Monitoring for Waste: Evidence from Medicare Audits *

Maggie Shi†

February 21, 2023

Abstract

This paper examines the effectiveness of monitoring for wasteful public spending. I study a large Medicare program that monitored for unnecessary healthcare spending, and consider its effect on government savings, provider behavior, and patient health. Every dollar Medicare spent on monitoring generated $24–29 in government savings. The majority of savings stem from the deterrence of future care, rather than reclaimed payments from prior care. The health of the marginal patient denied care is not harmed, indicating that monitoring is well-targeted and only reduces unnecessary care. Providers subject to monitoring face increased compliance costs. However, these costs are mostly incurred upfront and include investments in technology to assess the medical necessity and cost-effectiveness of care.

*This paper was previously circulated as "The Costs and Benefits of Monitoring Providers: Evidence from Medicare." I am grateful to Wojciech Kopczuk, Adam Sacarny, Pietro Tebaldi, and Michael Best for their input and support. I thank the editor and four anonymous referees for their helpful suggestions to improve the paper. I also thank Jetson Leder-Luis, Jon Skinner, Tal Gross, Ashley Swanson, Bentley MacLeod, Gautam Gowrisankaran, Cailin Slattery, Parker Rogers, Ben Chartock, Kelli Marquardt, Claudia Halbac, Melinda Pitts, Bernard Salanie, Aina Katsikas, Josh Gottlieb, Tim Layton, David Cutler, Angie Acquatella, Motaz Al-Chanati, and participants at various seminars and conferences who provided feedback on earlier versions of the project. Mohan Ramanujan, Daniel Feenberg, Elizabeth Adams, Jean Roth, Adrienne Henderson, and Ashley Badami provided excellent assistance in accessing and managing the data. Xiao Zheng provided helpful clinical expertise. I gratefully acknowledge fellowship support from the Agency for Healthcare Research and Quality (#R36HS027715-01) and the National Institute on Aging (#T32-AG000186). All errors are my own.

†NBER.

1 Introduction

Combating waste is a perennial problem for governments. In 2021, over seven percent of U.S. federal spending was wasted (Office of Management and Budget, 2022). Economic theory prescribes a straightforward solution: more effort should be devoted to monitoring and penalizing wasteful spending (Laffont and Tirole, 1986; Baron and Besanko, 1984). Yet many contend that monitoring is underutilized – by some estimates, over half of wasteful federal spending goes undetected (Cunningham et al., 2018; Office of the Inspector General, 2020). Policymakers may be hesitant to monitor aggressively because it is unclear whether it can successfully target waste, or if it just introduces needless regulatory costs. Despite the importance of this question, there is little empirical evidence on whether monitoring is a worthwhile endeavor, as both wasteful spending and the costs associated with monitoring for it are difficult to measure.

This paper considers the effectiveness of monitoring for waste in the context of Medicare, the federal insurance program for the elderly and disabled. On the one hand, the sheer magnitude of potential savings in this context makes increased monitoring an attractive policy tool. All Medicare expenditure is contracted out to healthcare providers, who then have considerable latitude over spending decisions. Perhaps, then, unsurprisingly, waste is widespread: up to 13 percent of Medicare spending goes to unnecessary or improperly billed care (Centers for Medicare and Medicaid Services, 2022).[1] At the same time, as healthcare becomes increasingly digitized, there has been significant progress in the development of technology to improve the efficiency of healthcare spending (Hillestad et al., 2005). Monitoring could therefore also serve as a useful tool to incentivize providers to seek out new ways of identifying waste and improving cost-effectiveness.

On the other hand, the social costs of excessive oversight are potentially high here as well. Poorly targeted monitoring could have dire implications for patient health (Doyle et al., 2015). Pressuring providers to cut back spending could deter *necessary* care, especially if it is unclear *ex ante* what services are necessary or not. Given the complexity of identifying unnecessary care,

[1]Medicare expenditure accounts for 15 percent of federal spending (Cubanski et al., 2019), so wasteful Medicare spending alone accounts for *2 percent* of total federal spending.

monitoring could also impose considerable compliance costs on providers. If these costs stem mostly from the "back and forth" of the monitoring process, then they pose a deadweight loss that adds to providers' already-high administrative burden (Cutler and Ly, 2011; Dunn et al., 2020). Thus, the effectiveness of monitoring in Medicare depends on the balance between the savings from reducing unnecessary care and the social costs that monitoring may impose on patients and providers.

I study Medicare's largest monitoring program, the Recovery Audit Contractor (RAC) Program. Through the RAC program, private auditing firms ("RACs") conduct manual reviews of individual Medicare claims ("audits") to identify and reclaim payments for unnecessary care. I focus on RAC auditing for unnecessary hospital stays. At the program's peak, four percent of all hospital admissions, Medicare's largest expenditure category, were audited, and one percent of all Medicare inpatient revenue was reclaimed through the RAC program.[2]

The rich data in this context offer a unique window to study the effectiveness of monitoring for waste. To estimate the savings from both the detection and deterrence effects of monitoring, I combine novel administrative data on RAC audits with claims data on hospital stays. To assess whether these savings stemmed from reductions in unnecessary care, I look to patient health outcomes for evidence of patient harm. In particular, I use emergency department (ED) discharge data that allow me to track patients' outcomes over time, even if they are denied a hospital stay. Then to characterize the effort hospitals put in to comply with RAC audits, I draw on measures of administrative costs and technology adoption from hospital cost reports and surveys.

To motivate the empirical analysis, I introduce a principal-agent model of monitoring and waste to understand how hospitals might respond to RAC audits. In the model, hospitals decide how many admissions to report to Medicare, where some of these admissions can be unnecessary. Hospitals also choose whether to adopt compliance technology that reduces unnecessary admissions in exchange for reduced scrutiny. The model delivers two key testable implications: increased auditing should reduce admissions, and these reductions should be only for unnecessary admissions.

[2] To put the size of the RAC program in context, consider the widely-publicized Hospital Readmissions Reduction Program (HRRP), which levied a mean penalty of 0.75 percent of hospital revenue (Gupta, 2021).

The effect of auditing on technology adoption is ambiguous, as it depends on the audit protection the technology provides relative to the cost of the technology.

I then test for the effects of auditing on admissions, patient health, compliance costs, and technology adoption in the data. I arrive at three core findings. First, RAC audits reduce Medicare spending on admissions, with a very high return – every dollar that Medicare spends on monitoring hospitals recovers *$24–29*. The vast majority of these savings stem from the deterrence of future spending, rather than the recovery of prior spending. Second, monitoring primarily deters *unnecessary* admissions. Hospitals are less likely to admit patients with higher audit risk, yet these patients were no more likely to return to the hospital due to a missed diagnosis. Third, RAC audits lead hospitals to invest in technology to assess whether admitting a patient is medically necessary. The majority of the compliance costs hospitals incur are such upfront costs. Taken together, the empirical results indicate that monitoring providers leads to large reductions in unnecessary spending, and one way it does so is by incentivizing providers to invest in technology that identifies waste.

The central challenge in identifying the causal effect of monitoring is that RAC audits are endogenous. RACs are private firms that are paid a contingency fee based on the payments they correct. So naturally, they target their audits at claims that are most likely to have an error. I address this endogeneity by leveraging two identification strategies: one compares hospitals subject to differentially aggressive RACs, and the other compares patient cohorts who face exogenously different audit likelihoods.

To understand hospital-level responses to RAC audits, I deploy a difference-in-difference specification comparing hospitals before and after a major expansion of the RAC program in 2011. I focus on hospitals subject to different RACs, leveraging sharp differences in auditing at the border between different RAC jurisdictions. In line with model predictions, hospitals subject to a more-aggressive RAC reduce their admissions -– a one percentage point (46 percent) increase in the share of admissions audited leads to a two percent drop in admissions. This effect persists even when auditing is scaled back in later years. 89 percent of the savings from the marginal audit stem from the deterrence of future admissions, and the remaining 11 percent are from the payments

RACs reclaim. This large deterrence effect is striking, especially given that policymakers only considered the recovered payments to assess the cost-effectiveness of the RAC program (Centers for Medicare and Medicaid Services, 2012). Extrapolating these effects to the overall hospital sample, I calculate that the RAC program led to upwards of \$9 billion in Medicare savings from 2011 to 2015.

Most of the savings from monitoring stem from deterred hospital admissions, and I find evidence that hospitals adopted technology in order to identify which patients to no longer admit. Hospitals subject to more audits were more likely to adopt "medical necessity checking" software, which cross-references electronic health records with payer (i.e., insurer) rules to provide guidance on the medical necessity of care in real time (3M, 2016; Experian Health, 2022; AccuReg, 2022). Accordingly, hospital administrative costs rise: for every \$1000 in Medicare savings in 2011–2015, hospitals incur \$178–218 in administrative costs. But these costs are mostly concentrated as a one-time spike that occurs immediately in 2011. This suggests that provider compliance costs comprise mostly of the fixed costs from investments like technology adoption, rather than the ongoing hassle costs of the monitoring process.

I then turn to investigating the second testable implication of the model — did the reductions stem from *unnecessary* stays? But because patient composition changes as hospital volume decreases, it is challenging to compare patient outcomes *across hospitals*. In light of this, I exploit a policy which generated exogenous variation in audit likelihoods *across patients* in the same hospital. In particular, I consider the "Two Midnights rule," which barred RACs from auditing patients whose time in the hospital crossed two or more midnights. For this rule, time in the hospital is measured from the point that the patient *arrives* at the ED. Visits that start right after midnight are less likely to reach two midnights than those that start right before. Therefore, patients who arrived at the ED after midnight were more likely to be audited than those who arrived before. I then use a difference-in-difference specification to compare admission rates and health outcomes for before- vs. after-midnight ED patients, pre- and post-Two Midnights rule.

Mirroring the hospital-level results, I find that once the Two Midnights rule is implemented, hospitals cut back on inpatient admissions for after-midnight patients. Turning to health outcomes, I do *not* find evidence that after-midnight patients were more likely to revisit a hospital within thirty days, a proxy for patient health that is observable in discharge data. Hospitals targeted their reductions to patients in the middle of the severity distribution, who faced up to a 25 percent reduction in admission likelihood. But even among these patients, there is no increase in revisit rates. This response is driven by hospitals with medical necessity checking software installed prior to the Two Midnights rule, illustrating how hospitals used this technology.

This paper contributes to the literature on government monitoring and enforcement. But compared to the large literatures on monitoring for tax compliance or adherence to environmental regulations, there is relatively little empirical work on monitoring for wasteful public spending.[3] This is in spite of the fact that the government *does* monitor for waste – for example, the Offices of Inspector General and the Government Accountability Office are devoted to uncovering waste. This kind of monitoring is likely understudied because what constitutes "waste" is often ambiguously defined and notoriously difficult to measure.[4] Defining waste is a more straightforward task in the healthcare setting: most would agree that spending on care which does not improve patient health is wasteful. My findings also highlight a mechanism through which monitoring combats waste: by incentivizing the adoption of technology to identify waste.

Beyond considering the government savings from monitoring, this paper also makes progress on measuring the various social costs it imposes on patients and providers. The private costs associated with public programs are often difficult to observe, so their existence is usually deduced indirectly – for example, by looking at how program participation changes when these costs change.[5] Health care presents a unique setting where two forms of social costs – provider admin-

[3]The baseline theoretical model relating tax enforcement with evasion comes from Allingham and Sandmo (1972), and subsequent extensions to this model and empirical work are surveyed by Andreoni et al. (1998) and Slemrod and Yitzhaki (2002). The empirical literature on environmental regulation has shown that increased monitoring leads to reductions in pollution (Magat and Viscusi, 1990; Hanna and Oliva, 2010; Duflo et al., 2018).

[4]For example, in Olken (2007), one of the few other papers to study the effects of monitoring on waste, measuring wasteful spending required assembling teams to take core samples from roads and compare the reported to actual materials used.

[5]Recent examples include Kopczuk and Pop-Eleches (2007); Deshpande and Li (2019); Finkelstein and Notowidigdo (2019); Meckel (2020); Zwick (2021); **?**.

istrative costs and patient health outcomes – *can* be observed more readily. Measuring provider administrative costs gives us insight into the magnitude of the private compliance costs that result from monitoring. Understanding the effects on patient health is crucial for assessing whether the savings from monitoring actually stemmed from reductions in wasteful spending.

Finally, this paper sheds further light on how healthcare providers respond to incentives. It has been well-documented that providers respond to financial incentives, either by changing the quantity and type of care provided or how they document care.[6] In contrast, less is known about how providers respond to non-financial incentives like monitoring, even though they are employed by both private and public insurers (Gottlieb et al., 2018). This paper contributes to a growing literature on how providers respond to various forms of non-financial incentives: pre-payment denials (**?**League, 2022), fraud enforcement (Nicholas et al., 2020; Howard and McCarthy, 2021; Leder-Luis, 2023), and prior authorization (**?**Roberts et al., 2021).

The rest of the paper proceeds as follows. Section 2 describes the policy context of the RAC program. Section 3 sets up the model and derives testable implications. Section 4.1 describes the data for the empirical analysis, Section 4.2 explains the hospital-level empirical strategy, and Section 4.3 explain the patient-level empirical strategy. Section 5 presents the empirical results and compares the findings across the two empirical strategies. Section 6 concludes.

2 Policy Context

Medicare spent $147 billion, or 19 percent of its total expenditure, on inpatient admissions in 2019 (Medicare Payment Advisory Commission, 2020). Medicare reimburses hospitals a fixed prospective payment per inpatient stay, where the payment depends on the severity-adjusted diagnosis category associated with the stay. Outside of a few exceptions,[7] the payment rate depends on

[6]Examples of the former include Cutler (1995); Ellis and McGuire (1996); Clemens and Gottlieb (2014); Einav et al. (2018); Eliason et al. (2018); Alexander and Schnell (2019); Gross et al. (2023); Gupta (2021). Examples of the latter include Silverman and Skinner (2004); Dafny (2005); Sacarny (2018); Gowrisankaran et al. (2019)

[7]One exception is that in "outlier" cases, the payment can depend on length of stay. Outlier stays account for 1.8 percent of overall Medicare hospital stays. Another exception is if an acute care hospital transfers a beneficiary to post-acute care, in which case Medicare pays a per diem rate (Office of the Inspector General, 2019).

the patient's diagnosis, their pre-existing health conditions, and procedures conducted during their stay. Importantly, it does not generally depend on the admission's length of stay.

Over time, policymakers became increasingly concerned with one area of perceived waste: unnecessary short (0–2 day) stays (Centers for Medicare and Medicaid Services, 2011b; US Department of Health and Human Services Office of Inspector General, 2013). The Medicare Payment Advisory Commission (MedPAC), a non-partisan government agency, contended that hospitals were admitting patients for these short inpatient stays because they were very profitable (Medicare Payment Advisory Commission, 2015): the payment-to-cost ratio for short stays was two times that of longer stays. Appendix Section A.1 describes the Medicare inpatient prospective payment system and short stays in greater detail.

To address this issue, in 2011 Medicare directed RACs to begin monitoring and reclaiming payments for unnecessary inpatient admissions. RAC audits are carried out by four private firms, each of which is in charge of conducting audits within its geographic jurisdiction, or "RAC region." Figure 1a illustrates these regions – they fall along state lines and, in the context of medical claims reviews, are unique to the RAC program.[8] RAC audits were introduced nationally in 2009 after a pilot program in select states. But RAC activity was fairly limited until 2011, when Medicare allowed them to begin auditing unnecessary inpatient stays. The total number of audits increased by *537 percent* from 2010 to 2012, which translated into a *1211 percent* increase in the value of payments reclaimed per hospital (Figure 1b).[9]

Ninety-five percent of inpatient stay RAC audits involve a manual review: the RAC first runs a proprietary algorithm on Medicare claims data to flag individual claims for issues such as missing documentation, incorrect coding, or – starting in 2011 – unnecessary care. A medical professional hired by the RAC, typically a nurse or a medical coder, then requests the documentation for the flagged claim from the provider and manually reviews it. The medical professional determines

[8] The RAC regions are also used by Durable Medical Equipment Medicare Administrative Contractors, who do not process claims for medical care, but rather claims for equipment and supplies ordered by healthcare providers. This includes, for example, oxygen equipment, wheelchairs, and blood testing strips.

[9] The total value of reclaimed payments across all hospitals increased from $229 million in 2010 to $3.15 billion in 2012.

whether Medicare made an overpayment or, in a small share of cases, an underpayment.[10] If they find an error, then they can demand that the provider repays Medicare (or vice versa). There is no additional penalty to the provider for each corrected payment, although RACs could refer violations that they suspected rose to the level of *fraud* to CMS or law enforcement (Centers for Medicare and Medicaid Services, 2015). The RAC firms are paid a negotiated contingency fee on the payments they correct: 9–12.5 percent, depending on the firm, of the reclaimed payment after appeals. Figure H1 illustrates the full process for claims auditing and appeals, including the remaining 5 percent of inpatient stay audits that do not involve a manual documentation review.

Figure 1b illustrates average per-hospital RAC activity, by year of audit (which is often *after* than the year the claim was originally paid). At the program's peak, RACs were reclaiming $1 million per hospital annually, or 3 percent of the average hospital's Medicare inpatient revenue of $32 million. By 2020, 96 percent of hospitals had at least one inpatient stay that was audited. RAC audits were then scaled back significantly by 2015, when Medicare paused the program to evaluate complaints made by hospitals and industry stakeholders (Foster and McBride, 2014). Appendix Section A.2 describes the RAC regions, RAC firms, audit process, and timeline of the RAC program in greater detail.

How could hospitals defend themselves from these audits? While they could not retroactively change previous admissions, they could re-optimize their admissions to reduce audits going forward. In a 2012 survey conducted by the American Hospital Association, the majority of hospitals reported that the RAC program increased their administrative costs. This increase was driven by activities like conducting training and education or purchasing tracking software (American Hospital Association, 2013). A particularly relevant type of software is "medical necessity checking software," which hospitals use to assess the medical necessity of the care they provide with respect to payer coverage rules (3M, 2016; Experian Health, 2022). This software informs providers in real-time about the medical necessity of care for each particular case, allowing them to make a more informed call about decisions like whether to admit a patient. Adoption of health IT like

[10] In 2011, 6 percent of inpatient stay audits resulted in an underpayment determination.

this is often touted as a way to reduce wasteful healthcare spending – in 2009, Congress passed the HITECH Act and devoted $20 billion to subsidize health IT adoption with the explicit goal of improving cost-effectiveness (Burde, 2011; Dranove et al., 2014). An additional benefit of this software to hospitals is that, because it is embedded hospital's electronic medical record system, it creates a "paper trail" in the documentation to support the provider's decision in the event of an audit.

I also leverage an additional policy within the RAC program which generated differences in audit risk across patients. Two years after expanding RAC scope to medical necessity, Medicare introduced a new rule in 2013 to clarify which admissions could be audited: the Two Midnights rule. Under this rule, Medicare counted the number of *midnights* during a patient's entire time in the hospital – including the time spent in the ED, in outpatient care, and in inpatient care.[11] If the patient's time in the hospital spanned two midnights, then the stay was presumed to be necessary and RACs could not audit for medical necessity. If the patient's stay *did not* span two midnights, then RACs could audit it (Centers for Medicare and Medicaid Services, 2017). So for the 73 percent of Medicare inpatient admissions that originate in the ED, the Two Midnights rule effectively increased audit likelihoods for patients who arrived after midnight relative to those who arrived before.

3 A Model of Waste, Monitoring, and Technology Adoption

Next, I adapt the model of costly auditing introduced in Baron and Besanko (1984) and use it to generate testable implications about hospital responses to RAC audits. A key modification I make to the model is to add a compliance technology adoption choice for the agent. I model this technology as operating in two main ways. First, it reduces the amount of revenue the principal

[11] Midnight cutoffs are surprisingly common in insurer billing rules; see the policies studied by Almond and Doyle (2011) and Rose (2020). A difference between the Two Midnights rule and the policies studied by Almond and Doyle (2011) and Rose (2020) is that the Two Midnights rule counts the number of midnights during a patient's entire stay in the hospital, starting from when they *arrive* at the hospital. In contrast, the rules studied by these two papers focus on how many midnights pass during a patient's hospital admission, starting from the *hospital admission hour* (that is, the hour that the patient is formally admitted for inpatient care or, in the case of newborns, born).

recoups from audits. Second, it reduces misreporting and waste by making it more costly for the agent to misreport. This differentiates *compliance* technology from technology that simply enables fraud – the audit protection must actually come with the tradeoff of reductions in waste. Then, as in Baron and Besanko (1984), the principal chooses the optimal audit rate, taking into account its own audit costs as well as the effects of auditing on the agent's behavior.[12]

This model of costly monitoring with technology adoption is relevant to settings where an agent reports a quantity to the principal, the principal can only ascertain if the agent is misreporting via costly monitoring, and the agent can purchase compliance technology. Take the medical necessity checking software, for example. There is cross-sectional evidence that the software protected hospitals from audit – hospitals with the software already installed in 2010 were less likely to face audits and denials later (Figures H10a and H10b). The additional cost to misreporting that the software imposes could be interpreted as provider altruism – providers who prefer to only admit medically necessary cases could find it costly to deviate from the software's recommendation. Or, it could be that the additional "paper trail" generated by the software in electronic health records makes it more costly for them to justify an admission when the software indicates it is unnecessary. Section B.3 discusses other contexts that this model could apply to, like an individual's choice to e-file taxes or a retailer's decision to adopt Electronic Benefit Transfer (EBT) systems for SNAP or WIC.

Hospital's Problem Each hospital observes q, the true number of patients needing admission, which is distributed uniformly between $[0, \overline{q}]$. Assume that hospitals admit patients in descending order of medical necessity – so the $(q-1)$-th admission is necessary, while the $(q+1)$-th admission is not (by Medicare's standards). Likewise, when a hospital reduces admissions, it goes in reverse order of medical necessity. Hospitals choose to admit and report Q patients. With each admission, the hospital incurs a per-treatment cost k and is reimbursed $P > k$ by Medicare. Hospitals inherently dislike over- or under-admitting patients – perhaps because they value patient welfare or because of reputation concerns. The cost of Q deviating from q is convex and scaled by

[12]For brevity, I only present the hospital's problem in the main text, and discuss Medicare's problem in Appendix Section B.

a factor $w > 0$.

Medicare only observes Q, the number of admitted patients. It can conduct costly audits to recoup a share $\beta \in [0, 1]$ of revenue for unnecessary admissions, $\beta P(Q - q)$.[13] Let Q_N be the number of admissions in the case with *no technology adoption*. The hospital's payoff is:

$$\Pi_N(Q_N) = \underbrace{(1-\beta)PQ_N}_{\text{net admission revenue}} - \underbrace{\frac{1}{2}(w(Q_N - q))^2}_{\text{over/under-admit cost}} - \underbrace{kQ_N}_{\text{treatment cost}} \tag{1}$$

Hospitals solve for the Q that maximizes this:

$$Q_N^* = \frac{1}{w^2}((1-\beta)P - k) + q \tag{2}$$

Hospitals can also decide whether to purchase compliance technology that shields them from some audits but also makes over- and under-admitting more costly. Technology adoption changes the hospital's payoff in three ways. First, it reduces the amount Medicare recoups and scales β by a factor $\gamma \in [0, 1)$. Second, it increases the cost of unnecessary admissions by $r > 0$: with technology, the cost of admitting $Q - q$ patients unnecessarily is equal to the no-adoption cost of admitting $Q - q + r$ patients unnecessarily. Third, the technology has a fixed cost $f > 0$. If Q_A is the number of admissions in the case with *technology adoption*, then a hospital's payoff when adopting technology is:

$$\Pi_A(Q_A) = \underbrace{(1-\gamma\beta)PQ_A}_{\text{net admission revenue}} - \underbrace{\frac{1}{2}(w(Q_A - q + r))^2}_{\text{over/under-admit cost}} - \underbrace{kQ_A}_{\text{treatment cost}} - \underbrace{f}_{\text{fixed tech. cost}} \tag{3}$$

The Q that maximizes this is:

$$Q_A^* = \frac{1}{w^2}((1-\beta\gamma)P - k) + q - r \tag{4}$$

We can use the solutions for Q_N^* and Q_A^* to generate the first testable implication:

[13]For simplicity, I call β the audit rate and in the model assume that it is equal to the denial rate: all audits result in a denial. In the RAC setting, β would correspond to the denial rate. But since the denial rate is monotonically increasing in the audit rate (Figure H2), the testable implications of the model should hold for both the audit and the denial rate.

Testable Implication 1. *Increasing the audit rate will decrease admissions.*

$$\frac{dQ_N^*}{d\beta} = \frac{-1}{w^2}P < 0,\ \frac{dQ_A^*}{d\beta} = \frac{-1}{w^2}\gamma P < 0$$

To ensure that $Q_N^* \geq 0$ and $Q_A^* \geq 0$ for all $q \in [0, \overline{q}]$, we need to assume that $(1-\beta\gamma)P - k \geq rw^2$, or that the additional misreporting cost r the technology imposes is not "too large." Then we can characterize the type of admissions affected by monitoring:

Testable Implication 2. *The marginal admission deterred by an increase in audit rate will be an unnecessary admission.*

Because $\frac{1}{w^2} > 0$ and $(1-\beta\gamma)P - k \geq rw^2$, Q will never fall below q.

$$Q_N^* = \frac{1}{w^2}((1-\beta)P - k) + q \geq q.$$

$$Q_A^* = \frac{1}{w^2}((1-\beta\gamma)P - k - rw^2) + q \geq q.$$

When β increases, hospitals reduce admissions in reverse order of medical need. In other words, they start by refusing to admit the last person they *would* have admitted under the lower β. Because both $Q_A, Q_N \geq q$, the marginal admission deterred by audits will have been ones admitted after q, meaning they were unnecessary.

Technology Adoption Decision. A hospital's decision to adopt technology can be characterized as depending its draw of q. In particular, it will adopt technology when $\Pi_A(Q_A^*(q)) - \Pi_N(Q_N^*(q)) > 0$. Let q^* be the q such $\Pi_A(Q_A^*(q)) - \Pi_N(Q_N^*(q)) = 0$. The derivative of $\Pi_A(Q_A^*(q)) - \Pi_N(Q_N^*(q))$ with respect to q is $P\beta(1-\gamma) > 0$. So because the value of adopting technology is increasing in q, hospitals that draw $q > q^*$ will adopt technology. To understand how β affects technology adoption, we need to know the sign of $\frac{dq^*}{d\beta}$:

$$\frac{dq^*}{d\beta} = \frac{f}{\beta^2 P(\gamma-1)} + \frac{P-k}{\beta^2 P(\gamma-1)} + \frac{P(\gamma+1)}{2w^2} \tag{5}$$

The effect of an increase in β on technology adoption is theoretically ambiguous – on the one hand, an increase in β means hospitals will lose more if they are audited, so they have more to gain from the audit protection technology adoption provides. On the other hand, because increasing β also causes hospitals to reduce admissions even in the absence of technology, there is less revenue at risk of audit. This is reflected in Equation 5, as the first term is negative while the second term is positive. In Figure H3, I depict the adoption decision for a simulation with parameters set to match the average hospital in 2010. In the simulation, $\frac{dq^*}{d\beta} < 0$, meaning more hospitals will adopt technology as the audit rate increases.

In Appendix Section B, I lay out Medicare's problem of setting β and extend the model to incorporate the case where Medicare considers subsidizing or purchasing the technology for hospitals. I also discuss other settings where this model could potentially apply. Solving for the optimal β^* requires calibrating for the model parameters, as well as taking a stance on the weight of each component in Medicare's objective function. This in and of itself is an interesting and open question which is left for further research.

4 Data and Identification Strategies

4.1 Data

The hospital-level analysis uses four main data sets. First, I use audit-level administrative data on the RAC program acquired through a Freedom of Information Act request. The data span 2010 to 2020 and include claim-specific information on 100 percent of RAC audits, such as characteristics of the audited claim (e.g., hospital, admission date, discharge date, diagnosis, Medicare payment) and of the audit (e.g., audit date, audit decision, amount of payment reclaimed or corrected, appeals). The dataset covers 4.5 million audits of inpatient stays.

Second, I use Medicare inpatient and outpatient claims data from 2007 to 2015. I merge the RAC audit data with the Medicare inpatient claims data (Medicare Provider Analysis and Review; MEDPAR) by matching on the following elements: provider, admission and discharge dates, diagnosis-related group, and initial payment amount. I am able to identify whether a claim was audited for 99.6 percent of Medicare inpatient claims between 2007 and 2015. I also conduct analyses using Medicare Outpatient claims to assess ED visit outcomes.

Third, I use hospital cost data from the Healthcare Cost Report Information System (HCRIS), which collects cost reports that hospitals submit to Medicare. In particular, HCRIS provides yearly measures of hospital administrative costs.

Fourth, I use data on IT adoption from the Healthcare Information and Management Systems Society (HIMSS) Analytics Database, which is a yearly survey of IT used by hospitals and other healthcare providers. HIMSS asks hospitals each year to report the types of IT they are planning to or have already installed. In particular, I focus on medical necessity checking software, which hospitals use to assess the medical necessity of care in real-time. Additionally, to study heterogeneity across hospital types, I also use hospital characteristics from the Medicare Provider of Services file and hospital group affiliations from Cooper et al. (2019).

Table I presents summary statistics by RAC region. Hospitals in Regions B (Midwest) and C (South) have much lower audit rates than hospitals in Regions A (Northeast) and D (West). Within each region, rural hospitals, small hospitals, non-profit hospitals, and hospitals with a higher share of short stay Medicare admissions are more likely to be audited (Figure H4). Appendix Section A.3 further explores the claim-level and hospital-level characteristics associated with auditing in further detail.

In the patient-level analysis of ED visits, I use the Florida State Emergency Department Database (SEDD) and State Inpatient Database (SID) between 2010 and 2015. I focus on Florida because it is the only state that reports ED arrival hour in the publicly available data for both the inpatient *and* emergency department datasets; Medicare's Inpatient and Outpatient files do not report this variable.[14] The most granular unit of time for ED arrival in my data is the hour. SEDD includes

[14]ED visits are known to be difficult to identify using claims data, as there is no standard method or definition. For example, whether a patient who receives an ED triage evaluation without emergency clinician professional services (e.g., evaluation by a primary care clinician) is considered an "ED visit" has been found to vary across different data sources (Venkatesh et al., 2017). Further, in my attempt to assemble a panel of ED visits using Medicare claims, I uncovered inconsistencies in the data that, after consulting with ResDAC, lead me to conclude that across-year and across-provider comparisons of ED visits are untenable using the Medicare claims (ResDAC, 2022).

discharge-level data on every outpatient ED visit, and SID includes every inpatient stay (and denotes whether the patient was admitted as inpatient from the ED). I combine the two to construct the universe of ED visits in Florida hospitals in this time period. I proxy for patient health after an ED visit by considering whether the patient revisits any hospital in Florida shortly after, either as an ED visit or an inpatient visit.[15] I use this proxy because mortality is not observable in hospital discharge data such as SID and SEDD. Table GI presents patient characteristics common across MEDPAR and SID/SEDD, and compares the overall inpatient sample (MEDPAR), border hospital inpatient sample (MEDPAR), inpatients admitted from a Florida ED (SID), and patients admitted from a Florida ED who arrived at the ED within 3 hours of midnight (SID). The samples are similar in terms of age, sex, race, and share with a recent inpatient stay.

