic	100	Yes
	English	100	Yes
	Mathematics	100	
	Accounting	150	Yes
	Administration	150	Yes
	Economics	100	
	Finance	100	Yes
	Finacial application	100	Yes
	Practical training	100	Yes
Agriculture	Islamic education	100	
(*Zera'i*)	Arabic	100	Yes
	English	100	Yes
	Farm management	100	Yes
	Chemistry	100	
	Biology	100	
	General agricultural sciences	100	Yes
	Private agricultural sciences	200	Yes
	Practical training	200	Yes
Manufacturing	Islamic education	100	
(*sena'i*)	Arabic	100	Yes
	English	100	Yes
	Mathematics	100	
	Physics	100	
	Industrial drawing	200	Yes
	Industry science	200	Yes
	Practical training	200	Yes

Notes. This table describes the list of study streams for students in the data. Within each stream (column 1), students sit exams for each subject in column (2). The maximum score for each subject depends on the stream (column 3). The overall score for students in that stream by summing the scores of the compulsory subjects (denoted by "yes" in column 4) and the highest two grades of the remaining two subjects.

C.2 Data Appendix

C.2.1 Summary of data sources used

The main analysis uses the following main data sources:

- **Education data**: This provides information on student exam scores from 2000–2006. We link this data to information on schools. This forms the basis of our analysis.
- **Mobility restrictions data**: The GIS data on the location of barriers over time is used to construct our barrier matrix (see Section 4) which serves as our main measure of student i's exposure to conflict infrastructure on their road to school. We also use this data to construct measures of the barriers within 10 km away from the school locality centroid. We match year t measures of mobility restrictions to year t education data.
- **Fatalities data**: This provides information on the number of fatalities per month in each locality. Lagged measures of fatalities are matched to the test score data.

In our regressions, we also several political factors that might influence educational outcomes and be correlated with mobility restrictions. All our regressions include controls for the population size of Israeli settlements within 10 km of school or home locality. In our robustness checks, we also include controls for the number of prisoners (in 1000s) in each locality, an indicator for at least one house demolitions in each locality, and the length of the separation wall (in km) going through the locality. Details on these data sources are as follow:

- **Settlements data**: This contains information on the location of Israeli settlements in the West Bank and its population size. Physical proximity between Israeli settlements and Palestinian localities may affect political attitudes and the intensity of conflict (Calì and Miaari, 2017). Furthermore, since physical barriers tend to be constructed near settlements, Palestinian localities situated near Israeli settlements tend to face more mobility restrictions than those that are located further away. This data allows us to calculate the road distance between the settle-

ment and various Palestinian localities. Using this information, we construct measures of the population size (in 1000s) of Israeli settlements within 10 km of the locality centre.

- **Prisoners data**: This contains information on the number of residents in each locality who are prisoners held in Israeli jails for security reasons in a given year. This data is provided by the Palestinian Ministry of Prisoners.
- **House demolitions**: This data provides information on house demolition orders for each Palestinian locality. Most house demolitions in the West Bank are enforced by the IDF as a counter-insurgency security measure.
- **Separation wall**: This data provides information on the length of the separation wall between Israel and the West Bank. The data is available for every other year; for years with missing data, we use linear interpolation to compute the length of the wall. The length of the separation wall varies over time for localities near the wall. For localities far from the wall, this variable remains at zero.

Lastly, our other sources of data include:

- **Wellbeing survey**: This is a survey on cognitive, emotional, and behavioural problems experienced by students in the West Bank. We use this to examine the impact of mobility restrictions on wellbeing.
- **Palestinian labour force survey**: This is a representative rotating quarterly household panel survey of Palestinians living in the West Bank and Gaza Strip. Households are surveyed four times over six quarters. The survey collects information on employment, school attendance, years of completed education of household members aged 15 and above. This is used to construct locality-level economic variables and to test for endogenous mobility (see Table C.8).
- **Palestinian Census**: This is used to check for endogenous mobility.

C.2.2 Data on delay factors at checkpoints

We collected data on the delay factors for all major internal and external checkpoints in the West Bank. In January 2017, there were 98 fixed checkpoints in the West Bank, 59 internal checkpoints located well within the West Bank and 39 external checkpoints that serve as the last inspection points before entering Israel.

