The Impact of Medicare Nonpayment:

A Quasi-Experimental Approach

Audrey Kornkven

Professor Charlie Becker, Faculty Advisor
Professor Frank Sloan, Faculty Advisor
Professor Grace Kim, Seminar Instructor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University
Durham, North Carolina
2020

Acknowledgements

This thesis would not have been possible without the support system and resources that extend across Duke and beyond, particularly when coronavirus threw a wrench into the spring semester. Thank you to my thesis advisor, Charlie Becker, for his support, guidance, and life advice, as well as Frank Sloan, for lending his expertise in health economics. Also thank you to the staff at the CDVS, particularly Mark Thomas and Joel Herndon, for patiently helping me solve thorny data issues. Brett Gall, a graduate student from the SSRI, was instrumental in helping me figure out valid strategies for matching and running a fuzzy regression discontinuity. Thank you to Stephen Burkhalter from the Economics IT staff, for setting up a virtual machine so that I could continue to access my data remotely, and to Tom Milledge, Andy Ingram, and Professor Connolly, for helping me access and learn to use the Duke Computing Cluster.

I also want to thank my seminar classmates for their useful comments and encouragement throughout the year, and my seminar instructor, Professor Kim, who was very helpful and supportive. To my friends and family, who have patiently listened as I dealt with the ups and downs of this thesis since August, thank you for being so supportive of me.

Abstract

In October 2008, a provision of the Deficit Reduction Act of 2005 known as Medicare "Nonpayment" went into effect, eliminating reimbursement for the marginal costs of preventable hospital-acquired conditions in an effort to correct perverse incentives in hospitals and improve patient safety. This paper contributes to the existing debate surrounding Nonpayment's efficacy by considering varying degrees of fiscal pressure among hospitals; potential impacts on healthcare utilization; and differences between Medicare and non-Medicare patient populations. It combines data on millions of hospital discharges in New York from 2006-2010 with hospital-, hospital referral region-, and county-level data to isolate the policy's impact. Analysis exploits the quasi-experimental nature of Nonpayment via difference-in-differences with Mahalanobis matching and fuzzy regression discontinuity designs. In line with results from Lee et al. (2012), Schuller et al. (2013), and Vaz et al. (2015), this paper does not find evidence that Nonpayment reduced the likelihood that Medicare patients would develop a hospital-acquired condition, and concludes that the policy is not likely the success claimed by policymakers. Results also suggest that providers may select against unprofitable Medicare patients when possible, and are likely to vary in their responses to financial incentives. Specifically, private non-profit hospitals appear to have been most responsive to the policy. These findings have important implications for pay-for-performance initiatives in American healthcare.

JEL Codes: I13, I18, H50, H51, D73

Keywords: Health economics, Medicare, Nonpayment, Hospital-acquired conditions, Pay-for-performance, Never event, New York

I. A. Introduction

In 2000, the Patient Safety Movement kicked off with the Institute of Medicine's publication of "To Err is Human: Building a Safer Health System". The report claimed that Americans were more likely to die from preventable medical errors than from car accidents or breast cancer, and that up to 100,000 patients were dying annually from such errors. The authors blamed faulty systems and processes, rather than a few bad apples. They called for a mix of regulatory and market-based incentives, aimed at reducing the occurrence of hospital-acquired conditions (HACs) like foreign objects retained after surgery or infections from catheters, also known as "never events" (Kohn, Corrigan, and Donaldson, 2000). To further advance the debate on pay-for-performance – or linking providers' payment to health outcomes – this paper aims to analyze whether the financial disincentives created by Medicare's 2008 "Nonpayment" policy significantly contributed to the declining rates of HACs in the United States. My approach will examine New York state data from 2006-2010 merged at the patient level, hospital level, hospital referral region (HRR) level, and county level; it will consider financial pressures and organizational variables, the potential for selection of profitable patients, as well as healthcare utilization and intensity.

Early momentum contributed to a budding conversation surrounding pay-for-performance (P4P) policies in healthcare. After "To Err is Human", the Office of Inspector General released a report that analyzed hospital discharge data from October of 2008, and suggested that the problem remained serious for Medicare beneficiaries. The report concluded that a shocking 13.5% of hospitalized Medicare patients experienced an "adverse event" during their stay, such as a HAC (as defined by policymakers), a medication error, or a delay in surgery due to equipment malfunction. 15,000 patients each month, or an estimated 1.5% of care-seeking

Medicare beneficiaries, experienced an event that led to death. Nearly half of these events – particularly infections – were deemed clearly or likely preventable. Despite the relatively low occurrence of HACs compared to other adverse events like medication issues, the ease of prevention represented a clear opportunity for improvement (Levinson, 2010).

Initially, pay-for-performance seemed like an ideal market-based solution. Economist Kenneth Arrow wrote that, under "ideal insurance", a patient would pay a provider based entirely on outcomes. However, due to information asymmetry in such a relationship, payment based on results proves nearly impossible (Arrow, 1963). It is costly for both patients and insurers to determine the necessity and quality of medical care, so patients are left to rely on a relationship of trust with the doctor (Town et al., 2000). The result, according to Arrow, is a departure from competitive equilibrium and a loss of welfare in the market for medical care (Arrow, 1963). However, measurement of health outcomes has become more sophisticated, given improvements in information systems. One might reason that these dilemmas are outdated, and that pay-for-performance may be a solution to high healthcare costs and poor quality, particularly for preventable errors (Eijkenaar, 2011). In recent years, the move from fee-for-service to P4P has aimed to correct perverse incentives that plague the principal-agent relationships between payers, providers, and patients, though moral hazard remains (Town et al., 2000). If these efforts persist, policymakers must consider both the ability of P4P to create meaningful change, and spillover effects from such policies.

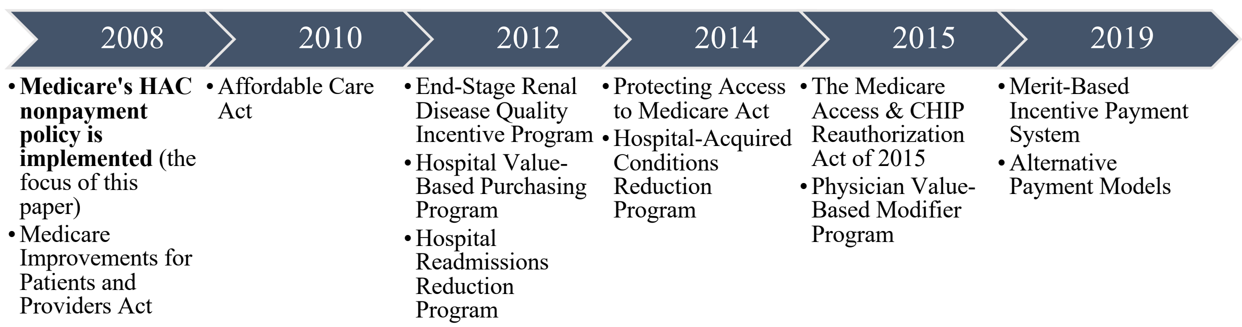
Medicare's hospital payment system, which pays a fixed amount derived from a patient's diagnoses and major procedures, has been a prime area for reform. The system is based on diagnosis-related groups, or DRGs, which become more lucrative as a patient experiences complications. By compensating hospitals for the treatment of sicker patients, Medicare aims to prevent providers from selecting only healthy patients. Prior to 2008, hospitals that improved patient safety and reduced their HAC rate were reimbursed for less complex DRGs than would have been the case otherwise, actually experiencing a loss of revenue and profit from improved quality (Rosenthal, 2007). In response to this perverse incentive, the Deficit Reduction Act of 2005 introduced one of the earliest large-scale P4P programs: in October of 2008, a Medicare "nonpayment" policy went into effect (hereafter referred to as "Nonpayment"). This policy eliminated reimbursement for the marginal cost of ten high-cost or high-volume hospital-acquired conditions (HACs) that were deemed preventable through the application of evidence-based guidelines. Some serious issues were excluded, like medication errors, which may be more challenging to prevent (Levinson, 2010). According to the Centers for Medicare and Medicaid Services (CMS), the hospital-acquired conditions listed in the original policy are as follows:

1) Foreign object retained after surgery
2) Air embolism
3) Blood incompatibility
4) Stage III and IV pressure ulcers
5) Falls and trauma
6) Manifestations of poor glycemic control
7) Catheter-associated urinary tract infection
8) Vascular catheter-associated infection
9) Certain surgical site infections
10) Deep vein thrombosis/pulmonary embolism

The rule linked financial disincentives to performance, theoretically incentivizing hospitals to better prevent simple but potentially fatal mistakes. Nonpayment also mandated that hospitals report a "present on admission", or POA code, for diagnoses that were detected upon admission (McNair et al., 2009). This change in coding requirements enabled Medicare to penalize hospitals that harmed patients during their stay. According to the CMS website, certain hospitals, such as critical access hospitals (which are located in rural areas and provide limited care), long-term care hospitals, and Maryland waiver hospitals, are exempted from Nonpayment, since they are not subject to the Inpatient Prospective Payment System (IPPS) used by Medicare.

Since 2008, Congress has pursued a wave of “value-based” programs. Federal efforts to link payments to provider performance remain at the forefront of policy discussions. The following graphic is adopted from the Centers for Medicare and Medicaid Services:

FIGURE 1: Pay-for-Performance Policy Timeline



In 2010, the Affordable Care Act ushered in a wave of value-based programs based on P4P, which were implemented beginning in 2012. These policies include the 2014 Hospital Acquired Conditions Reduction Program, which financially penalizes hospitals with high rates of HACs by reducing payments to the worst-performing quartile by 1%, according to the CMS website. At first glance, these efforts seem successful: the Agency for Healthcare Research and Quality, run by the Department of Health and Human Services, claims that HACs declined by 17% from 2010-2014, saving nearly 87,000 lives and close to $20 billion. The agency credited Medicare payment incentives, among other federal efforts (“Saving Lives and Saving Money”, 2016).

However, there may be some potential unintended consequences of these policies. For example, providers’ intrinsic motivation may be crowded out by financial incentives, diluting the policy’s desirable impact (Eijkenaar, 2011). One study models physician behavior under different payment systems, and finds that complex behavior-based payments indeed crowd out motivation, resulting in relatively poor quality of care (Green, 2013). Similarly, providers may focus only on the elements of care that are easily measured and rewarded, to the detriment of many patients. If they are able, providers may seek out patients who are healthy and compliant, while avoiding those that are at higher risk of poor outcomes (Eijkenaar, 2011).

Today, patients are less likely to die from hospital-acquired conditions, and policymakers appear ready to attribute such progress to financial incentives ("AHRQ National Scorecard"). Despite this enthusiasm, economists remain, at best, skeptical of P4P's ability to improve healthcare quality and reduce costs (Mullen, Frank, and Rosenthal, 2010). The empirical literature also remains split regarding the impact of Nonpayment. Before charging forward with P4P efforts, it is imperative that researchers examine the efficacy of existing programs, as well as consider any unintended consequences. My approach aims to isolate the impact of Medicare-specific financial incentives from other influences that might indirectly lower the HAC rate, such as patient selection, organizational factors, and healthcare utilization.

Section II reviews the existing literature on Nonpayment, including a summary of the data and methods used by previous researchers. Section III provides an overview of the relevant economic theory. It also justifies variables of interest, and predicts their influence on HAC likelihood. Section IV considers the available data, including strengths and weaknesses of particular datasets; it also summarizes my merging strategy and explains how I managed my data. Section V describes my empirical methodology. Via difference in differences with matching and fuzzy regression discontinuity designs, I isolate the impact of Nonpayment on the Medicare population. I find little evidence that the policy was successful, though certain characteristics appear to have influenced hospitals' responses. I conclude with a discussion of results, implications for researchers and policymakers, and suggestions for further research.