Table II reports summary statistics for before- and after-midnight arrivals before the Two Midnights rule, before and after the rule was in effect. Figure 2 plots the quarterly share of before- and after-midnight Medicare ED arrivals who are admitted as inpatient. Prior to the Two Midnights rule, after-midnight arrivals are more likely to be admitted as inpatient, but this gap closes once the Two Midnights rule is implemented in 2013Q3. After-midnight ED arrivals tend to be older, less likely to be white, less likely to be female, and sicker (i.e., more chronic condition, more likely to have had a recent hospital visit, and higher predicted admission likelihood) than before-midnight arrivals. This pattern is consistent in both the pre-policy and post-policy periods, which supports making a parallel trends assumption about the before- and after-midnight arrivals.

4.2 Identifying the Effect of Monitoring on Hospital Outcomes

The aim of the first, hospital-level identification strategy is to understand how hospital behavior responds to audits. To understand the causal effect of auditing, we need to focus on the year medical necessity audits begin: 2011. I leverage variation only in the *first year* of the expansion because audit rates in subsequent years are endogenous. Hospitals may respond to audits by adjusting their

[15] Hospital inpatient readmission rates are a widely used measure of hospital quality (Krumholz et al., 2017). Reducing hospital readmissions was the focus of the Hospital Readmissions Reduction Program, one of the value-based purchasing programs introduced as part of the Affordable Care Act.

behavior, which then affects RACs' willingness to audit down the line. There is also a mechanical negative relationship between the number of claims previously audited and the number of claims yet to be audited. The pool of eligible claims may vary across the different regions so the speed with which they are exhausted may differ, which will affect how audit rates evolve over time.

I focus on on hospitals close to the RAC border and compare hospitals who are subject to a more-aggressive RAC to their neighbors who are subject to a less-aggressive one. I then look at how their behavior changes after 2011 using a difference-in-difference specification, with two modifications. First, I include local fixed effects to compare hospitals that are neighbors to each other. Second, I instrument for a hospital's audit rate using a measure of how aggressively its RAC audits *other* hospitals.

Border Hospital Sample: Figure 1a illustrates the sharp changes in audit intensity at the border between RAC regions. The changes across the RAC borders are twice as large as the changes across state borders *within* each RAC region. I focus on the sample of hospitals close to the border, where I define "close" as being within one hundred miles of it. By focusing on this subset of hospitals, this research design requires a weaker parallel trends assumption relative to one incorporating all hospitals. Here, I only need to assume that *geographically proximate* hospitals are not on differential trends. Table I columns 1 and 2 compare the border hospital sample to the overall sample. Border hospitals tend to be smaller, more rural, and more likely to be non-profit than the overall sample. Because these characteristics correlate with audit rate, border hospitals have a higher 2011 audit rate than the overall sample. Additionally, a larger share of border hospitals come from RAC regions B and C.

Neighbor Comparison Groups: To ensure that I am comparing hospitals that are close *to each other*, and not just hospitals that are close to the border, I identify a unique set of neighbors for each hospital and call this its "neighbor comparison group."[16] I define a hospital's neighbor comparison group to be hospitals on the *other* side of the border within 100 miles. I then include a fixed effect for each group interacted with a year indicator in my specification. With these fixed

[16]In identifying a unique set of neighbors for each hospital, I follow Dube et al. (2010), whose state border-county identification strategy allows individual counties to be paired with unique sets of adjacent counties.

effects, I effectively "stack together" local comparisons of hospitals to their neighbors across the border. Table GII reports the correlations between 2010 hospital and stay characteristics with audit rates in the two samples. Comparing *within* neighbor comparison groups for the border hospital sample, the 2011 audit rate is uncorrelated or weakly correlated with 2010 hospital characteristics. In contrast, these correlations are statistically significant and larger in magnitude in the overall sample.

Figure H5 illustrates how I construct a hospital's neighbor comparison group. The hospital in question is on the Oklahoma side of the border (RAC Region C) and has an audit rate of 1.44 percent. The members in its neighbor comparison group are the hospitals on the other side of the border within a hundred miles – in this case, that would be hospitals in Kansas (RAC Region D) that face a much higher average audit rate of 5.42 percent. Together, the Oklahoma hospital and its neighbors in Kansas form the neighbor comparison group for the Oklahoma hospital.

Including these group-year fixed effects improves upon a specification with just border fixed effects (or border-year fixed effects) in two ways. First, it accounts for local geographic trends in utilization and spending. Prior work in the healthcare literature has documented substantial geographic variation in Medicare spending (Skinner, 2011; Finkelstein et al., 2016). Each RAC border spans hundreds of miles. A specification with just border fixed effects would therefore end up comparing hospitals that are close to the border, but possibly far from *each other*; this may not adequately account for local trends. Second, constructing these neighbor comparison groups allows me to include hospitals at the intersection of multiple borders. In a specification with border fixed effects, I would have to either arbitrarily assign these hospitals to one of their adjacent borders, or exclude them from the analysis.

Because a hospital can be a member of multiple neighbor comparison groups, the sample includes repeated hospital observations which will have correlated errors. To account for this, I divide the border into segments and cluster at the border segment level. Figure H6 illustrates the border segments used for clustering, with each segment in a different color. Each border segment is a hundred miles except for segments that cross state lines, which are split at the state border.

Event Study Specification: The event study specification of interest for the hospital-level strategy is:

$$Y_{ht} = \sum_{\tau=2007}^{2015} \mathbb{1}[t = \tau] \times X_h^{2011} \beta^\tau + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \ . \tag{6}$$

In Equation 6, Y_{ht} is an outcome for hospital h in year t, X_h^{2011} is the hospital's 2011 audit rate, $\phi_{g(h)t}$ is a hospital's neighbor comparison group $g(h)$-times-year fixed effect, and ψ_h is a hospital fixed effect. The main results are in the form of an event study to allow for dynamic responses, so there is a β^τ for each year τ between 2007 and 2015, omitting 2010. β^τ can be interpreted as the effect of a one percentage point increase in 2011 audit rate on a hospital outcome in year τ, relative to 2010.

Audit Rate Instrument: One concern with estimating Equation 6 directly is the endogeneity of a hospital's 2011 audit rate X_h^{2011} – that is, that $E[\varepsilon_{ht}|X_h^{2011}] \neq 0$. This could arise if hospitals that are targeted by RACs were on a differential trend relative to their neighbors – for example, if RACs target lower-quality hospitals and admissions at lower-quality hospitals were already on a downward trend. To isolate variation driven by the *RAC* and not by the hospital, I consider how aggressively the RAC audits *other hospitals* under its jurisdiction. In practice, I instrument for a hospital's 2011 audit rate with the audit rate of other hospitals in the same state. For each hospital, I calculate the "leave-one-out state audit rate," which is the state average *excluding* that hospital. It is defined as:

$$Z_h^{2011} = \frac{1}{n_{s(h)} - 1} \sum_{h' \in \ s(h) \setminus h} X_{h'}^{2011} \ , \tag{7}$$

where $X_{h'}^{2011}$ is the 2011 audit rate for hospital h' that is in the same state $s(h)$ as hospital h. Because RAC borders fall along state lines, hospital h' is subject to the same RAC as hospital h. There are $n_{s(h)}$ total hospitals in the state.

The reduced form event study specification is:

$$Y_{ht} = \sum_{\tau=2007}^{2015} \mathbb{1}[t = \tau] \times Z_h^{2011} \gamma^\tau + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \ . \tag{8}$$

In order to interpret the coefficients as the effect of a one percentage point increase in the 2011 audit rate (as in Equation 6), I scale the γ^{τ} coefficients in Equation 8 by the correlation between X_h^{2011} and Z_h^{2011} (after accounting for hospital-group fixed effects).[17]

I also report results that pool the post-2011 effects into a single coefficient:

$$Y_{ht} = \mathbb{1}[t \geq 2011] \times X_h^{2011} \beta^{post} + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \ . \tag{9}$$

In this case, the reduced form specification is:

$$Y_{ht} = \mathbb{1}[t \geq 2011] \times Z_h^{2011} \beta^{post} + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \ . \tag{10}$$

Identification Assumptions and Checks: The identification strategy relies on three underlying premises: first, that the changes in audit rate at the border are driven by RACs (*exogeneity*); second, that neighboring hospitals are "comparable" to each other (*parallel trends* and *homogeneous treatment effect*); and third, that the leave-one-out state audit rate is a valid instrument for the hospital audit rate (*exclusion restriction* and *monotonicity*).

First, suppose that the sharp changes in audit rate at the border in Figure 1a were *not* driven by variation across RACs. If they were instead driven by hospital or patient characteristics (or a policy that is correlated with them) we would expect to see similarly sharp variation at the border in these characteristics as well. But as shown in Table GII, there is little correlation between audit rate and hospital and patient characteristics within neighbor comparison groups along the border.

On each side of the border, RACs face the same incentives to audit and presumably similar local labor costs. So what could be driving these sharp differences in audit rate across the RAC border? One explanation could be that because each RAC comes from a different industry background,[18]

[17]In particular, I generate eight instruments, each of which is an interaction of Z_h^{2011} with a year indicator, and combine them to instrument for the interactions of X_h^{2011} with a year indicator. For example, I use $\sum_{\tau=2007}^{2015} \mathbb{1}[t = \tau] \times Z_h^{2011}$ to instrument for $\mathbb{1}[t = 2007] \times X_h^{2011}$, and the coefficient is equal to the correlation between X_h^{2011} and Z_h^{2011} when $\tau = 2007$,and zero for $\tau \neq 2007$. I repeat this for all 8 years between 2007 and 2015. I implement this in a two-stage procedure to allow for clustering in the estimation of standard errors.

[18]For example, the RAC in Region A is primarily a debt collection agency, while the RAC in Region C is a healthcare data analysis company.

this variation in prior experience translates into differences in how RACs approach auditing. These differences would be especially pronounced in 2011, as it is the first year that RACs were allowed to conduct medical necessity audits. Another explanation could be that RACs set their audit strategies at the regional, rather than local, level. For example, this would be the case if RACs combined data from all hospitals in its region to train a single algorithm to flag claims, so a hospital's audit rate would reflect within-region spillovers via the algorithm. Or, it could be that RACs set their audit rates based on the average *regional* labor cost of hiring auditors, rather than the local labor cost.

Second, the border hospitals must be "comparable" to each other. Note that I do not need to assume there are *no differences* in hospitals across the RAC border – this would be clearly violated by the fact that hospitals on opposite sides of the border are in different states. Instead, I need to make weaker assumptions: that hospitals on each side of the border have parallel trends and homogeneous treatment effects. With the inclusion of group-year fixed effects, for the parallel trends assumption we only need that neighboring hospitals on opposite sides of the border do not differentially deviate from local trends. While this assumption is in principle untestable, a lack of preexisting differential trends in the event study would support it.[19] The parallel trends assumption could be violated if the results are actually driven by state policies changing over time. In robustness tests I show that the results are robust to omitting individual states, suggesting that they are not driven by any individual state's policy changes. If, however, states developed policies in response to their RACs' aggressiveness (i.e., they make Medicaid denials more aggressive in response to a less aggressive RAC), then the results would reflect both a response to RAC auditing *and* these state policy responses. But it appears that in this time period, there was little transparency on the overall aggressiveness of different RACs – CMS did not release statistics about it until much later, as evidenced by the AHA's push to survey its members on their RAC experiences.

Since a hospital's audit rate is continuous and therefore "fuzzy," I also need to assume that hos-

[19]Restricting the comparison to *border* hospitals allows me to make a weaker parallel trends assumption than a comparison of *all* hospitals. Figure H7f shows the results from an alternate specification that includes all hospitals; there is evidence of differential pretrends when comparing across all hospitals.

pitals in the border sample have homogeneous treatment effects (de Chaisemartin and D'HaultfŒuille, 2018). One concern is that if hospitals on opposite sides of the border are very different at baseline, then they may also have heterogeneous responses to auditing. But as shown in Table GII, within neighbor comparison groups, hospitals that are subject to different audit rates are still relatively similar by other measures.

Finally, to justify using the leave-one-out state audit rate as an instrument, I need the exclusion restriction as well as a monotonicity assumption. The exclusion restriction requires that the leave-one-out audit rate only affects a hospital's outcomes via its own audit rate. To violate this, time-varying confounders like changes in state policies would have to be consistent across multiple states and occur simultaneously in 2011. Non-time-varying confounders like existing state policies are absorbed by the hospital fixed effect in the difference-in-difference specification. The exclusion restriction could also be violated by reverse causality – if, say, the leave-one-out audit rate reflects a given hospital's spillovers onto other hospitals in the same state. This could be true if a given hospital has a large market share, or if hospitals in the same chain have spillovers on each other. To address this concern, I run robustness tests that instrument using the average audit rate of hospitals in the same state but in other markets, as well as hospitals in the same state but not in the same chain. The results from using each of these instruments are similar to the main results (Figure H8). Additionally, note that we need to make an assumption about monotonicity in audit intensity across RACs – that a given hospital would be subject to more audits under a more-aggressive RAC, and fewer audits under a less-aggressive RAC (Imbens and Angrist, 1994).

4.3 Identifying the Effect of Monitoring on Patient Outcomes

I next turn to the patient-level identification strategy that leverages the Two Midnights rule. Here, we can test for the second testable implication of the model: monitoring only deters unnecessary admissions. I split ED visits by whether the patient arrived before midnight (lower audit risk) or after midnight (higher audit risk), and then compare them pre– and post–policy in a difference-in-difference specification.

Specification: The event study specification is:

$$Y_v = \sum_{\tau=2010}^{2016} \mathbb{1}[y=\tau] \times \mathbb{1}[t \geq 00{:}00]\beta^\tau + \boldsymbol{W_v'}\boldsymbol{\gamma} + \lambda_{hy} + \phi_{ht} + \varepsilon_v \ , \tag{11}$$

where ED visit v occurs in fiscal year y at hospital h, and the ED arrival hour of the visit is $t \in [21{:}00, 03{:}00)$ (that is, between 9PM and 3AM). Y_v is the outcome of interest, such as an indicator for whether the ED visit resulted in an inpatient admission or whether the patient revisited a hospital within thirty days. $\mathbb{1}[q=\tau]$ is an indicator for whether the visit occurred in fiscal year τ, omitting 2013. $\mathbb{1}[t \geq 00{:}00]$ is an indicator for whether the patient arrived at the ED after midnight. λ_{hy} is a hospital-year fixed effect, and ϕ_{ht} is a hospital–ED-arrival-hour fixed effect. W_v are controls for patient characteristics, including patient age, race, Hispanic, point of origin, an indicator for whether last ED visit was within 30 days, number of chronic conditions, and average income in patient's zip code. β^τ is the coefficient of interest and can be interpreted as the effect of the increased audit likelihood on after-midnight ED arrivals in year τ, relative to 2013.

Equation 12 pools the event study into a single post-policy coefficient β:

$$Y_v = \mathbb{1}[q \geq 2013\text{Q3}] \times \mathbb{1}[t \geq 00{:}00]\beta + \boldsymbol{W_v'}\boldsymbol{\gamma} + \lambda_{hq} + \phi_{ht} + \varepsilon_v \ . \tag{12}$$

Here $\mathbb{1}[q \geq 2013\text{Q3}]$ is an indicator for whether the quarter of the visit occurs after the Two Midnights rule is implemented in 2013Q3, and λ_{hq} is a hospital-quarter fixed effect.

Identifying Assumption and Checks Interpreting β and β^τ as the causal effects of auditing requires two assumptions. First is the standard parallel trends assumption – that absent the Two Midnights rule, before- and after-midnight patients would have trended similarly. To substantiate this, I check that there are no differential pre-trends between the two groups in the event study figures.

The second assumption is that there is no manipulation of the ED arrival hour. This would be violated if, for example, hospitals misreported after-midnight ED arrivals as arriving before midnight. If this were the case, we would expect to see bunching of ED arrivals right before midnight once the policy is implemented (that is, an increase in the share of patients reported arriving between 11:00 PM and midnight). Figure H9 plots the share of patients by ED arrival hour, pre- and post-policy – bunching before midnight does not appear post-policy. I test this empirically in Table GIII by looking at whether there is a higher share of patients arriving in the hour before midnight (column 1) or a lower share of patients arriving after midnight (column 2) post-policy. Neither of these measures changes after the Two Midnights rule is implemented.

Practically speaking, it may be difficult for hospitals to manipulate the ED arrival hour in response to the Two Midnights rule. The arrival hour is recorded as soon as the patient walks in to the ED, which makes it more difficult to manipulate than a measure that is recorded later on. Additionally, to game the Two Midnights rule, hospitals would have to make after-midnight arrivals look like before-midnight ones. This would require them to actively *move up* a patient's ED arrival hour to an earlier time, rather than a more passive form of misreporting by "dragging their feet" to record a later arrival hour, in contrast to other contexts where this kind of behavior has been found (e.g., Chan (2016); Jin et al. (2018)).

We may also be concerned that hospitals respond to the Two Midnights rule by simply extending all stays to span two midnights. This would not be a threat to identification per se; instead we would simply see no effect of the Two Midnights rule on inpatient admission likelihood. Due to patient confidentiality reasons in the discharge data, I cannot directly observe how long a patient's entire stay in the hospital spanned. However, I do not find evidence that after-midnight patients have additional charges, diagnoses, or procedures after the rule is implemented (Table GIV), suggesting that hospitals did not respond to the Two Midnights rule by extending stay duration.

5 Results

5.1 Hospital Outcomes: Admissions, Revenue, Costs, and IT Adoption

Figure 3 plots a binscatter of the cross-sectional relationship between the instrument, the leave-one-out state audit rate, and hospital audit rates in the border hospital sample. The leave-one-out audit rate explains 74 percent of the variation in the actual audit rate, with a coefficient of 1.04. There is a positive linear relationship between the two and it is not driven by outliers, which supports using a linear specification.

Figure 4 presents the first set of main results: the event study coefficients on hospital-level outcomes, scaled by the cross-sectional correlation between the audit rate and the leave-one-out audit rate in Figure 3. Table III reports the yearly coefficients for 2011 to 2015 (for brevity, the pre-2011 coefficients are estimated but not reported in the table). Figures 4a and 4b plot the results for log Medicare admissions and log Medicare inpatient revenue, where inpatient revenue is defined as the sum of all Medicare inpatient payments. Prior to 2011, hospitals with higher audit rates do not seem to be on differential trends relative to their neighbors across the border. Starting in 2011, there is a decline and then a plateau in Medicare admissions and inpatient revenue among hospitals subject to a more-aggressive RAC. A one percentage point increase in the 2011 audit rate results in a 1.1 percent decrease in admissions in 2011, which increases in magnitude to a 1.9 percent decrease by 2012 and 2013. Similarly, a one percentage point increase in the 2011 audit rate results in a 1.0 percent decrease in inpatient revenue in 2011, and then a 1.8 percent decrease by 2012 and a 2.8 percent decrease by 2013. These results are consistent with the first testable implication of the model, that admissions will decrease in response to higher audit rates. Extrapolating to the overall hospital sample (albeit under fairly strong assumptions, as discussed in Appendix Section E) indicates that RAC audits saved the Medicare program $9.28 billion between 2011 and 2015.

I next turn to the administrative burden RAC audits place on hospitals. Figure 4 and Table III columns 5-6 present results on two dimensions of this burden: hospital administrative costs and IT adoption. Figure 4c plots estimates of the effect on log administrative costs, as reported in hospital cost reports. A one percentage point increase in RAC auditing in 2011 results in an immediate 1.5 percent uptick in administrative costs, but this increase lasts for only about a year. This result corroborates the findings of a 2012 AHA survey in which 76 percent of hospitals reported that RAC audits increased their administrative burden (American Hospital Association, 2012).

Investments into technology to improve compliance could be one driver of higher administrative costs and lower admissions. But as shown in the model, the effect of audits on technology adoption is theoretically ambiguous – thus, this is an empirical question. Figure 4d presents the event study results for whether a hospital reported installing medical necessity checking software in a given year. In response to a one percentage point increase in the 2011 audit rate, hospitals were 2.2 percentage points *more* likely to report that they were installing or upgrading this software in 2012 (a 3.7 percent increase relative to the 59 percent of hospitals that had this software installed in 2010). This is also in line with the findings in the 2012 AHA survey: a third of hospitals reported responding to RACs by installing tracking software (American Hospital Association, 2012).

The increased take up of this software suggests that the effect of the RAC program on admissions was mediated in part by technology adoption. I also provide four additional pieces of cross-sectional evidence that support this mechanism. First, hospitals with the technology already installed in 2010 were less likely to be audited and less likely to have a payment reclaimed (Figures H10a and H10b), which is consistent with the notion that this software protects hospitals from audits. Second, larger hospitals were more likely to already have the software installed in 2010, which is consistent with the model prediction that hospitals with larger draws of q should be more willing to adopt technology (Figure H10c). Third, among hospitals without the software, the ones that lost more money to RAC audits in 2011 were more likely to adopt it in later years (Figure H10d). If hospitals face similar costs of technology adoption, then hospitals that risk greater losses in the future should be more willing to make this investment. Fourth, Table GV shows that hospitals that adopted the software after 2011 had larger reductions in admissions from 2010 to 2012, which is consistent with hospitals installing the software with the goal of reducing unnecessary admissions.

To estimate the total savings from RAC audits, Figure H11 plots the results for the payments directly reclaimed by RACs. A one percentage point increase in audit rate in 2011 is associated with $314,115 in demanded payments in 2011 per hospital. There are additional demands in subsequent years as well, although the magnitude diminishes over time. Comparing the savings from deterred admissions to reclaimed payments, I calculate that 89 percent of government savings from the RAC program are due to deterrence. RAC auditing brings in $24 in Medicare savings per dollar spent to run the program.[20] I can also use the estimates on administrative costs to compare Medicare's savings to the burden the RAC program imposed on hospitals. For every $1,000 in savings between 2011 and 2015, hospitals spent $218 in compliance costs.[21]

Next, I explore the effects on different types of admissions to understand where the deterrence savings stem from. Given policymakers' concerns about short stays being the primary driver of unnecessary stays, Figure 5 splits admissions by their length of stay. The effect is driven in large part by a reduction in short stays – that is, admissions with length of stay less than or equal to two days, which comprised 31 percent of stays on average in 2010. A one percentage point increase in the audit rate results in a 4.6 percent decrease in short stay admissions and a 4.6 percent decrease in revenue from these stays by 2012 (Table III). In contrast, there is a much smaller and statistically insignificant decrease in longer stay admissions.

Figure 6 then splits admissions by diagnosis. Specifically, I categorize diagnoses by the degree of payment errors associated with each Medicare Severity Diagnosis Related Group (MS-DRGs, also referred to as DRGs). I use the ranking of base DRGs[22] by payment error calculated by the Comprehensive Error Rate Testing Program (CERT) in 2010, which randomly samples Medicare claims to calculate improper payment rates (Centers for Medicare and Medicaid Services, 2011b). The purpose of the CERT program is to measure payment error rates across different Medicare

[20]For a one percentage point increase in 2011 audit rate, the government costs by 2015 are $88k, savings from reclaimed payments are $232k, and the total Medicare savings are $2.08 million. These numbers are calculated under the assumption that CMS *returned* 68 percent of reclaimed payments to hospitals. I assume this because in August 2014, Medicare announced a one-time option to return part of the reclaimed payments in exchange for hospitals dropping their appeals. See Section for more details on the settlement. Under the assumption that hospitals do *not* settle and Medicare keeps all the payments they demand, the savings by 2015 from reclaimed payments are $721k, and total government savings are $2.57 million. Thus in this case, RAC audits save $29 per dollar of monitoring costs, and deterred admissions account for 72 percent of the savings.

[21]The value of compliance costs by 2015 is $455k, compared to the total government savings of $2.08 million. Under the assumption that a hospitals do not settle and CMS does not return reclaimed payments to hospitals, the total government savings are $2.57 million, so the ratio between compliance costs and savings is $177 in hospital compliance costs per $1000 in Medicare savings.

[22]DRGs can be grouped into groups of 1-3 DRGs called "base DRG groups" where the underlying diagnosis is the same but the different DRGs represent different levels of severity. For example, the heart failure base DRG group comprises of three DRGs: heart failure with major complication/comorbidity (291), heart failure with complication (292), and heart failure without complication/comorbidity or major complications/comorbidity (293).

claim types, and RACs did not participate in this program. Figure H12b plots the audit rates for the top 20 base DRGs. Figures 6a and 6b plot the event study results, and show slightly larger and more sustained reductions in admissions for the top 20 base DRGs compared to DRGs outside of the top 20 (0.8 percent reduction), consistent with hospitals focusing on the types of diagnoses that policymakers considered most likely to be medically unnecessary. However, the difference between high- and low-error diagnosis groups is smaller than the difference between short and long stays. This is likely because policymakers framed the unnecessary admissions problem as a length of stay issue, rather than a diagnosis-specific issue (Centers for Medicare and Medicaid Services, 2013; Medicare Payment Advisory Commission, 2015).

The list of top 20 base DRGs includes both emergent (i.e., arising from an emergency) and non-emergent diagnoses – they range from major joint replacement, where only 13% of stays originate in the ED, to chest pain, where 83% of stays originate in the ED. Emergent and non-emergent stays differ both in the potential health risks a deterred stay poses for a patient, but also in terms of the tactics hospitals can use to reduce each type of admission. Thus, in Figures 6c and 6d I then split the top 20 base DRG groups into emergent (i.e., arising from an emergency) and non-emergent diagnoses.[23] There are reductions in both emergent and non-emergent cases, with a larger effect (but noisier) for non-emergent stays of 5.1 percent after 2015 compared to a 2.1 percent decrease among emergent stays.

The fact that both emergent and non-emergent admissions decrease indicates that the overall reduction in admissions was not attained through adopting medical necessity checking software *alone*. The software is most useful for emergent cases, as its purpose is to relay information to providers as they make care decisions in real time. But the decisions to reduce non-emergent admissions can be made at a higher level – say, if a hospital decides shift major joint replacements from the inpatient to the outpatient setting. Hospitals also reported hiring utilization management consultants and undergoing training and education in the AHA survey (American Hospital Association, 2012), which could reflect efforts to support decisionmakers in setting these hospital-wide

[23]The event studies begin in 2008 because of a 2007 reform to DRG categories Gross et al. (2023).

policies. In contrast to software adoption, these activities are not easily observed in the data. But we can infer from the reduction in non-emergent admissions and the survey evidence that these other efforts likely contributed to the uptick in administrative costs as well.

The event studies in Figures 4, 5, and 6 illustrate the dynamics of hospital responses. Admissions and revenue decline steadily between 2011 and 2012. The fact that this happened over two years rather than immediately likely reflects two factors. First, some of the 2011 admissions occurred before hospitals knew how aggressively they would be audited by RACs. Second, it may have taken time to implement policies or technology to reduce unnecessary admissions. After 2012, admissions remained at their decreased levels – even in 2014 and 2015, when audit activity slowed down significantly. In contrast, there was an immediate but short-lived increase in hospital administrative costs in 2011. The timing of this effect suggests that the bulk of hospital compliance costs were *fixed*, rather than *variable*, costs. If the costs were primarily variable costs, like the paperwork associated with responding to audits, then we would expect to see elevated costs for several years, since audits continued until 2015 (Figure 1b). Instead, the one-time spike in administrative costs is consistent with hospitals making upfront investments like installing software to improve compliance.

Importantly, the dynamic effects should be interpreted as capturing hospitals' responses to a combination of the exogenous audit rate in 2011 and all the (possibly endogenous) audit rates in subsequent years. As shown in Figure H13, the high-audit regions' audit rates decrease over time (relative to their highest point in 2012) and while low-audit regions' audit rates continue to increase. Thus the estimates likely *understate* what we would see if RAC audit rates persisted over time. If high-audit hospitals anticipated that their audit rate would decrease, then they may not have pulled back as much on admissions or made investments to improve compliance. Likewise, if low-audit hospitals anticipated that their audit rate would increase, they may have decreased admissions or made investments in anticipation.

The dynamic effects also suggest that prior to 2011, the high rate of unnecessary admissions was not solely due to hospitals *knowingly* admitting them.[24] The event studies reveal that the full effect on admissions took two years to materialize – in contrast, other work has found that spending drops almost immediately in response to efforts to clamp down on Medicare fraud (Howard and McCarthy, 2021; Roberts et al., 2021; Leder-Luis, 2023).

Table GVI pools the post-2011 years of the main results into a single post-2011 coefficient, as in Equation 10. Given the dynamics of the results, the pooled coefficients are noisily estimated. Averaging across 2011 to 2015, there is a 1.5 percent reduction in overall admissions (although not statistically significant) and a 2.2 percent reduction in short stay admissions relative to the pre-period. Table GVII considers heterogeneity in the effect by hospital characteristics. The results point to rural, for-profit, smaller, and non-chain hospitals as being more responsive to audits.[25] Reassuringly, the increase in medical necessity checking software seems to be driven by hospitals that do not have the software installed in 2010.

In Appendix Section C, I check that the results are robust to instrumenting for the share of claims that are denied rather than just audited, using varying bandwidths to define the hospital sample, excluding hospitals that are very close to the border, using alternative instruments for audit rate, removing individual states or neighbor comparison groups, using varying border segment lengths for clustering, and running a placebo test using state borders in the interior of each RAC region. In Appendix D, I consider whether RAC audits affected rural hospital closure rates in subsequent years. If hospitals lost enough revenue from auditing that it caused them to close, then this would have important implications for patient access to care. I find that border hospitals subject to more auditing were no more likely to close in subsequent years, mitigating concerns about this channel.

[24] For the sake of simplicity, in Section 3 I model hospitals as having perfect information about q, and the technology acts by making it more costly for them to deviate from q. Another way to model compliance technology would be to assume that hospitals observe a noisy measure of q, and the technology is a costly way to improve the precision of their observation.

[25] The larger policy response by for-profit hospitals is in line with other work which has found that for-profit hospitals tend to be more responsive to Medicare policy changes (Silverman and Skinner, 2004; Dafny, 2005; Gross et al., 2023).

5.2 Patient Outcomes: Inpatient Admission Likelihood and Revisit Likelihood

I next turn to the results from the patient-level analysis. Figure 7 plots the event studies of the patient-level analysis of ED visits in Equation 11. There is no clear trend in the pre-policy coefficients, which supports making the parallel trends assumption. Immediately after the Two Midnights rule is implemented, there is a drop in the share of after-midnight ED arrivals that result in an inpatient admission. There is a symmetric increase in the share of patients who are not admitted, but are placed into observation.

Table IV reports the β coefficient from Equation 12. In columns 1 and 2, the coefficients on the inpatient indicator and observation indicator are symmetric in opposite directions. After the Two Midnights rule goes into effect, after-midnight arrivals are 0.7 percentage points (1.7 percent) less likely to be admitted as inpatient and 0.7 percentage points (14 percent) more likely to be placed in observation. There is no change in the share of patients who are sent home directly from the ED ("Not Admitted"). This indicates that for ED patients who are on the margin for being admitted as an inpatient, hospitals still preferred to keep them in the hospital rather than sending them home directly.