For the internal checkpoints, 70 public transport drivers selected from across the West Bank collected GPS data on a daily basis between January 2018 and June 2018. Drivers then emailed their smartphone GPS data. Using information on when the car entered the checkpoint and when it exited the checkpoint, we construct the delay times at each internal checkpoint. In particular, we calculate the average time it takes cars to travel from 750 meters before the checkpoint to 750 meters after the checkpoint (a total distance of 1500 meters). We take the average delay for each month and then the average over the 6 months. The GPS data recorded the location of the car, time, and speed every five seconds during the survey period. These cars covered 47 of the 59 existing internal checkpoints. Checkpoints in the district of Hebron were not accessed due to the restrictions on vehicular movement across these checkpoints. The delay time at the remaining internal checkpoints were evaluated by using the closest similar checkpoint.

For the remaining external checkpoints, the GPS method could not be used since Palestinian vehicles are restricted from crossing these checkpoints from the West Bank to Israel. Instead, a random sample of 600 Palestinian labourers and students crossing these checkpoints were interviewed over a course of a week between 4am and 8am (before work). The respondents were asked to provide information on (a) the starting location of the trip, (b) the time she/he began the trip, (c) the arrival time at the checkpoints, (d) the time required to cross the checkpoint, and (e) the time and distance required to reach the destination. 11 external checkpoints were covered, including the two external checkpoints which students in the education data cross (Qalandiya and Gilo 300). At these two checkpoints, the average delay times over our survey period were 81 and 34 minutes respectively.

Survey data on wellbeing

The *wellbeing survey* used in Section 6 is more formally called the "Impact of the Israeli Measures on the Well-being of the Palestinian Children, Women and the Palestinian Household (2001)". The well-being survey was conducted by the Palestinian Central Bureau of Statistics. The target population consists of all Palestinian households that usually reside in the West Bank and Gaza Strip, excluding persons living in institutions such as prisons. We focus on households residing in the West Bank to be consistent with the rest of our data. This section provides some information about the survey.

- Sample size: A random stratified cluster sample composed of 3,393 households of which 2,301 in the West Bank and 1,092 in Gaza Strip was selected to represent the target population. The sample included enumeration areas close to clashes, settlements, and Israeli checkpoints. It also covered areas close to military exposure (shelling, shooting, uprooting of trees, land drifting, etc.)
- Data collection: Data collection took place between April 11 2001 to May 15 2001. Therefore, we define exposure to fatalities at the locality level a year before the survey as fatalities occurring between February 2000 to March 2001.
- Response rate: The number of completed interviews was 90% in the West Bank and 89% in the Gaza Strip.

The questions we use to measure cognitive, psychological, and behavioural problems are as follows:

- Cognitive: Does X suffer from a lack of concentration?
- Emotional: Which of the following psychological/emotional problems suffered by X? (a) Feelings of hopelessness and frustration, (b) Anger and nervous breakdown, (c) None of the above.
- Behavioural: Which of the following behavioural problems suffered by X? (a) Tendency to violence, (b) Tendency to screaming, (c) Tendency to beat/imprecate others, (d) None of the above.

C.3 Additional Tables

Table C.2: Summary statistics for traversing vs. non-traversing students

	Traverse (1)	**Do not traverse** (2)
Age	18.21	18.20
Female	0.51	0.53
Muslim	0.99	0.99
Art	0.64	0.70
Science	0.27	0.24
Vocational	0.08	0.06
Pass	0.73	0.73
Average	64.63	63.98
Observations	46,810	100,132

Notes. Variable means for traversing and non-traversing students. Traversing students are those who live and study in different localities.

Table C.3: Predictors of growth in number of schools within localities

	Dep. Var.: **Number of schools in locality**	
	(1)	(2)
Population aged 17–19 (in 1000s)	1.2601***	0.2570***
	(0.0305)	(0.0634)
$\geq$ 1 CP within 10 km	0.0680	0.0682
	(0.0656)	(0.0477)
Hourly wage	-0.0002	-0.0002
	(0.0031)	(0.0017)
Unemployment rate	-0.4353*	-0.0457
	(0.2361)	(0.1671)
Israeli settler population (in 1000s)	-0.0052***	-0.0025
	(0.0009)	(0.0081)
Year FE	Yes	Yes
Locality FE		Yes
Observations	1,011	993

Notes. Regression results for equations examining the determinants of school growth.