I. B. Background on Medicare Payment

Medicare was established in 1965 by the Social Security Act, and now provides health insurance to patients over the age of 65 as well as the permanently disabled. Inpatient hospital services are covered under a fee-for-service model by Medicare Part A ("Medicare Primer"). Historically, hospitals have been disconnected from the market mechanisms that determine financial performance in other industries. Regulation, along with the non-profit or government ownership status of (most) hospitals, somewhat insulates providers from the pressures of capital markets[1]. When Medicare was enacted, payments were based on "reasonable costs" faced by hospitals. This system suppressed price competition since Medicare would reimburse hospitals with high charges; consequently, healthcare costs skyrocketed. Fearing that Medicare Part A would become insolvent, policymakers in the 1970s and 80s aimed to slow the rapid growth in Medicare spending, which was aligned with ballooning costs. So, Medicare reform changed hospitals' profit incentives. Following cost containment in the 70s, the Inpatient Prospective Payment System (IPPS) made it possible for hospitals to collect a profit or make a loss, since public payers now reimbursed hospitals a fixed amount based on each inpatient stay rather than the costs incurred by each patient (Hodgkin and McGuire, 1994). Hospitals' bottom lines were now impacted by the difference between a diagnosis-related group (DRG) rate and the cost to treat that patient. Facing profit incentives, hospitals were expected to actively tamp down on costs, such as reducing the length of hospital stays or intensity of care.

The Inpatient Prospective Payment System, implemented in 1983, set a prospectively determined payment amount for each type of discharge based on the average cost for all hospitals. CMS adjusts these rates based on the hospital's type and location, its case mix, and the weight associated with a discharge's severity-diagnosis related group. Medicare makes

[1] However, private hospitals do receive funding from external sources, particularly as private equity grows its focus on hospital chains.

additional payments in cases of extraordinary events, to teaching hospitals, and to "disproportionate share" hospitals, which provide a higher volume of care to low-income patients ("Medicare Primer"). Some diagnoses are more profitable than others, resulting in a system that, in theory, financially incentivizes hospitals to prioritize some discharges (patients) over others (Hodgkin and McGuire, 1994). With this perverse incentive in mind, my analysis will uniquely adjust for potential changes in patient selection and intensity.

II. Literature Review

The Prospective Payment System

There is substantial economics literature evaluating Medicare's move to the Prospective Payment System in the 1980s, which also altered financial incentives for providers in an effort to reduce healthcare costs. Dranove (1987) incorrectly hypothesized that hospitals would specialize, choosing to focus on the most profitable DRGs. Efficient hospitals may be willing to treat patients from less efficient hospitals, resulting in transfers as a means of patient selection. However, Newhouse (1989) does not find that unprofitable cases were likely to be transferred out of hospitals. Instead, he shows that about one quarter of all unprofitable DRG cases were taken by last-resort hospitals post-PPS, demonstrating evidence of selection. Newhouse concludes that selection was occurring, but not via transfers. Hospitals may not always be able to turn patients away; instead, they may reduce attention and effort directed toward less profitable patients.

Other researchers discovered that the quantity of care, measured by length of stay (LOS), fell among hospitals that were subject to the new payment system and experienced declining marginal reimbursement rates. Some research suggests that quantity of care increased among hospitals that were exempt from the system, suggesting that providers may have selected a different patient mix in response to the change (Newhouse and Byrne, 1988). Ellis and McGuire (1996) study a similar change among New Hampshire Medicaid payments, and attribute a decline in quantity of care to moral hazard and changes in the patient mix among hospitals. Indeed, Hodgkin and McGuire (1994) find that the market share of Medicare patients treated at exempt hospitals increased following the system's implementation. Hospital admissions decreased as well; the volume of admissions at non-exempt short-stay hospitals declined by 11% in the first eight years of the system, with most of the decline occurring in the first two years.

These volume trends were particularly striking for the Medicare population, compared to non-Medicare patients, suggesting that providers have the ability to discriminate based on payer. After the payment change, length of stay and volume trends reversed direction for Medicare patients. The magnitude of these utilization changes were substantially greater for Medicare patients than those who were covered by other payers (Hodgkin and McGuire, 1994). Hadley et al. (1989) find that hospitals facing greater financial pressure (as measured by the index shown on page 23) were likely to have lower occupancy rates, and thus responded to fiscal pressure by increasing their daily census (number of admissions) while reducing costs by lowering LOS. Staiger and Gaumer (1990) find evidence of another adverse trend in their analysis of financial pressure on hospitals: a reduction in payment was associated with increased mortality rates, with the impact most significant for small hospitals and government facilities.

Despite economic analyses of the Prospective Payment System and subsequent price shocks throughout the '80s and '90s, only the medical field has examined the impact of Medicare Nonpayment. This may be because the payment shock it triggered was relatively small; the policy was overshadowed by the 2010 Affordable Care Act, which seemed more interesting to economists; or because the policy attracted medical researchers rather than economists. Given its medical nature, the existing literature has failed to account for quantity or intensity of care, or fiscal pressure, which are typically considered in health economics applications. To address this gap, I will include new variables in my analysis, detailed in the "Data Overview" section.

Medicare Nonpayment

Whether Medicare's Nonpayment policy was actually effective remains very much in question, with around half of papers concluding that the financial disincentives caused a statistically significant decline in HACs. The existing research also varies in terms of HACs studied, geographic focus, consideration of hospital characteristics, timeline of data collection, and datasets used. In 2012, Lee et al. published an initial analysis, which finds that the policy had no impact. However, their sample is small and geographically skewed. The researchers also assume that Medicare patients would not be differentially impacted by the policy, despite warnings by economists that providers may "teach to the test" by focusing on patients directly linked to financial incentives (Lee et al., 2012; Eijkenaar, 2011). Several subsequent papers have used data from the Healthcare Cost and Utilization Project (HCUP), which offers a more representative sample of hospitals at state- and national- levels. In my paper, I use the New York State Inpatient Database (SID) from HCUP, which captures nearly all discharges in the state.

Schuller et al. (2013) use the HCUP National Inpatient Sample (NIS), a representative weighted national sample, to study the impact of Nonpayment across hospitals with different characteristics. Though they conclude that there was no overall reduction in the single HAC they studied, catheter-associated urinary tract infections, the researchers note significantly higher infection rates in large, urban, and Southern hospitals. This analysis demonstrates the importance of controlling for a variety of hospital traits. This paper will also control for patient and hospital characteristics, especially variables with an organizational impact, using data from the SID and Centers for Medicare and Medicaid Services (CMS); this is described further in the following section.

Schuller et al. (2013) also suggest that a longer study period may be appropriate, since hospitals may take time to respond to incentives. Similarly, findings from Gidwani & Bhattacharya (2015) suggest that providers focused most on reform one quarter after the policy was implemented, perhaps after reviewing quarterly revenue. Despite the importance of timeline selection, many studies inexplicably lack either pre-implementation or post-policy panel data. Thirukumaran et al. (2017) study the policy's impact over the longest period of time, from 2005 until 2012. However, a longer timeline is not necessarily more robust, since P4P policies implemented in 2012 may confound the impact of Nonpayment on HAC rates. To ensure that enough pre- and post- policy data are included in the analysis, but other changes (like reforms associated with the Affordable Care Act) are excluded, my paper will examine data from 2006-2010. This limits the impact of hospital consolidation in New York, since only hospitals that were open during the entire period are included in my final sample, as shown in Appendix I.

Other researchers have raised concerns that Nonpayment may widen disparities by negatively impacting funding for hospitals that are both likely to treat the poor and ill-equipped to reduce HAC rates. One study finds that hospitals with low financial margins, which are more likely to serve as safety-net hospitals that treat the uninsured and underinsured, experience higher rates of HACs on average. If hospitals that serve complex patients have fewer resources, they might be less capable of responding to financial incentives (McHugh et al., 2011). However, Vaz et al. (2011) do not find evidence of disparities, measured by infection rates, resulting from Nonpayment in their comparison of safety net and non-safety net hospitals. To address the possibility that hospitals that serve the poor might be less responsive to incentives, several papers include a measure of patient complexity. One paper includes hospitals' Medicare case mix, or the sum of diagnosis-related group weights divided by total discharges (Waters et al., 2015).

Because more complex conditions usually have higher DRG weights, a high case mix indicates that a hospital may be treating an underserved population. CMS includes case mix in the base payment rate for a hospital as well. Thirukumaran et al. (2017) include a hospital's transfer-adjusted case mix to account for potential transfers. In addition, they include disproportionate patient percentage, a measure defined and used by Medicare to measure the proportions of uninsured and underinsured patients in a hospital. These variables are important measures of a hospital's existing resource level, and I incorporate them into my analysis. It's important to note that Thirukumaran et al. (2017) neglect to include teaching status, another indicator of resource level, which I find to be significant.

Assumptions Made in Previous Work

The literature uses various types of a quasi-experimental design, assuming that patients in a chosen control group were not impacted by the policy. This approach raises concerns of data selection bias. Controls include groups of patients who were not covered by Medicare, based on opposite assumptions used by Lee et al. (Gidwani & Bhattacharya, 2015), as well as groups of patients treated for conditions that were not named in the policy (Lee et al., 2012; Vaz et al., 2015). However, economic theory implies that spillover effects, both positive and negative, may impact these types of control groups. If providers "teach to the test", conditions not included in the policy may actually *worsen*, as providers are incentivized to shift focus toward preventing events that will result in a financial penalty. Even worse, providers may intentionally turn away patients that are at high risk of acquiring a HAC. Evidence of provider selection has been documented in other initiatives meant to improve healthcare quality. When some states mandated that hospitals publish quality report cards detailing coronary artery bypass graft surgery mortality rates, providers selected healthier patients for treatment. As a result, surgery was avoided for sicker patients, who were treated with less intensive methods and experienced higher mortality (Dranove et al., 2003). I will account for provider selection by conditioning on weekend admission and controlling for transfer-adjusted case mix, explained further in Section III.

On the other hand, incentivizing better quality control may lead to positive spillover, improving outcomes for patients who are not covered by Medicare but benefit from an overall increase in attention to quality (Eijkenaar, 2011). Gidwani & Bhattacharya (2015) examine two control groups, concluding that spillover effects were more likely to impact non-Medicare patients above 65 than slightly younger patients who were ineligible for Medicare. Their findings also suggest evidence of welfare shifting, meaning that negative spillover from the policy worsened outcomes for younger, non-Medicare patients. In my paper, I use a matched group of non-Medicare patients as a control group, then conduct an RDD to further account for any selection issues.

Research also differs in its choice of patients selected for study. Around half of studies look at the incidence of HACs across all patients (Lee et al., 2012; Waters et al., 2015; Vaz et al., 2015). Other researchers limit study to population to elderly adults with Medicare (Schuller et al., 2013). In some papers, the choice of HAC being studied influences the cohort. Gidwani and Bhattacharya (2014) analyze the impact of Nonpayment on pulmonary embolism and deep vein thrombosis, which are frequent complications following hip or knee replacement surgeries. Their difference-in-differences model thus compares Medicare and non-Medicare patients receiving hip or knee surgery. Thirukumaran et al. (2017) similarly look at particular diagnosis groups, rather than all inpatient stays, to analyze a broader group of HACs with the goal of minimizing heterogeneity among patients. They only consider elderly Medicare patients with a primary diagnosis of acute myocardial infarction, congestive heart failure, pneumonia, and stroke, as these patients may be at greater risk of hospital-associated mortality. It is important to choose cohorts that limit heterogeneity, but restricting analysis to a few diagnosis groups may drastically reduce the size of the sample and limit the generalizability of the study. Instead, I address heterogeneity by matching patient populations on characteristics that influence HAC likelihood.

Financial Incentives

Few papers account for the magnitude of the financial disincentives created by Nonpayment. Lee et al. (2012) and Vaz et al. (2015) do not include data on patients' primary payer, limiting the researchers' ability to isolate the impact on Medicare patients. Data from HCUP have remedied this issue in subsequent papers, allowing researchers to restrict their sample to Medicare patients. Studies that have used other datasets have attempted to control for financial measures in other ways. For example, Waters et al. (2015) include hospital-level data on payer market shares, Medicare case mix, and total profit margin. Beyond controlling for payer, most existing research has not focused on the magnitude of financial disincentives introduced by Nonpayment, except for Thirukumaran et al. (2017).

Previous research has focused on various subsets of the ten HACs named in the policy, without clear justification for why certain HACs are of more interest than others. The financial impact of each HAC has been estimated using data on reimbursement costs and prevalence, and shows a wide range of incentive sizes. CMS data were analyzed by Kavanagh (2011) to estimate ranges of total cost for each condition. Conditions like stage III and stage IV pressure ulcers appeared to cost over $11 billion per year; infections, which were more common but significantly less expensive to treat, typically came in under $1 billion. I look at all HACs in my analysis, rather than a subset.