Next, I consider whether the reduction in inpatient admissions harmed patients. Panel 7d plots the event study results for an indicator of whether a patient revisited a hospital within thirty days of her ED visit, and column 4 in Table IV reports the pooled coefficient. Despite their reduced inpatient admission rate, there was no increase in revisits for after-midnight patients. This finding is consistent with the second testable implication in the model, that the marginal admission deterred by audits will be unnecessary. It is also in line with other work which has found that the marginal hospitalization has no effect on mortality (Currie and Slusky, 2020).

However, because only a small subset of patients should be on the margin of an admission, this null average effect may be masking heterogeneity across patients. Patients in the middle of the severity distribution should be more likely to be denied admission as a result of RAC audits, so one would also expect any health effects to be concentrated among these patients as well. To explore this heterogeneity, I predict a patient's severity based on information available at the outset of an

ED visit. Using data on ED visits between 9:00 AM and 3:00 PM (that is, a time window outside of that used for the main results), I estimate a logistic regression predicting whether a patient is admitted within thirty days of the visit, based on information available during an ED visit.[26] I then apply this prediction to the main sample to create a measure of predicted patient severity, and split patients into deciles of this measure. I reestimate the specification in Equation 12, interacting β with an indicator for each severity decile.

Figure 8 plots the heterogeneity by severity results for inpatient status and for revisits within thirty days, and Table GVIII reports the coefficients. The Two Midnights rule has no effect on admission rates for patients at the bottom and top severity deciles. Instead, the reduction in admissions is coming primarily from the middle of the severity distribution. There is a 5 percentage point, or 25 percent, decrease in admissions for patients in the fifth predicted decile. However, I do not see this pattern for revisits, as the coefficient on revisits is statistically insignificant at all risk deciles. Thus, the overall null effect on revisits is not masking heterogeneity by patient severity. Even among patients with the highest likelihood of being denied admission, there is no increase in revisits. I also restrict to particularly vulnerable patient populations as defined by age, number of chronic conditions, race, and income in Table GIX, and likewise find no effect on revisits on these subpopulations.

Table GX reports heterogeneity of the patient-level effect by hospital characteristics. Urban, teaching, for-profit, and smaller hospitals are more responsive to the rule. Notably, the response is driven by hospitals with the medical necessity checking software in 2012. This adds to the cross-sectional evidence that this software mediates hospitals' admission reductions – here, it likely helps hospitals decide whether to bill as an inpatient or observation stay by notifying them of the Two Midnights rule. Appendix Section C shows that the results are robust to varying the time window to define before- and after-midnight ED arrivals, the period used to measure hospital revisits, changing the prediction model training sample, as well as a falsification test on non-

[26]This includes patient demographics such as age-bin, sex, race, a Hispanic indicator, a point-of-origin indicator, and mean zip code income. It also includes hospital and quarter fixed effects; the number of visits, inpatient stays, or length of stay in the last month or last year; and any diagnoses and procedures recorded for stays within the last month or last year.

Medicare patients, who should not be directly affected by the Two Midnights rule.

5.3 Comparing the Hospital-level and Patient-level Outcomes

With both the hospital-level and patient-level approaches, I find that hospitals respond to higher audit risk by decreasing inpatient admissions. However, there do exist some differences across the two samples and results that warrant further discussion and investigation.

The first is the difference in the patient population considered in each approach. The hospital-level results captures all Medicare inpatient stays, regardless of admission source. The patient-level approach focuses on a much more narrow sample: patients who enter the ED in Florida around midnight. Table GI shows that the patient characteristics across the two samples are similar and that a large majority (73 percent) of Medicare admissions originate in the ED. But there is still the key difference that the patient-level sample consists *only* of emergent cases. We may therefore be concerned about the externality validity of extrapolating the null patient health effect from the patient-level sample to the overall hospital-level sample, where we see reductions in both emergent and non-emergent stays.

But something that points to the external validity of the patient-level results is that patient health tends to be worse among emergent stays relative to non-emergent stays. Figure H14 shows that 30-day mortality is greater for DRGs where a greater share of stays originate in the ED. Because emergent cases are most risk at harm, we might expect that any effect on patient harm in the patient-level sample should be *larger* than the effect for the overall sample. On the contrary, the patient-level results show no negative health effects for these high-risk patients.

Alternatively, the hospital-level specification can also be extended to look at ED visits by incorporating the Medicare Outpatient file into the analysis. I find evidence at the hospital level that is consistent with increased observation usage and a null effect on 30-day revisits and mortality among ED visits (Figure H15). However, note that due to known reliability issues with measuring emergency department visits in Medicare claims data, these results should be interpreted with caution. The results and the data reliability issues are explained in further detail in Appendix Section

F.

The second difference between the two sets of results is about what happens when audit rates *decrease*. The hospital-level results show that once the RAC program is scaled back in 2014, admissions do not “bounce back.” In contrast, in the patient-level results, before midnight arrivals’ inpatient admission rate increases after the Two Midnights rule reduces their audit rate (Figure 2). There is a a key difference between the two policy environments that could explain this discrepancy: the level of confidence hospitals had in whether they could be retroactively punished in the future.

With the Two Midnights rule, hospitals could be fairly confident that their admissions would be protected by the rule from future audits. Even if the rule was later reversed, hospitals could plausibly argue that at the time of the decision, they made it when the Two Midnights rule was active. In fact, the Two Midnights rule is still in effect today. However with the 2014 pause, hospitals could not be confident that auditing wouldn’t ramp back up in later years. Medicare made no indication that it was planning on pausing audits indefinitely – they actually announced several dates for restarting the audits, but kept pushing it back. The pause began at the end of 2014Q1 and was originally meant to end in 2014Q3. After several quarters of announced and then delayed resumption dates, inpatient RAC audits finally resumed in 2015Q4, although they were subject to significant limitations to reduce the administrative burden on providers. Since RACs had a lookback period of 3 years, admission decisions made in 2014 and 2015 could still be audited and denied up until 2017 and 2018, respectively.

6 Conclusion

In this paper, I consider the efficacy of monitoring for wasteful public spending by studying a large Medicare program that audited for unnecessary hospital admissions. I introduce a model to understand how hospital admissions, patient health, and technology adoption will respond to increased monitoring. According to the model, as audit rate increases, hospital admissions decrease, and these reductions will comprise of unnecessary admissions. The effect on technology adoption is theoretically ambiguous. In the empirical analysis, I first compare hospitals subject to differentially aggressive RACs and find that auditing has a large deterrence effect on hospital admissions. Then looking at patient-level outcomes, I find that the reduction in admissions stemmed primarily from *unnecessary* care, indicating that RAC audits targeted wasteful spending. I also find evidence that technology adoption mediates this response: as auditing increases, hospitals adopt software to detect the medical necessity of care. In all, these results show that monitoring can be a powerful tool to combat waste and improve the cost-effectiveness in public spending.

References

3M, "3M APCfinderTM Software with Medical Necessity Validation," May 2016.

AccuReg, "EngageCare Provider Medical Necessity," 2022.

Alexander, Diane and Molly Schnell, "The Impacts of Physician Payments on Patient Access, Use, and Health," Working Paper w26095, National Bureau of Economic Research July 2019.

Allingham, Michael G. and Agnar Sandmo, "Income tax evasion: a theoretical analysis," *Journal of Public Economics*, November 1972, *1* (3), 323–338.

Almond, Douglas and Joseph J. Doyle, "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays," *American Economic Journal: Economic Policy*, August 2011, *3* (3), 1–34.

American Hospital Association, "Exploring the Impact of the RAC Program on Hospitals Nationwide: Results of AHA RACTrac Survey, 1st Quarter 2012," May 2012.

_ , "Exploring the Impact of the RAC Program on Hospitals Nationwide: Results of AHA RACTrac Survey, 4th Quarter of 2012," March 2013.

_ , "Exploring the Impact of the RAC Program on Hospitals Nationwide: Results of AHA RACTrac Survey, 1st Quarter 2014," May 2014.

Andreoni, James, Brian Erard, and Jonathan Feinstein, "Tax Compliance," *Journal of Economic Literature*, 1998, *36* (2), 818–860.

Baron, David P. and David Besanko, "Regulation, Asymmetric Information, and Auditing," *The RAND Journal of Economics*, 1984, *15* (4), 447–470. Publisher: [RAND Corporation, Wiley].

Barosso, Gerri, "How to Identify Hospital Claims for Emergency Room Visits in the Medicare Claims Data," July 2015.

Burde, Howard, "THE HITECH ACT: An Overview," *AMA Journal of Ethics*, March 2011, *13* (3), 172–175.

Carroll, C., "Impeding Access or Promoting Efficiency? Effects of Rural Hospital Closure on the Cost and Quality of Care," March 2019.

Centers for Medicare and Medicaid Services, "Implementation of Recovery Auditing at the Centers for Medicare and Medicaid Services FY2010 Report to Congress," Technical Report 2011.

_ , "Medicare Fee-For-Service 2010 Improper Payment Report," Technical Report 2011.

_ , "Recovery Auditing in the Medicare and Medicaid Programs for Fiscal Year 2011 FY 2011 Report to Congress," Technical Report 2012.

— , “2 Midnight Inpatient Admission Guidance &Patient Status Reviews for Admissions on or after October 1, 2013,” Technical Report 2013.

— , “Hospital Appeals Settlement Process,” 2014.

— , “Statement of Work for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1- 4,” June 2015.

— , “Clarifying Medical Review of Hospital Claims for Part A Payment,” MLN Matters MM10080 May 2017.

— , “Improper Payment Rates and Additional Data,” Technical Report 2022.

Chan, David C., “Teamwork and Moral Hazard: Evidence from the Emergency Department,” *Journal of Political Economy*, June 2016, *124* (3), 734–770.

Clemens, Jeffrey and Joshua D. Gottlieb, “Do Physicians’ Financial Incentives Affect Medical Treatment and Patient Health?,” *American Economic Review*, April 2014, *104* (4), 1320–1349.

Cooper, Zack, Stuart V Craig, Martin Gaynor, and John Van Reenen, “The Price Ain’t Right? Hospital Prices and Health Spending on the Privately Insured*,” *The Quarterly Journal of Economics*, February 2019, *134* (1), 51–107.

Cubanski, Juliette, Tricia Neuman, and Meredith Freed, “The Facts on Medicare Spending and Financing,” August 2019.

Cunningham, Susan, Mark McMillan, Sara O’Rourke, and Eric Schweikert, “Cracking down on government fraud with data analytics,” October 2018.

Currie, Janet and David Slusky, “Does the Marginal Hospitalization Save Lives? The Case of Respiratory Admissions for the Elderly,” Working Paper w26618, National Bureau of Economic Research January 2020.

Cutler, David, “The Incidence of Adverse Medical Outcomes under Prospective Payment,” *Econometrica*, 1995, *63* (1), 29–50.

Cutler, David M. and Dan P. Ly, “The (Paper)Work of Medicine: Understanding International Medical Costs,” *Journal of Economic Perspectives*, June 2011, *25* (2), 3–25.

Dafny, Leemore S., “How Do Hospitals Respond to Price Changes?,” *American Economic Review*, December 2005, *95* (5), 1525–1547.

de Chaisemartin, C and X D’HaultfŒuille, “Fuzzy Differences-in-Differences,” *The Review of Economic Studies*, April 2018, *85* (2), 999–1028.

Deshpande, Manasi and Yue Li, “Who Is Screened Out? Application Costs and the Targeting of Disability Programs,” *American Economic Journal: Economic Policy*, November 2019, *11* (4), 213–248.

Doyle, Joseph J., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner, "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns," *Journal of Political Economy*, February 2015, *123* (1), 170–214.

Dranove, David, Chris Forman, Avi Goldfarb, and Shane Greenstein, "The Trillion Dollar Conundrum: Complementarities and Health Information Technology," *American Economic Journal: Economic Policy*, November 2014, *6* (4), 239–270.

Dube, Arindrajit, T. William Lester, and Michael Reich, "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties," *The Review of Economics and Statistics*, 2010, *92* (4), 945–964.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "The Value of Regulatory Discretion: Estimates From Environmental Inspections in India," *Econometrica*, 2018, *86* (6), 2123–2160.

Dunn, Abe, Joshua D. Gottlieb, and Adam Hale Shapiro, "Administration above Administrators: The Changing Technology of Health Care Management," *AEA Papers and Proceedings*, May 2020, *110*, 274–278.

Einav, Liran, Amy Finkelstein, and Neale Mahoney, "Provider Incentives and Healthcare Costs: Evidence From Long-Term Care Hospitals," *Econometrica*, 2018, *86* (6), 2161–2219.

Eliason, Paul J., Paul L. E. Grieco, Ryan C. McDevitt, and James W. Roberts, "Strategic Patient Discharge: The Case of Long-Term Care Hospitals," *American Economic Review*, November 2018, *108* (11), 3232–3265.

Ellis, R. P. and T. G. McGuire, "Provider behavior under prospective reimbursement. Cost sharing and supply," *Journal of Health Economics*, June 1986, *5* (2), 129–151.

_ and _ , "Hospital response to prospective payment: moral hazard, selection, and practice-style effects," *Journal of Health Economics*, June 1996, *15* (3), 257–277.

Experian Health, "Product Sheet: Medical Necessity," 2022.

Finkelstein, Amy and Matthew J Notowidigdo, "Take-Up and Targeting: Experimental Evidence from SNAP*," *The Quarterly Journal of Economics*, August 2019, *134* (3), 1505–1556.

_ , Matthew Gentzkow, and Heidi Williams, "Sources of Geographic Variation in Health Care: Evidence From Patient Migration*," *The Quarterly Journal of Economics*, November 2016, *131* (4), 1681–1726.

Foster, Anne and B. Scott McBride, "CMS Announces Changes to RAC Program; Temporary Pause in Document Requests," *JD Supra*, March 2014.

Gottlieb, Joshua D., Adam Hale Shapiro, and Abe Dunn, "The Complexity Of Billing And Paying For Physician Care," *Health Affairs*, April 2018, *37* (4), 619–626. Publisher: Health Affairs.

Gowrisankaran, Gautam, Keith A. Joiner, and Jianjing Lin, "How do Hospitals Respond to Payment Incentives?," Working Paper 26455, National Bureau of Economic Research November 2019. Series: Working Paper Series.

Gross, Tal, Adam Sacarny, Maggie Shi, and David Silver, "Regulated Revenues and Hospital Behavior: Evidence from a Medicare Overhaul," *Review of Economics and Statistics*, 2023, *Forthcoming.*

Gujral, Kritee and Anirban Basu, "Impact of Rural and Urban Hospital Closures on Inpatient Mortality," Working Paper 26182, National Bureau of Economic Research August 2019. Series: Working Paper Series.

Gupta, Atul, "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program," *American Economic Review*, April 2021, *111* (4), 1241–1283.

Hanna, Rema and Paulina Oliva, "The impact of inspections on plant-level air emissions," *The B.E. journal of economic analysis & policy*, 2010, *10* (1).

Hillestad, Richard, James H. Bigelow, Katya Fonkych, Anthony G. Bower, Constance Fung, Jason Wang, Roger Taylor, Federico Girosi, Robin Meili, and Richard Scoville, "Health Information Technology: Can HIT Lower Costs and Improve Quality?," Technical Report, RAND Corporation September 2005.

Howard, David H. and Ian McCarthy, "Deterrence effects of antifraud and abuse enforcement in health care," *Journal of Health Economics*, January 2021, *75*, 102405.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Internal Revenue Service, "IRS E-File: A History," Technical Report FS-2011-10 June 2011.

Iwashyna, Theodore J., Jeremy M. Kahn, Rodney A. Hayward, and Brahmajee K. Nallamothu, "Interhospital Transfers Among Medicare Beneficiaries Admitted for Acute Myocardial Infarction at Nonrevascularization Hospitals," *Circulation: Cardiovascular Quality and Outcomes*, September 2010, *3* (5), 468–475.

Jin, Ginger Zhe, Ajin Lee, and Susan Feng Lu, "Medicare Payment to Skilled Nursing Facilities: The Consequences of the Three-Day Rule," Working Paper w25017, National Bureau of Economic Research September 2018.

Kopczuk, Wojciech and Cristian Pop-Eleches, "Electronic filing, tax preparers and participation in the Earned Income Tax Credit," *Journal of Public Economics*, 2007, *91* (7-8), 1351–1367.

Krumholz, Harlan M., Kun Wang, Zhenqiu Lin, Kumar Dharmarajan, Leora I. Horwitz, Joseph S. Ross, Elizabeth E. Drye, Susannah M. Bernheim, and Sharon-Lise T. Normand, "Hospital-Readmission Risk — Isolating Hospital Effects from Patient Effects," *New England Journal of Medicine*, September 2017, *377* (11), 1055–1064.

Laffont, Jean-Jacques and Jean Tirole, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 1986, *94* (3), 614–641.

League, Riley, "Administrative Burden and Consolidation in Health Care: Evidence from Medicare Contractor Transitions," 2022.

Leder-Luis, Jetson, "Can Whistleblowers Root Out Public Expenditure Fraud? Evidence from Medicare," *Review of Economics and Statistics*, 2023, *Forthcoming*, 60.

Lopez, Eric, Gretchen Jacobson, Tricia Neuman, and Larry Levitt, "How Much More Than Medicare Do Private Insurers Pay? A Review of the Literature," April 2020.

Magat, Wesley A. and W. Kip Viscusi, "Effectiveness of the EPA's Regulatory Enforcement: The Case of Industrial Effluent Standards," *The Journal of Law and Economics*, October 1990, *33* (2), 331–360.

Meckel, Katherine, "Is the Cure Worse Than the Disease? Unintended Effects of Payment Reform in a Quantity-Based Transfer Program," *American Economic Review*, June 2020, *110* (6), 1821–1865.

Medicare Contractor Management Group, "Medicare Administrative Contractor Workload Transition Handbook," October 2017.

Medicare Payment Advisory Commission, "Hospital Short-Stay Policy Issues," Technical Report June 2015.

_ , "July 2020 Data Book: Health Care Spending and the Medicare Program," Technical Report July 2020.

Nicholas, Lauren Hersch, Caroline Hanson, Jodi B. Segal, and Matthew D. Eisenberg, "Association Between Treatment by Fraud and Abuse Perpetrators and Health Outcomes Among Medicare Beneficiaries," *JAMA internal medicine*, January 2020, *180* (1), 62–69.

Office of Management and Budget, "Annual Improper Payments Datasets," Technical Report November 2022.

Office of the Inspector General, "Hospitals' Use of Observation Stays and Short Inpatient Stays for Medicare Beneficaries," Technical Report OEI-02-12-00040 July 2013.

_ , "Medicare Improperly Paid Acute-Care Hospitals $54.4 Million for Inpatient Claims Subject to the Post-Acute-Care Transfer Policy," Technical Report A-09-19-03007 November 2019.

_ , "HHS-OIG Strategic Plan 2020-2025," Technical Report May 2020.

Olken, Benjamin A., "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, April 2007, *115* (2), 200–249.

ResDAC, "Anomaly in MedPAR/inpatient claims when counting emergency department visits," May 2022.

Roberts, James, Ryan McDevitt, Paul Eliason, Jetson Leder-Luis, and Riley League, "Enforcement and Deterrence of Medicare Fraud: The Case of Non-emergent Ambulance Rides," Working Paper w29491, National Bureau of Economic Research 2021.

Rose, Liam, "The Effects of Skilled Nursing Facility Care: Regression Discontinuity Evidence from Medicare," *American Journal of Health Economics*, January 2020, *6* (1), 39–71.

Sacarny, Adam, "Adoption and learning across hospitals: The case of a revenue-generating practice," *Journal of Health Economics*, July 2018, *60*, 142–164.

Silverman, Elaine and Jonathan Skinner, "Medicare upcoding and hospital ownership," *Journal of Health Economics*, March 2004, *23* (2), 369–389.

Skinner, Jonathan, "Causes and Consequences of Regional Variations in Health Care," Handbook of Health Economics, Elsevier 2011.

Slemrod, Joel and Shlomo Yitzhaki, "Tax avoidance, evasion, and administration," Handbook of Public Economics, Elsevier 2002.

Span, Paula, "In the Hospital, but Not Really a Patient," June 2012.

US Department of Agriculture, "How FNS Fights SNAP Fraud, Waste, and Abuse," Technical Report March 2013.

US Department of Health and Human Services Office of Inspector General, "Hospitals' Use of Observation Stays and Short Inpatient Stays for Medicare Beneficiaries," Technical Report OEI-02-12-00040 July 2013.

Venkatesh, Arjun K., Hao Mei, Keith E. Kocher, Michael Granovsky, Ziad Obermeyer, Erica S. Spatz, Craig Rothenberg, Harlan M. Krumholz, and Zhenqui Lin, "Identification of Emergency Department Visits in Medicare Administrative Claims: Approaches and Implications," *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, April 2017, *24* (4), 422–431.

Zwick, Eric, "The Costs of Corporate Tax Complexity," *American Economic Journal: Economic Policy*, May 2021, *13* (2), 467–500.

7 Tables

Table I. Hospital Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	Sample		*RAC Region*			
	Overall	Border	A	B	C	D
			A. Hospital Characteristics			
2011 audit rate	2.16	2.23	3.01	1.79	1.36	3.33
	(2.03)	(2.08)	(2.29)	(1.21)	(1.18)	(2.73)
Share region A	0.17	0.08				
Share region B	0.19	0.36				
Share region C	0.42	0.37				
Share region D	0.22	0.18				
Beds	202.16	177.41	238.22	198.04	194.41	193.59
	(177.33)	(171.06)	(194.54)	(170.28)	(186.64)	(146.62)
Share urban	0.72	0.55	0.83	0.70	0.64	0.82
Share non-profit	0.63	0.70	0.88	0.79	0.46	0.63
Share for-profit	0.19	0.16	0.05	0.09	0.29	0.19
Share government	0.18	0.14	0.07	0.12	0.24	0.18
Total cost (million $)	199.23	160.96	271.89	211.01	154.97	218.05
	(250.93)	(247.87)	(336.29)	(270.04)	(204.40)	(221.91)
Admin costs (million $)	29.17	24.25	36.00	33.38	22.24	33.47
	(36.63)	(37.59)	(40.83)	(44.18)	(29.48)	(36.18)
			B. Medicare Inpatient Characteristics			
Admissions	3465.75	3151.42	4264.70	3845.22	3262.61	2928.68
	(3205.86)	(3069.49)	(3591.67)	(3383.92)	(3260.47)	(2399.90)
Mean payment ($)	8617.36	7366.40	9349.37	8177.97	7578.76	10393.64
	(3179.31)	(2349.10)	(3461.79)	(2433.87)	(2663.76)	(3501.44)
Total payments (million $)	34.00	27.51	45.75	36.03	29.15	32.65
	(39.96)	(35.80)	(53.88)	(40.65)	(35.72)	(32.25)
Short stay share	0.31	0.32	0.28	0.32	0.31	0.33
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)
Top 20 error share	0.51	0.54	0.50	0.51	0.52	0.50
	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)
Predicted 2011 audit rate	0.02	0.02	0.02	0.02	0.02	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	2960	510	489	571	1237	663
N border hospitals	510	510	41	184	191	94

This table presents 2010 summary statistics of hospital characteristics and Medicare inpatient admissions by sample and RAC region. Standard deviation is in parentheses. Bed size, urban status, and profit type status come from the Medicare Provider of Services file. Total and administrative costs come from HCRIS. Medicare admissions and inpatient stay characteristics are from MEDPAR. Mean inpatient characteristics are defined as the average of each hospital's average (i.e., weighted by hospitals rather than claims). Short stay share is the share of Medicare admissions with length of stay ≤ 2. Top 20 error share is the share of Medicare admissions with a top 20 error rate MS-DRG, as identified in the 2010 CMS Improper Payments Report (Centers for Medicare and Medicaid Services, 2011b). "Predicted 2011 audit rate" is a claim-level prediction in 2011 audit rate using solely stay characteristics (but not hospital, state, or RAC characteristics) trained on 2007-2009 claims. The border sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group.

Table II. Patient Summary Statistics by ED Arrival Hour, Pre- and Post-Two Midnights Rule

	(1)	(2)	(3)	(4)
	ED Arrival Hour			
	Pre-Policy		*Post-Policy*	
	Before MN	*After MN*	*Before MN*	*After MN*
share inpatient	0.40	0.42	0.41	0.41
	(0.49)	(0.49)	(0.49)	(0.49)
share observation	0.05	0.05	0.04	0.05
	(0.21)	(0.22)	(0.20)	(0.22)
average charges ($)	24171	26068	25757	26572
	(43629)	(49564)	(47944)	(52421)
average age	68.32	68.55	68.40	68.47
	(17.22)	(17.19)	(17.06)	(17.07)
share white	0.79	0.77	0.79	0.79
	(0.41)	(0.42)	(0.40)	(0.41)
share hispanic	0.12	0.11	0.11	0.10
	(0.32)	(0.31)	(0.32)	(0.30)
share female	0.57	0.54	0.57	0.54
	(0.50)	(0.50)	(0.50)	(0.50)
average n of chronic conditions	3.98	4.20	4.21	4.31
	(3.59)	(3.67)	(3.59)	(3.59)
share inpatient in last 30 days	0.13	0.14	0.14	0.15
	(0.33)	(0.35)	(0.34)	(0.36)
share hospital visit in last 30 days	0.28	0.30	0.29	0.32
	(0.45)	(0.46)	(0.45)	(0.47)
average predicted admission likelihood	0.49	0.52	0.50	0.52
	(0.37)	(0.36)	(0.36)	(0.36)
share hospital visit in next 30 days	0.28	0.29	0.29	0.29
	(0.45)	(0.45)	(0.45)	(0.45)
share hospital visit in next 60 days	0.38	0.39	0.39	0.40
	(0.49)	(0.49)	(0.49)	(0.49)
share hospital visit in next 90 days	0.45	0.46	0.46	0.46
	(0.50)	(0.50)	(0.50)	(0.50)
Observations	31419	17690	32420	17637

This table presents summary statistics of characteristics of traditional Medicare patients in Florida who arrived in the ED within 3 hours of midnight in 2013Q2 (“pre-policy”) and in 2014Q2 (“post-policy”). Standard deviation is in parentheses. “Share inpatient” is the share of ED patients admitted to inpatient (this includes patients who could have initially been placed in observation and eventually admitted). “Share observation” is the share of patients who are placed in outpatient observation only. “Average predicted admission likelihood” is the predicted admission likelihood from estimating a logit using ED visits between 9:00AM and 3:00PM of an indicator for being admitted within 30 days of an ED visit on patient demographics, current ED visit information, and information on any prior visits in the last 365 days. Data: HCUP SID/SEDD.

Table III. Event Studies of Effect of 2011 Audit Rate on Hospital Outcomes, 2011-2015 Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		LOS ≤ 2		Admin Costs	Software Installation
	Log Adm.	*Log Rev.*	*Log Adm.*	*Log Rev.*	*Log Costs*	*Medical Necc.*
2011 audit rate $\times$ 2011	-0.0115**	-0.0102**	-0.0145*	-0.0120***	0.0154***	0.0037
	(0.0044)	(0.0044)	(0.0074)	(0.0039)	(0.0053)	(0.0088)
2011 audit rate $\times$ 2012	-0.0192***	-0.0177*	-0.0457***	-0.0460***	0.0068	0.0217**
	(0.0051)	(0.0093)	(0.0111)	(0.0056)	(0.0080)	(0.0079)
2011 audit rate $\times$ 2013	-0.0191**	-0.0280**	-0.0282***	-0.0364***	0.0034	0.0225*
	(0.0089)	(0.0129)	(0.0082)	(0.0103)	(0.0092)	(0.0129)
2011 audit rate $\times$ 2014	-0.0113	-0.0216	-0.0241**	-0.0329**	0.0054	0.0225*
	(0.0114)	(0.0157)	(0.0092)	(0.0120)	(0.0096)	(0.0110)
2011 audit rate $\times$ 2015	-0.0193	-0.0285	-0.0208*	-0.0282**	-0.0014	0.0090
	(0.0148)	(0.0182)	(0.0109)	(0.0107)	(0.0107)	(0.0123)
Hosp FE	X	X	X	X	X	X
Nbr group FE	X	X	X	X	X	X
N Hosp	510	510	510	510	510	506
Obs	52139	52139	52139	52118	52107	36906
F	12.5	12.5	12.5	13.36	12.45	13.87

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors are in parentheses and are clustered at the state and border segment level. This table reports the coefficients of the reduced form event study in Equation 8, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. For brevity, the pre-2011 coefficients are estimated but not reported in the table. Omitted year is 2010. Columns 1 and 2 report the effect on the log number of Medicare inpatient admissions and log Medicare inpatient revenue from the MEDPAR data, and columns 3 and 4 report the effect on short stay admissions and revenue. Column 5 reports the effect on log net administrative costs from HCRIS data. Net administrative costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Column 6 reports the effect on an indicator for installing medical necessity software application, which is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in the HIMSS data. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group.

Table IV. After-Midnight ED Arrival Hour Difference-in-Difference Coefficients on Patient Status and Revisits

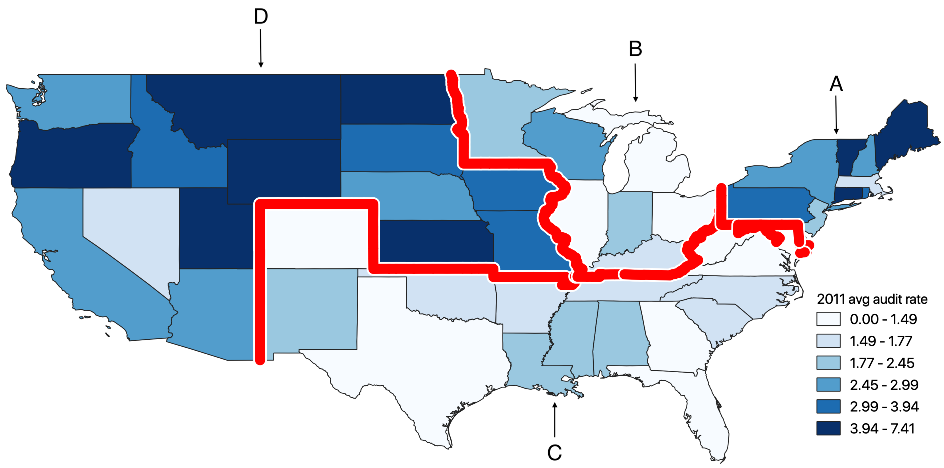
	(1)	(2)	(3)	(4)	(5)
	Medicare				Non-Medicare
	Inpatient	*Observation*	*Not Admitted*	*Revisit 30d*	*Inpatient*
β	-0.007***	0.007***	0.000	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Pre-reform mean	0.420	0.042	0.538	0.259	0.126
Estimate as % of mean	1.67	16.67	0.00	0.39	0.79
Observations	1254857	1254857	1254857	1254857	7428583

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors are in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[q \geq 2013\text{Q3}] \times \mathbb{1}[T_v \geq 00{:}00]$ of the specification in Equation 12, where $\mathbb{1}[q \geq 2013\text{Q3}]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "Inpatient" is an indicator for whether the patient was eventually admitted as inpatient from the ED. "Observation" is an indicator for whether the patient was placed in observation status and was never admitted. "Not Admitted" is an indicator equal to one when a patient is neither admitted nor placed in observation status. "Revisit within 30 days" is an indicator for whether the patient had another ED visit or inpatient stay within 30 days of the ED visit. Sample for columns 1-4 consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. The sample for column 5 consists of all non-Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Data: HCUP SID/SIDD.