Table C.4: Impact of crossing one or more checkpoints on educational performance (with additional controls)

	Pass	Overall score	Maths	English
	(1)	(2)	(3)	(4)
Mean	0.73	64.19	63.37	57.79
SD	0.44	20.79	25.88	22.55
A. Mobility restrictions				
Cross checkpoint	-0.0263***	-0.0571***	-0.0622***	-0.0467**
	(0.0100)	(0.0204)	(0.0223)	(0.0197)
B. Other conflict variables				
Fatalities in school locality	-0.0005**	-0.0011**	-0.0015*	-0.0009*
	(0.0002)	(0.0005)	(0.0009)	(0.0005)
Fatalities in student locality	-0.0001	-0.0001	0.0002	0.0002
	(0.0002)	(0.0005)	(0.0006)	(0.0005)
C. Other variables				
Prisoners in school locality (in 1000s)	-3.209	-11.046***	-13.706***	-2.804
	(2.0790)	(3.8610)	(4.5050)	(3.6800)
Prisoners in student locality (in 1000s)	-1.295	-4.588	-4.122	-7.051*
	(1.8820)	(3.8190)	(4.5350)	(3.7620)
$\geq$ 1 House demolition in school locality	-0.0122	-0.0226	-0.0158	-0.0313
	(0.0089)	(0.0197)	(0.0211)	(0.0213)
$\geq$ 1 House demolition in student locality	0.0015	0.0128	-0.0023	0.0170
	(0.0087)	(0.0192)	(0.0199)	(0.0210)
Separation wall length (km) in school locality	-0.0003	0.0002	0.0009	0.0007
	(0.0010)	(0.0027)	(0.0032)	(0.0025)
Separation wall length (km) in home locality	-0.0005	-0.0038	-0.0057*	-0.0024
	(0.0011)	(0.0028)	(0.0033)	(0.0026)
# Student-Years	123,097	123,097	122,498	123,097
Adjusted R^2	0.107	0.257	0.164	0.328

Notes. This table presents estimated coefficients from equation (19) where the vector P is extended to include additional controls. Scores expressed in standard deviations. Mean and standard deviation of untransformed dependent variables (e.g. exam scores) presented in the first two rows. All regressions include the following fixed effects: school locality fixed effects, home locality fixed effects, and academic year fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of home and school locality, additional policy variables (Palestinian prisoners (1000s), indicator for at least one house demolition, length of the separation wall going through the locality), student indicators (gender, religion, calender year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, and gender of school). Standard errors in parentheses, clustered at the home-school locality pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: Impact of crossing one or more checkpoints on educational performance (inc. academic-year specific fixed effects)

	Pass		Overall score		Maths		English	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Mobility restrictions								
Cross Checkpoint	-0.0279***	-0.0249***	-0.0818***	-0.0701***	-0.0794***	-0.0669***	-0.0680***	-0.0608***
	(0.0092)	(0.0087)	(0.0193)	(0.0180)	(0.0206)	(0.0191)	(0.0191)	(0.0180)
B. Other conflict variables								
Fatalities in school locality	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–
Fatalities in home locality	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–
Home locality-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School locality-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE		Yes		Yes		Yes		Yes
# Student-Years	146,942	146,942	146,942	146,942	146,268	146,268	146,942	146,942
Adjusted R^2	0.115	0.129	0.266	0.288	0.179	0.202	0.338	0.357

Notes. This table presents estimated coefficients from equation (19) where time-invariant home and school locality fixed effects are replaced with academic-year specific home and school locality fixed effects $\rho_{ht} + \sigma_{lt}$. Scores expressed in standard deviations. All regressions include the following fixed effects: home-by-academic year fixed effects and school-by-academic year. Results in even columns also include school fixed effects. All regressions include the following controls: student indicators (gender, religion, year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, gender of school). Standard errors in parentheses, clustered at the home-school locality pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.6: Impact of having to cross checkpoints on educational performance (inc. home-school locality pair fixed effects)

	Pass		Overall score		Maths		English	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Mobility restrictions								
Cross Checkpoint	-0.0308**	-0.0304**	-0.0764***	-0.0764***	-0.0773***	-0.0779***	-0.0671***	-0.0680***
	(0.0134)	(0.0134)	(0.0260)	(0.0253)	(0.0280)	(0.0274)	(0.0237)	(0.0232)
B. Other conflict variables								
Fatalities in school locality	-0.0006***	-0.0008***	-0.0012**	-0.0016***	-0.0016*	-0.0020***	-0.0010**	-0.0015***
	(0.0002)	(0.0002)	(0.0005)	(0.0005)	(0.0009)	(0.0008)	(0.0004)	(0.0005)
Fatalities in home locality	0.0006	0.0002	-0.0029	-0.0045	-0.0061	-0.0067	-0.0007	-0.0029
	(0.0017)	(0.0017)	(0.0047)	(0.0046)	(0.0053)	(0.0052)	(0.0045)	(0.0043)
Home-by-locality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE		Yes		Yes		Yes		
# Student-Years	146,942	146,942	146,942	146,942	146,268	146,268	146,942	146,942
Adjusted R^2	0.109	0.122	0.258	0.279	0.168	0.190	0.329	0.347