However, it is not clear whether the financial disincentives created by the policy were sufficiently large to have an impact. McNair et al. (2009) simulate the hospital payment process using cost weights, cost-to-charge ratios to measure profitability, and hospital-specific prices. This process estimates the size of the financial disincentive for California hospitals. Ultimately, the researchers conclude that Nonpayment did not create a meaningful financial disincentive, contrary to estimates from the Centers for Medicare and Medicaid Services, which are available on their website. However, the analysis is limited to California data from 2006 and only includes data for six of the HACs outlined in the policy. Despite these limitations, the departure from CMS estimates in McNair et al. (2009) is certainly cause for concern. In a study of pay-for-performance, Gneezy and Rustichini (2000) find that small financial incentives could actually be counterproductive, since payment could crowd out powerful social norms. The study concludes that payments must be substantial to actually change behavior in the intended direction.

Of the Nonpayment literature, Thirukumaran et al. (2017) is the most recent and robust analysis of the policy. They introduce an economic lens to the problem by including a hospital's "Medicare Utilization Ratio" (MUR) to explicitly account for incentive strength. MUR measures the proportion of a hospital's inpatient days financed by Medicare. The researchers study the impact of Nonpayment by MUR quartile, concluding that the impact of the program was more significant for hospitals that used more Medicare. To extend this approach, my paper will include a fiscal pressure index that interacts a hospital's MUR with its cost-to-charge ratio, a measure of profitability from treating Medicare patients. Below, I summarize the existing papers that analyze the impact of Nonpayment on various hospital-acquired conditions that were included in the policy:

TABLE 1: Summary of Nonpayment Literature

Date	Study	Data Source	Cohort(s) Studied	Method	Policy Impact
Oct. 2012	Lee et al.	Centers for Disease Control and Prevention's National Health Safety Network	All patients	Interrupted time series with comparison series (conditions not named in the policy)	None (only looked at infections)
Jan. 2013	Schuller et al.	Healthcare Cost and Utilization Project's (HCUP) Nationwide Inpatient Sample	Adults 65 and older with Medicare	Interrupted time series with Poisson regression growth curve	None (only looked at catheter-associated urinary tract infections)
Dec. 2014	Gidwani & Bhattacharya	Healthcare Cost and Utilization Project's (HCUP) Nationwide Inpatient Sample	Medicare patients receiving hip or knee surgery compared to non-Medicare patients receiving hip or knee surgery	Difference-in-differences estimation with control group and hierarchical regression models	Statistically significant (only looked at pulmonary embolism and deep-vein thrombosis)
Jan. 2015	Waters et al.	National Database of Nursing Quality Indicators	All patients	Negative- and β- binomial models	Statistically significant for infections
March 2015	Vaz et al. (similar to original Lee team)	Centers for Disease Control and Prevention's National Health Safety Network	All patients	Interrupted time series with comparison series, comparing safety net to non-safety net hospitals	None (only looked at infections)
May 2017	Thirukumaran et al.	Healthcare Cost and Utilization Project's (HCUP) State Inpatient Database for New York	Medicare stays for acute myocardial infarction, congestive heart failure, pneumonia, and stroke in New York state	Difference-in-differences estimation with logistic regression models, compared across Medicare Utilization Ratio quartiles	Statistically significant for hospitals with a high Medicare load

My paper will seek to improve the existing literature in several ways. First, I further integrate economic theory of pay-for-performance policies into my analysis. I build on Thirukumaran et al. (2017) by considering the magnitude of financial incentives more carefully, specifically by introducing a fiscal pressure index based on economics papers that analyze other Medicare payment shocks. I also analyze a more representative cohort of patients, in contrast to Thirukumaran et al. (2017), whose results are only generalizable to patients with four primary diagnoses that account for less than one sixth of Medicare hospitalizations. My approach studies a shorter timeline to better isolate Nonpayment's effect from other policy trends.

Because of issues like free riding and other organizational factors, group size can influence whether incentives are impactful for providers. To better address this, I plan on considering hospital size, measured by number of beds, as well as other hospital-level variable like ownership and teaching status (Eijkenaar, 2011). Other hospital-specific considerations, such as a proxy for average occupancy rate, will also be uniquely considered by my paper.

Lastly, I will account for hospital "intensity" in my model, by controlling for length of stay. Increasing intensity results in higher costs for hospitals, but may be a valid strategy for attracting profitable patients (Hodgkin and McGuire, 1994).

III. Theoretical Framework

To formally account for how hospitals balance intensity and profit, Hodgkin and McGuire (1994) estimate a model. The following demonstrates providers' theoretical volume and intensity responses to a payment change:

(1) $U = U(\pi, I)$
(2) $\pi = R - TC + Y$
(3) $R = pX$
(4) $p = \alpha + \beta c$
(5) $TC = cX$
(6) $c = c(I)\ where\ c' > 0$
(7) $X = X(I)\ where\ X' > 0$

In this model, providers draw utility from profit, π, and intensity of care, *I,* assuming hospitals are concerned about quality of care (1). Profit is the sum of revenue, R, and outside income Y (such as donations), minus total costs (2). Revenue depends on the volume of admissions, *X,* and the average price per admission (3). This price p has a fixed component, α, as well as a variable component, β, which depends on cost (4). Total cost then depends on the cost per discharge, *c,* times the volume of discharges (5). Cost per discharge increases with intensity (6); higher intensity may in turn attract more patients, according to equation (7). This set of equations suggests a fundamental relationship between fiscal pressure and healthcare provision; the two should be considered in tandem, motivating my choice to include a fiscal pressure index.

However, this model is simplistic. It assumes one payer, one hospital, and a single type of discharge. While it acknowledges hospitals' competition for patients within a market, it does not explain how market share might impact a provider's behavior. It also assumes that hospitals are always paid enough to operate (Hodgkin and McGuire, 1994). In reality, a hospital's cost per discharge is not exogenous, and the volume of admissions depends on more than just intensity. While the relationship between costs and admissions is an important fundamental insight, the hospital market is far more complex than this model allows.

To capture potential unobservable changes in provider selection, quality of care, and intensity, I consider length of stay, which has been significantly impacted by responses to Medicare policies in the past. While it's not clear that this reduction in length of stay meant that patients were discharged quicker and sicker, hospitals may be motivated to prioritize simpler cases. Competitive hospitals may decrease the length of stay for high-cost, complex patients by reducing the intensity of care (Ellis and McGuire, 1996). This spurs a demand-side effect of lower admissions for patients who can expect lower intensity. This theory holds even without the assumption that hospitals act as profit-maximizers (Hodgkin and McGuire, 1994). If the financial disincentives created by Nonpayment were powerful, they may result in a shorter LOS for patients in the treatment group. However, this relationship has not held true empirically for hospitals with low occupancy rates, which face greater fiscal pressure and tend to source demand from all patient types (Hadley et al., 1989). To capture this nuance, I will include a measure of hospital occupancy, calculated as average daily census divided by total number of beds.

In my analysis, I also include patient- and hospital-level control variables, guided by theories of hospital behavior. For example, it has been shown that private for-profit, private not-for-profit, and government hospitals operate under different budget constraints. Though decision-makers across these types are all motivated by profit, government hospitals face softer financial incentives since local governments may alter funding in response to changing payments, diluting a policy's impact. The 1990 California Disproportionate Share Program, which allocated greater funding to hospitals that helped the indigent, produced significant responses from private hospitals and no response from public hospitals (Duggan, 2000). Therefore, it is important that studies of financial incentives control for hospital ownership type. I also proxy a patient's socioeconomic status with the state-level median household income quartile for the patient's county, and include data on county-level income inequality. Given results from Chen and Lakdawalla (2019), which show that providers facing increases in Medicare reimbursement boost utilization by 10% more for richer patients, socioeconomic status may influence both provider actions and patient severity. This may be because richer patients tend to request more care. Other patient demographic variables are standard.

To address selection of profitable patients, I condition on weekend admission. Patients admitted on the weekend tend to be sicker, as those who have choice in their admission tend to go to the hospital during the week, since care is more accessible and convenient. Theoretically, hospitals also are less able to turn away patients who are more severe and arrive on weekends (Barnett et al., 2002). My paper is the first in the Nonpayment literature to address this.

Hospitals also face various degrees of competition for patients. Market concentration can be calculated using the hospital referral region-level Herfindahl-Hirschman Index (HHI). This index is calculated by using a hospital's average daily census to compute its market value (Jacobson et al., 2017). In a model that assumes physicians care about their own utility as well as patient welfare, Jacobson et al. (2017) show that market concentration impacts patients through two measures: 1) the level of healthcare provided, and 2) the selection of patients into treatment. Their model predicts that more concentrated markets will have less elastic supply curves as a consequence of market power, implying that hospitals in competitive areas will respond more to changes in payment. These hospitals are also more able to turn away unprofitable patients.

When analyzing the impact of financial incentives on an agent's behavior, it is also important to consider organizational variables. Modeling provider behavior is challenging because agents engage in multidimensional tasks, meaning they balance multiple responsibilities. Financial incentives allocate risk, as well as encourage agents, like nurses and doctors, to shift attention between their various duties (Holmstrom and Milgrom, 1991). Incentives may be least effective, or even harmful, in large hierarchies, since agents are able to shift their attention from unmeasured outcomes and efforts (or team efforts), to measured outcomes (or individual efforts) with low risk of detection. For these reasons, the policy's impact may be influenced by job design, or how tasks are grouped for each agent (Holmstrom and Milgrom, 1991). Addressing the consequences of variation in job design is, of course, impossible, but can be attempted by including variables like bed size and teaching status.

Economic analyses of hospital payment changes typically include a "bite" variable that estimates the degree of fiscal pressure each hospital faces (Hodgkin and McGuire, 1994). This is not simply a measure of Medicare utilization, as used by Thirukumaran et al. (2017). Rather, it involves the interaction of each hospital's approximate payment rate with a measure of its dependence on Medicare (Hodgkin and McGuire, 1994). For example, Hadley et al. (1989) construct the following fiscal pressure index (FPI) for rate changes between 1984 and 1985.

$$FPI_{i,t} = \frac{\left[\left(PPSRT_{i,t} - MCPC_{i,t-1}\right) * MCRDCH_{i,t-1}\right]}{TOTEXP_{i,t-1}}$$

In this index, PPSRT = payment rate per case; MCPC = Medicare cost per case; MCRDCH = total Medicare discharges; and TOTEXP = total expenses. These values are measured as percent changes from 1984 and 1985. However, its exact structure is less pertinent to my analysis, as I am looking for changes over a period of time rather than annual fluctuations. Staiger and Gaumer (1990) define a simpler "bite" variable, calculated by multiplying the payment rate times Medicare's share of hospital costs. They find that this variable had far more explanatory power than the payment level on its own. Given my data and the objective of my analysis, I will estimate a similar fiscal pressure index of my own by interacting a hospital's cost-to-charge ratio with its Medicare utilization ratio. This formula aggregates a hospital's Medicare profitability with its dependence on Medicare, rather than estimating each separately. Broken into its components, this index will be estimated as follows:

$$FPI_i = \frac{Total\ Medicare\ Costs_i}{Total\ Medicare\ Charges_i} * \frac{Inpatient\ Days\ Financed\ by\ Medicare_i}{Total\ Inpatient\ Days_i}$$

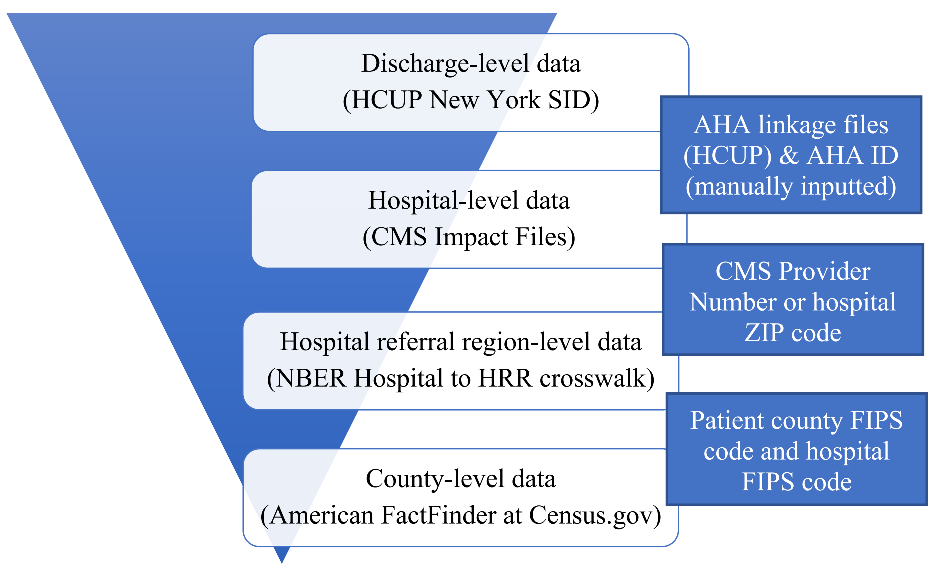
The first component of this FPI represents Medicare profitability for provider i, while the second component measures reliance on Medicare admissions. As previously explored, Medicare profitability and volume may influence each other, necessitating a variable that interacts them. This is more sophisticated than the simple Medicare utilization ratio used by Thirukumaran et al.