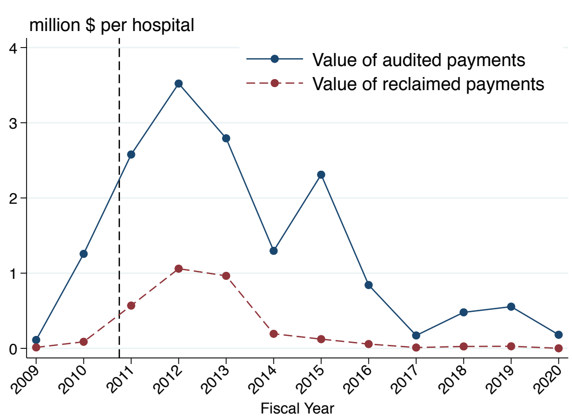
8 Figures

Figure 1. RAC Audit Activity

(a) Average 2011 Hospital Audit Rates by State and RAC Regions

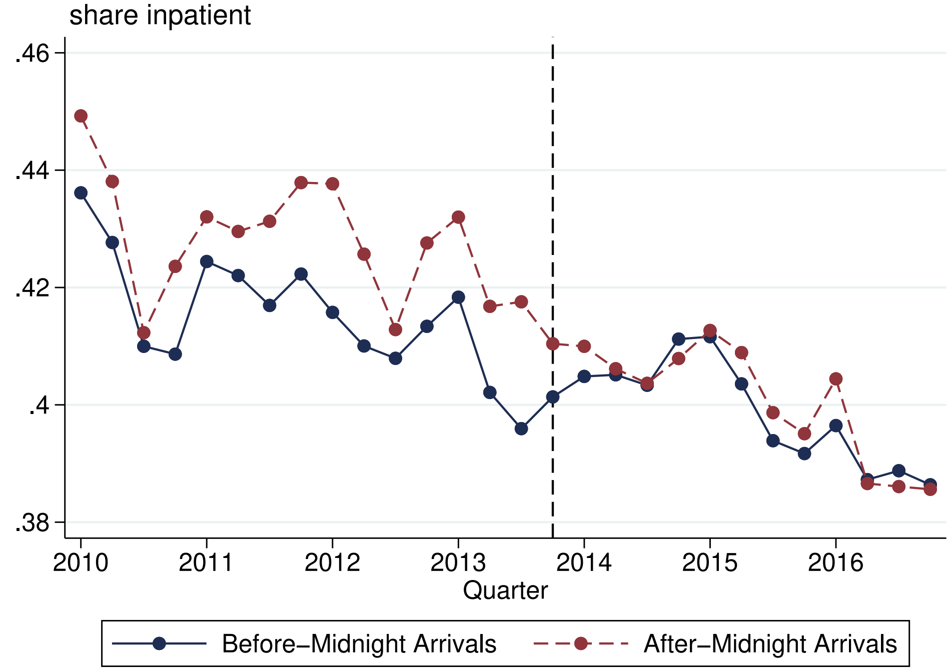


(b) Value of Audited Inpatient Payments and Net Reclaimed Payments per Hospital, by Year of Audit



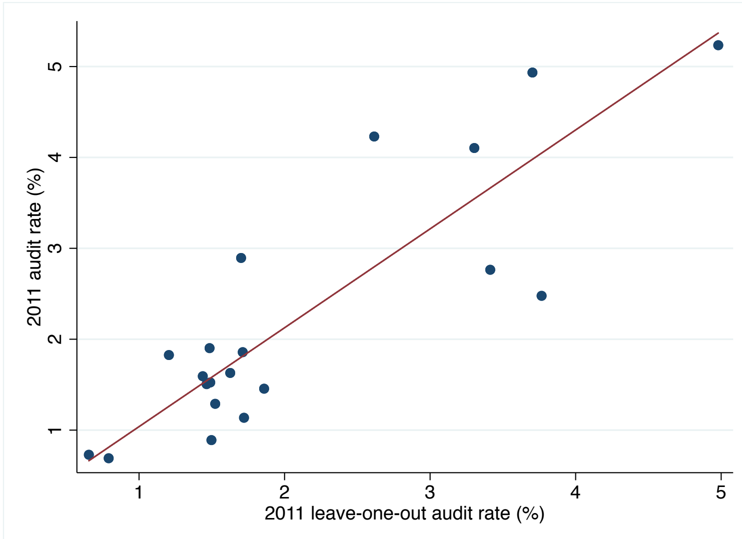
Panel (a) plots the 2011 average state audit rates, where audit rate is defined as the share of a hospital's 2008-2011 claims that were audited by RACs. The RAC regions are Region A (Northeast), Region B (Midwest), Region C (South), and Region D (West). Darker shades denote a higher audit rate. The red line demarcates RAC regions. Panel (b) plots the average per-hospital value of inpatient payments audited by RACs and the net reclaimed payments, by year of audit. Net reclaimed payments are defined as the sum of reclaimed payments from overpayments minus refunded payments from underpayments. These values are based on RACs' original reclaimed or refunded payments at the time of audit. Data: MEDPAR claims and CMS audit data.

Figure 2. Inpatient Admission Rates from ED, Before vs. After-Midnight ED Arrivals in Florida



This figure plots the share of traditional Medicare patients admitted as inpatient from the emergency department, among Florida patients who arrived within three hours before midnight (9:00-11:59PM), in the blue solid line, and three hours after midnight (12:00-2:59AM), in the red dashed line. The dashed vertical line denotes 2013Q3, which is when the Two Midnights rule is implemented. Data: HCUP SID/SEDD.

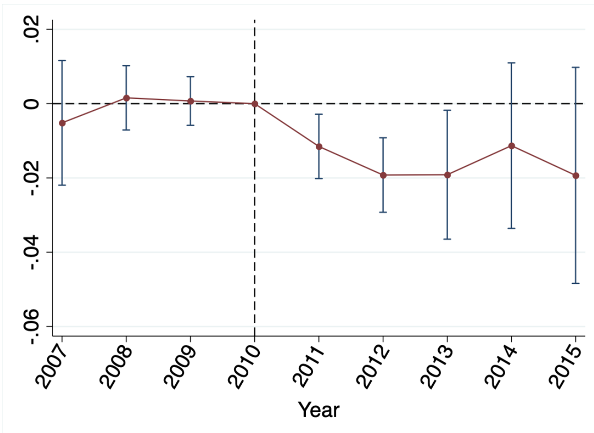
Figure 3. Binscatter of 2011 Leave-One-Out State Audit Rate and 2011 Hospital Audit Rate, Border Hospital Sample



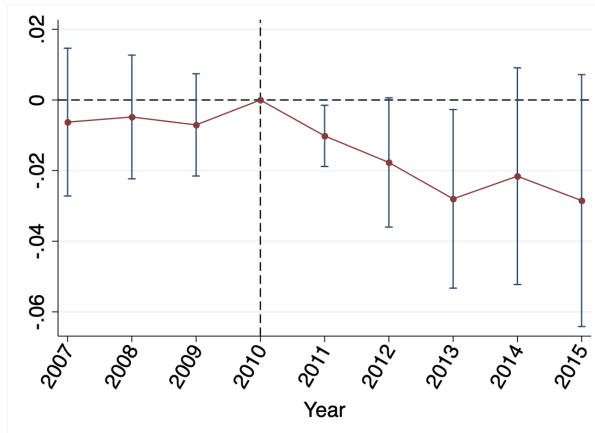
This figure plots a binscatter of the 2011 hospital audit rate compared to the 2011 leave-one-out state audit rate. The 2011 audit rate is defined as the share of 2008-2011 inpatient claims that were audited by RACs in 2011. The leave-one-out state audit rate is defined as the average audit rate of all other hospitals in the same state as a given hospital. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR claims and CMS audit data.

Figure 4. Event Studies on Effect of 2011 Audit Rate on Hospital Outcomes

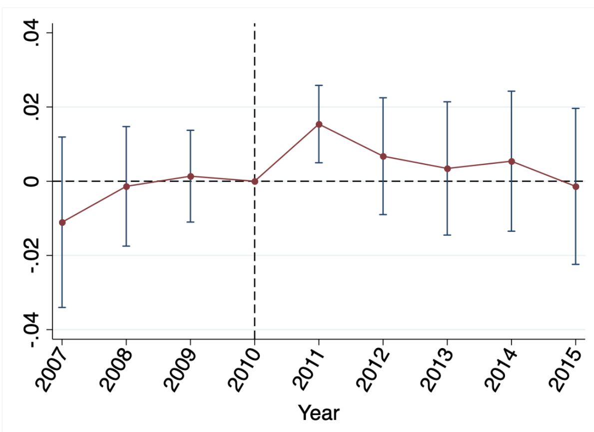
(a) Log Medicare admissions



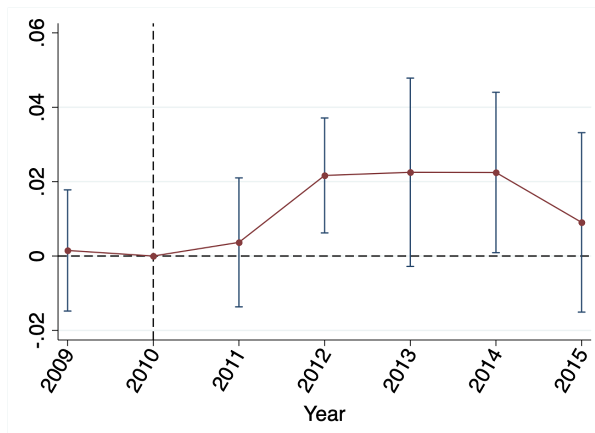
(b) Log Medicare inpatient revenue



(c) Log hospital administrative costs



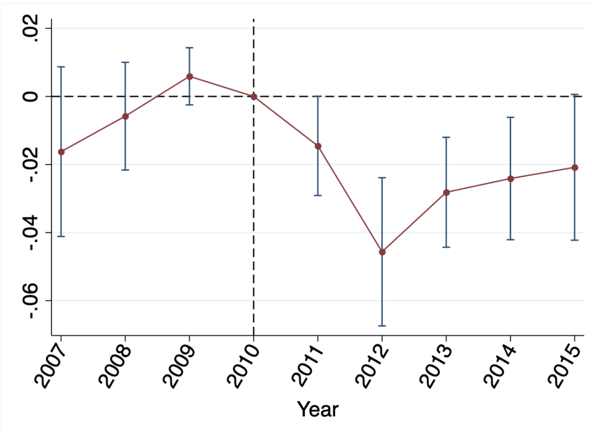
(d) Indicator for installing medical necessity software



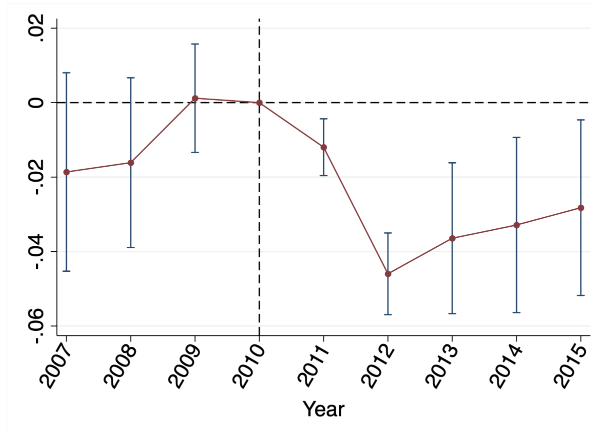
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 8, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. Medicare admissions and revenue are from MEDPAR. Inpatient revenue is the sum of all Medicare inpatient payments. Net administrative costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in HIMSS. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group.

Figure 5. Event Studies on Effect of 2011 Audit Rate on Medicare Admissions and Revenue, by Length of Stay

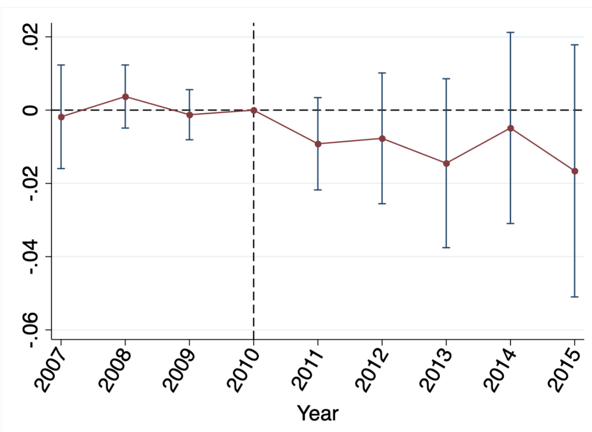
(a) Log Medicare admissions, LOS ≤ 2



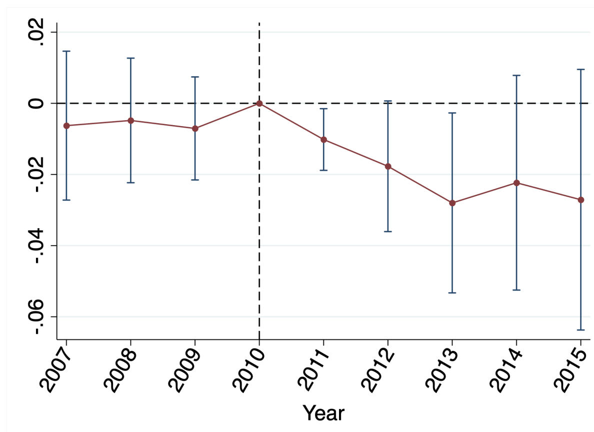
(b) Log Medicare inpatient revenue, LOS ≤ 2



(c) Log Medicare admissions, LOS > 2



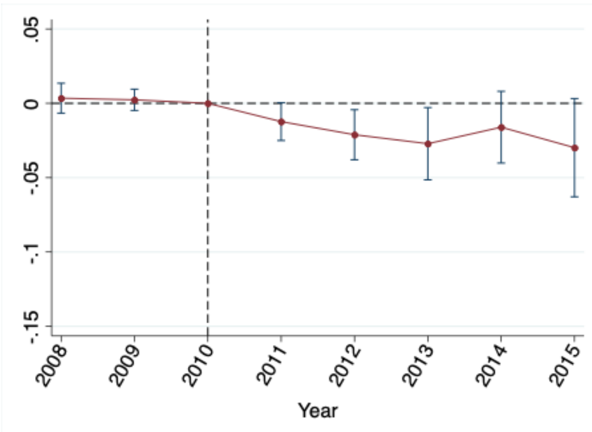
(d) Log Medicare inpatient revenue, LOS > 2



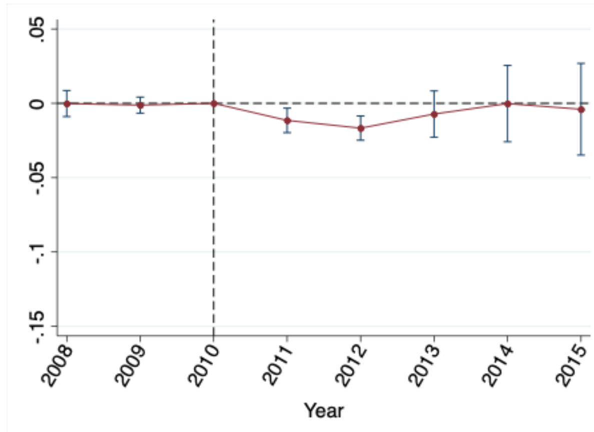
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 8, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. Medicare volume and revenue of short stay admissions and longer admissions are from MEDPAR. Length of stay is counted as the difference in days between the admission and discharge date. Inpatient revenue is the sum of all Medicare inpatient payments. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group.

Figure 6. Event Studies on Effect of 2011 Audit Rate on Medicare Admissions, Diagnosis Error Rates

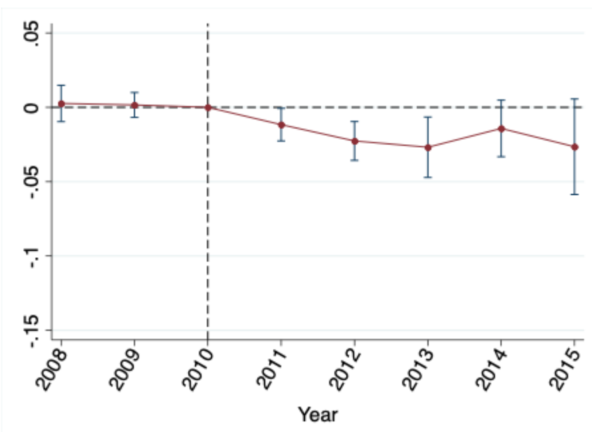
(a) Log Medicare admissions,
all top 20 error base DRG groups



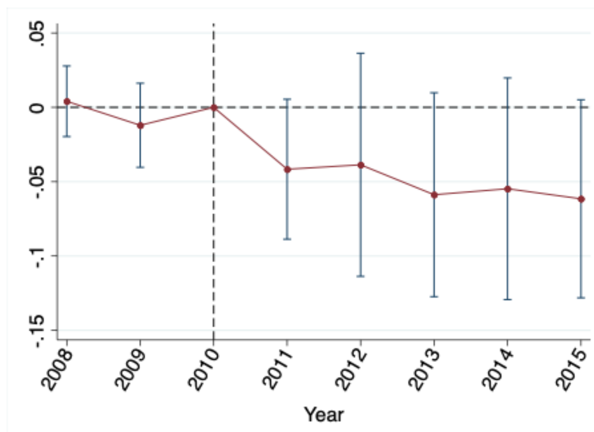
(b) Log Medicare admissions,
non-top 20 base DRG groups



(c) Log Medicare admissions,
top emergent MS-DRGs

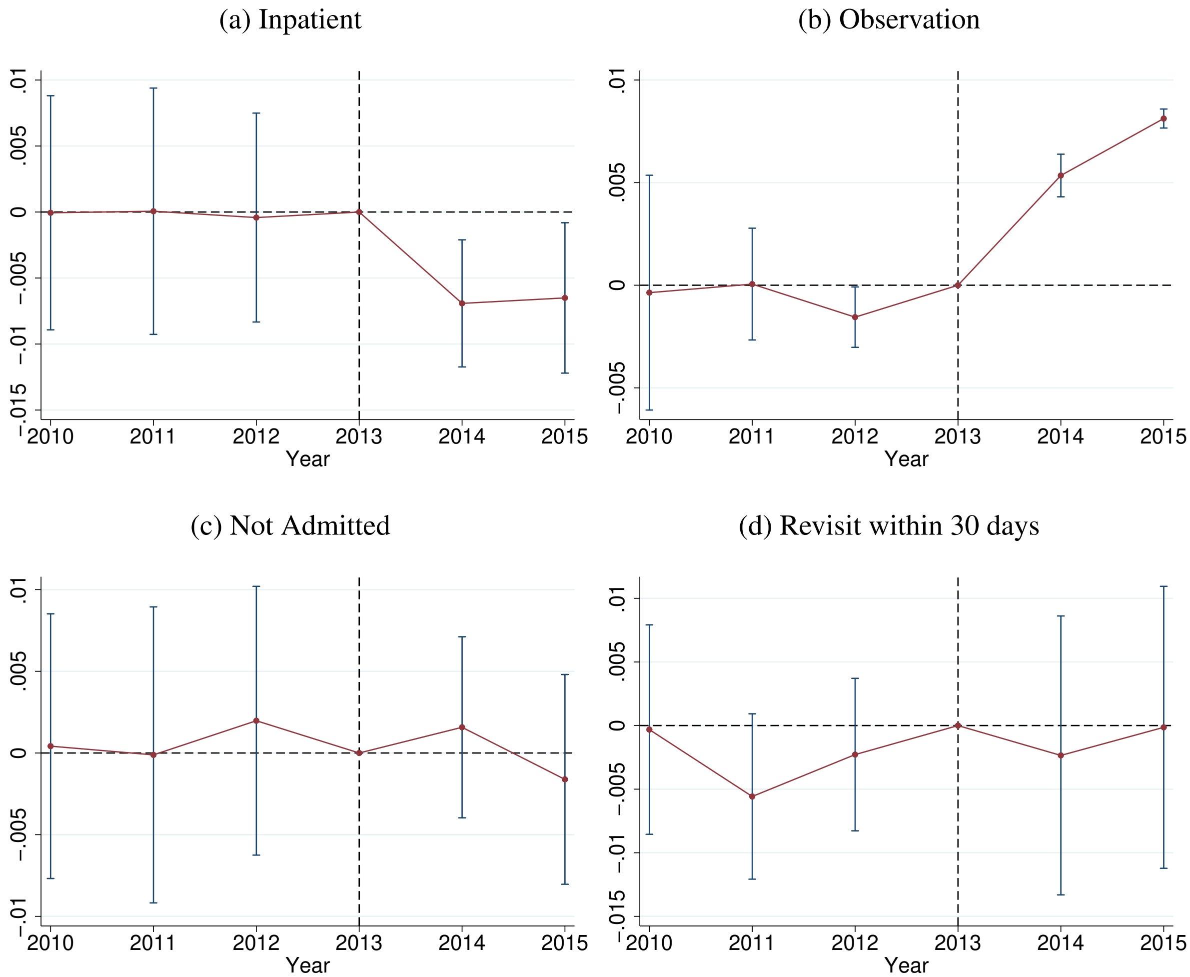


(d) Log Medicare admissions,
top non-emergent MS-DRGs



This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 8, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. Panel (a) plots admissions for the top 20 groups of MS-DRGs with the largest errors, according to the 2010 CERT Improper Payments report (Centers for Medicare and Medicaid Services, 2011b). Panel (b) plots admissions for the non-top-20 MS-DRGs. Panel (c) plots admissions for the 14 emergent MS-DRGs with the highest payment errors: sepsis (MS-DRG 871-872; ED rate 79%), chest pain (313; 83%), GI hemorrhage (377-379; 74%), respiratory infections (177-179; 71%), esophagitis and misc digestive disorders (391-392; 71%), kidney and UTI (689-690; 69%), nutritional and metabolic (640-641; 68%), renal failure (291-293; 67%), syncope and collapse (312; 78%), heart failure and shock (291-293; 69%), cardiac arrhythmia (308-309; 69%), pneumonia and pleurisy (193-195; 65%), acute myocardial infarction (280-282; 77%), chronic obstructive pulmonary disease (190-192; 69%), hip and femur except major joint (480-482; 82%), and intracranial hemorrhage or cerebral infarction (064-066; 76%). Panel (d) plots admissions for the remaining 6 non-emergent MS-DRGs among the top 20: major joint replacement (MS-DRGs 469-470, ED rate 13%), permanent cardiac pacemaker (242-244, 57%), drug-eluting stents (242-244, 42%), major bowel procedures (329-331, 38%). The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR and CMS audit data.

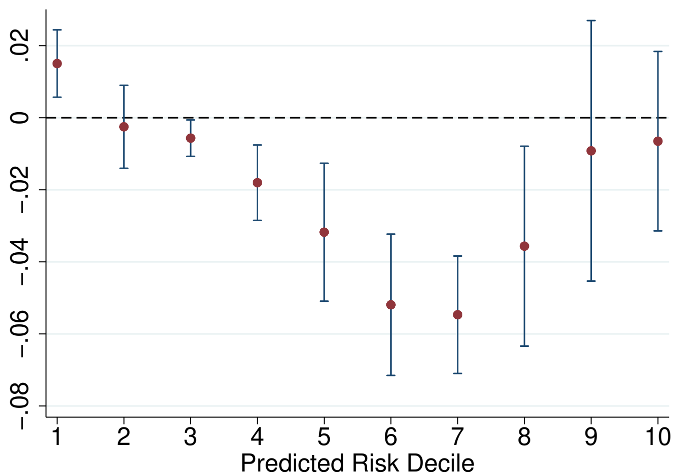
Figure 7. Event Studies on Effect of After-Midnight ED Arrival on Patient Status and Outcomes



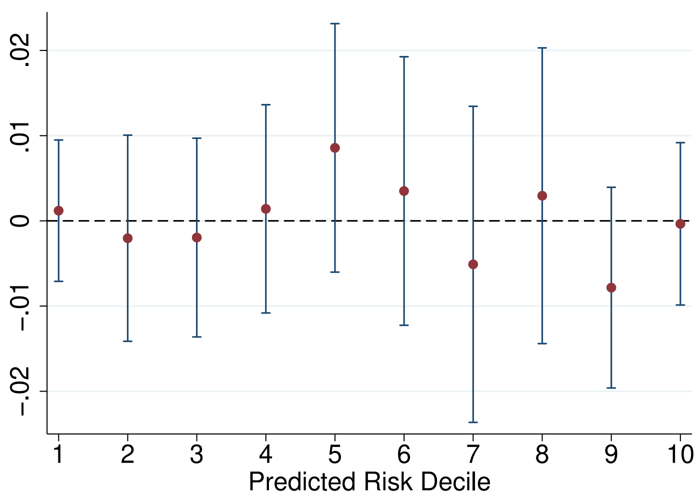
This figure plots the coefficients and 95% confidence intervals for β^{τ} on $\mathbb{1}[y = \tau] \times \mathbb{1}[T_v \geq 00{:}00]$ of the specification in Equation 12, where $\mathbb{1}[y = \tau]$ is an indicator for whether the visit occurred in fiscal year τ (i.e., October year $\tau - 1$ through September year τ), and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. The results are clustered at the ED arrival hour and year level. The omitted year is 2013. "Inpatient" is an indicator for whether the patient was eventually admitted as inpatient from the ED. "Observation" is an indicator for whether the patient was placed in observation status and was never admitted. "Not Admitted" is an indicator equal to one when a patient is neither admitted nor placed in observation status. "Revisit within 30 days" is an indicator for whether the patient had another ED visit or inpatient stay within 30 days of the ED visit. Sample consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-year, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Data: HCUP SID/SEDD.

Figure 8. Heterogeneity of After-Midnight ED Arrival Coefficient by Patient Severity

(a) Inpatient



(b) Revisit in 30 days



This figure plots estimates and 95% confidence intervals of the β coefficient in Equation 12, interacted with an indicator for predicted severity decile. β is the coefficient on $\mathbb{1}[q \geq 2013Q3] \times \mathbb{1}[T_v \geq 00{:}00]$, where $\mathbb{1}[q \geq 2013Q3]$ is an indicator for whether the visit occurred after 2013Q3, and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. The top panel plots results for an indicator for whether the patient was admitted as inpatient from the ED, and the bottom panel plots results for an indicator for whether the patient revisited any hospital in Florida within 30 days of the ED visit. The results are clustered at the ED arrival hour and quarter level. Patient risk is predicted by estimating a logit using ED visits between 9:00AM and 3:00PM of an indicator for being admitted within 30 days of an ED visit on patient demographics, current ED visit information, and information on any prior visits in the last 365 days. Demographics include age-bin, sex, race, Hispanic indicator, point of origin indicator, and mean zip code income. Information on current visit includes hospital, quarter, and the AHRQ CCS category for the patient's first diagnosis code. Information on previous visits includes the number of visits/inpatient stays/length of stay in the last month or last year, as well as any diagnoses and procedures recorded in stays within the last month or last year. Figure H16 plots the mean outcomes for each decile. Data: HCUP SID/SEDD.

A Appendix: Additional Policy Context

A.1 Medicare Inpatient Prospective Payment System and Short Stays

Medicare pays for inpatient hospital admissions through the inpatient prospective payment system (IPPS), in which Medicare pays a fixed amount per inpatient stay within broad categories of diagnoses called Medicare Severity Diagnosis Related Groups (MS-DRGs, also referred to as DRGs). The prospective payment system was introduced in 1983 with the intent of incentivizing providers to reduce healthcare costs (Ellis and McGuire, 1986). Hospitals keep the difference between the DRG payment and the costs to treat the patient, so they have an incentive to keep costs low. The payment rate for each DRG reflects the national average cost of treating a patient across all cases, and it is revised each year based on claims data in the last two years. The per-stay payment is adjusted based on a patient's pre-existing chronic conditions in order to account for the patient's diagnosis severity. It is also adjusted by hospital-specific factors such as a hospital's wage index, teaching status, share of low-income patients, and number of unusually costly outlier cases. The prospective payment system generally works well to keep inpatient hospital spending relatively low for the Medicare program (Lopez et al., 2020).

However, one persistent issue with IPPS that has been noted by policymakers is the high number of short stays. A CMS report found that "a large percentage of medically unnecessary [payment] errors are related to hospital stays of short duration... these services should have been rendered at a lower level of care" (Centers for Medicare and Medicaid Services, 2011b). One less intensive alternative to an inpatient stay is an outpatient observation stay, which consists of short-term (often diagnostic) services provided at the hospital while a physician decides whether to formally admit a patient as inpatient or send them home. Observation stays typically last less than forty-eight hours and are billed as an outpatient service (Medicare Payment Advisory Commission, 2015).

From the patient's point of view, it is often difficult to differentiate between an observation stay and a short inpatient stay (Span, 2012). Thus, a hospital's costs for an observation stay are likely similar to the costs for a short inpatient stay. However, hospitals earn much more from Medicare for admitting a patient for a short inpatient stay rather than for an outpatient observation stay: among DRGs common to both inpatient and observation stays, Medicare payments for inpatient stays were two to three times higher than payments for observation stays (Medicare Payment Advisory Commission, 2015).

Policymakers considered various alternative solutions to address unnecessary short stays before settling on RAC audits. They were wary of reducing the payment rate for short stays or penalizing high rates of short stays, due to concerns that hospitals would simply keep patients for longer to evade these policies (Medicare Payment Advisory Commission, 2015). There is evidence that hospitals delay discharging patients if they have an incentive to do so (Jin et al., 2018). Additionally, short stays constitute almost a third of inpatient stays; their prevalence suggests that not all short stays are unnecessary, and cutting payments for short stays across the board would reduce payments for some necessary stays.

Medicare enacted other monitoring and education programs to measure and mitigate unnecessary inpatient stays. They measured payment errors across different discharge and service types through the Comprehensive Error Rate Testing Program (CERT) in 2010, which randomly samples Medicare claims to calculate improper payment rates (Centers for Medicare and Medicaid Services, 2011b). The CERT reports then informed provider education programs, like the "Targeted Probe and Educate" program, which involves claim reviews and one-on-one education sessions for providers, as well as the PEPPER and Comparative Billing Reports (CBR) programs which distributed provider-specific reports on which discharges and services were most vulnerable to improper payments. See the CMS websites for the TPE program (link) and the PEPPER and CBR programs (link).