Notes. This table presents estimated coefficients from equation (19) where the additive home and school locality fixed effects are replaced with home locality-school locality pair fixed effects. Scores expressed in standard deviations. All regressions include the following fixed effects: home-by-school fixed effects and academic year fixed effects. Results in even columns also include school fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of home and school locality, student indicators (gender, religion, year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, gender of school). Standard errors in parentheses, clustered at the home-school locality pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: Impact of crossing one or more checkpoints on educational performance (multi-way clustering)

	Pass	Overall score	Maths	English
	(1)	(2)	(3)	(4)
Mean	0.73	64.19	63.37	57.79
SD	0.44	20.79	25.88	22.55
A. Mobility restrictions				
Cross checkpoint	-0.0305***	-0.0714***	-0.0727***	-0.0529**
	(0.0089)	(0.0243)	(0.0267)	(0.0222)
B. Other conflict variables				
Fatalities in school locality	-0.0007**	-0.0012**	-0.0017*	-0.0010**
	(0.0003)	(0.0006)	(0.0009)	(0.0005)
Fatalities in home locality	0.0001	0.0005	0.0010	0.0005
	(0.0002)	(0.0005)	(0.0008)	(0.0004)
# Student-Years	146,942	146,942	146,268	146,942
Adjusted R^2	0.107	0.256	0.165	0.327

Notes. This table presents estimated coefficients from equation (19). Scores expressed in standard deviations. Mean and standard deviation of untransformed dependent variables (e.g. exam scores) presented in the first two rows. All regressions include the following fixed effects: school locality fixed effects, home locality fixed effects, distance bin fixed effects, and academic year fixed effects. All regressions include the following controls: population size (in 1000s) of Israeli settlements within 10 km of home and school locality, student indicators (gender, religion, year of birth, study branch) and school controls (number of classrooms, total number of students, total number of teachers, gender of school). Standard errors in parentheses, multi-way clustered at the home and school locality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.8: Sample selection and endogenous mobility tests

	Attend School	**Drop out**	**Move next yr**
Data source	PLFS	Test score	PLFS
Dep. Var. Mean	0.927	0.031	0.002
	(1)	(2)	(3)
A. Mobility restrictions			
$\geq$ 1 CP within 10 km	-0.0105	0.0019	-0.0003
	(0.0119)	(0.0016)	(0.0003)
B. Other conflict/policy variables			
Fatalities in locality	-0.0112	0.0014	-0.0005
	(0.0081)	(0.0009)	(0.0004)
Israeli settler population	-0.0085	0.0014	0.0000
within 10 km (in 1000s)	(0.0276)	(0.0011)	(0.0000)
Locality FE	Yes		Yes
School FE		Yes	
Year FE	Yes	Yes	Yes
Observations	1,716	2,865	169,606
Adjusted R^2	0.510	0.569	0.058

Notes. Column (3) includes additional controls for gender and age of respondent in the PLFS. Standard errors in parentheses, clustered at the locality (columns 1, 3, and 4) or school (column 2) level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.9: Reverse causality test

	Dep. Var.: **Number of CP within 10 km**			
	(1)	(2)	(3)	(4)
A. Educational outcomes				
	Contemporaneous		Lagged	
	Pass	Overall score	Pass	Overall score
	0.416	0.201	0.452	0.225
	(0.695)	(0.931)	(0.747)	(0.772)
B. Other potential determinants (Lagged)				
Fatalities in school locality	0.009	0.009	-0.008	-0.009
	(0.025)	(0.025)	(0.029)	(0.029)
Israeli settler population	0.057*	0.057*	0.047	0.048
within 10 km (in 1000s)	(0.033)	(0.033)	(0.034)	(0.034)
Observations	1,589	1,589	1,309	1,309
Adjusted R^2	0.647	0.647	0.670	0.670

Notes. All regressions include school locality and year fixed effects. Standard errors in parentheses, clustered at the locality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.