(2017). By introducing a fiscal pressure index, my paper will be only the second in the literature on Nonpayment to consider how a provider's financial situation impacts its policy response. My analysis will be the first in the literature to take an economic (rather than medical) approach. Given the division in the literature on whether Nonpayment was influential, introducing this approach offers a new perspective toward a pressing patient safety problem. After considering a variety of variables, I will analyze my primary dependent variable: a HAC indicator. This is distinct from most previous researchers, who were more concerned with specific types of HACs.

IV. Data Overview

I draw from several datasets, whereas earlier papers have lacked patient- or hospital-level detail. After manually inputting the American Hospital Association (AHA) identifier and FIPS county code for each hospital, I can merge discharge-level, hospital-level, hospital referral region-level, and county-level data across more than 11 million observations between 2006-2010. The graphic below shows how my chosen datasets were combined, and by which criteria they were merged (indicated by the blue boxes).

FIGURE 2: Data Flow Chart



As the literature has progressed, data have become more detailed and representative. Lee et al. (2012) rely on a dataset from the Centers for Disease Control (CDC) that lacks information on payer and appears skewed toward hospitals that were located in the Northeast. Importantly, data from the CDC do not include any information on payer, making it impossible to distinguish Medicare patients from non-Medicare patients beyond age indications. This data limitation forces the assumption that hospitals will treat Medicare and elderly non-Medicare patients the same, despite differing financial incentives (Lee et al., 2012). Similarly, the National Database of Nursing Quality Indicators (NDNQI), used by Waters et al. (2015), is limited across time, and is now proprietary. One strength of the NDNQI is that it may be less susceptible to coding manipulation than claims data, since hospitals self-report to the NDNQI. However, by comparing Medicare patients aged 65-69 to privately insured patients in the same age group, Gidwani and Bhattacharya (2014) show that no coding changes seemed to arise after the policy was implemented. These findings indicate that government-mandated data may be no less reliable than self-reported datasets. Data for certain conditions of interest, such as central-line associated bloodstream infections and catheter-associated urinary tract infections, were not added to the NDNQI until late 2007, limiting pre-implementation analysis to only three quarters of data. Overall, the NDNQI lacks detail compared to larger datasets, and is not used in this paper.

In contrast to papers that use the NDNQI or CDC data, Thirukumaran et al. (2017) employ a merging strategy, similar to mine, that aims to capture the magnitude of a hospital's financial incentives. Data from HCUP, a family of databases sourced by a partnership between the Agency for Healthcare Research and Quality and statewide data organizations, are larger and more detailed than other datasets. HCUP databases constitute the largest uniform collection of all-payer inpatient administrative data. For my purposes, The National Inpatient Sample (NIS) and State Inpatient

Database (SID), two databases maintained by HCUP, are most relevant. These datasets provide discharge records from hospitals at the patient, provider, market, and state levels. The NIS approximates a representative sample of 20% of U.S. community hospitals, and is used to estimate a policy's national impact, although it does not include data from every state, and some important variables are missing across states (Steiner et al., 2002). To remedy issues of data availability in the NIS, my paper will use the New York SID. This smaller dataset also allows for manual collection of certain variables and computationally complex methods like covariate matching.

Discharge-Level Data

From 2006-2010, the New York SID includes data for over 13 million discharges and over 200 hospitals. Of these, about 80% are community hospitals that would have been subject to the Nonpayment rule. The NY SID has a few advantages compared to other states. First, it includes the present-on-admission indicator for several years before 2008. This allows me to confirm whether a condition was hospital-acquired, avoiding a potential overestimate of HACs (Houchens et al., 2008). Second, New York data include a variable for median household income quartile for patient ZIP code; this is important for proxying a patient's socioeconomic status, particularly for patients that may travel to receive medical care. Lastly, New York data is far more complete than the NIS or other state databases, with relatively few missing values for variables of interest.

These discharge records contain demographic information about the patient, length of stay, payer status, comorbidities that might complicate care, condition(s), and treatment(s) received. Substantial pre- and post-implementation data is available. The NY SID also includes up to fifteen diagnoses per discharge, allowing HACs to be identified by their ICD-9-CM diagnosis codes. Along with the SID, the supplemental AHA linkage files are necessary for merging hospital data with discharge data, via a common hospital identifier.

Hospital-Level Data

Second, I use hospital-level data from the CMS Impact Files, provided each fiscal year. Thirukumaran et al. (2017) merge the SID with data from CMS to include Medicare utilization ratio, transfer-adjusted case mix, disproportionate patient percentage, and operating profit margin. Similarly, Waters et al. (2015) use CMS cost reports and Impact Files to collect data on Medicare case mix and total profit margin. These variables are all available for free download in Stata format via the NBER. The manageable size of state-level hospital data allows me to manually input several variables of interest that were not originally included, as detailed in Section IV.

Hospital Referral Region-Level Data

Third, a hospital to hospital referral region (HRR) crosswalk is available via the NBER, allowing me to calculate the Herfindahl-Hirschman Index, a measure of market concentration, at the HRR level. Hospital referral regions are geographic boundaries defined by the Dartmouth Atlas of Healthcare and used by CMS and other health data researchers. They are bounded based on referral patterns for Medicare beneficiaries; must have a population of at least 120,000; and must receive at least 65% of hospitalizations within the region. Because they are based on referrals, HRRs define the market that a hospital competes within.

In the crosswalk, each unique provider number is matched to an HRR, enabling a many-to-one merge. HRR crosswalks were available for the years 2005, 2007, and 2010. In case hospital market consolidation had impacted HRRs during the period, I used the most recent HRR file for each fiscal year. Since consolidation occurs continuously, this approach is imperfect, but it is the most granular approach possible given the available data. Matches were not made for just under 10% of hospitals in each year. To remedy this, I used NBER's ZIP code crosswalks to manually match each hospital to its HRR based on its city and ZIP code for the appropriate year. All hospitals were eventually matched.

The HHI was calculated by squaring each hospital's market share percentage, then summing the resultant values for each hospital referral region. 13 hospitals incorrectly appeared to have very high market concentrations, since their HRR was primarily located in bordering states like Pennsylvania or Vermont (ArcGIS). To correct this issue, I coded a dummy variable, HRR_NY, coded 1 if the hospital referral region was located in New York and 0 otherwise. In my empirical specification, I interact this dummy with each hospital's HHI. This effectively drops the HHI for hospitals that operate in regional markets outside of New York state.

County-Level Data

To capture a patient's socioeconomic status, the New York SID includes the variable MEDINCSTQ; this measures the median household income state quartile for the patient's ZIP code. This quartile measure, however, is weak – there are only four possible values. To strengthen this measure, I include the county-level Gini Coefficient using PSTCO2, or patient county FIPS code. This measure of income inequality is interacted with the median household income state quartile to more comprehensively account for socioeconomic status. I obtain the 2006 Gini Coefficient for each U.S. county via American FactFinder, a tool provided by Census.gov. The 2006 values were chosen to avoid any unusual skew resulting from the financial crisis and short-term spikes in unemployment. Then, I merge on patient FIPS code. The average county-level Gini Coefficient in New York state was 0.432. New York County, or Manhattan, was the least equal, with a Gini Coefficient of 0.599; St. Lawrence County, which borders Canada, was most equal, with a Gini Coefficient of 0.387. Gini Coefficients were obtained for all U.S. counties, rather than just New York state, because some (less than 10% in this sample) patients travel across state lines for hospital care. Some datapoints were missing for small counties. Ultimately, 95.2% of observations were matched with a Gini Coefficient.

I also merge in county-level median household income for hospital ZIP codes, obtained from the U.S. Census website. Due to data availability, I am restricted to data from 2010. The lowest median household income is $32,568 for Bronx County, compared to $91,204 in Nassau County, part of Long Island. The median household income is $50,640, with a mean of $54,453.

Collecting and Managing Data

For the years I study, SID data is stored on discs as ASCII text files. Stata load programs were only available for SID datasets beginning in 2014. To solve this problem, I manually recoded the 2014 Core Stata load program five times, for years 2006-2010. I used file specifications for each year, along with historical SPSS load programs, to account for any changes between 2014 and the years of interest. These changes included various variable names, addition of new variables and deletion of old ones, and ordering of the approximately 225 variables. The load programs accomplished three things: 1) naming and specifying variables and values in Stata, 2) labeling each variable, and 3) recoding missing or irrelevant values for each variable. (For example, birthweight was only recorded for babies)[2].

Next, I prepared the data for merging. Data in the SID can be linked using a hospital identifier. HCUP includes three identification variables for each provider: 1) a unique HCUP hospital identifier that is used to link supplemental files, like the cost-to-charge ratio files or AHA Linkage Files, 2) American Hospital Association (AHA) identifier, and 3) the state data source's hospital identifier. In contrast, datasets from CMS include each hospital's name and "provider number". I recoded five years of load programs for the AHA Linkage Files, which was necessary for merging the core SID files with hospital-level data, via the AHA identifier.

[2] These files are available upon request.

Thirukumaran et al. (2017) merge the SID and CMS datasets using the free online hospital look-up provided by Health Forum, an affiliate of the American Hospital Association. This tool allowed the user to look up a hospital's address based on its AHA identifier. From there, the online American Hospital Directory can be used to look up each provider number and manually link the SID to data from CMS. However, the Health Forum tool is no longer available, forcing me to look elsewhere. I eventually uncovered an online directory provided by the American College of Surgeons and the Commission on Cancer, which displays state, city, facility name, and AHA ID, described as "facility identification number" (CoC Datalinks). Focusing on the New York portion of the directory, I manually matched hospitals in the CMS dataset to their AHA ID using hospital name. Where there were naming ambiguities (i.e., four variations of St. Luke's Hospital), I used the facility's city to confirm a match based on county. I matched remaining hospitals based on address, using the U.S. News and World Report website, which contains the AHA ID in the URL of each hospital's profile page. The American Hospital Directory, which summarizes information on all New York community hospitals, allowed me to search hospitals by name, confirm with the provider number, and match by address. This process resulted in successful AHA ID matches for all hospitals included in my final sample. Next, I dropped all hospitals outside of New York from the FY2007-FY2011 Impact Files, and merged the AHA IDs from the cleaned 2006 dataset.

Finally, I merged the fiscal year Impact Files with the calendar year SID files. Based on the quarter of each discharge, DQTR, I created a "fyear" variable in each SID file to match the Impact Files, since only the first three quarters of a calendar year correspond with that same fiscal year. From 2008-2010, the discharge quarter was missing for about 3% of discharges, resulting in those observations being dropped. This process resulted in a two-way merge on AHA ID and fiscal year. After merging, I dropped observations from non-community hospitals that would not have been covered by the Medicare policy (~20% of state discharges). Some hospitals that were present in the Impact Files did not successfully match to the SID; this is likely because the Impact Files seem to lag hospital operations. For example, some hospitals that announced closing before 2006 were still present in the fiscal year file, despite an average daily census of zero, meaning they were no longer operating. Some hospitals also closed, merged, or were added to the IPPS from 2006-2010. I provide a list of these 18 hospitals in Appendix I. They do not appear substantially different from the final sample. Once I finished merging, my final dataset contained 155 community hospitals.