A.2 RAC Program Details

RAC Regions In the context of medical claims processing and reviews, the jurisdictions used for RAC regions are unique. Medicare Administrative Contractors (MACs) are contractors who process Medicare claims *before* payment; they operate in different, smaller regions than RAC regions. The RAC regions do align with the regions of Durable Medical Equipment MACs. However, they only process payments for durable medical equipment like prosthetics, orthotics, and other devices, and they do not process claims for medical services (Medicare Contractor Management Group, 2017). To hire RAC firms for each region, Medicare posts a separate contract solicitation for each region, and firms submit separate bids.

RAC Firms The four firms originally contracted to conduct RAC audits in 2010 were Health Data Insight, Cotiviti, CGI, and Performant Recovery (Centers for Medicare and Medicaid Services, 2011a). Some firms focus on healthcare (for example, Health Data Insight, Cotiviti), while others serve other government agencies and corporations as well (for example, CGI, Performant Recovery). Other clients of the RAC firms include state tax authorities, student loan companies, private health insurance companies, the Internal Revenue Service, the National Health Service in the UK, and Public Health England.

RAC Audit Process RACs conduct postpayment reviews to identify and correct overpayments or underpayments for claims for inpatient care, outpatient care, long-term care, and durable medical equipment in the last three years. Figure H1 illustrates the claims auditing and appeal process, using 2011 inpatient audits as an example. Each RAC develops and runs its own proprietary algorithm on claims data to identify claims with potential payment errors. In 2011, RACs' auditing scope for inpatient claims included incorrect or incomplete coding, DRG validation, and medical necessity reviews. Five percent of audits were "automated reviews," which rely solely on claims data to make a determination based on clearly outlined Medicare policies. The rest of the audits were "complex reviews," in which a medical professional (for example, coder, nurse, or therapist) employed by the RAC submits a medical record request and manually reviews all documentation associated with an inpatient stay. It is up to the medical professional to determine whether an overpayment or underpayment was made. Once the complex review is finished, RACs send a demand letter to providers that outlines whether a payment error was identified, the amount of overpayment or underpayment demanded, and references supporting the decision. Fifty-seven percent of complex reviews in 2011 resulted in no finding, 37 percent resulted in an overpayment demand (in which providers must return payment back to Medicare), and 6 percent resulted in an underpayment demand (in which Medicare returns payment to the provider). Providers can appeal demands by first requesting a redetermination by the RAC and then escalating it to higher levels of appeals – for example, by requesting that a separate contractor reconsider the case, requesting a hearing by an administrative law judge, or escalating it to a review by the Medicare Appeals Council.

Timeline of the RAC Program The RAC program was first proposed as part of the Medicare Modernization Act of 2003. After an initial pilot demonstration from 2005 to 2008 in select states, the RAC program was implemented nationally in 2010 (Centers for Medicare and Medicaid Services, 2011a). At first, RACs were authorized only to audit claims with complex coding issues and for DRG validation. Each year, Medicare expanded the scope of RAC audits, and in 2011 it expanded the scope to include medical necessity reviews of inpatient claims (Centers for Medicare and Medicaid Services, 2012). As shown in Figure 1b, RAC audit activity peaked in 2011–13, then dropped precipitously in 2014. The peak corresponds with the period in which RACs were authorized to audit inpatient claims for medical necessity.

In the face of a sudden rise in auditing and overpayment demands, hospitals began mounting a campaign to fight back. Hospitals started appealing high volumes of RAC determinations, and some hospital systems worked with the American Hospital Association (AHA) to file lawsuits and complaints against Medicare over RAC audits.[27] Between 2011 and 2013, the number of appeals that reached the administrative-judge level of the appeals process increased by 500 percent, and by mid-2014 there was a backlog of 800,000 appeals at that level (Medicare Payment Advisory Commission, 2015). The AHA also began tracking the effect of RAC activity on its own through the quarterly RACTrac Survey of hospitals. Many hospitals reported that RAC audits imposed significant administrative burdens on them; for example, 11 percent of hospitals reported costs associated with managing the RAC program of over $100,000 (American Hospital Association,

[27] See the AHA website for a list of all past and ongoing litigation: https://www.aha.org/legal/past-litigation (link).

2014).

Hospitals and industry stakeholders filed several complaints with Medicare stating that RAC audits were overly aggressive. As a result, in 2014 Medicare paused almost all RAC audits by significantly limiting their scope (Foster and McBride, 2014). Other Medicare contractors such as MACs picked up additional review responsibilities after the RAC audits were paused.[28] Medicare maintained that the pause on RAC audits was temporary and would resume at previous levels, but it is clear from Figure 1b that RAC auditing never returned to its peak level after the pause. The pause began at the end of 2014Q1 and was originally meant to end in 2014Q3. After several quarters of delayed resumption, inpatient RAC audits finally resumed in 2015Q4, although they were subject to limitations to reduce the administrative burden on providers. In August 2014, Medicare announced a one-time option to settle appeals by offering hospitals 68 percent of each appealed denied inpatient claim, in exchange for hospitals dropping all of their appeals rather than settling them one by one. As a result, hospitals dropped almost 350,000 appeals in exchange for $1.5 billion in settled denials (Centers for Medicare and Medicaid Services, 2014).

A.3 Characteristics of Audits and Audited Hospitals

Given Medicare policymakers' focus on short stays as the main source of unnecessary admissions, I examine audit frequency as a function of an admission's length of stay in Figure H12. Admissions with a length of stay of two or fewer days have much higher rates of auditing than longer admissions. The majority of audits of short stays result in the full payment being reclaimed (Figure H17). The majority of audits of short stays result in the full payment being reclaimed (Figure H17). I also consider audit frequencies by diagnosis. I defer the explanation of how the diagnoses are categorized to the discussion of the results by diagnosis type in Section 5.

I next consider hospital-level characteristics and their correlation with audit rate in Figure H4. The RAC region a hospital is in is highly correlated with its audit rate. Within each region, rural

[28]For example, MACs conducted a program called "Teach, Probe, and Educate" in which they targeted hospitals with high payment errors and conducted education sessions. If hospitals failed to improve their payment accuracy sufficiently after three rounds of education sessions, then they were referred to Medicare for further remediation.

hospitals, small hospitals, non-profit hospitals, and hospitals with a higher share of short stay Medicare admissions are more likely to be audited.

Although almost every hospital was subject to an audit by 2020, in any given year there is a substantial portion of hospitals that do not face any audits. In 2011, 15 percent of hospitals had an audit rate of 0 percent. The share of hospitals with no audits varies across RAC regions from 2 to 23 percent (Figure H18).

B Appendix: Model (cont'd)

Next, I build on the model in Section 3 and present Medicare's problem of picking β, an extension to subsidies, and discuss other settings the model could apply to.

B.1 Medicare's Problem

Medicare can conduct audits to recoup revenue for unnecessary stays, but conducting audits is costly for Medicare. Medicare picks β to *minimize* a combination of three priorities: (1) over/under admissions, (2) what it reimburses hospitals, net of revenue recouped from audits, and (3) the cost of auditing. Each priority is scaled by c_1, c_2, and c_3, respectively. The first term could be interpreted as capturing Medicare caring about patient welfare. Because hospitals will always admit $Q_N, Q_A \geq q$, I do not include a term in Medicare's payoff to capture patient welfare for the q patients who truly need admission, since they are always admitted. Medicare's payoff depends on how many hospital adopt technology, and how many hospitals that adopt or do not adopt technology admit:

$$\begin{cases} G_N(\beta) = \underbrace{c_1(Q_N - q)^2}_{\text{over/under-admit cost}} + \underbrace{c_2 P(1-\beta)Q_N}_{\text{net reimbursement}} + \underbrace{c_3(\beta)^2 Q_N}_{\text{audit cost}} & \text{no tech} \iff q < q^* \\ G_A(\beta) = c_1(Q_A - q)^2 + c_2 P(1-\gamma\beta)Q_A + c_3(\beta)^2 Q_A & \text{adopt tech} \iff q \geq q^* \end{cases} \tag{13}$$

As a hospital's admission and technology decisions both depend on their draw of q, Medicare's payoff also depends on q, which it cannot observe. Instead, Medicare knows that q is distributed uniformly between $[0, \overline{q}]$. So, it picks β to minimize:

$$\underset{\beta}{min}\ G(\beta) = \int_0^{q^*} G_N(\beta)(1/q)dq + \int_{q^*}^{\overline{q}} G_A(\beta)(1/q)dq \tag{14}$$

Medicare will set the optimal audit rate β^* to be the β that minimizes $G(\beta)$, or where $\frac{dG}{d\beta}$ crosses 0. In Figure H19, I solve for β^* numerically in a model with parameters set to match the average acute hospital in 2010 (Table GXII). Figure H20 then shows graphically how $\frac{dG}{d\beta}$ (and therefore β^*) changes with respect to model parameters. β^* increases as γ and r increase – as the technology worsens, Medicare must audit at a higher rate to induce technology takeup and to deter waste. β^* increases as c_2, the weight on provider payments, increases. As paying providers becomes costlier, Medicare wants to recoup more revenue through audits. In contrast, as c_3, the weight on the cost of conducting audits, increases, β^* decreases. Finally, β^* does not change as f or c_1 (the weight on unnecessary admissions) change. This is because as β changes, providers will re-optimize and pick combinations of technology adoption and admissions to maximize their payoffs.

B.2 Extension: Should Medicare Subsidize the Technology?

We can also extend the model to consider an alternative policy – whether Medicare should purchase the technology itself on behalf of hospitals. This would capture policies like the HITECH Act, which gave healthcare providers financial incentives to encourage health IT adoption (Burde, 2011). I will consider the case where Medicare would fully subsidize the technology and incur the full cost f, but this could be easily adapted to the partial subsidy case. And in exchange for subsidizing the technology, Medicare would be guaranteed that all hospitals report Q_A.[29]

If Medicare purchases the technology, its payoff becomes:

$$\int_0^{\overline{q}} G_S(\beta^{*'})(1/q)dq = \int_0^{\overline{q}} [(c_1(Q_A - q)^2 + c_2(P(1-\gamma\beta^{*'})Q_A + f) + c_3(\beta^{*'})^2 Q_A)](1/q)dq \tag{15}$$

[29]Note that Medicare will only find it worthwhile to subsidize technology if $Q_A^* < Q_N^*$. So the question of whether to subsidize is only interesting when $Q_N^* - Q_A^* > 0$, which is when $r > \frac{1}{w^2}\beta(1-\gamma)P$.

where $\beta^{*'}$ is the β which minimizes $\int_0^{\overline{q}} G_S(\beta)(1/q)dq$. Note that in the second term, Medicare now incurs an additional cost f in addition to its reimbursement. Because Medicare cares about *minimizing* its payoff, it will choose to subsidize when:

$$\int_0^{\overline{q}} G_S(\beta^{*'})(1/q)dq < \int_0^{q^*} G_N(\beta^*)(1/q)dq + \int_{q^*}^{\overline{q}} G_A(\beta^*)(1/q)dq \tag{16}$$

Figure H21 characterizes how Medicare's subsidy decision depends on f. Medicare will subsidize when the "Subsidize" payoff is less than than the "No subsidize" payoff, which occurs when the technology cost f is low. This result is fairly intuitive: if the technology is inexpensive, then it may be worth it for Medicare to fund the purchase itself and require all hospitals to install it, rather than allowing hospitals to voluntarily choose whether to purchase it.

B.3 Other Model Applications

The model can also be applied to other contexts where waste, monitoring, and technology adoption intersect. For example in filing taxes, an individual or firm can purchase software to e-file to reduce their audit rate – e-filed returns have lower audit rates compared to paper returns. But the tradeoff is that it may be more difficult to misreport earnings when e-filing compared to filing by paper. E-filed returns have an error rate of 1 percent whereas paper returns have an error rate of 20 percent (Internal Revenue Service, 2011). In the context of welfare programs administered by private vendors, the vendors may be willing to adopt automated systems to avoid scrutiny from regulators. For example, retailers that administer WIC (Special Supplemental Nutrition Program for Women, Infants, and Children) and SNAP (Supplemental Nutrition Assistance Program) may voluntarily adopt a EBT system instead of using paper vouchers to mitigate "undercover buys" by state and federal agencies (US Department of Agriculture, 2013). But adopting EBT makes setting different prices for program beneficiaries and non-beneficiaries, which is illegal, infeasible (Meckel, 2020). Finally, one could adapt the model to procurement settings where a supplier is reimbursed based on the *costs* it reports to a purchaser. Here, the supplier could use a third-party accounting firm to compile its financial statements instead of doing it in-house. While this makes misreporting more difficult, the purchaser may subject the supplier to less scrutiny if it knows that the cost-reporting is handled by a third party.

C Appendix: Robustness and Placebo Tests

Hospital-Level Analysis As a robustness test, in Figure H22 I regress on a hospital's denial rate – the share of claims for which a denial is made after audit – rather than its audit rate. Equation 17 defines the relationship between denial rate and audit rate.

$$Denial\ Rate_{ht} = \underbrace{P(Audit)}_{Audit\ Rate}{}_{ht} \times \underbrace{P(Demand|Audit)}_{Demand\ Rate}{}_{ht} \tag{17}$$

Since 41 percent of audits in 2011 resulted in a demand in the main sample and denial rate is monotonically increasing in audit rate (Figure H2), one would expect that a hospital's response to a one-percentage point increase in the denial rate should be about twice the response to one percentage point increase in the audit rate. Indeed, this is what the results reflect; for example, hospitals reduced admissions by 2.5 percent in 2012 in response to a one-percentage point increase in the 2011 audit rate, and they reduced admissions by 5.7 percent in 2012 in response to a one-percentage point increase in the denial rate.

In Figure H7, I show that the results are robust to alternative sample definitions. Figure H7a reproduces the event study from the main specification for the outcome of log Medicare admissions, in which the sample is defined as all hospitals within 100 miles of the RAC border and the coefficient is scaled by the correlation between a hospital's audit rate and its leave-one-out state audit rate. This is robust to changing the sample to all hospitals within 50 miles (Figure H7b) or 150 miles (Figure H7c) of the border, although the results are noisier with a shorter distance. One concern with boundary discontinuity identification strategies is the potential for spillovers among hospitals that are close to the border. On the one hand, if patients were redirected from a hospital near the border in a high-audit rate state to a nearby hospital in a low-audit rate state, then this would bias the coefficients to be larger in magnitude. On the other hand, if hospitals on the low-audit side internalize their high-audit neighbors' audit rates in making their admission decisions, this would bias the coefficients to be smaller in magnitude. These spillovers would be less of a concern as the distance from the border increases or if the hospitals closest to the border are excluded.

Figure H7d shows similar results when restricting the sample to hospitals that are at least 10 miles away from the border, demonstrating that the result is not driven by such spillovers. Finally, Figure H7e shows that the results are similar when restricting the sample to hospitals with audit rates greater than 0 percent, meaning that the results are driven by variation in auditing across hospitals on the intensive, rather than the extensive, margin.

Figure H8 shows that the results are robust to using alternative instruments to scale the reduced form effect. The main specification instruments for a hospital's audit rate using the leave-one-out state audit rate in order to capture the variation in audit intensity that is unrelated to the hospital's own behavior. Figure H8a plots the results of using the state audit rate (which includes the hospital) as an instrument. Figure H8c shows that the results using the leave-one-out RAC region audit rate, rather than the state audit rate, are similar.

While using the leave-one-out audit rate strips away the direct effects of a hospital's own behavior, it still includes other hospitals surrounding a given hospital, whose audit rates may still reflect that hospital's behavior. This can be the case if, for example, a given hospital has a large market share. To address this, in Figures H8b and H8d I consider using the audit rate of other hospitals in the same state or RAC region in *other* markets, which I define using hospital referral regions. This instrument leverages hospitals whose behavior should not be affected by a given hospital's behavior since they are much farther away in different markets. Similarly, one might be concerned that a hospital's audit rate is correlated with the behavior and audit rates of other hospitals in the same hospital system, as they share a common owner. Figure H8e uses the audit rate of hospitals in the same state but different hospital systems in 2010. The results are robust to using these hospitals to instrument for a hospital's audit rate.

Because neighbor comparison groups can overlap, they could potentially span multiple border segments. Thus, clustering at the border segment-level may not capture the correlated errors across border segments, which would bias the standard errors. Given how the neighbor comparison groups are defined, there is no way to set border segments that eliminates this problem. However, it should be less of a concern for longer border segments. Figure H23 plots the event studies from using 50- and 150-mile border segments. While the the standard errors do increase as the segments become longer, the coefficients remain statistically significant.

Finally, to confirm that the results are not driven by a single state or hospital comparison group, Figure H24 plots the distribution of coefficients when one state or one hospital comparison group is removed from the sample. The coefficients are always negative and the distribution is centered around the main effect.

Finally, I consider a falsification test using state borders in the *interior* of each RAC region. In the interior of each region, there is no change in RAC identity at state borders, so comparing hospitals across these interior borders does not capture exogenous variation driven by different audit strategies across RACs. Figure H25a illustrates the interior borders and the sample of hospitals within one hundred miles of the interior border (excluding hospitals that are within one hundred miles of the RAC border). The falsification test shows no effect on admissions on the "high-audit side" of the interior border (Figure H25b), in contrast to the main results, which show a drop in admissions on the high-audit side of the RAC border.

Patient-Level Analysis In Table GXI, I show that the Two Midnight rule difference-in-difference results are robust to varying the sample to include patients who arrive between one and five hours of midnight. Table GIV shows that, in addition to a null effect on revisits within thirty days, there is no effect on revisits within sixty or ninety days.

In column 5 of Table IV, I consider whether there is an effect on non-Medicare patients, who are not directly affected by the Two Midnights rule. I find that after-midnight, non-Medicare ED arrivals do not face a reduction in admissions after the rule is implemented. This indicates that there were no spillovers from the Two Midnights rule onto populations not covered by the rule.

In Figure H26, I show that the heterogeneity by severity results are robust to training the prediction model on the pre-period before-midnight arrivals.

D Appendix: Rural Hospital Closures

The main results show that RAC audits decrease hospital revenue and increase their costs. This raises the concern that RAC auditing may have driven hospitals into financial distress and, given the prevalence of hospital closures in recent years, led them to close. Hospital closures are associated with decreases in access to care and increases in patient mortality (Carroll, 2019; Gujral and Basu, 2019). To study whether RAC auditing led to hospital closures, I use data from the Sheps Center for Health Services Research on rural hospital closures between 2005 and 2022.[30] I adapt the main specification for the hospital-level analysis to study rural hospital closures. In the border hospital sample, no hospitals closed before 2012 – this is by definition, since the hospital had to be open in 2011 to be audited. Therefore there is no variation in the pre-2010 period to use a difference-in-differences framework. Instead, I run the following specification separately for each year Y in the post period:

$$Close_h^Y = X_h^{2011}\beta^Y + \phi_{g(h)} + \varepsilon_h \tag{18}$$

which regresses a dummy for whether a rural hospital has closed in year Y, $Close_h^Y$, on its (instrumented) audit rate X_h^{2011}, after taking into account the hospital's neighbor comparison group. Figure H27 plots the β^Y coefficients for years where there is variation in closures among rural hospitals in the border sample (i.e., excluding 2012, 2017, and 2021). The results indicate that higher RAC auditing did not cause rural hospitals to close.

[30]Data available at https://www.shepscenter.unc.edu/programs-projects/rural-health/rural-hospital-closures/. Last accessed March 2022.

E Appendix: Extrapolation to Overall Hospital Sample

This section describes the calculation to extrapolate the savings estimates from the border hospital sample to the overall RAC program. This calculation rests on fairly strong assumptions, but nonetheless may be of interest for gauging the magnitude of overall savings from the RAC program. First, we must assume that the savings scale linearly with audit rate, so that the effects estimated from a marginal increase in audit rate can be extrapolated beyond the support to a wide range of audit rates. Second, we must assume homogeneous treatment effects across hospitals in the border sample and overall hospitals. Note that while hospitals on opposite sides of the border are similar to each other (Table GII), the border hospital sample differs from the overall sample. Hospitals in the border sample are smaller, more rural, more likely to be non-profit and disproportionately from the Midwest RAC region, Region B, (Table I). Third, this calculation assumes that even at high levels of auditing, there is still no effect on other outcomes that may affect welfare, like patient health or hospital closures.

Under these assumptions, I can calculate the extrapolated savings by multiplying the 2011-2015 event study coefficients on Medicare inpatient revenue (Figure 4b) and payments demanded (Figure H11) by each hospital's 2011 audit rate. Since the estimates are based on the logarithm of inpatient revenue and represent a percent change relative to the baseline in 2010, I multiple these coefficients by the hospital's 2010 inpatient revenue. Figure H28 plots the extrapolated savings for each hospital-year, compared to the actual changes in Medicare inpatient revenue and actual payments demanded. For both types of savings, the extrapolated and actual savings are positively correlated. This indicates that in the overall sample, hospitals subject to higher audit rates reduced their Medicare inpatient revenue and were subject to more audit demands in subsequent years. Summing up the extrapolated savings across all hospitals from 2011 to 2015 implies that the RAC program saved the Medicare program \$9.28 billion between 2011 and 2015, compared to the actual \$11.74 billion in savings from reductions in inpatient spending and audit demands in this period. Note, however, the relatively low R^2 from the regression between extrapolated and actual savings, indicating that much of the variation in savings is not explained by variation in 2011 audit rate.

F Appendix: Hospital-level Emergency Department Visit Analysis

In addition to looking at inpatient admissions, I can also use the Medicare Outpatient file to extend the hospital-level analysis to ED visits, mirroring the patient population in the patient-level analysis. I focus in particular on three outcomes: the share of ED visits that are associated with an observation or "suspected" observation stay, the 30-day ED revisit rate, and the 30-day mortality rate. However, because there are known data reliability issues with measuring emergency department visits in the Medicare claims, these results should be interpreted with caution. Below, I lay out the potential reliability concerns with these measures and then discuss the results.

I use the MEDPAR (Inpatient) and Outpatient files to identify all ED visits by Medicare beneficiaries at the hospitals in my sample. Note that the ED outcomes I consider are *shares*, rather than counts. I use shares because of concerns about inconsistencies in the reporting of ED visit count across different data sources, different providers, and different time periods. Venkatesh et al. (2017) counts ED visits in Medicare claims in one year using four different definitions, and finds differences up to 17 percent across the different measures. Additionally, in attempting to construct a measure of ED visits across multiple years, I also found surprising data anomalies that CMS's Research Data Assistance Center (ResDAC) confirmed were likely due to reporting errors.[31]

The first outcome I consider is the share of ED visits that also include outpatient observation services. I define this as the share of outpatient claims with ED services that also list observation services *or* outpatient visits that span two days (what I call "suspected observation stays"). I include the latter to capture cases where a hospital provides observation services but does not code for it. According to a report by the Office of the Inspector General (2013), many payers do not always paid separately for observation stays, so some hospitals have little incentive to code for observation services. However, they found that many multi-day outpatient visits have similar diagnoses as claims that include observation services, and hospitals vary widely in their tendency

[31] Specifically, I found that over 40% of hospitals in Kansas saw a 200% or higher increase in inpatient stays with ER charges between 2007 and 2008. This anomaly unique to Kansas and 2007-2008 – only 3 percent of other hospitals saw this large of an increase. This anomaly was reproduced by analysts at ResDAC, but they could not identify a reason for why it occurred (link).

to report these as observation stays or simply multi-day outpatient visits. Failing to count multi-day outpatient claims undercounted "suspected" observation stays by almost half. Thus, following their definition, I also include multi-day outpatient visits in my measure.

It is challenging to determine whether an ED visit resulted in an inpatient stay using the MEDPAR (Inpatient) and Outpatient files. This is because the inpatient stays with ED charges only capture a portion of all inpatient stays associated with an ED visit. A portion of ED visits that result in an inpatient stay are located in the MEDPAR file, while the rest are in the Outpatient file with no direct linkage to the associated inpatient stay. ResDAC cautions that "although one can assume ER patients found in the inpatient data *were* admitted to the hospital, one *cannot* assume ER patients found in the outpatient data were *not* admitted to the hospital...some patients are transferred to a different hospital for admission and some hospitals bill ER and inpatient services separately" (Barosso, 2015). A substantive share of Medicare patients undergo inter-hospital transfer, especially for diagnoses that are prevalent in the ED – for example, up to 50% of patients with heart attacks are transferred Iwashyna et al. (2010). Thus, I cannot discern in the Medicare claims which ED visits resulted in an inpatient stay.

Turning to health outcomes, I use the MEDPAR and Outpatient files to construct a measure of the share of ED visits where the patient revisited the ED within 30 days. A slight difference between this outcome and the revisit rate in the patient-level analysis is that I do not count revisits that are direct inpatient admissions without ED charges. This is to avoid accidentally capturing inpatient stays in the MEDPAR file that are actually the result of an outpatient ED claim, as discussed above. I also merge in the patient date of death from the Master Beneficiary file to construct the share of ED visits where the patient died within 30 days of their discharge date (discharged from ED, from inpatient, or died in the hospital).

Figure H15 shows the event studies from Equation 6 on the ED visit outcomes. The results from this analysis are largely consistent with the patient-level results. Among hospitals with a higher 2011 audit rate, the share of ED visits with outpatient observation services increases after 2011. However, ED visits at these hospitals do not seem to result in greater revisits or mortality.

There is a slight (but statistically insignificant) increase in revisit rates after 2012, but its magnitude (0.25%) is very small relative to the 2010 mean, 15%.

G Appendix Tables

Table GI. Summary Statistics of 2010 Inpatient Characteristics, by Sample

	(1)	(2)	(3)	(4)
	MEDPAR Sample		SID/SEDD Sample	
	All	*Border (100 mile)*	*FL ED*	*ED, 3 hr*
average age	73.04	73.35	74.10	72.59
	(14.03)	(13.66)	(14.19)	(15.12)
share female	0.56	0.56	0.55	0.54
	(0.50)	(0.50)	(0.50)	(0.50)
share white	0.82	0.87	0.83	0.81
	(0.39)	(0.33)	(0.38)	(0.39)
share inpatient last 30d	0.16	0.16	0.15	0.16
	(0.37)	(0.36)	(0.36)	(0.37)
Observations	11919671	2681021	602059	88027

This table presents 2010 summary statistics of traditional Medicare beneficiaries receiving inpatient stays in the following samples: all hospitals (column 1), hospitals within 100 miles of the border (column 2), patients admitted as inpatient from a Florida ED (column 3), and patients admitted as inpatient from a Florida ED who arrived at the ED within 3 hours of midnight (column 4). Data: MEDPAR and HCUP SID/SEDD.

Table GII. Correlation between 2010 hospital characteristics and 2011 audit rate

	(1) beds	(2) urban	(3) for profit	(4) non-chain	(5) total costs (millions)	(6) admin costs (millions)	(7) Medicare admissions	(8) inpatient revenue (millions)	(9) short stay share	(10) predicted 2011 audit rate
					Panel A: Border Sample					
2011 audit rate	-3.82	-0.02**	-0.02	-0.00	1.53	-0.43	-120.16	-0.88	0.00*	0.00
	(4.33)	(0.01)	(0.01)	(0.01)	(5.66)	(0.70)	(71.29)	(0.70)	(0.00)	(0.00)
Nbr group FE	X	X	X	X	X	X	X	X	X	X
Mean	178.81	.57	.13	.41	166.98	23.44	3128.15	26.51	.31	.02
N Hosp	510	510	510	510	510	510	510	510	510	496
					Panel B: Overall Sample					
2011 audit rate	-12.82***	-0.02**	-0.03***	0.03***	-7.52*	-0.66	-241.67***	-2.38***	0.01***	0.00**
	(2.93)	(0.01)	(0.01)	(0.01)	(3.79)	(0.62)	(50.90)	(0.52)	(0.00)	(0.00)
Mean	202.16	.72	.19	.38	212.16	29.17	3465.75	34	.31	.02
N Hosp	2960	2960	2960	2758	2960	2960	2960	2960	2960	2873

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the state level. Panel A reports the coefficients from regressing the 2011 audit rate on an outcome variable in 2010 in the border sample, with neighbor comparison group fixed effects. The border sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Panel B reports the coefficients from regressing the 2011 audit rate on an outcome variable in 2010 in the overall sample. Bed size, urban status, and profit type status come from the Medicare Provider of Services file. Non-chain status comes from hospital merger data via Cooper et al. (2019). Total and administrative costs come from HCRIS. Medicare admissions and inpatient stay characteristics are from MEDPAR. Mean inpatient characteristics are defined as the average of each hospital's average (i.e., weighted by hospitals rather than claims). Short stay share is the share of Medicare admissions with length of stay ≤ 2. "Predicted 2011 audit rate" is a claim-level prediction using solely stay characteristics (but not hospital, state, or RAC characteristics) trained on 2007-2009 claims. The prediction specification is a regression of the likelihood of being audited in 2011 on admission month, major diagnostic category, admission source, and length of stay for each hospital's 2007-2009 claims. Data: MEDPAR, Medicare Provider of Services File, Cooper et al. (2019) merger data, and HCRIS.

Table GIII. ED Arrival Hour Manipulation Tests

	(1) $[23{:}00 \leq T_v \leq 23{:}59]$	(2) $\mathbb{1}[T_v \geq 00{:}00]$
$\mathbb{1}[q \geq 2013\text{Q3}]$	-0.001	-0.003
	(0.001)	(0.002)
Observations	1511606	1511606

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered by the ED arrival hour and quarter. This table reports estimates and standard errors of the coefficient on $\mathbb{1}[q \geq 2013\text{Q3}]$, an indicator for whether the ED visit occurred after the Two Midnights rule was implemented in 2013Q3. $[23{:}00 \leq T_v \leq 23{:}59]$ is an indicator equal to 1 if a patient's ED arrival hour is between 11:00PM and midnight, and 0 otherwise. $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether at patient's ED arrival hour was after midnight. Regression includes hospital fixed effects. Sample consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Data: HCUP SID/SEDD.

Table GIV. After-Midnight ED Arrival Coefficient on Stay Characteristics and Patient Outcomes

	(1) Total Charges ($)	(2) ED Charges ($)	(3) N Diagnoses	(4) N Procedures	(5) OR Procedure	(6) Revisit 60d	(7) Revisit 90d
β	42.707	-22.58	-0.003	-0.005	-0.001	0.002	0.000
	(254.406)	(15.67)	(0.013)	(0.009)	(0.001)	(0.002)	(0.002)
Observations	1252735	1254857	1254857	1254857	1254857	1254857	1254857

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[q \geq 2013\text{Q3}] \times \mathbb{1}[T_v \geq 00{:}00]$ of the specification in Equation 12, where $\mathbb{1}[q \geq 2013\text{Q3}]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "OR procedure" is an indicator for whether a patient received an OR procedure during their stay. "Revisit within 60/90 days" is an indicator for whether the patient had another ED visit or inpatient stay within 60/90 days of the ED visit. Sample comprises traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Data: HCUP SID/SEDD.

Table GV. Cross-sectional Correlations with Medical Necessity Checking Software

	2010-2012 Difference in Admissions	
Install MN App, 2011-2015	-48.631**	-44.858**
	(21.918)	(21.907)
RAC FE		X
Hosp	2960	2960

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors are in parentheses. This table reports the coefficients from regressing a dummy variable for whether a hospital installs medical necessity checking software in 2011-2015 on the 2010-2012 change in admissions, with and without RAC region fixed effects. Data: HIMSS and CMS audit data.