To determine a hospital's teaching status, I created a new indicator variable, hosp_teach, and set it equal to 1 if resbed > 1, and 0 otherwise, since only teaching hospitals employ residents[3]. I created a second variable, hosp_owner, to capture ownership type; in my final regressions, I include an indicator for private nonprofit hospitals, and a second indicator for government hospitals. (It is important to note that New York state does not have for-profit hospitals, though one appeared in the data, potentially grandfathered in). I manually filled in this variable using the 2008 American Hospital Association guide to the Healthcare Field, which lists each hospital's provider number and ownership type ("AHA Guide"). For the eight hospitals that were not open at the time of the Guide's publication, older documents from CMS or the American Hospital Directory were used to determine ownership type. For each discharge, I also calculated the Charlson Comorbidity Index (CCI), which reflects the likelihood of one-year mortality based on the severity of a patient's comorbidities, or pre-existing conditions (for example, diabetes or liver disease). I downloaded a Stata module from RePEc, a database of collaborative tools for economics research, that used the ICD-9-CM diagnosis codes in the SID to identify the seventeen comorbidities necessary to calculate the weighted Charlson Index (Stagg 2006).

[3] Note that this is a somewhat broad definition of "teaching" hospital. There is literature that addresses differences between "major" and "minor" teaching hospitals, which is not addressed in this paper.

Lastly, I estimated the hospital-level fiscal pressure index. I had two potential estimates of each provider's cost-to-charge ratio: the first, provided in supplemental files that accompany the SID, was the all-payer inpatient cost-to-charge ratio (APICC). However, the SID excludes a value if cost information is missing or if it is considered an outlier, resulting in about 15% of APICC values coded as missing from my data. Also, this variable is calculated across all payers; my analysis is specific to Medicare. So, instead, I chose to use the ratio of Medicare operating costs to Medicare covered charges, as estimated in the Impact Files. To calculate the fiscal pressure index, I interact this measure of Medicare profitability with a measure of Medicare utilization: the proportion of Medicare days as a percent of total inpatient days. This ratio was also provided in the Impact Files.

HAC Identification

I identified HACs using diagnosis codes that were named in the policy, which I summarize below. HACs are identified using ICD-9-CM codes. The first three digits indicate diagnostic category, while digits following the decimal point further detail the condition. For example, codes beginning with 590 indicate a kidney infection, while subsequent digits might indicate whether the infection is chronic or acute. The SID data include up to fourteen secondary diagnoses per patient. I ultimately excluded certain surgical site infections that specified a principal diagnosis of 278.01 or 998.59 following certain procedures, since identifying this HAC would have required me to know whether a patient had undergone the procedure during a previous inpatient stay.

TABLE 2: Hospital-Acquired Conditions Diagnosis Codes ("Medicare Learning Network")

Hospital-Acquired Condition	International Classification of Diseases ICD-9-CM Diagnosis Code(s)
Air embolism	999.1
Blood incompatibility	999.60, 999.61*, 999.62*, 999.63*, 999.69
Foreign object retained after surgery	998.4, 998.7
Stage III and IV pressure ulcers	707.23, 707.24
Falls and trauma	800–839, 850–854, 925–929, 940–949, 991–994
Manifestations of poor glycemic control	250.10–250.13, 250.20–250.23, 251.0, 249.10–249.11, 249.20–249.21

Catheter-associated urinary tract infection	996.64, 112.2, 590.10, 590.11, 590.2, 590.3, 590.80, 590.81, 595.0, 597.0, 599.0
Vascular catheter-associated infection	999.31
Certain surgical site infections (following certain procedures)	519.2 following any procedure 36.10–36.19; 996.67, 998.59 following any procedure 81.01–81.08, 81.23, 81.24, 81.31–81.38, 81.83, or 81.85; principal diagnosis of 278.01, 998.59 following any procedure 44.38, 44.39, or 44.95
Deep vein thrombosis / pulmonary embolism (following certain orthopedic procedures)	415.11, 415.19, 453.40–453.42 following any procedure 00.85-00.87, 81.51-81.52, or 81.54

*Zero occurrences in the NY data

According to these guidelines from CMS, hundreds of combinations of procedures and secondary diagnoses may result in a HAC included in the policy. I generated a new indicator variable, HAC, coded "1" if any of the 14 potential secondary diagnoses matched the ICD-9-CM codes listed above and was not coded as "present on admission", as indicated by a dummy variable for each secondary diagnosis. I only code a HAC for surgical site infections and deep vein thrombosis/pulmonary embolism if the patient had undergone any of the specified procedures, as determined by PR1-PR15, which coded up to 15 procedures per discharge.

The present-on-admission indicator, necessary for accurately identifying whether a condition was truly hospital-acquired, was coded as a numeric variable with values 0 or 1 and named "DXADADMIT1-15" in 2006; however, it was recoded as a string variable with values "N" and "Y" and named "DXPOA1-15" in subsequent years of data. For the 2006 observations, I replaced DXPOA1-15 was with the appropriate values to match the rest of the data. These conditions were combined to code my primary dependent variable, "HAC". [4]

Variables of Interest

TABLE 3: Patient- and Hospital-Control Variables

Continuous Variable	Rationale	Data Source	Predicted Impact on HAC Rate
Age (AGE), sex (FEMALE), and race (RACE) of patient	Standard demographic variables	New York SID	Age: +
Charlson Comorbidity Index (charlindex)	Measure of patient illness / severity	New York SID	+

[4] The code for HAC identification in the SID is available upon request.

		+ Stata module from Stagg (2006)	
Patient-level median county income quartile for the state (MEDINCSTQ), interacted with Gini Coefficient (GINI)*	Proxies a patient's socioeconomic status	New York SID and American FactFinder	-
Length of stay, in days (LOS)*	May change as a result of provider selection (expected to fall for Medicare patients)	New York SID	+
Categorical Variable	**Rationale**	**Data Source**	**Predicted Impact on HAC Rate**
Payer status (PAY1)	Allows identification of Medicare patients	New York SID	Higher for Medicare
Weekend admission (AWEEKEND)*	Measure of ability to select patients and patient severity	New York SID	Higher for weekend admissions
Hospital ownership status (hosp_owner)	Differential financial incentives	American Hospital Association Guide (manually inputted)	Private hospitals will experience greater change
Teaching status (hosp_teach)	Standard hospital-level control variable	CMS Impact Files (imputed based on resident-to-bed ratio)	Not clear
Large urban, other urban, or rural (urgeo1, urgeo2, urgeo3), recoded as binary variables and omitting one due to collinearity	Standard hospital-level control variable	CMS Impact Files	Rural hospitals will have higher rates

TABLE 4: Organizational Variables

Continuous Variable	Rationale	Data Source	Predicted Impact on HAC Rate
Number of beds (beds) interacted with average daily census (adc)	Measures hospital size and resource level	CMS Impact Files	-
Disproportionate Patient Percentage (dshpct)	Accounts for a provider's resource level	CMS Impact Files	+
Transfer-Adjusted Case Mix (tacmiv)[5]	Measures a provider's reliance on Medicare	CMS Impact Files	+
Fiscal Pressure Index (FPI)*	Estimates a provider's sensitivity to payment changes	Calculated using variables from CMS Impact Files	+ pre-policy - post-policy
Hospital referral region-level Herfindahl-Hirschman Index (HHI), interacted with border indicator (HRR_NY)	May impact response to fiscal pressure	CMS Impact Files and NBER HRR crosswalk	+

* Indicates a new variable that has not previously been considered by any Nonpayment analysis

[5] According to the Centers for Medicare and Medicaid Services, the transfer-adjusted case mix index "represents the average diagnosis-related group (DRG) relative weight for that hospital. It is calculated by summing the DRG weights for all Medicare discharges and dividing by the number of discharges". In other words, it proxies overall patient severity at the provider-level.

Summary Statistics

Below, I summarize data for the 155 community hospitals in my final sample, which represent around 80% of state discharges, at both the patient- and hospital- level. Hospital-level statistics are calculated across all six fiscal years of data, for a total of 930 hospital observations.

TABLE 5: Patient Descriptive Statistics

Variable	Obs.	Mean	SD	Median	Min	Max
Age[6]	11,275,207	48.42	27.85	52	0	119
Charlson Index	11,275,375	1.15	1.82	0	0	20
Median Income Quartile for Patient ZIP	10,506,700	2.54	1.20	2	1	4
Length of Stay	11,275,020	5.49	8.88	3	0	365

Variable	Frequency
Weekend Admission[7] (*N = 11,275,375)*	19.49%
Female[8] (*N = 11,275,050)*	57.03%
Race Categories (*N = 11,110,536)*	
White	56.90%
Black	18.11%
Hispanic	13.55%
Asian, Native Hawaiian, or other Pacific Islander	3.60%
Native American	0.65%
Other	7.18%
Admission Type[9] (*N = 11,259,190)*	
Emergency	60.84%
Urgent	9.93%
Elective	19.45%
Newborn	9.77%
Trauma Center	0.01%
Primary Payer (*N = 11,275,375)*	
Medicare	36.15%
Medicaid	24.78%
Private Insurance	31.52%
Self-Pay	5.33%
Other	2.15%
Hospital-Acquired Condition, coded 0 or 1[10] (*N = 11,275,375)*	1.11%

[6] While a maximum age of 119 seems improbably high, HCUP does not report any “known data issues” with the AGE variable in any year of New York data.
[7] Note that, if randomly distributed, the proportion of weekend admissions would be 28.57%. The relatively low percentage observed in the data is consistent with the idea that weekend admissions tend to be truly emergent.
[8] The high proportion of female patients is presumably because women spend time in the hospital during childbirth and live longer. More than 75% of hospital stays for patients in this sample between the ages of 20 and 35 are women, and nearly 67% of hospital stays for patients over 80 are women. Excluding these age groups eliminates the skew toward female patients.
[9] Emergent admissions are classified as more threatening than urgent admissions, while trauma admissions involve a designated trauma center.
[10] This rate is in line with what previous researchers have found.

TABLE 6 : Hospital Descriptive Statistics

Variable	Obs.	Mean	SD	Median	Min	Max
Bed Size	930	257.11	225.86	195	18	1,800
Average Daily Census	930	190.02	200.33	129.9	3	1,549
Disproportionate Patient Share	930	33.80%	.253	24.50%	2.08%	83.20%
Transfer-Adjusted Case Mix Index	930	1.360	.264	1.320	.424	2.348
Ratio of Medicare Operating Costs to Medicare Charges	930	.431	.143	.423	.122	1.148
HHI (non-border)	930	242.88	530.56	13.365	0	3,747
FPI	930	.199	.102	.176	.050	.868

Variable	**Frequency**
Geographic Designation (*N = 930)*	
Large Urban	64.52%
Other Urban	19.03%
Rural	16.45%
Hospital Ownership Status (*N = 930)*	
Not-for-profit	87.10%
Government	12.26%
For-profit	0.65%
Teaching Hospitals (*N = 930)*	53.55%

The typical inpatient stay in New York from 2006-2010 involved a white, middle-aged female patient, covered by private insurance or Medicare; she likely stayed for less than one week at a large, urban, teaching, not-for-profit hospital that earned a profit from Medicare, as indicated by the less-than-one mean value for the ratio of Medicare operating costs to Medicare charges. Some of the above variables may act together to impact patient severity or hospital fiscal pressure. Here, I summarize correlations between my continuous patient- and hospital- level variables, calculated using all years of data.

TABLE 7: Variable Correlation Matrix

	Age	*Charlson Index*	*Median Income Quartile for Patient ZIP*GINI*	*Length of Stay*
Age	1			
Charlson Index	0.4008	1		
Median Income Quartile for Patient ZIP*GINI	0.0872	-.0179	1	
Length of Stay	0.1206	0.1479	-0.0068	1

	Bed Size	*Average Daily Census*	*Disproportionate Patient Share*	*Transfer-Adj. Case Mix Index*	*Occupancy rate proxy*	*HHI*HRR on border*	*FPI*
Bed Size	1						
Average Daily Census	0.9832	1					
Disproportionate Patient Share	0.1960	0.1979	1				
Transfer-Adjusted Case Mix Index	0.5929	0.6156	-0.0685	1			
Occupancy rate proxy	0.5100	0.6035	0.2304	0.6240	1		
HHI*HRR on border	0.1145	0.1005	-0.0688	0.1028	0.0110	1	
FPI	-0.4865	-0.4934	-0.4388	-0.4117	-0.5670	0.0672	1

Some interesting observations emerge here. Unsurprisingly, older and poorer patients tend to be sicker, as measured by the Charlson Index. They also tend to have longer hospital stays. At the provider level, larger hospitals see higher volume, and treat a higher proportionate of Medicare and Medicaid patients (indicated by the disproportionate patient share). Despite this, Medicare is more profitable for larger hospitals, which tend to have much lower fiscal pressure indices. As expected, hospitals that tend to have a high ratio of average daily census to beds, captured by the occupancy rate proxy variable, experience much lower fiscal pressure from Medicare.

It may seem intuitively strange to see a negative correlation between disproportionate patient share and the transfer-adjusted case mix index, which are also both associated with a lower fiscal pressure index. The former is a measure of how many Medicare and Medicaid patients use the hospital, and the latter measures the overall severity of patients. However, the high positive correlation between bed size and transfer-adjusted case mix index may indicate that larger, resource-rich hospitals attract patients with severe cases; these patients may be wealthier and less likely to be on Medicare or Medicaid. CMS also factors these variables into its payments for hospitals, which is likely responsible for negative correlations with the FPI.

Though not displayed in Table 7, it should also be noted that geographic classification (large urban, other urban, or rural) strongly predicts fiscal pressure index. No rural hospitals fall into the first quartile of FPI, and very few fall into the second. Large urban hospitals, however, overwhelmingly fall into the first two quartiles.

V. Empirical Specifications

Mahalanobis Matching and Difference-in-Differences

The quasi-experimental nature of the Nonpayment policy enables a pre-post research design. As discussed in Section II, previous research has revealed that hospitals can and do discriminate based on payer when Medicare implements a reimbursement change. If Nonpayment meaningfully altered hospitals' prevention of HACs, there is likely to be a difference in HAC rate between the non-Medicare (control) and Medicare (treatment) populations following the policy's implementation. I will explore two empirical approaches to examine whether this is true. First, I estimate a difference-in-differences (diff-in-diff) model using a control group of non-Medicare patients who are matched to Medicare patients based on similar characteristics. However, there may be selection issues that arise from comparing patients that are younger, on average, than their Medicare counterparts. To address this, I also estimate a fuzzy regression discontinuity, which only compares Medicare and non-Medicare patients within a narrow age range. My paper is the first Nonpayment analysis to incorporate a regression discontinuity approach.

A diff-in-diff design requires me to assume that time trends in HAC outcomes were parallel for non-Medicare and Medicare patients prior to October 2008. A quick assessment of quarterly HAC rates suggests that HACs were worsening for the Medicare population more quickly than they were worsening for non-Medicare patients, probably because Medicare patients spend more time in the hospital, and are therefore most likely to be impacted by poor quality standards.

The primary challenge when working with discharge-level data is addressing such heterogeneity, since patients differ between populations and over time in unobservable ways. To identify a comparable non-Medicare sample, I use covariate matching, a common technique used to address heterogeneity in health policy applications. This approach will provide robustness against a potential violation of the parallel time trends assumption. I choose Mahalanobis matching due to potential issues with propensity score matching, and evidence that Mahalanobis matching may provide a more balanced, less biased sample (King and Nielson, 2019).

This method employs an algorithm defined by Leuven and Sianesi (2003) that calculates each observation's Mahalanobis distance. Essentially, each discharge is assessed based on observable factors other than age, and assigned a short distance if it looks like a Medicare patient. This "nearest neighbor" approach matches Medicare patients to their most closely matched non-Medicare counterpart. An illustration of Mahalanobis matching is provided in Appendix II. After comparing HAC trends based on various variables that might impact patient severity, I choose four criteria for matching: race, sex, length of stay in days (LOS), and Charlson Index (charlindex). To make the race variable binary, I create three new indicator variables: one indicating white, another indicating black, and a third indicating all other races[11]. Next, I rescale the values for LOS and Charlson Index to fall between 0 and 1, so that differences in variable scaling do not impact matching. I redefine PAY1 as a binary variable, called MEDICARE. I also condition on weekend admission (AWEEKEND), since these admissions tend to be emergent, rather than elective. I choose to condition on weekend admission rather than labeled emergent admissions, since 1) admission type is more likely to be biased by child births, and 2) elderly patient admissions are more often categorized as "emergent" simply based on patient age[12]. This restricts analysis to 2,197,913 observations, of which 36.55% are Medicare patients.

[11] This strategy ensures that there is enough variation within each matching criterion. Otherwise, small groups like Alaskan Natives are very difficult to match well.
[12] The proportion of Medicare admissions classified as "emergent" is more than double the proportion of Medicare patients in the sample, indicating that the emergent classification may be biased by age.

This procedure results in 100% common support, meaning all observations are matched. About 75% of control observations are matched to several treated observations. To check the balancing assumption – meaning matching criteria are balanced between the treatment and control group – I compare differences in the mean value for each matching variable. If the groups are balanced, differences in the means are not statistically significant, and bias is negligible for all criterion.

TABLE 8: Mahalanobis matching balancing test

Variable	Mean			t-test		V(T) / V(C)
	Treated	Control	% Bias	t	p > | t |	
LOS (scaled)	.01859	.01857	0.1	0.14	0.887	1.01
Charlson Index (scaled)	.09866	.09856	0.1	0.20	0.844	1.00
Female	.57764	.57763	0.0	0.01	0.992	--
White	.6793	.67931	-0.0	-0.01	0.996	--
Black	.15164	.15164	0.0	0.00	1.000	--
Other Non-White	.15921	.15919	0.0	0.01	0.995	--

Based on this test, none of the matching criteria is significantly different between the treatment and control group, meaning the resultant sample is balanced. I also test the sample based on age. Because age mostly predicts Medicare "treatment", there is no way the mean value between these two groups can be insignificant. However, the age distribution should be similar enough to show that Medicare patients are not being matched to very young patients. I find that the average age in the non-Medicare matched sample is 49. Although these age differences are necessarily significant, matching does not skew toward extremely young patients.

A diff-in-diff approach also requires me to assume that there are no spillover effects between the two groups, and that the composition of both groups is stable over the entire time period. In other words, any "common shocks" in the post-intervention period affected both groups equally[13] (Columbia University). I examine the latter assumption by comparing patient descriptive statistics for discharges before October 2008 to discharges in later quarters. There are no meaningful

[13] This is further justification for choosing a study time period that ends in 2010, before Medicare-specific policies from the Affordable Care Act went into effect.

differences, other than a small shift toward Medicare and Medicaid in the post-treatment period (35.75% to 36.56% and 24.56% to 25.01%, respectively). However, potential spillover effects may limit the explanatory power of this approach. If hospitals overall improve quality as part of the effort to reduce HACs, then my results will be biased toward zero. For robustness, I run the same model with a dependent variable that should not be impacted by Nonpayment. Although Lee et al. (2012) and Vaz et al. (2015) compare policy-defined HACs to ventilator-associated pneumonia, instances of the latter condition appear to be absent from the New York data prior to 2009, perhaps reflecting changes in reporting requirements. Instead, I consider the likelihood of hospital-acquired *clostridium difficile,* a common bacterial gut infection that sometimes leads to death (Enoch and Aliyu, 2010). In my data, the rate of *C. difficile* is 0.28%.

I generate a new variable, POST, for observations beginning in the third quarter of 2008, when Nonpayment took effect. I run the model with robust standard errors in case there is autocorrelation among pre-post observations from the same hospital. Because the outcome variable HAC is binary, I estimate a logit model. This specification allows for fatter tails, making it an appropriate model when fractions of the sample have probabilities close to 0 or 1; i.e., extreme events are likely. Given the spread of variables that are likely to predict the development of a HAC, like length of stay and age – which, as shown in the summary statistics, have large standard deviations – it is likely that many observations have a probability very close to zero or one. This suggests that a logit model is appropriate for my data (Caliendo and Kopeinig, 2005).

I exclude patient-level controls because a) they are considered in the matching process, and b) they may be impacted by Medicare treatment or Nonpayment, making them bad controls. To consider hospital-level variation in HAC likelihood, I run my preferred specification with provider fixed effects and hospital referral region fixed effects. Due to computation limits, I am restricted to running matching and diff-in-diff estimates using a random 50% subset of my sample. I estimate the following diff-in-diff model, restricted to Medicare patients and their matched controls:

$$HAC_i = \beta_0 + \beta_1 MEDICARE_i + \beta_2 POST_t + \beta_3 MEDICARE_i * POST_i + \beta_4 * provider + \beta_5 * HRR + \varepsilon_i$$

Fuzzy Regression Discontinuity Design

In additional to a difference-in-differences approach, I conduct a fuzzy regression discontinuity design (RDD). This method better addresses the selection issues inherent in a diff-in-diff, though its results are not as generalizable. An RDD allows me to isolate the policy's impact by comparing non-Medicare patients in their early sixties to Medicare patients who have just passed the age threshold of 65; these groups are likely quite similar after controlling for other variables. The key assumption in this RDD is that all patients within the age range are statistically exchangeable, besides their eligibility for Medicare. I also assume that there are no behavioral responses to Medicare eligibility once a patient turns 65. If there is a discontinuity, or jump, in HAC rates when a patient turns 65 and begins to use Medicare, then the size of the jump is the causal impact of Nonpayment (Thomas, 2018). My design is "fuzzy" because age does not perfectly predict Medicare assignment. There are older patients who use other types of insurance ("no-shows"), and younger patients who are on Medicare due to disabilities ("crossovers").

An RDD approximates the *local* impact of turning 65 and becoming eligible for Medicare on HAC rate, and does not require me to assume parallel time trends for both the control and treatment groups. Another strength of using an RD in this context is that participants cannot manipulate their age in relation to the cut-point. First, I determine an optimal model for the relationship between the rating variable and the outcome variable (HAC). I use a nonparametric estimate, since my sample size is large and minimizing bias can be favored over precision. Unlike a parametric approach, which estimates the right model to fit the data, a nonparametric approach estimates the functional form itself; in other words, the model is constructed according to information derived from a small neighborhood (bandwidth) of data to the right and left of the cut-point. As the sample size gets large, a parametric approach may remain biased, while the bias of a nonparametric model approaches zero. Bias cannot always be completely eliminated, though, since local linear regressions may be biased due to their functional form (Jacob et al., 2012).

Because this RDD is fuzzy, I run it as a two-stage IV regression. In the first stage, I predict the likelihood of treatment. I instrument with a dummy variable, D, that indicates whether age is above the cut-point of 65. This is appropriate since the cut-point indicator impacts whether or not a patient develops a HAC, but only through Medicare treatment (for the narrow chosen age range), and this method is standard for fuzzy RDDs. I estimate robust standard errors to account for heteroskedasticity in the first stage. This two-stage approach adjusts "sharp" intent-to-treat estimates, allowing me to estimate the causal impact of Medicare treatment. The first stage equation is carried out as follows:

$$
\begin{aligned}
Medicare_i = {} & \alpha_0 + \alpha_1 D_i + \alpha_2(AGE - cutoff) + \alpha_3(AGE - cutoff) * D_i + \alpha_4(patient) \\
& + \alpha_5(hospital) + \alpha_6(organizational) + \delta_i
\end{aligned}
$$

First, I test relevance, and find that the F statistic for how well D predicts Medicare (within each of my chosen age ranges) is very high, with a p-value of zero. This confirms that D is indeed a strong instrument. I cannot empirically test excludability using an overidentification test because I do not have more instruments than endogenous variables.

In my RD design, the dependent variable, HAC, is binary. A limitation of the IV fuzzy RD approach is that it cannot be extended to non-linear models. The results would be inconsistent, resulting in a "forbidden regression" (Lee and Lemieux, 2010). For these reasons, I am limited to a standard linear model, which should not have a major impact on my results. In the second stage, I estimate the following nonparametric, local linear regression using data before October 2008; for the two quarters following October 2008 (adjustment period); and for the post-adjustment period:

$$HAC_i = \beta_0 + \beta_1(AGE - cutoff) + \beta_2\widehat{MEDICARE_i} + \beta_3(AGE - cutoff) * D_i + \beta_4(patient) + \beta_5(hospital) + \beta_6(organizational) + \varepsilon_i$$

Here, I estimate the HAC likelihood for a range of patient ages. The Medicare variable uses the predicted values from my first stage regression. Before the policy was implemented, there should be no difference in HAC rate at the cut point (Jacob et al., 2012). After the third quarter of 2008, though, the HAC rate may be discontinuous at the age threshold for Medicare.