Table GVI. Across-Hospital Post-2011 Coefficient

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		LOS ≤ 2		Admin Costs	Software Installation
	Log Adm.	*Log Rev.*	*Log Adm.*	*Log Rev.*	*Log Costs*	*Medical Necc.*
2011 audit rate	-0.0154	-0.0166	-0.0227**	-0.0234***	0.0087	0.0153*
$\times$ post-2011	(0.0092)	(0.0136)	(0.0096)	(0.0056)	(0.0100)	(0.0081)
Hosp FE	X	X	X	X	X	X
Nbr group FE	X	X	X	X	X	X
Hosp	510	510	510	510	510	506
N	52139	52139	52139	46437	52107	36906
F	104.98	104.98	104.98	104.61	104.68	84.15

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the state and border segment level. This table reports the coefficients of the reduced form event study in Equation 10, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome after 2011. Columns 1-2 report two stage least squares outcomes for the number of and revenue from Medicare admissions overall, columns 3-4 report outcomes for the number of and revenue from Medicare admissions with length of stay ≤ 2, column 5 reports the outcomes for log net administration costs, and column 6 reports the outcomes for an indicator for installation of medical necessity software. Length of stay is counted as the difference in days between the admission and discharge date. Inpatient revenue is the sum of all Medicare inpatient payments. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in HIMSS. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, CMS audit data, HCRIS, and HIMSS.

Table GVII. Heterogeneity of Across-Hospital Post-2011 Coefficient

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		LOS ≤ 2		Admin Costs	Software Installation
	Log Adm.	*Log Rev.*	*Log Adm.*	*Log Rev.*	*Log Costs*	*Medical Necc.*
Panel A: Urban						
2011 audit rate × post-2011	-0.0410***	-0.0226	-0.0513***	-0.0215*	-0.0042	0.0130
	(0.0131)	(0.0145)	(0.0130)	(0.0113)	(0.0096)	(0.0082)
2011 audit rate × post × Urban	0.0367***	0.0086	0.0410***	-0.0017	0.0185**	0.0034
	(0.0090)	(0.0069)	(0.0109)	(0.0108)	(0.0083)	(0.0064)
Panel B: Teaching						
2011 audit rate × post-2011	-0.0195**	-0.0200	-0.0254**	-0.0235***	0.0042	0.0154
	(0.0082)	(0.0135)	(0.0105)	(0.0081)	(0.0104)	(0.0100)
2011 audit rate × post × Teaching	0.0195	0.0162	0.0131	0.0037	0.0217***	-0.0008
	(0.0131)	(0.0112)	(0.0177)	(0.0153)	(0.0069)	(0.0147)
Panel C: Hospital Profit Type						
2011 audit rate × post-2011	-0.0100	-0.0136	-0.0164*	-0.0199***	0.0116	0.0136*
	(0.0104)	(0.0143)	(0.0092)	(0.0069)	(0.0097)	(0.0073)
2011 audit rate × post × For-Profit	-0.0357*	-0.0386**	-0.0517**	-0.0539**	-0.0318	0.0169
	(0.0182)	(0.0162)	(0.0217)	(0.0256)	(0.0216)	(0.0114)
2011 audit rate × post × Gov't	-0.0258*	-0.0098	-0.0279	-0.0041	-0.0103	0.0030
	(0.0147)	(0.0130)	(0.0181)	(0.0178)	(0.0159)	(0.0075)
Panel D: Chain vs. non-chain						
2011 audit rate × post-2011	-0.0079	-0.0148	-0.0071	-0.0167*	0.0119	0.0193***
	(0.0140)	(0.0162)	(0.0110)	(0.0082)	(0.0094)	(0.0061)
2011 audit rate × post × Non-chain	-0.0150	-0.0037	-0.0312**	-0.0121	-0.0063	-0.0067
	(0.0122)	(0.0097)	(0.0143)	(0.0107)	(0.0044)	(0.0083)
Panel E: Bed Size						
2011 audit rate × post-2011	-0.0364***	-0.0260*	-0.0433***	-0.0231*	0.0015	0.0090
	(0.0104)	(0.0140)	(0.0126)	(0.0131)	(0.0110)	(0.0139)
2011 audit rate × post × Above Avg Beds	0.0419**	0.0187	0.0410**	0.0009	0.0144	0.0133
	(0.0165)	(0.0124)	(0.0173)	(0.0182)	(0.0090)	(0.0147)
Panel F: Medical Necessity Software Installed in 2010						
2011 audit rate × post-2011	-0.0172	-0.0210	-0.0188	-0.0204**	0.0187	0.0258***
	(0.0156)	(0.0177)	(0.0121)	(0.0093)	(0.0115)	(0.0051)
2011 audit rate × post × Med. Necc. (2010)	0.0035	0.0081	-0.0070	-0.0042	-0.0183	-0.0164***
	(0.0131)	(0.0103)	(0.0136)	(0.0099)	(0.0127)	(0.0051)
Hosp	510	510	510	510	510	506
N	52139	52139	52139	52118	52107	36906

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors are in parentheses and are clustered at the state and border segment level. This table reports the coefficients of the reduced form event study in Equation 10, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome after 2011. Columns 1-2 report two stage least squares outcomes for the number of and revenue from Medicare admissions overall, columns 3-4 report outcomes for the number of and revenue from Medicare admissions with length of stay ≤ 2, column 5 reports the outcomes for log net administration costs, and column 6 reports the outcomes for an indicator for installation of medical necessity software. Length of stay is counted as the difference in days between the admission and discharge date. Inpatient revenue is the sum of all Medicare inpatient payments. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted," "installation in process," and "to be replaced" in the HIMSS data in 2012. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Omitted year is 2010. Data: MEDPAR, CMS audit data, HCRIS, HIMSS, Medicare Provider of Services, and Cooper et al. (2019) merger data.

Table GVIII. After-Midnight ED Arrival Difference-in-Difference Coefficient, Heterogeneity by Patient Severity

	(1) Inpatient	(2) Revisit 30d
$\beta \times$ (Risk Decile 1)$_v$	0.015***	0.001
	(0.003)	(0.003)
$\beta \times$ (Risk Decile 2)$_v$	-0.006**	-0.002
	(0.002)	(0.005)
$\beta \times$ (Risk Decile 2)$_v$	-0.018***	0.001
	(0.004)	(0.005)
$\beta \times$ (Risk Decile 3)$_v$	-0.018***	0.009
	(0.007)	(0.006)
$\beta \times$ (Risk Decile 4)$_v$	-0.052***	0.004
	(0.008)	(0.006)
$\beta \times$ (Risk Decile 6)$_v$	-0.055***	-0.005
	(0.006)	(0.007)
$\beta \times$ (Risk Decile 7)$_v$	-0.036**	0.003
	(0.011)	(0.007
$\beta \times$ (Risk Decile 8)$_v$	-0.009	-0.008
	(0.014)	(0.005)
$\beta \times$ (Risk Decile 9)$_v$	-0.007	-0.000
	(0.010)	(0.004)
$\beta \times$ (Risk Decile 10)$_v$	-0.003	-0.002
	(0.004)	(0.005)
Observations	1236048	1236048

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors are in parentheses and are clustered by the ED arrival hour and quarter. This table reports the $\beta \times$ (Risk Decile 1)$_v$ coefficient on $\mathbb{1}[q \geq 2013Q3] \times \mathbb{1}[T_v \geq 00{:}00]$ of the specification in Equation 12, interacted with an indicator for the predicted risk decile of visit v. $\mathbb{1}[q \geq 2013Q3]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. Patient risk is predicted by estimating a logit using ED visits between 9AM and 3PM of an indicator for being admitted within 30 days of an ED visit on patient demographics, current ED visit information, and information on any prior visits in the last 365 days. Demographics include age-bin, sex, race, Hispanic indicator, point of origin indicator, and mean zip code income. Information on current visit includes hospital and quarter. Information on previous visits includes the number of visits/inpatient stays/length of stay in the last month or last year, as well as any diagnoses and procedures recorded in stays within the last month or last year. Data: HCUP SID/SEDD.

Table GIX. After-Midnight ED Arrival Difference-in-Difference Coefficient on Vulnerable Subsamples

	(1)	(2)	(3)	(4)
	Patient Sample			
	Top 25% age	Top 25% n cc	Non-white	Bottom 25% income
	Panel A: Inpatient			
β	-0.009*	-0.006***	-0.007*	-0.008**
	(0.004)	(0.000)	(0.003)	(0.003)
	Panel B: Revisit within 30 days			
β	-0.004	-0.001	0.009	-0.000
	(0.005)	(0.004)	(0.005)	(0.004)
Observations	321649	377451	250824	381927

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors are in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient for different patient subsets on $\mathbb{1}[q \geq \text{2013Q3}] \times \mathbb{1}[T_v \geq \text{00:00}]$ of the specification in Equation 12, where $\mathbb{1}[q \geq \text{2013Q3}]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq \text{00:00}]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "Inpatient" is an indicator for whether the patient was eventually admitted as inpatient from the ED. "Revisit within 30 days" is an indicator for whether the patient had another ED visit or inpatient stay within 30 days of the ED visit. The sample consists of Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Column 1 comprises the subset of patients in the top quartile of age, column 2 comprises patients in the top quartile of numbers of chronic conditions, column 3 comprises non-white patients, and column 4 comprises patients living in zip codes with the lowest quartile income. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Data: HCUP SID/SIDD.

Table GX. After-Midnight ED Arrival Coefficient, Heterogeneity by Hospital Chars.

	(1)	(2)	(3)	(4)	(5)	(6)
			Inpatient			
β	0.011*	-0.005**	-0.004*	-0.008***	-0.007***	0.002
	(0.005)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)
× Urban	-0.019**					
	(0.005)					
× Teaching		-0.006*				
		(0.003)				
× For-profit			-0.007*			
			(0.003)			
× Gov't			-0.003			
			(0.006)			
× Non-chain				0.003		
				(0.006)		
× Above Avg. Beds					0.010**	
					(0.003)	
× Med. Necc. App						-0.013***
						(0.003)
Observations	1246862	1246856	1246862	1222485	1246862	1203528

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[q \geq 2013\text{Q3}] \times \mathbb{1}[T_v \geq 00{:}00]$ of the specification in Equation 12, interacted with hospital characteristics. $\mathbb{1}[q \geq 2013\text{Q3}]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "Inpatient" is an indicator variable for whether the patient was eventually admitted as inpatient from the ED (HCUP SID/SEDD). The sample consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Urban/rural, teaching/non-teaching, for-profit/government/non-profit, and bed size come form the Medicare Provider of Services file. Non-chain status come from Cooper et al. (2019). Medical necessity application is an indicator which is equal to one if medical necessity checking application is listed as "live and operational," "contracted" "installation in process," or "to be replaced" in the HIMSS data in 2012. Data: MEDPAR, CMS audit data, HCRIS, HIMSS.

Table GXI. Robustness Test: Sample of Patients by ED Arrival Relative to Midnight

	(1)	(2)	(3)	(4)	(5)
	Patient Sample				
	Within 1 Hour	Within 2 Hours	Within 3 Hours	Within 4 Hours	Within 5 Hours
	Panel A: Inpatient				
β	-0.007	-0.007**	-0.007***	-0.008***	-0.007***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
	Panel B: Revisit within 30 days				
β	-0.002	0.000	0.001	-0.000	0.000
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	394222	809058	1254857	1740915	2267496

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[q \geq \text{2013Q3}] \times \mathbb{1}[T_v \geq \text{00:00}]$ of the specification in Equation 12, where $\mathbb{1}[q \geq \text{2013Q3}]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq \text{00:00}]$ is an indicator for whether the ED arrival hour for the visit was after midnight. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. The samples comprise of traditional Medicare patients who arrive at the ED in a Florida hospital within 1 hour of midnight (11PM-12:59AM; column 1), within 2 hours of midnight (10PM-1:59AM; column 2); within 3 hours of midnight (9PM-2:59AM; column 3); within 4 hours of midnight (8PM-3:59AM; column 4); and within 5 hours of midnight (7PM-4:59AM; column 5). Data: HCUP SID/SEDD.

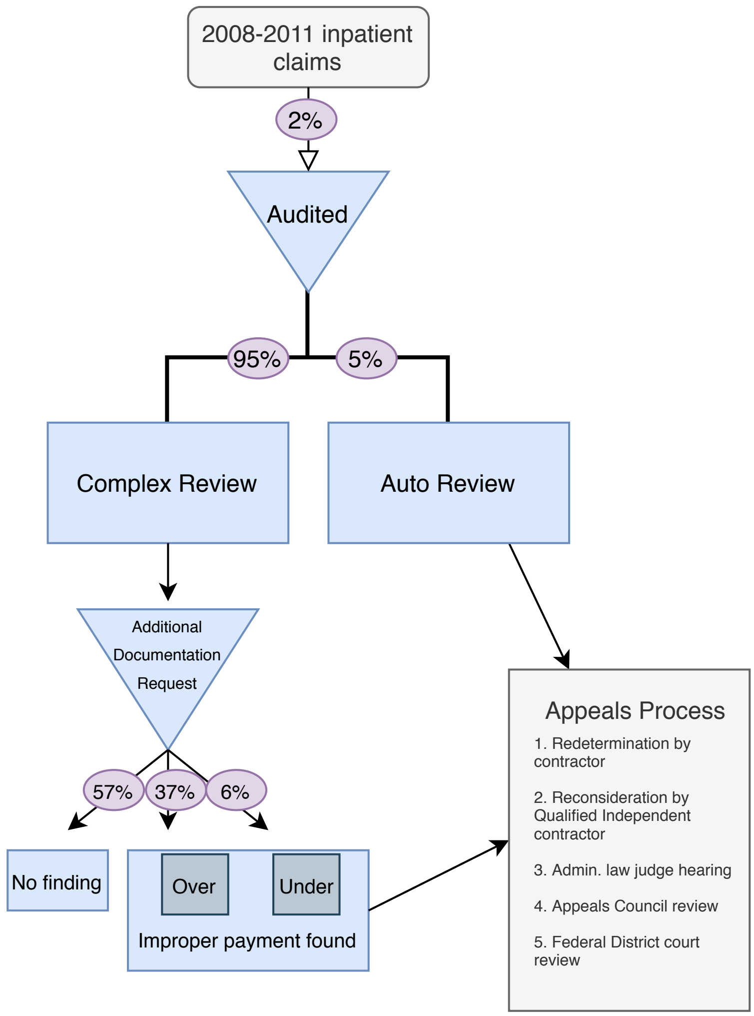
Table GXII. Model Parameters

Parameter	**Value**	**Definition**	**Source**
		Parameters Matched from Data	
β	0.022	audit rate	2008-2011 Medicare claims + RAC data
P	6104	price/stay	2010 Medicare claims, $LOS < 2$
q	5020	actual # pts needing admission	2010 Medicare claims, $LOS < 2$ $\times 5$
$\overline{q}$	21840	max # pts needing admission	99th pctile 2010 Medicare claims, $LOS < 2$ $\times 5$
		Other Parameters	
k	$\frac{p}{1.55}$	treatment cost/stay	Medicare Payment Advisory Commission (2015) Table 7-1
f	50000	fixed technology cost	American Hospital Association (2012)
$1-\gamma$	0.5	tech reduction in audit rate	
w	3	hospital misreporting cost	
r	100	tech misreporting cost	
		Medicare Weights	
c_1	1	misreporting weight	
c_2	1	expenditure weight	
c_3	1	audit cost weight	

This table lists the parameters and assumptions depicted in the numerical solutions to the regulator's problem in Figures H19, H20, and Figure H21. Annual patient count values q and $\overline{q}$ are multiplied by 5 to represent a 5-year horizon.

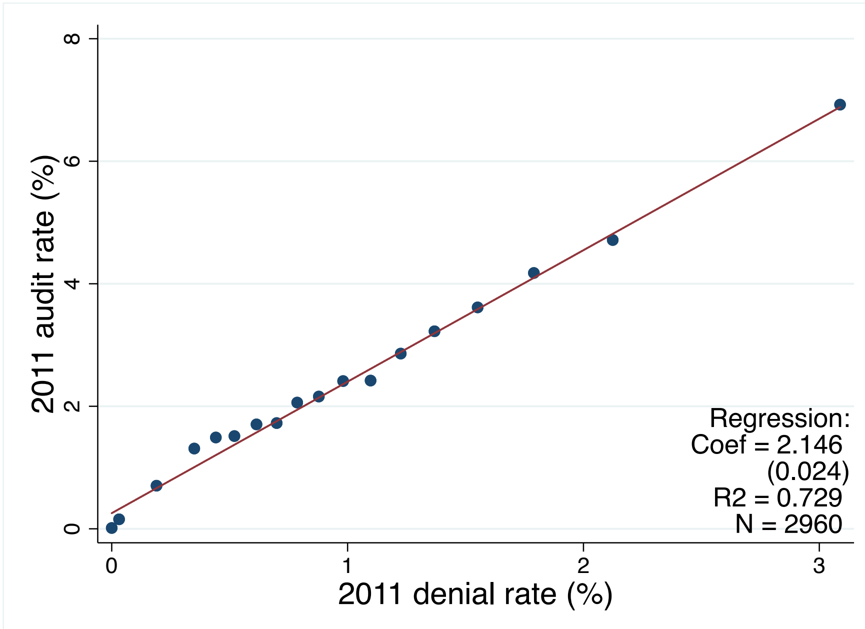
H Appendix Figures

Figure H1. RAC Inpatient Claims Auditing and Appeals Process, 2011 Audits



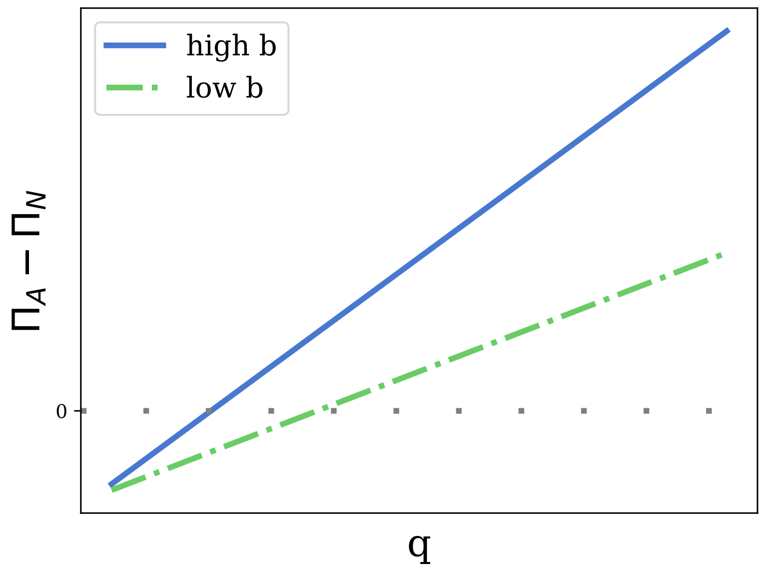
This figure illustrates the stages of the claims auditing and appeals process. The percentages in ovals denote the percent of claims that, conditional on reaching a given stage in the process, reach the next stage. The percentages are calculated based on audits in 2011 of inpatient claims between 2008 and 2011. Data: CMS audit data.

Figure H2. 2011 Audit Rate vs. 2011 Denial Rate



This figure plots a binscatter of 2011 hospital audit rate (share of claims subject to an audit) against the 2011 hospital denial rate (share of claims with reclaimed payment because of audit). Data: CMS audit data and MEDPAR.

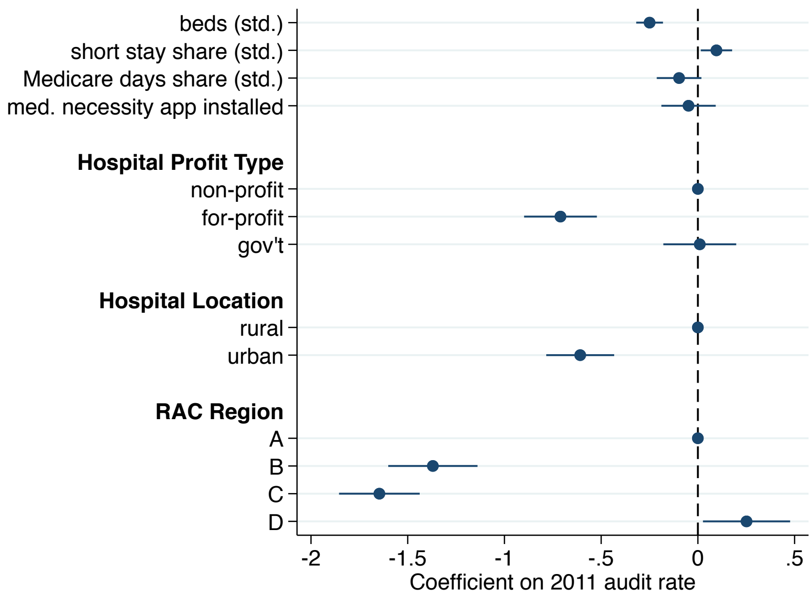
Figure H3. Solving Hospital's Technology Adoption Problem



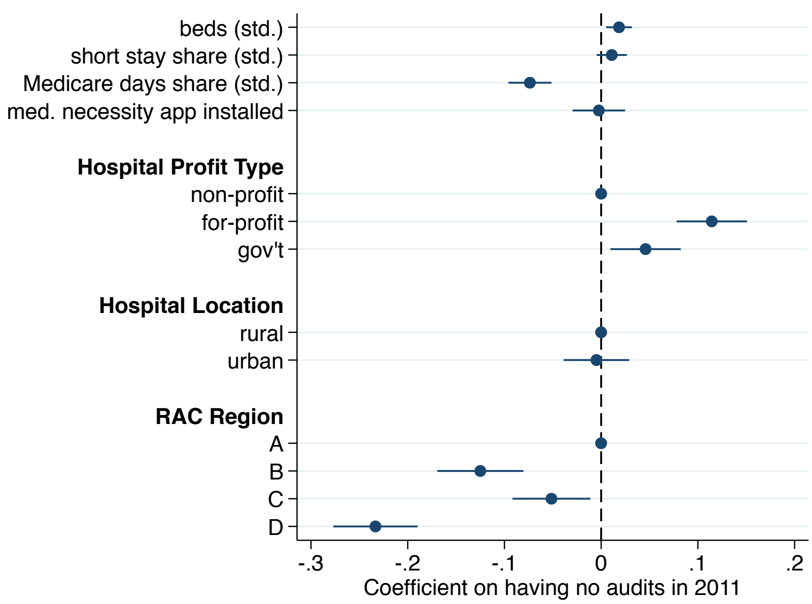
This figure plots the relationship between the difference between the technology adoption payoff and no technology payoff ($\Pi_A - \Pi_N$) and true patient count q, with a high β and low β. The model parameters used are described in Table GXII.

Figure H4. Correlation between Hospital Characteristics on 2011 Audit Rate and No Audit

(a) Outcome: 2011 hospital audit rate

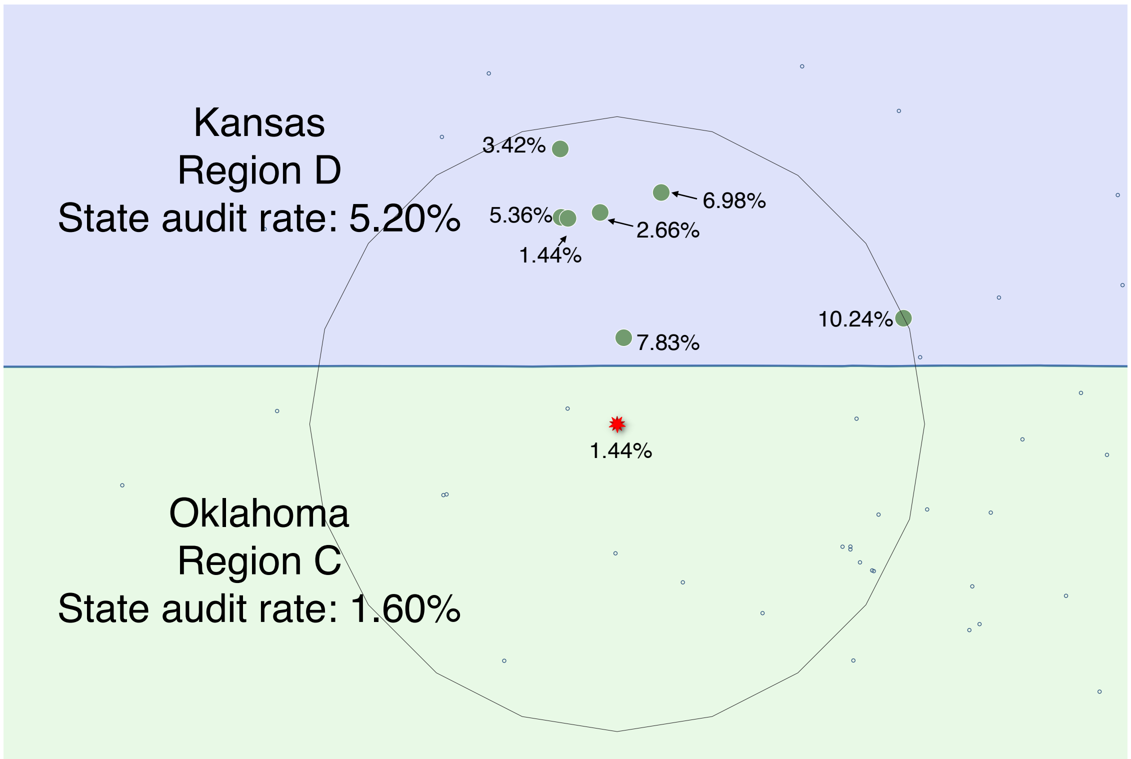


(b) Outcome: no audits at hospital in 2011



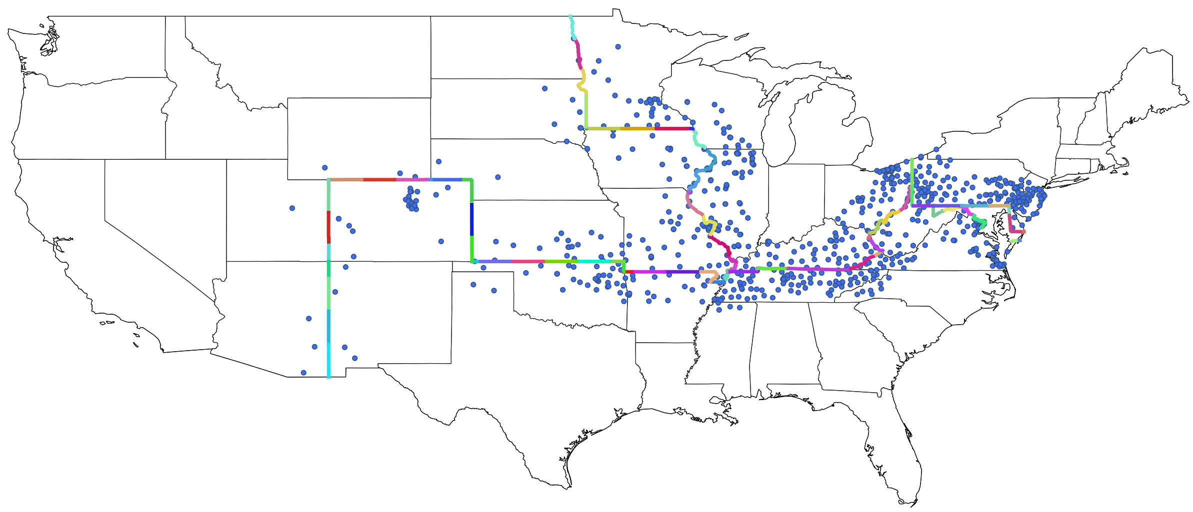
These figures plot coefficients from a regression of (a) a hospitals 2011 audit rate and (b) an indicator variable for whether a hospital was not audited in 2011 on 2010 hospital characteristics. Short stay share is the share of 2010 Medicare admissions with lengths of stay 0-2. Medicare days share is percent of hospital days that are Medicare. Beds, short stay share, and Medicare days share are standardized relative to the mean. Data: MEDPAR, CMS audit data, and Medicare Provider of Services file.

Figure H5. Example of Border Hospital and Neighbor Comparison Group Definition



This figure illustrates how a "neighbor comparison group" is identified for each border hospital in the across-hospital empirical strategy. Neighboring hospitals are all hospitals within a 100 mile radius of a hospital, on the opposite side of the RAC border. In this example, the green circle hospitals in Kansas are considered neighboring hospitals to the red spiked hospital in Oklahoma.

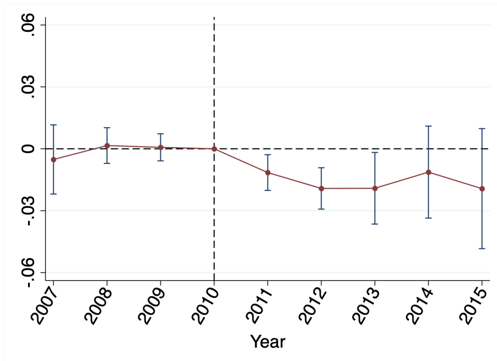
Figure H6. RAC Border Segments and Hospitals Within 100 Miles



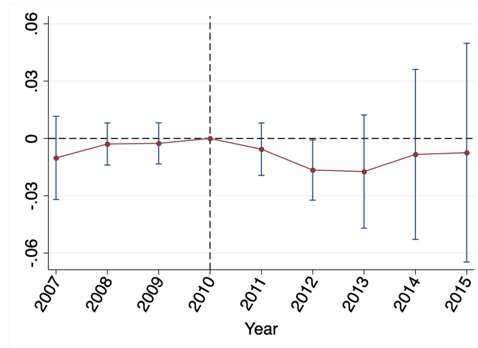
This figure shows how the RAC border is divided into one-hundred mile segments that do not cross state borders, and all hospitals within one-hundred miles of the RAC border. These border segments are used for clustering in Equation 6.