Next, I run robustness checks. I alter the bandwidth (originally ages 62-68) by several years on either end to see if the same relationship still exists. I also look for evidence of a discontinuity between the outcome variable and control variables (for example, demographic traits), which should not exist.

VI. Results and Discussion

Between 2006 and 2010, the overall HAC rate at community hospitals in the state of New York was approximately 1.11%[14]. Age significantly increased the likelihood that a patient would develop a HAC, and Medicare patients were nearly twice as likely (2.08% overall) to develop such a condition. Simple logit regressions show that length of stay, Charlson Index, and being female are each correlated with a higher probability of a HAC. Just as previous researchers have found, I notice that overall HAC rates appear to increase in the quarters leading up to the policy's implementation, then modestly decline afterwards. Interestingly, the decline begins *after* the fourth quarter of 2008, supporting findings from Gidwani & Bhattacharya (2015), who speculate that

[14] The majority of HACs are catheter-associated urinary-tract infections, reflecting the common use of catheters among inpatient stays. This finding is similar to other researchers' results and findings detailed in a 2010 report from the Department of Health and Human Services.

hospitals implement reform after reviewing quarterly revenue. In my RDD analysis, I allow for an adjustment period for the two quarters following Nonpayment's implementation.

First, I run a diff-in-diff with provider and quarterly fixed effects to see whether there is evidence of a post-policy decline in HAC likelihood for Medicare patients.

TABLE 9: Difference-in-Differences Results with Binary Dependent Variable HAC

	Matching With Replacement *Provider and Hospital referral region fixed effects*	
	Covariate	Marginal Effect
Constant	-145.032*** (.000)	--
Medicare	-1.381*** (.000)	-2.90%***
POST	.289*** (.007)	0.61%***
Medicare*POST	-.225** (.041)	-0.47%**
Observations	406,063	
Pseudo R^2	.0063	

***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

These results show that, compared to a similar group of non-Medicare patients, Medicare patients were actually *less* likely to develop a HAC prior to Nonpayment, potentially because otherwise comparable non-Medicare patients tend to be younger and may be more unusual if they are similar to a Medicare patient. After October 2008, non-Medicare patients were slightly more likely to develop a HAC than before. The marginal effect of the Medicare*POST term – the difference-in-differences estimate – indicates that, following Nonpayment, Medicare patients were about 0.5% less likely to develop a HAC than they would have been had the policy not been implemented, at a 5% significance level. However, the safety of these patients hardly improved, since they were still 0.14% more likely to develop a HAC post-Nonpayment (0.61% - 0.47%). Other specifications of this diff-in-diff are reported in Appendix III, but do not drastically change results. When looking only at hospitals in the top half or top quartile of fiscal pressure, Nonpayment's impact is no longer significant. Contrary to Thirukumaran et al. (2017), this finding suggests that hospitals under greater fiscal pressure may have been less able to respond to policy incentives. Perhaps these hospitals are worse managed in some unobservable way, or do not have the resources to effectively improve patient safety.

For robustness, I run the same diff-in-diff with a hospital-acquired condition that was excluded from the policy, *C. difficile*. I find that Medicare patients in this sample were initially 0.88% less likely to develop the condition, at a 1% significance level. Similar to the results for HAC, patients were overall slightly more likely to develop *C. difficile* post-Nonpayment (0.20%), though there were no significant differences in *C. difficile* trends between Medicare and non-Medicare patients. These results imply that providers may have implemented measures to improve HAC rates for Medicare patients, but only for conditions included in the policy.

However, the low difference-in-differences estimate makes it difficult to hail Nonpayment a success based on these results. At best, Nonpayment slightly blunted the impact of worsening quality for Medicare patients. To address potential selection bias in the diff-in-diff, I run a local linear fuzzy RDD to determine whether there is a jump in HAC rate upon Medicare treatment, and find no evidence of any discontinuity:

TABLE 10: Fuzzy Regression Discontinuity Results with Binary Dependent Variable HAC

	Bandwidth: Ages 62-68; Weekend Admissions Only		
	Pre-Policy FY 2006-2008	Adjustment Period Q1FY09 - Q2FY09	Post-Policy Q3FY09 - FY2011
Constant	-.0231** (.045)	-.0455 (.121)	-.0204 (.147)
Medicare	-.0097 (.138)	.0103 (.530)	.0074 (.350)
Age (running variable)	.0015* (.073)	-.0007 (.769)	-.0004 (.707)
Age*D	-.0009 (.297)	.0015 (.514)	.0005 (.611)
Female	.0073*** (.000)	.0078*** (.000)	.0069*** (.000)
Race			
Black	.0010 (.432)	.0018 (.606)	-.0030* (.051)
Other Non-White	.0005 (.700)	.0049 (.142)	-.0023 (.115)

Charlson Index	.0009*** (.000)	.0008 (.186)	.0004 (.160)
Length of stay in days			
Non-Medicare	.0023*** (.000)	.0034*** (.000)	.0037*** (.000)
Medicare	.0004 (.326)	-.0002 (.855)	-.0004 (.480)
Median income state quartile	.0040 (.255)	-.0027 (.759)	-.0042 (.331)
Gini Coefficient	.0215 (.316)	.0153 (.781)	-.0267 (.313)
Median income state quartile*Gini Coefficient	-.0068 (.344)	.0081 (.655)	.0088 (.320)
Teaching hospital	.0026* (.065)	-.0002 (.962)	.0008 (.613)
Hospital ownership			
Private nonprofit	-.0009 (.559)	-.0017 (.707)	-.0069*** (.005)
Geographic Classification			
Large Urban	-.0004 (.797)	-.0055 (.152)	.0014 (.435)
Rural	.0022 (.346)	-.0029 (.542)	.0012 (.641)
Bed size	.0001** (.035)	.0000 (.421)	.00000 (.311)
Disproportionate Patient Share percent	-.0104*** (.001)	-.0060 (.451)	.0039 (.338)
Transfer-Adjusted Case Mix Index	.0097*** (.001)	.0073 (.284)	.0099*** (.006)
Occupancy rate proxy	- .0037 (.360)	.0121 (.185)	-.0002 (.957)
HHI*HHR on state border	.00001 (.180)	-.0000 (.758)	.0000* (.076
FPI quartile	.0002 (.662)	.0001 (.960)	-.0003 (.651)
Median household income for hospital county	-.0000 (.256)	.0000 (.426)	.0000***[15] (.002)
Observations	75,910	14,866	54,778
R^2	.0474	.0653	.0681

***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

This fuzzy RDD shows that there is no discontinuity in HAC likelihood at the threshold for Medicare eligibility, before or after Nonpayment was implemented. There is still no discontinuity when looking only at discharges in the top quartile or even top decile of fiscal pressure indices or certain quartiles of median household income. The strongest predictors of HAC are 1) being

[15] Looking at HAC likelihood by median household income for hospital county, broken into quartiles, similarly shows that hospitals in wealthier areas are significantly (~0.24%) more likely to develop HACs in the post-adjustment period. This suggests that hospitals with wealthier patients may be less pressured to respond to financial incentives created by Medicare.

female, 2) a long length of stay, and 3) being discharged from a hospital with a high transfer-adjusted case mix index, as expected. Since I control for length of stay, these results are not muddled by changes in intensity for Medicare patients. Conditioning on weekend admission further controls for the possibility that providers may select profitable patients.

It is interesting that, following the adjustment period, private nonprofit hospitals are significantly less likely than government hospitals to discharge patients with HACs post-implementation; perhaps this is further evidence to support Duggan (2000), who finds that private hospitals respond significantly to a policy-driven payment change, while public hospitals do not respond at all. Teaching hospitals were no longer significantly more likely to treat patients with HACs following Nonpayment implementation, suggesting that their focus on teaching proper prevention techniques may have made them a prime target for reform.

Length of stay significantly increases HAC likelihood for non-Medicare patients. This coefficient is larger following Nonpayment's implementation, possibly because quality improvement efforts mostly impact patients with short-term stays. For Medicare patients, longer time spent in the hospital does not increase HAC likelihood. It's possible that providers decrease intensity for Medicare patients, discharging them earlier than an equally severe, but more profitable, non-Medicare patient.

For robustness, I run a series of regressions and confirm that there are no discontinuities for any of the control variables at the age threshold of 65, except for fiscal pressure index. If I scale FPI to lie between 0 and 100, Medicare treatment is associated with an FPI .55 higher with a p-value of 0.044, suggesting that hospitals under greater fiscal pressure are significantly more likely to treat Medicare patients, even after controlling for weekend admission. Other than this discontinuity, these robustness checks strongly support my assumption that all discharges within the chosen bandwidth are statistically exchangeable. I also confirm that there is still no evidence of a discontinuity if I adjust the bandwidths to plus or minus two years, four years, or five years, meaning my RDD model is robust to bandwidth choice.

VII. Conclusion

In line with results from Lee et al. (2012), Schuller et al. (2013), and Vaz et al. (2015), I do not find evidence that Nonpayment caused a meaningful improvement in HAC rates for Medicare patients. While my diff-in-diff estimate suggests that the policy somewhat protected Medicare patients from worsening quality by lowering HAC likelihood by 0.47%, the lack of any discontinuity in my RDD results imply that the diff-in-diff may face selection issues. Nonpayment's lack of success may be because the policy was not implemented on a large enough scale, perhaps supporting results from Gneezy and Rustichini (2000), who find that financial incentives must be large to desirably change behavior. In an analysis of California data, McNair et al. (2009) find that the financial incentives created by Nonpayment were likely to be small. While it is possible that spillover effects equally improved conditions for both Medicare and non-Medicare patients, the data suggest that HAC likelihood did not significantly fall for anyone, even when analysis is restricted to hospitals under high fiscal pressure. This finding contradicts results from Thirukumaran et al. (2017), who use a longer time period and fail to consider variables such as hospital teaching status or measures of healthcare utilization and intensity.

There is some evidence that Nonpayment produced a response at certain hospitals. In the post-adjustment period, private non-profit hospitals were significantly less likely than government hospitals to treat patients that developed HACs. Teaching hospitals and larger hospitals (measured by bed size) were no longer more likely to produce HACs following Nonpayment, suggesting that such hospitals were better able to respond to policy incentives. These findings imply that policymakers may want to target P4P reform more narrowly, perhaps focusing on private hospitals. To improve quality at less responsive government hospitals, perhaps other types of incentives may be more powerful. I also notice that hospitals under greater fiscal pressure, measured by the index I develop on page 24, are significantly more likely to treat Medicare patients. This suggests that reformers must be wary of providers selecting toward profitable non-Medicare patients.

As researchers continue to study Nonpayment and other pay-for-performance initiatives in healthcare, they would surely benefit from cooperation between the medical and economics fields. Though it is beyond the scope of this paper, a better approach might consider diagnoses, perhaps using the clinical classification software provided by HCUP to group patients with similar diagnoses. Researchers with significant time and resources might want to look into using Medicare claims data, which would enable better understanding of the financial incentives created by Nonpayment. If future researchers wish to compare hospitals impacted by the policy to hospitals that were not impacted, it might be useful to compile data from exempt hospitals, and restrict analysis to a very small group of patients. For example, research could compare exempt Maryland hospitals in the Baltimore area to community hospitals in a comparable American city. However, these results would be less generalizable.

Better aligning healthcare costs with outcomes remains a great challenge in American healthcare. However, it is not clear whether pay-for-performance is the appropriate policy solution. Researchers must acknowledge that such policies may have undesirable or counterintuitive impacts on patient selection, healthcare utilization, and responses to fiscal pressure, and that such factors must be considered before declaring a policy successful.

Works Cited

AHA Guide to the Healthcare Field (2008 Edition). (2007). Health Forum.