Figure H7. Robustness to Sample Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

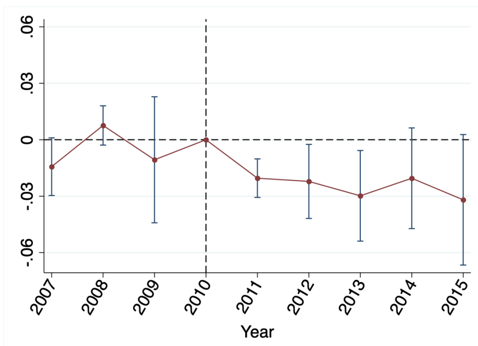
(a) Sample: within 100 miles (main)



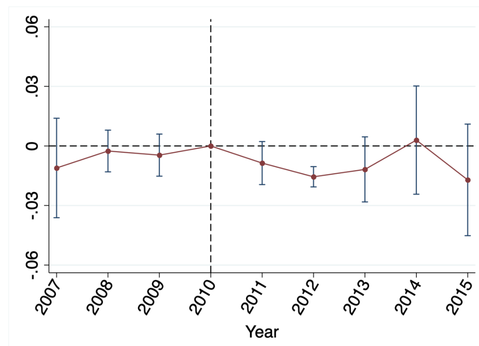
(b) Sample: within 50 miles



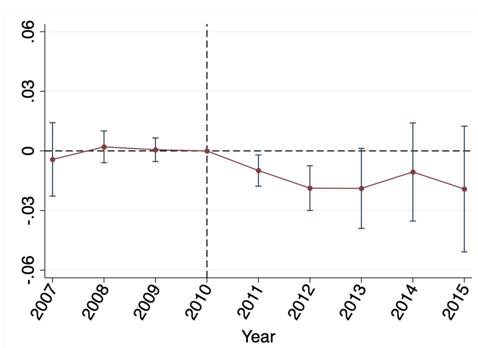
(c) Sample: within 150 miles



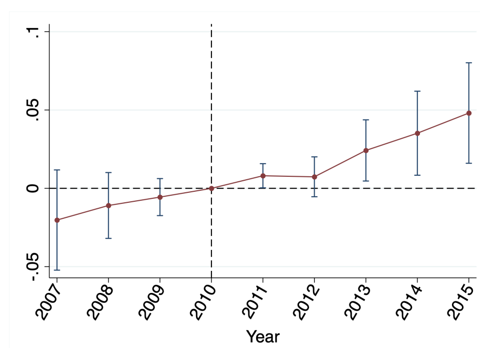
(d) Sample: w/in 100 miles, outside 10



(e) Sample: within 100 miles, audit rate > 0



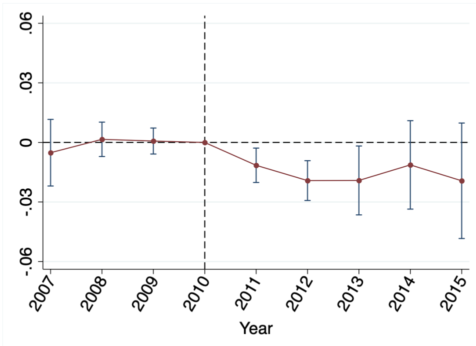
(f) Sample: non-border, all hospitals



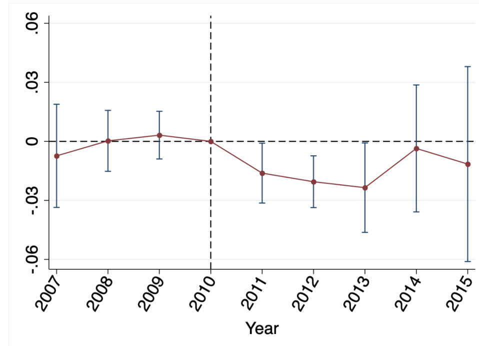
This figure plots robustness analysis event studies of the scaled reduced form coefficients and 95% confidence intervals of the specification in Equation 8, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient estimates the effect of a one percentage point increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different definitions of the border sample: (a) reproduces the main result and defines the border sample to be all hospitals within 100 miles of the RAC border; (b) defines the border sample to be all hospitals within 50 miles of the RAC border, (c) defines the border sample to be all hospitals within 150 miles of the RAC border, (d) defines the border sample to be all hospitals within 100 miles of the RAC border, excluding hospitals within 10 miles of the border, and (e) uses the 100 mile border sample and restricts to hospitals with 2011 audit rate greater than 0. Panel (f) plots the results for all hospitals (N=3014), in a specification where the hospitals audit rate is instrumented using the leave-one-out RAC region rate and includes hospital and year fixed effects. Data: MEDPAR and CMS audit data.

Figure H8. Robustness to Instrument Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

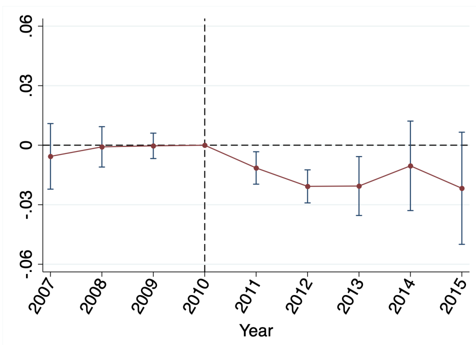
(a) Instrument: state audit rate



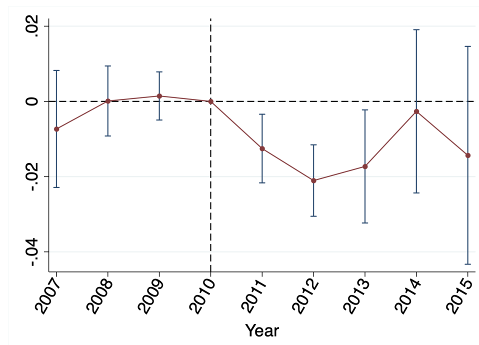
(b) Instrument: state rate, outside HRR



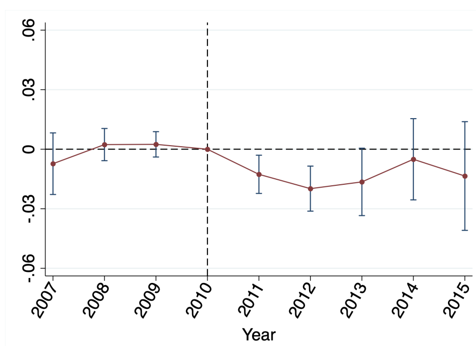
(c) Instrument: leave-one-out RAC region rate



(d) Instrument: RAC rate, outside HRR

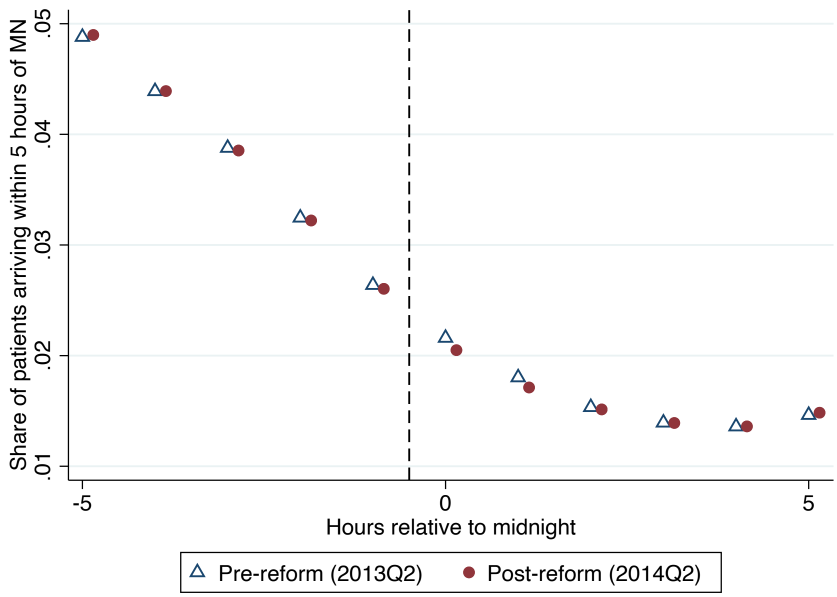


(e) Instrument: state rate, outside system



This figure plots robustness analysis event studies of the reduced form coefficients and 95% confidence intervals of the specification in Equation 8, scaled by the correlations between the instruments and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different instruments for a hospital's 2011 audit rate. Panel (a) uses 2011 state audit rate and panel, (b) uses 2011 audit rate among hospitals in the same state but in different hospital referral regions (HRR) as the hospital, (c) uses the 2011 audit rate of other hospitals in the same RAC region, (d) uses the 2011 audit rate of other hospitals in the same RAC region but in different HRRs, and (e) uses the 2010 audit rate of other hospitals in different hospital systems in 2010. Data: MEDPAR, CMS audit data, and hospital systems from Cooper et al. (2019).

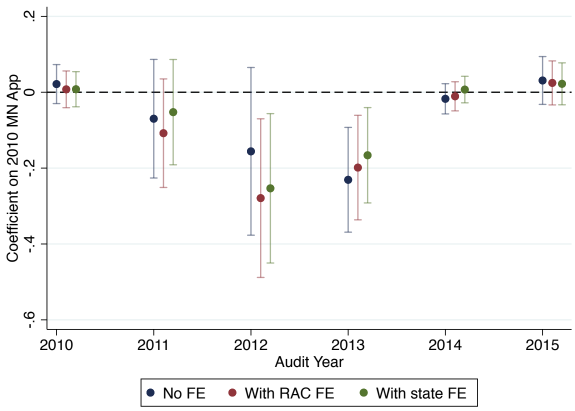
Figure H9. Share of Medicare ED Patients By Hour of ED Arrival



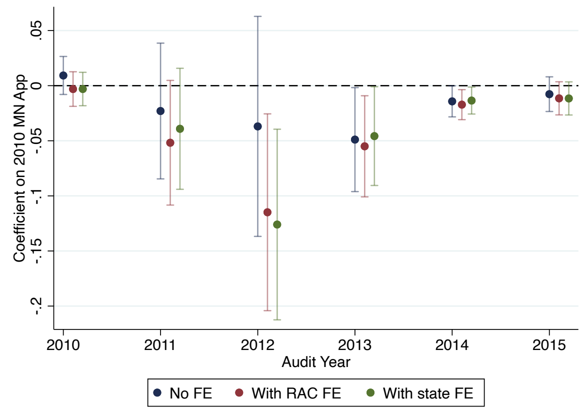
This figure plots the share of Medicare patients that arrive at the ED at each hour (relative to midnight) pre- and post-reform, among traditional Medicare patients who arrived in the ED within 5 hours of midnight in Florida. Data: HCUP SID/SEDD.

Figure H10. Cross-sectional Correlations with Medical Necessity Checking Software

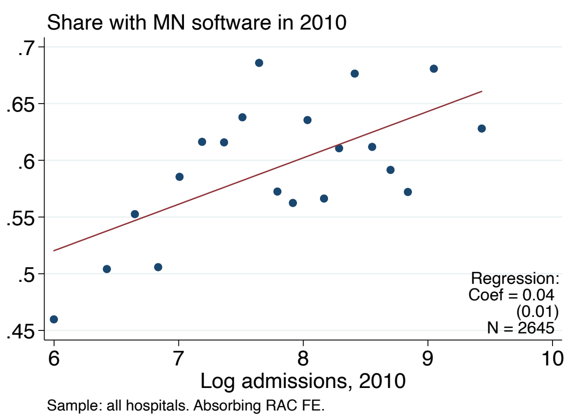
(a) Audit rate and MN installed in 2010



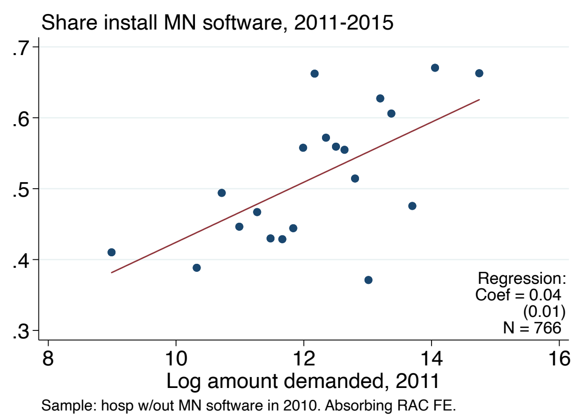
(b) Denial rate and MN installed in 2010



(c) Binscatter of 2010 admissions and dummy for having MN installed in 2010

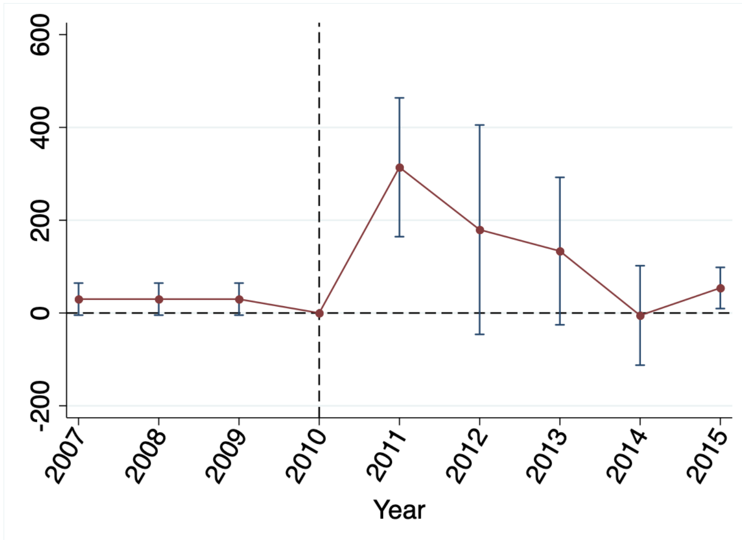


(d) Binscatter of 2011 demanded amount and dummy for install MN in 2011-2015



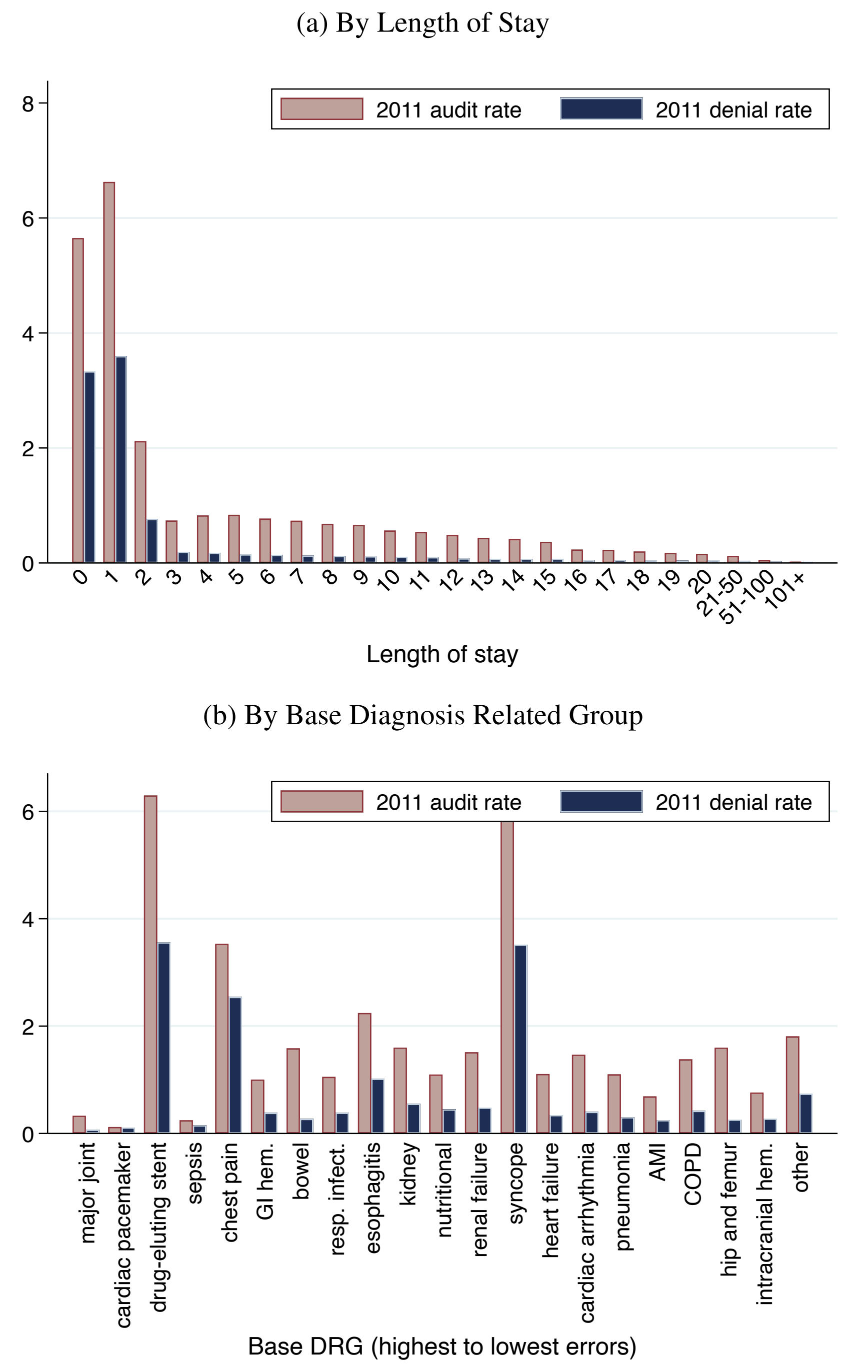
Panels (a) and (b) plot the coefficients of the correlation between a dummy variable for whether a hospital has medical necessity checking software installed in 2010 with 2010-2015 audit rates and denial rates, respectively. The first specification has no fixed effects, the second specification has RAC region fixed effects, and the third specification has state fixed effects. Panel (c) plots a binscatter of 2010 log admissions and a dummy variable for whether a hospital has medical necessity checking software installed in 2010. Panel (d) plots a binscatter of 2011 demanded amount (log \$) and a dummy variable for whether a hospital installs medical necessity checking software in 2011-2015, restricting to hospitals that do not have it installed in 2010. Both the specifications in panels (c) and (d) absorb RAC fixed effects. Data: HIMSS and CMS audit data.

Figure H11. Event Study on Effect of 2011 Audit Rate on Payment Demanded ($1000s) from RAC Audits



This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 8, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. The outcome is the amount of payment demanded initially from RAC audits of inpatient stays, by year of audit. Data: CMS audit data.

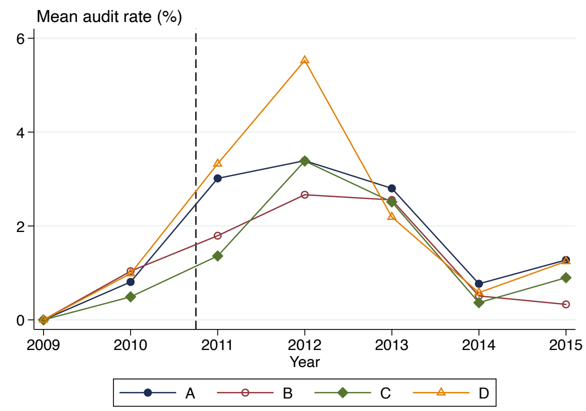
Figure H12. 2011 Audit and Denial Rates by Stay Characteristics



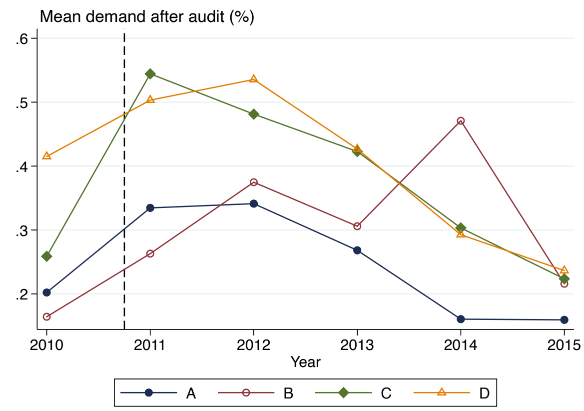
This figure plots the count of 2011 audit and denial rates by (a) an admission's length of stay and (b) its base DRG. The graph shows the top 20 base DRGs with highest improper payments identified in the 2010 CERT report, in descending order of estimated improper payments (Centers for Medicare and Medicaid Services, 2011b), compared to other DRGs. Data: MEDPAR and CMS audit data.

Figure H13. Audit Outcomes and Admissions by RAC Region and Year

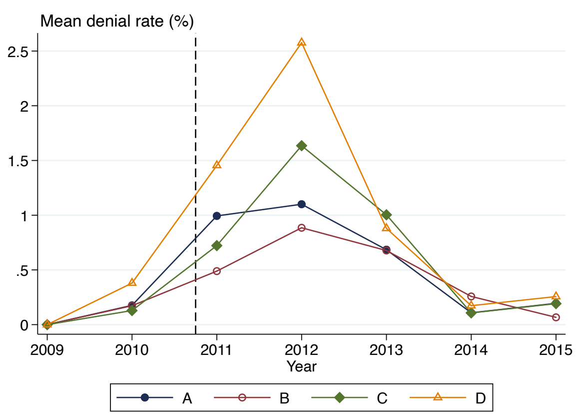
(a) Mean Audit Rate



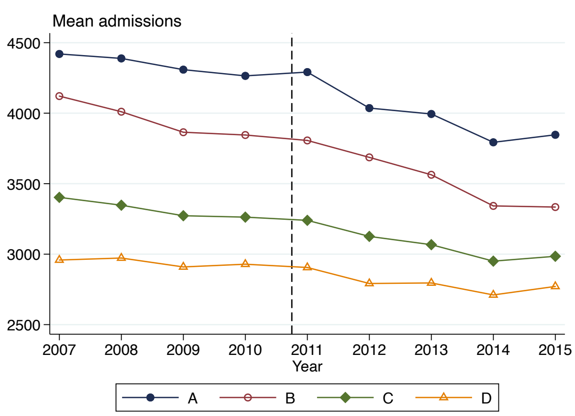
(b) Mean Demand After Audit Rate



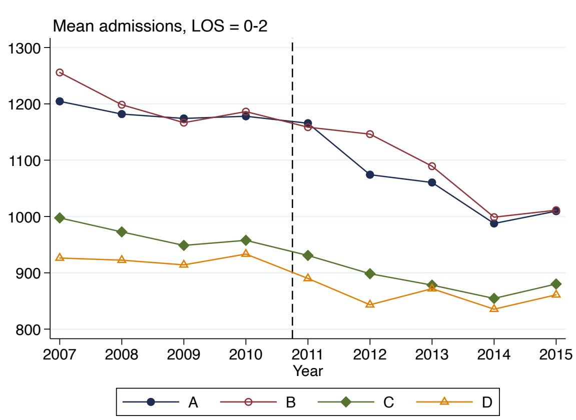
(c) Mean Denial Rate



(d) Mean Admissions

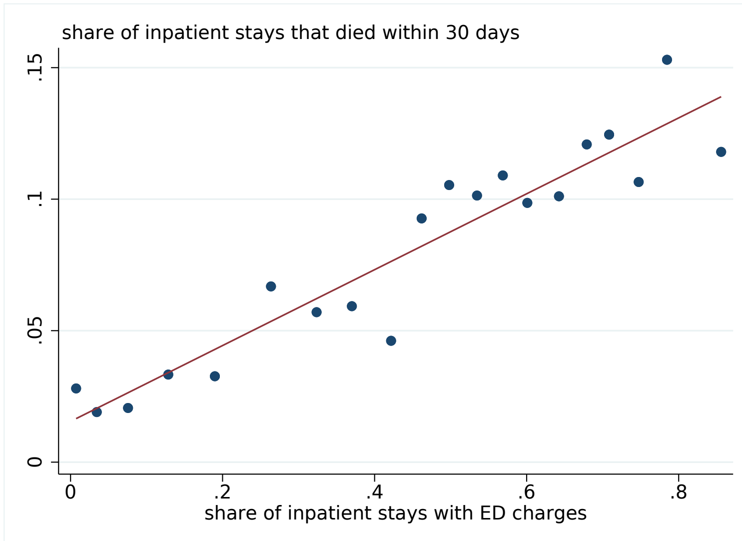


(e) Mean Admissions (LOS = 0-2)



These figures plot over time for each RAC region the (a) mean hospital audit rate, (b) demand after audit rate, (c) denial rate, (d) mean number of admissions, and (e) mean number of admissions with LOS ≤ 2. Denial and demand after audit rate are defined as follows: $Denial\ Rate_{ht} = P(Audit)_{ht}\ \times\ P(Demand|Audit)_{ht}$, where $P(Audit)_{ht}$ is the audit rate and $P(Demand|Audit)_{ht}$ is the demand after audit rate. Data: MEDPAR and CMS audit data.

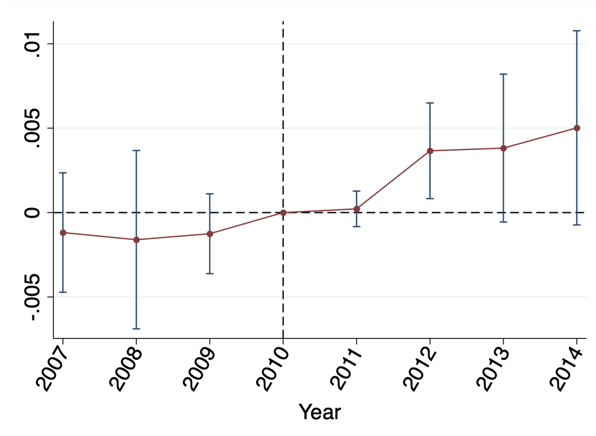
Figure H14. DRG ED Visit Share vs. Died Share



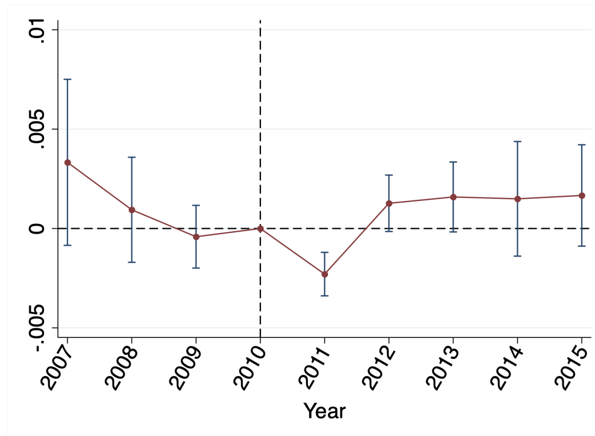
This figure plots a binscatter between the share of a DRG's admissions with ED charges and the share with a death within 30 days, among 2010 Medicare inpatient stays. Data: MEDPAR.

Figure H15. Event Studies on Effect of 2011 Audit Rate on Medicare ED Visits

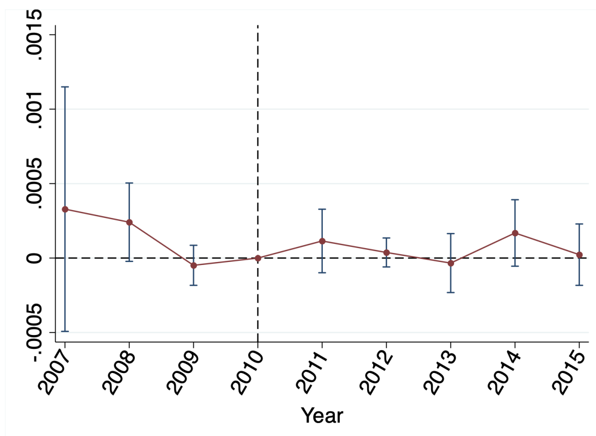
(a) Share with Outpatient Obs or "Suspected" Obs



(b) Share Revisit in 30 Days



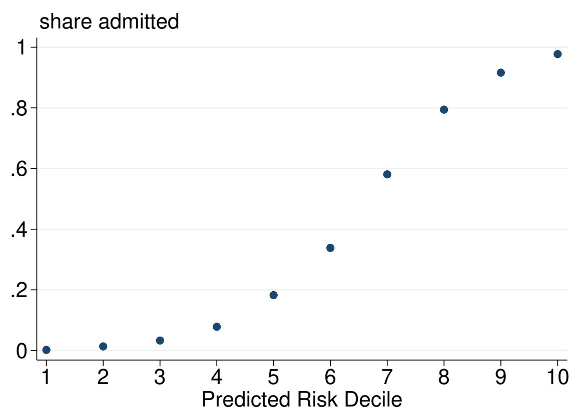
(c) Share Died in 30 Days



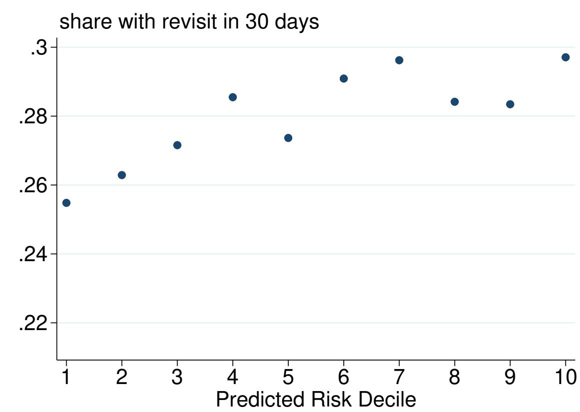
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 8, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. Panel (a) shows the share of Medicare ED visits that report outpatient observation payment or where the outpatient stay spans two days ("suspected" observation stay). The 2010 mean was 12%. Panel (b) shows the share of Medicare ED visits with a revisit to the ED within 30 days (2010 mean: 15%). Panel (c) shows the share of visits with a beneficiary death within 30 days of visit (2010 mean: 1.4%). A Medicare ED is defined as an inpatient or outpatient claim with ED charges. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, Outpatient file, Beneficiary file, and CMS audit data.