"AHRQ National Scorecard on Hospital-Acquired Conditions: Updated Baseline Rates and Preliminary Results 2014–2017." Agency for Healthcare Quality and Research, January 2019.

Table B19083: Gini Index of Income Inequality. (2006). American FactFinder. Census.gov

Table B19013: Median Household Income in the Past 12 Months. (2010). American FactFinder. Census.gov

Hospital Referral Regions. (2015, October 26). ArcGIS. arcGIS.com

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review*, *53*(5).

Barnett, M. J., Kaboli, P. J., Sirio, C. A., & Rosenthal, G. A. (2002). Day of the Week of Intensive Care Admission and Patient Outcomes: A Multisite Regional Evaluation. *Medical Care*, *40*(6).

Caliendo, M., & Kopeinig, S. (2005). *Some Practical Guidance for the Implementation of Propensity Score Matching: Discussion Paper No. 1588*. Institute for the Study of Labor.

Centers for Medicare and Medicaid Services. "Affected Hospitals." www.cms.gov. (accessed September 20, 2019).

Centers for Medicare and Medicaid Services. "CMS' Value-Based Programs." www.cms.gov. (accessed September 20, 2019).

Centers for Medicare and Medicaid Services. "Hospital-Acquired Conditions." www.cms.gov. (accessed October 20, 2019).

Chen, A., & Lakdawalla, D. (2019). Healing the Poor: The Influence of Patient Socioeconomic Status on Physician Supply Responses. *Journal of Health Economics*, *64*.

Columbia University Mailman School of Public Health. "Difference-in-Difference Estimation." mailman.columbia.edu. (accessed February 28, 2020).

Commission on Cancer Datalinks. Accessed January 2, 2020. datalinks.facs.org.

Dranove, D. (1987). Rate-Setting by Diagnosis Related Groups and Hospital Specialization. *RAND Journal of Economics*, *18*(3).

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2003). Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy*, *11*(3).

Duggan, M. G. (2000). Hospital Ownership and Public Medical Spending. *Quarterly Journal of Economics*, *115*(4).

Eijkenaar, F. (2011). Key Issues in the Design of Pay for Performance Programs. *European Journal of Health Economics*, *14*(1), 117–131.

Ellis, R. P., & McGuire, T. G. (1996). Hospital Response to Prospective Payment: Moral Hazard, Selection, and Practice-Style Effects. *Journal of Health Economics*, *15*(3).

Enoch, David A., and Sani H. Aliyu. "Is Clostridium Difficile Infection Still a Problem for Hospitals?" *Canadian Medical Association Journal* 184, no. 1 (January 2012).

Fang, H., Keane, M. P., & Silverman, D. (2008). Sources of Advantageous Selection: Evidence from the Medigap Insurance Market. *Journal of Political Economy*, *116*(2).

Gaynor, M., & Gertler, P. (1995). Moral Hazard and Risk Spreading in Partnerships. *RAND Journal of Economics*, *26*(4).

Gidwani, R., & Bhattacharya, J. (2015). CMS Reimbursement Reform and the Incidence of Hospital-Acquired Pulmonary Embolism or Deep Vein Thrombosis. *Journal of General Internal Medicine*, *30*(5).

Gneezy, U., & Rustichini, A. (2000). Pay Enough or Don't Pay At All. *The Quarterly Journal of Economics*, *115*(3).

Green, E. P. (2014). Payment Systems in the Healthcare Industry: An Experimental Study of Physician Incentives. *Journal of Economic Behavior & Organization*, *106*.

Hadley, J., Zuckerman, S., & Feder, J. (1989). Profits and Fiscal Pressure in the Prospective Payment System: Their Impacts on Hospitals. *Inquiry*, *26*(3).

Hodgkin, D., & McGuire, T. G. (1994). Payment Levels and Hospital Response to Prospective Payment. *Journal of Health Economics*, *13*(1).

Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, *7*.

Houchens, R. L., Elixhauser, A., & Romano, P. S. (2008). How Often are Potential Patient Safety Events Present on Admission? *The Joint Commission Journal on Quality and Patient Safety*, *34*(3).

Jacob, R., Zhu, P., Somers, M.-A., & Bloom, H. (2012). *A Practical Guide to Regression Discontinuity*. MDRC.

Jacobson, M. G., Chang, T. &., Earle, C. C., & Newhouse, J. P. (2017). Physician Agency and Patient Survival. *Journal of Economic Behavior & Organization*, *134*.

Kavanagh, K. T. (2011). Financial Incentives to Promote Health Care Quality: The Hospital Acquired Conditions Nonpayment Policy. *Social Work in Public Health*, *26*(5).

King, G., & Nielson, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, *27*(4).

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To Err is Human: Building a Safer Health System*. National Academies Press.

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, *48*(2).

Lee, G. M., Kleinman, K., Soumerai, S. B., Tse, A., Cole, D., Fridkin, S. K., Horan, T., Platt, R., Gay, C., Kassler, W., Goldmann, D. A., Jernigan, J., & Jha, A. K. (2012). Effect of Nonpayment for Preventable Infections in U.S. Hospitals. *The New England Journal of Medicine*, *367*(15), 1428–1437.

Leuven, E., & Sianesi, B. (2003). *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*. RePEc.org

Levinson, D. R. (2010). *Adverse Events in Hospitals: National Incidence Among Medicare Beneficiaries*. Department of Health and Human Services - Office of Inspector General.

"Medicare Learning Network Matters: Fiscal Year (FY) 2009 Inpatient Prospective Payment System (IPPS), Long Term Care Hospital (LTCH) PPS, and Inpatient Psychiatric Facility (IPF) PPS Changes." Centers for Medicare & Medicaid Services, October 2008.

"Medicare Primer." Congressional Research Service, August 2019.

McHugh, Megan, Timothy C. Martin, John Orwat, and Kevin Van Dyke. "Medicare's Policy to Limit Payment for Hospital-Acquired Conditions: The Impact on Safety Net Providers." *Journal of Health Care for the Poor and Underserved* 22, no. 2 (May 2011).

McNair, P. D., Luft, H. S., & Bindman, A. B. (2009). Medicare's Policy Not to Pay For Treating Hospital-Acquired Conditions: The Impact. *Health Affairs*, *28*(5).

Mullen, K. J., Frank, R. G., & Rosenthal, M. B. (2010). Can You Get What You Pay For? Pay-For-Performance and the Quality of Healthcare Providers. *The Rand Journal of Economics*, *41*(1).
Newhouse, J. P. (1989). Do Unprofitable Patients Face Access Problems? *Health Care Financing Review*, *11*(2).
Newhouse, J. P., & Byrne, D. J. (1988). Did Medicare's Prospective Payment System Cause Length of Stay to Fall? *Journal of Health Economics*, *7*(4).
Patient Safety Movement. "Patient Safety Movement History." www.PatientSafetyMovement.org. (accessed September 20, 2019).
Rosenthal, M. B. (2007). Nonpayment for Performance? Medicare's New Reimbursement Rule. *The New England Journal of Medicine*, *357*(16).
Rothschild, M., & Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics*, *90*(4).
"Saving Lives and Saving Money: Hospital-Acquired Conditions Update: Final Data From National Efforts To Make Care Safer, 2010-2014." Agency for Healthcare Research and Quality, December 2016.
Stagg, V. (2006). *CHARLSON: Stata Module to Calculate Charlson Index of Comorbidity* [Statistical Software Components S456719]. RePEc.org
Staiger, Douglas, and Gary L. Gaumer. "The Impact of Financial Pressure on Quality of Care in Hospitals: Post-Admission Mortality Under Medicare's Prospective Payment System." Abt Associates Inc (1990).
The Healthcare Cost and Utilization Project. "NIS Description of Data Elements." www.hcup-us.ahrq.gov. (accessed October 15, 2019).
The National Bureau of Economic Research. "Dartmouth Atlas Data." www.nber.org. (Accessed October 20, 2019).
Thirukumaran, C. P., Glance, L. G., Temkin-Greener, H., Rosenthal, M. B., & Li, Y. (2017). Impact of Medicare's Nonpayment Program on Hospital-acquired Conditions. *Medical Care*, *55*(5).
Thomas, D. (2018, December). *Lecture 27*.
Town, R., Wholey, D. R., Kralewski, J., & Dowd, B. (2004). Assessing the Influence of Incentives on Physicians and Medical Groups. *Medical Care Research and Review*, *61*(3).
Vaz, L. E., Kleinman, K. P., Kawai, A. T., Jin, R., Kassler, W. J., Grant, P. S., Rett, M. D., Goldmann, D. A., Calderwood, M. S., Soumerai, S. B., & Lee, G. M. (2015). Impact of Medicare's Hospital-Acquired Condition Policy on Infections in Safety Net and Non–Safety Net Hospitals. *Infection Control and Hospital Epidemiology*, *36*(6).
Waters, T. M., Daniels, M. J., Bazzoli, G. J., Perencevich, E., Dunton, N., Staggs, V. S., Potter, C., Fareed, N., Liu, M., & Shorr, R. I. (2015). Effect of Medicare's Nonpayment for Hospital-Acquired Conditions: Lessons for Future Policy. *JAMA Internal Medicine*, *175*(3).

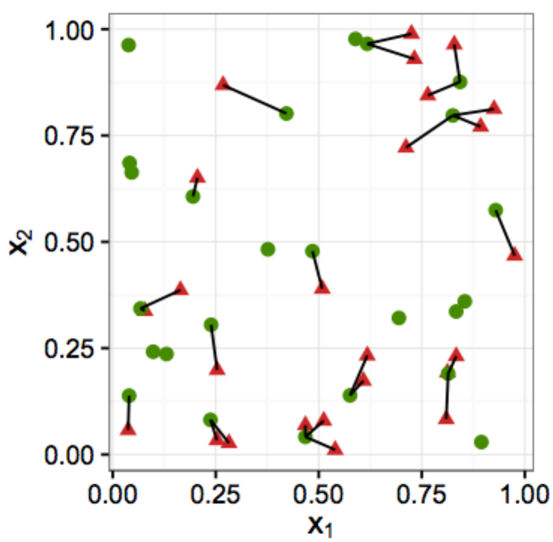
Appendix I: Excluded Hospitals

In case hospital consolidation somehow impacted HAC rates – for example, if poor-performing hospitals tended to close over time – hospitals that were not operating under the Inpatient Prospective Payment System (IPPS) from 2006-2010 were dropped:

Hospital	FY Removed from IPPS
Victory Memorial Hospital	2011
Manhattan Eye Ear Throat Hospital	2011
St. Clare's Hospital	2011
Cabrini Medical Center	2011
Our Lady of Mercy Medical Center	2011
St. Vincent's Midtown Hospital	2010
Orthopaedic Hospital	2009
Bellevue Woman's Hospital	2009
NY United Hospital Medical Center	2008
Brunswick Hospital	2008
The Hospital (Delaware County)	2008
Schuyler Hospital	2007
Soldiers & Sailors Memorial Hospital	2007
St. Luke's Hospital	2007

Hospital	FY Added to IPPS
Monroe Community Hospital	2008
Amsterdam Memorial Hospital	2008
Sunnyview Hospital	2007
Helen Hayes Hospital	2007
Winifred Masterson Burke Rehabilitation Hospital	2007

Appendix II: Illustration of Mahalanobis Nearest Neighbor Matching with Replacement



Treatment observations in red are matched with control observations in green, based on covariates x1 and x2. This process results in all treated observations being matched to the nearest control, leaving some control observations to be dropped.

Appendix III: Diff-in-Diff Alternate Specifications

	Matching With Replacement *No controls or fixed effects*
Constant	-2.487*** (.000)
Medicare	-1.378*** (.000)
POST	.289*** (.007)
Medicare*POST	-.225** (.041)
Observations Pseudo R^2	406,063 .0061

	Matching With Replacement *With hospital and organizational controls*
Constant	-3.770*** (.000)
Medicare	-1.344*** (.000)
POST	.254** (.019)
Medicare*POST	-.209* (.058)
Observations Pseudo R^2	406,063 .0144

	Matching With Replacement *Provider and Hospital referral region fixed effects* *Top 50% of fiscal pressure*
Constant	-211.767*** (.000)
Medicare	-1.438*** (.000)
POST	.275 (.116)
Medicare*POST	-.212 (.234)
Observations Pseudo R^2	223,996 .0050

***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.