Figure H16. Average Outcomes by Patient Severity

(a) Inpatient

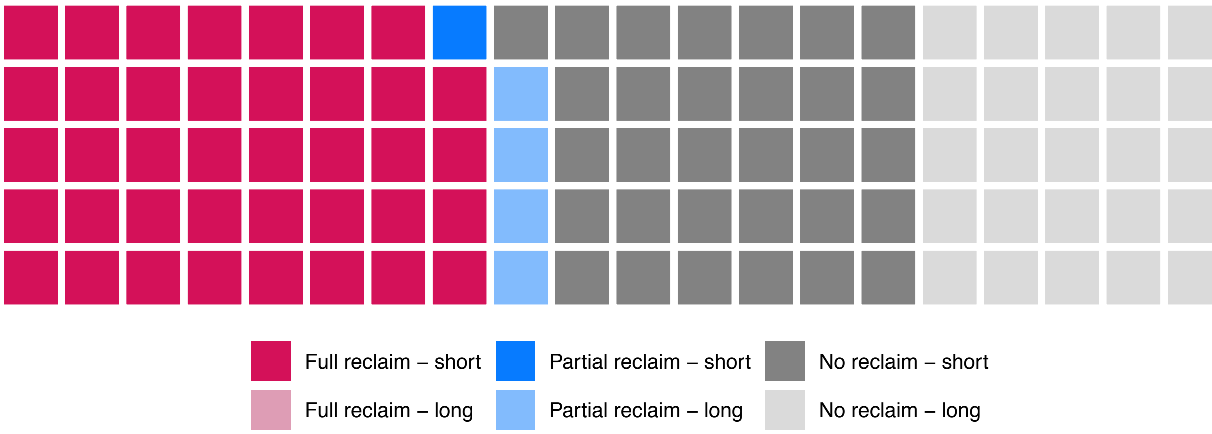


(b) Revisit in 30 days



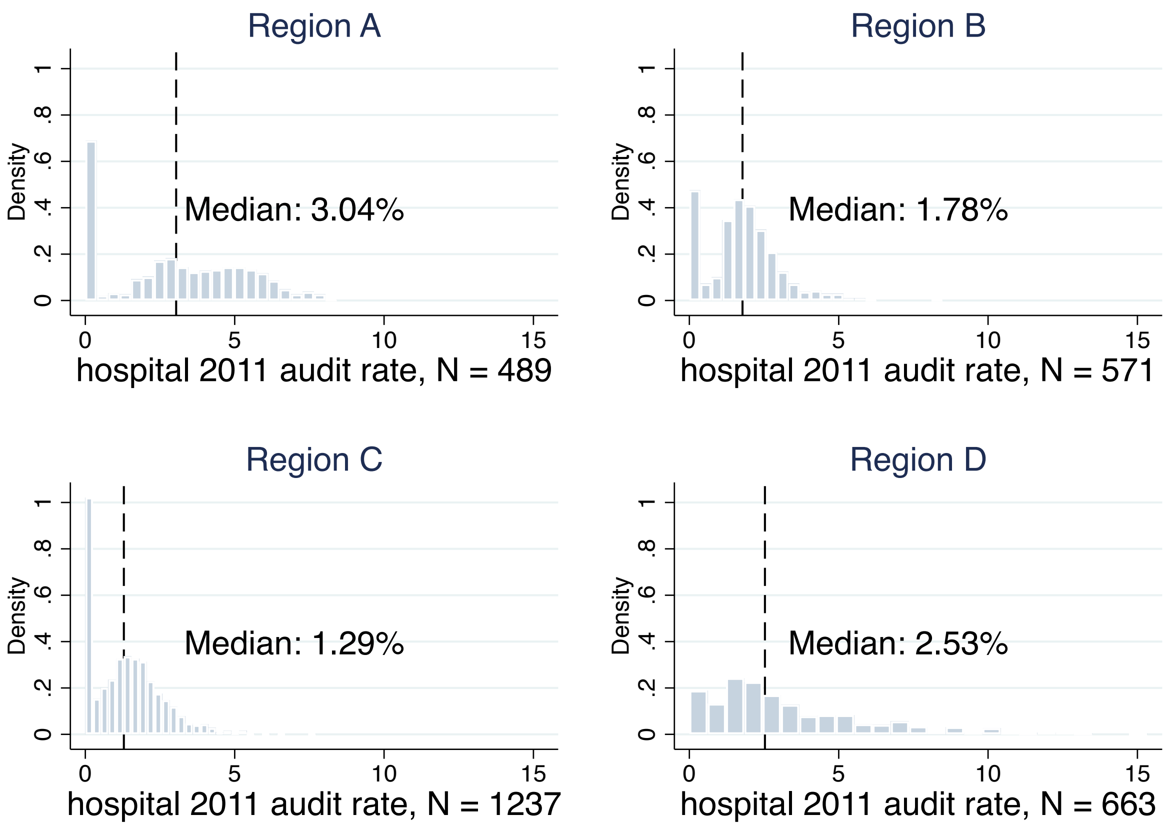
This figure plots (a) the share of patients admitted as inpatient from the ED and (b) the share of patients with a revisit within 30 days by predicted severity decile, in 2013Q2. Patient risk is predicted by estimating a logit using ED visits between 9:00AM and 3:00PM of an indicator for being admitted within 30 days of an ED visit on patient demographics, current ED visit information, and information on any prior visits in the last 365 days. Demographics include age-bin, sex, race, Hispanic indicator, point of origin indicator, and mean zip code income. Information on current visit includes hospital and quarter. Information on previous visits includes the number of visits/inpatient stays/length of stay in the last month or last year, as well as any diagnoses and procedures recorded in stays within the last month or last year. Data: HCUP SID/SEDD.

Figure H17. 2011 Audit and Denial Characteristics



This figure is a waffle plot of 2011 audits of inpatient stays in 2008-2011, where each box represents one percent of total audits. The dark shaded boxes of each color denote audits of inpatient stays. The red and blue colored boxes denote audits that result in the full payment being reclaimed or a partial payment being reclaimed, respectively. The figure plots the following shares of 2011 inpatient stay audits: 39 percent of audits are for short stays where the full payment is reclaimed, less than 1 percent of audits are for long stays where the full payment is reclaimed, one percent of audits are for short stays where a partial payment is reclaimed, 4 percent of audits are for long stays where a partial payment is reclaimed, 31 percent of audits are for short stays where there is no payment reclaimed, and 25 percent of audits are for long stays where there is no payment reclaimed. Data: MEDPAR and CMS audit data.

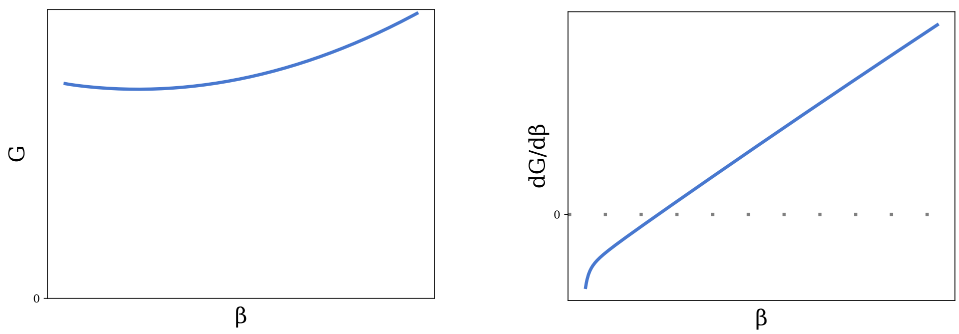
Figure H18. Histogram of 2011 Hospital Audit Rates by RAC Region



This figure plots the histogram of 2011 hospital audit rates by RAC region, where audit rate is defined as the percent of a hospital's 2008-2011 claims that were audited by RACs. Data: MEDPAR and CMS audit data.

Figure H19. Solving Medicare's Problem: Optimal β^*

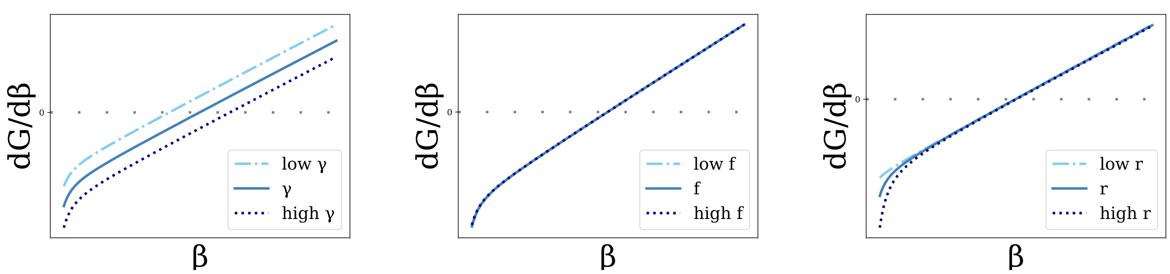
(a) G as function of β (b) $\frac{dG}{d\beta}$ as function of β



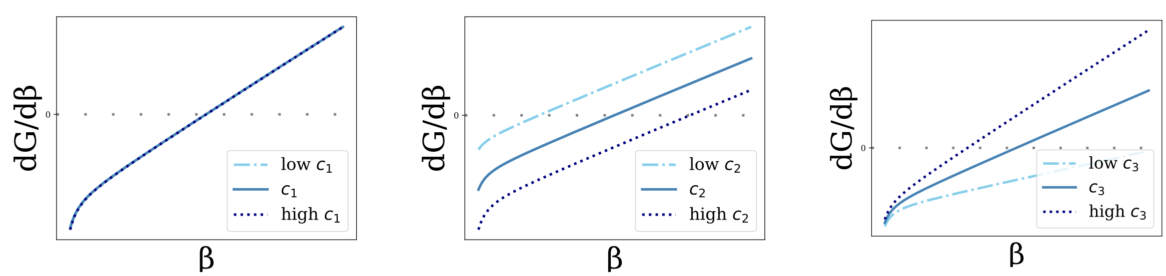
This figure plots the relationship between the (a) Medicare's payoff G and (b) the first order condition with respect to $\beta \in [0.01, 0.5]$. The remaining model parameters used are described in Table GXII.

Figure H20. Characterizing Optimal Audit Rate β^*

(a) γ (b) f (c) r

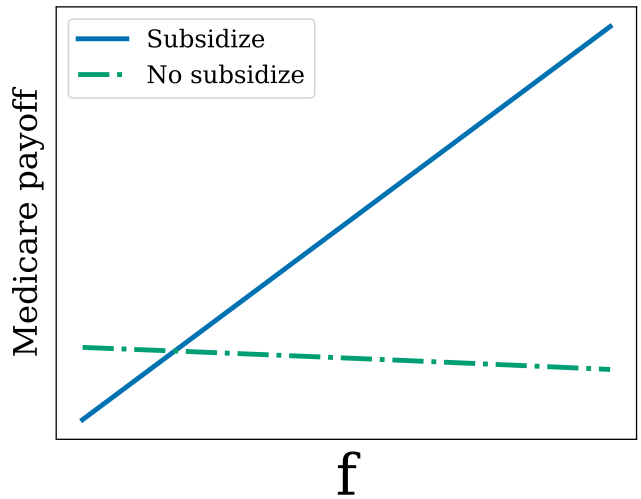


(d) c_1 (e) c_2 (f) c_3



This figure plots the relationship between $\frac{dG}{d\beta}$ and β, at different values of $\gamma \in \{0.4, 0.5, 0.6\}$, $f \in \{40000, 50000, 60000\}$, $r \in \{50, 100, 150\}$, $c_1 \in \{0.5, 1, 1.5\}$, $c_2 \in \{0.5, 1, 1.5\}$, $c_3 \in \{0.5, 1, 1.5\}$. The intersection of $\frac{dG}{d\beta}$ and 0 represents the optimal β^*. The remaining model parameters used are described in Table GXII.

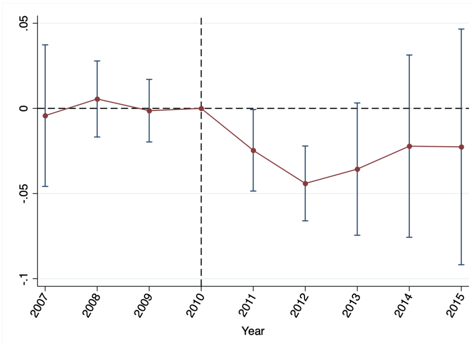
Figure H21. Solving Medicare's Subsidy Problem



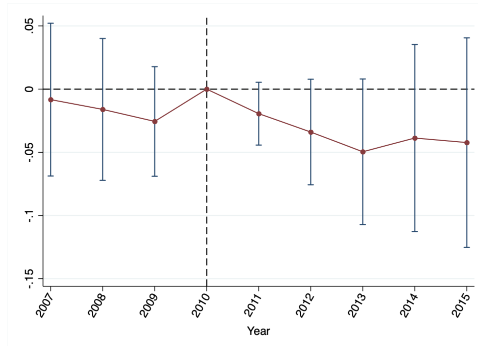
This figure plots the relationship between Medicare's payoff in each scenario and $f \in [1000, 60000]$, where a lower payoff is better. The model parameters used are described in Table GXII.

Figure H22. Event Studies on Effect of 2011 Denial Rate on Medicare Admissions and Revenue, and Administrative Burden

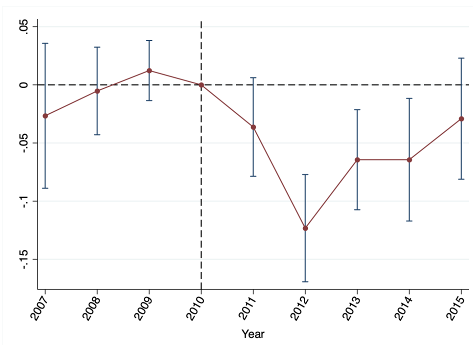
(a) Log Medicare admissions



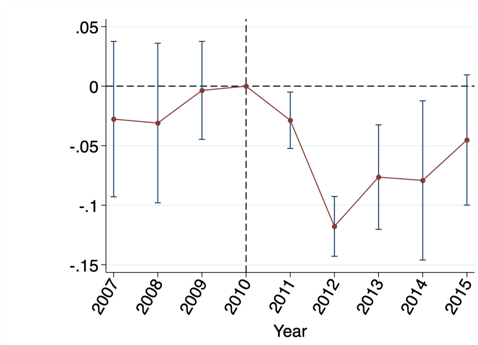
(b) Log Medicare inpatient revenue



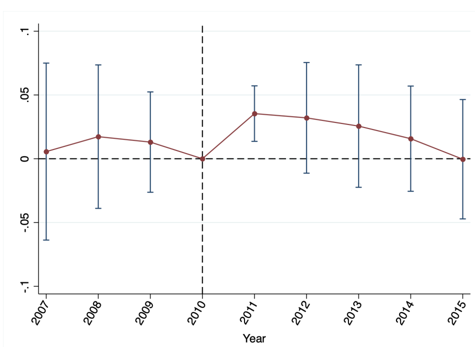
(c) Log Medicare admissions, LOS ≤ 2



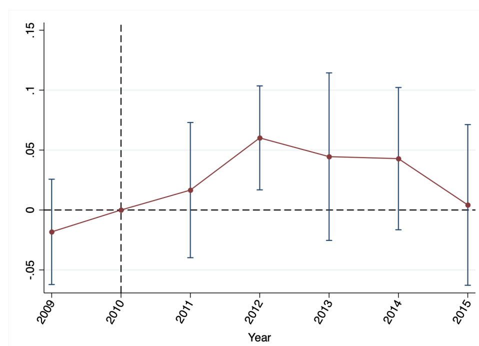
(d) Log Medicare inpatient revenue, LOS ≤ 2



(e) Log net administration costs

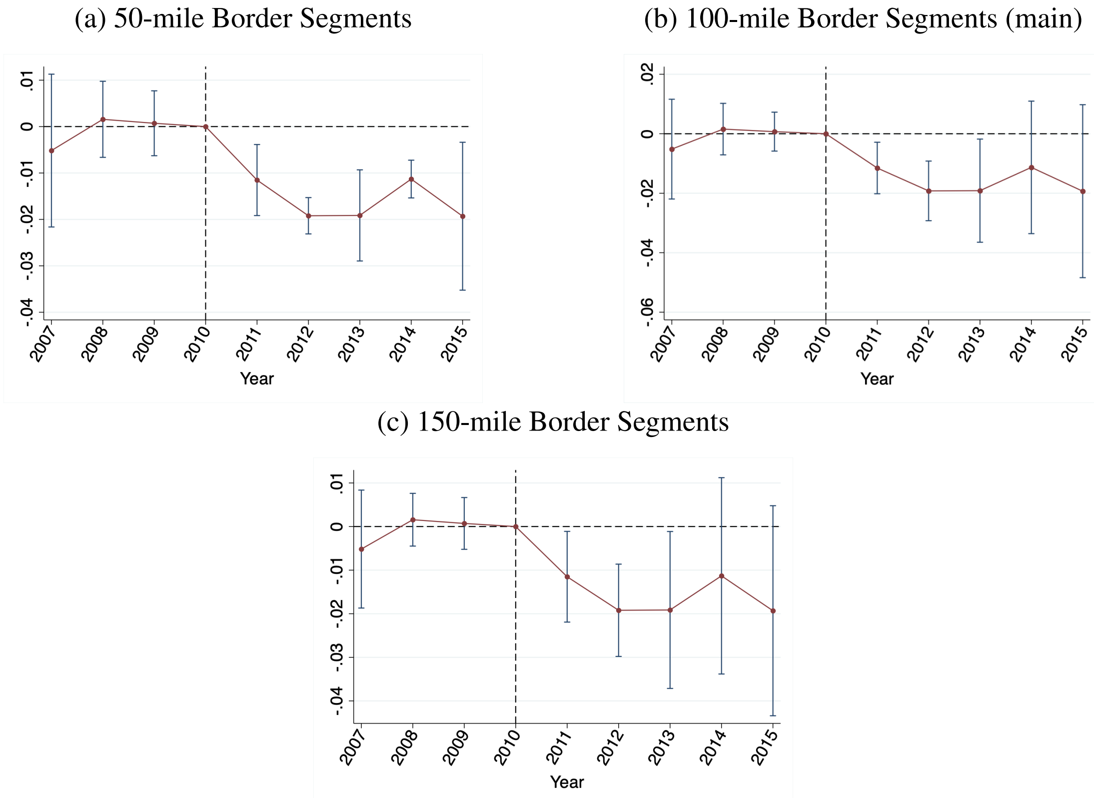


(f) Indicator for installing medical necessity software



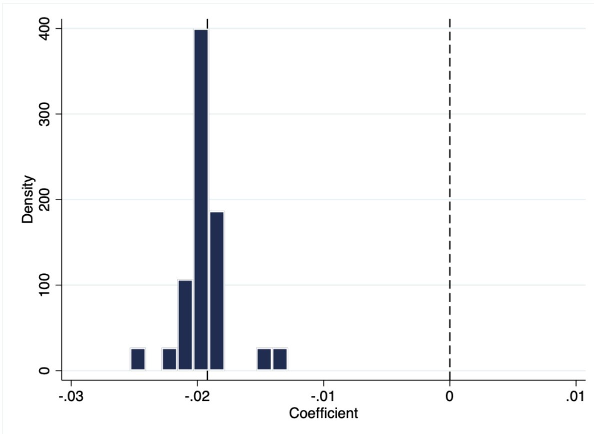
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 8 (using the denial rate rather than the audit rate), scaled by the correlation between the leave-one-out 2011 denial rate and the actual 2011 denial rate in the weighted border hospital sample. Denial rate is the share of claims that are audited and result in an overpayment demand or repayment for an underpayment. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 denial rate rate on a hospital-level outcome. Medicare admissions and revenue are from MEDPAR. Inpatient revenue is the sum of all Medicare inpatient payments. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in HIMSS. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, CMS audit data, HCRIS, HIMSS.

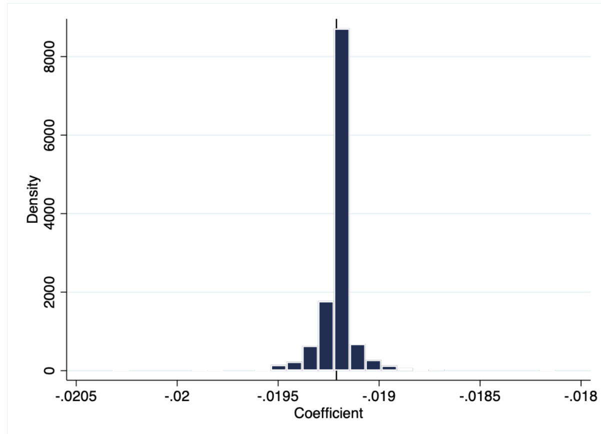
Figure H23. Robustness to Border Segment Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions



This figure plots robustness analysis event studies of the reduced form coefficients and 95% confidence intervals of the specification in Equation 8, scaled by the correlations between the instruments and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different border segment lengths used for clustering. The segments are defined such that they do not cross state lines. Panel (a) shows the results for 50-mile segments, (b) shows the main results for 100-mile segments, and (c) shows the main results for 150-mile segments. Data: MEDPAR and CMS audit data.

Figure H24. Robustness Test: Leave-One-Out Coefficients of 2012 Effect of 2011 Audit Rate on Log Medicare Admissions

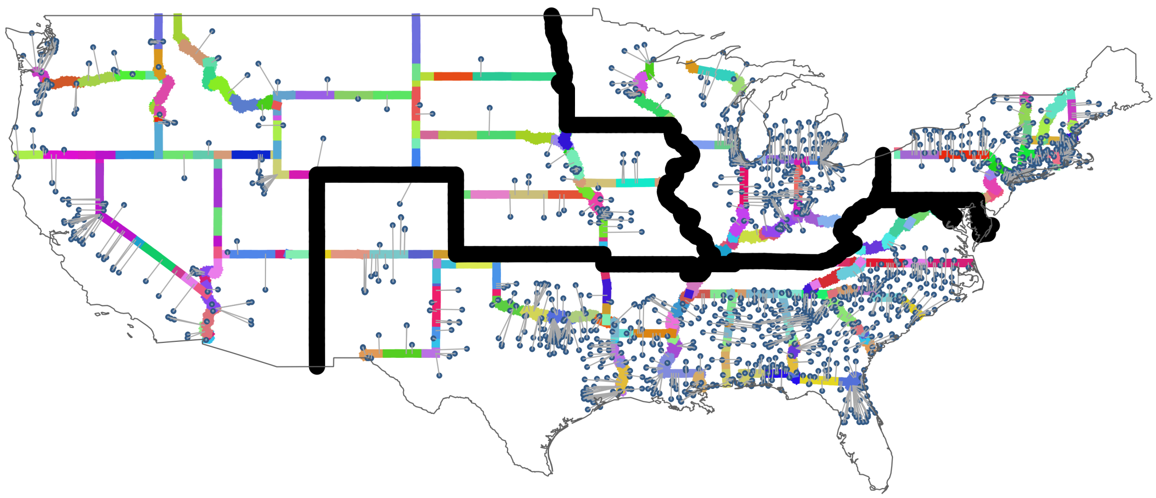




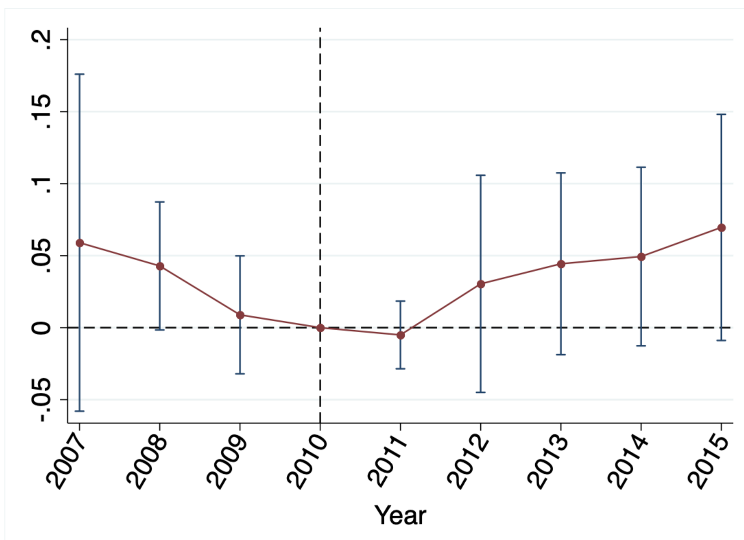
This figure plots distributions of the 2012 coefficient of the reduced form event study specification in Equation 8 on log Medicare admissions, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample the outcome. Panel (a) plots the distribution of the coefficient when leaving one state out at a time, and panel (b) plots the distribution of the coefficient when leaving one hospital neighbor comparison group out at a time. Data: MEDPAR and CMS audit data.

Figure H25. Falsification Test: Interior State Borders

(a) Falsification Test Border Segments and Hospitals Within 100 Miles



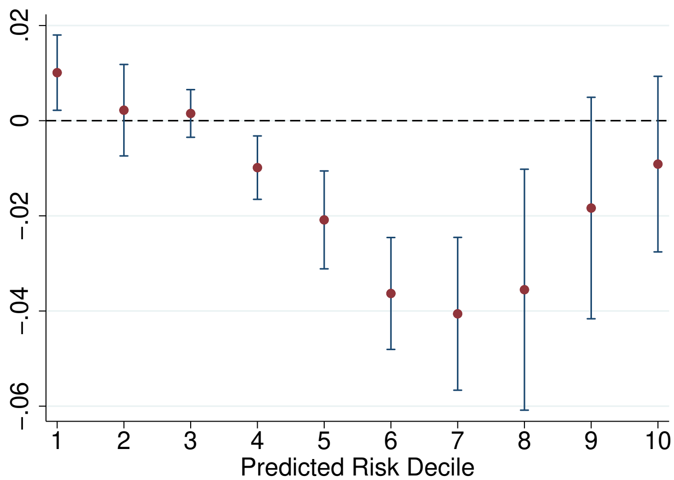
(b) Event Study on Effect of 2011 Audit Rate on Log Medicare Admissions



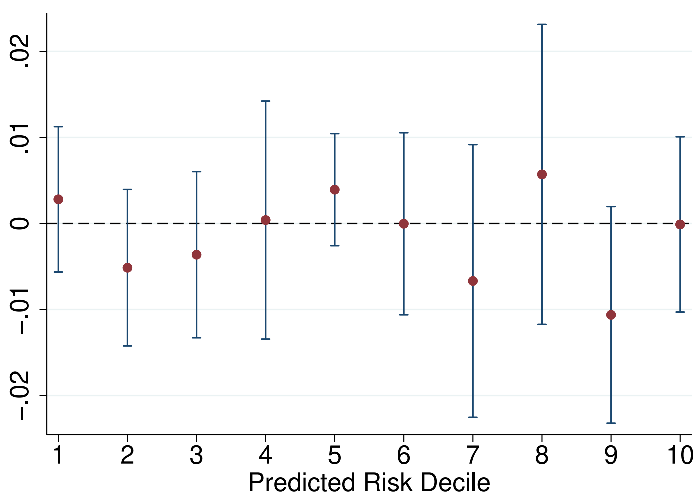
The top panel of this figure plots a map of state borders on the interior of RAC regions, divided into 100-mile segments that do not cross state borders. The RAC border is the thick black line. Each dot represents a hospital within 100 miles of the interior state borders, excluding hospitals that are in the main sample (within 100 miles of the RAC border). The line between the hospital and the interior state border denotes the closest interior state border to that hospital. The bottom panel plots the reduced form coefficient and 95% confidence interval of the specification in Equation 8 (scaled by correlation between 2011 audit rate and 2011 leave-one-out audit rate in the interior border hospital sample), where the outcome variable is log Medicare admissions (MEDPAR). Sample is comprised of hospitals within 100 miles of the state interior border with at least 1 hospital in their "neighbor hospital comparison group" and are clustered at the state and border segment level. Data: MEDPAR and CMS audit data.

Figure H26. Heterogeneity of After-Midnight ED Arrival Coefficient by Patient Severity (Trained on Before-Midnight Arrivals Pre-Policy)

(a) Inpatient

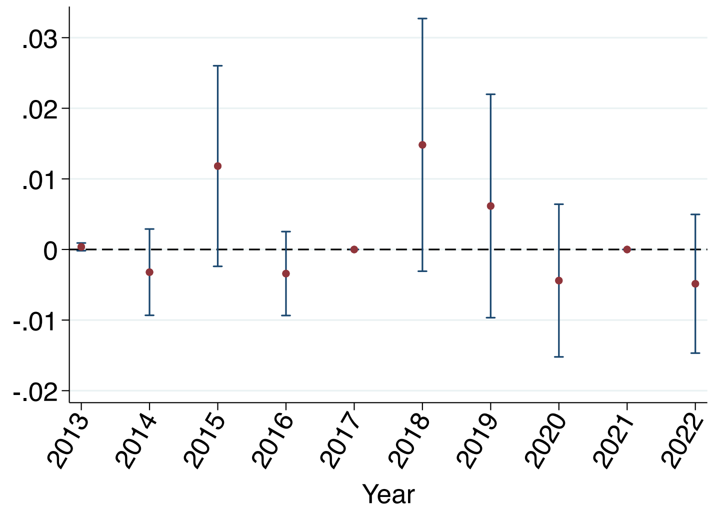


(b) Revisit in 30 days



This figure plots estimates and 95% confidence intervals of the β coefficient in Equation 12, interacted with an indicator for predicted severity decile. β is the coefficient on $\mathbb{1}[q \geq 2013Q3] \times \mathbb{1}[T_v \geq 00{:}00]$, where $\mathbb{1}[q \geq 2013Q3]$ is an indicator for whether the visit occurred after 2013Q3, and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. The top panel plots results for an indicator for whether the patient was admitted as inpatient from the ED, and the bottom panel plots results for an indicator for whether the patient revisited any hospital in Florida within 30 days of the ED visit. The results are clustered at the ED arrival hour and quarter level. Patient risk is predicted by estimating a logit using before-midnight ED visits (within 3 hours of midnight) that occurred prior to 2013Q3 on an indicator for being admitted within 30 days of an ED visit on patient demographics, current ED visit information, and information on any prior visits in the last 365 days. Demographics include age-bin, sex, race, Hispanic indicator, point of origin indicator, and mean zip code income. Information on current visit includes hospital, quarter, and the AHRQ CCS category for the patient's first diagnosis code. Information on previous visits includes the number of visits/inpatient stays/length of stay in the last month or last year, as well as any diagnoses and procedures recorded in stays within the last month or last year. Figure H16 plots the mean outcomes for each decile. Data: HCUP SID/SEDD.

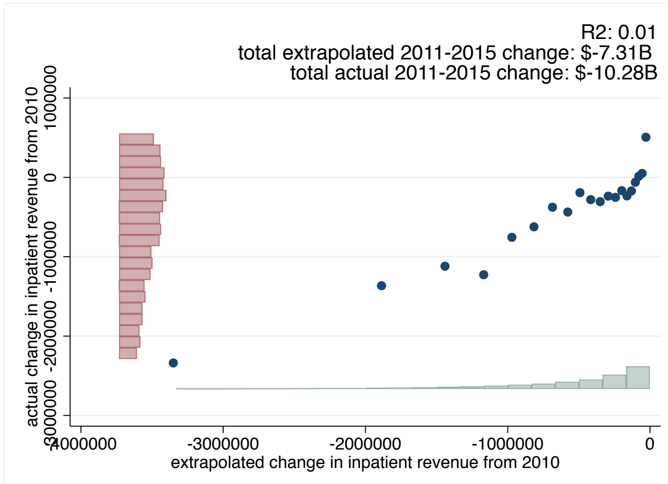
Figure H27. Coefficients of Effect of 2011 Audit Rate on Rural Hospital Closure in a Given Year



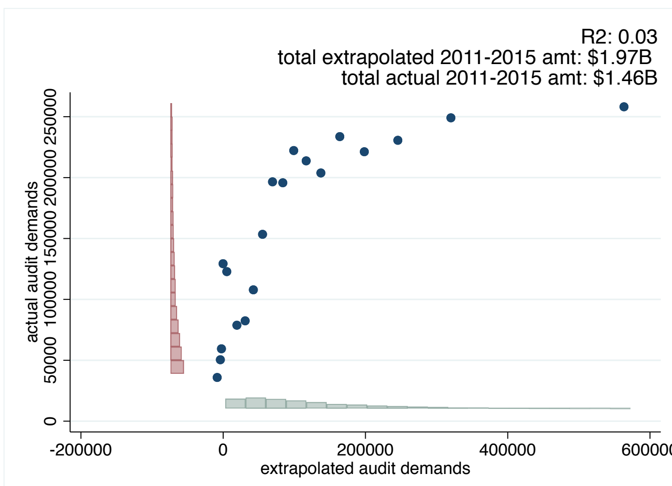
This figure plots the coefficients from individual regressions of the instrumented 2011 audit rate on a dummy for whether a hospital closed in a given year, for rural hospitals in the border sample. There are no closures prior to 2013 and no closures in 2017 and 2021 in the border hospital sample. Data: Sheps Center for Health Services Research and CMS audit data.

Figure H28. Extrapolation Exercise: Actual vs. Extrapolated Savings

(a) Savings from changes in Medicare inpatient revenue



(b) Savings from audit demands



This figure plots binscatters of the actual versus extrapolated savings between 2011 and 2015 from (a) the reductions in Medicare inpatient revenue and (b) the payments demanded from audits. Actual changes in Medicare inpatient revenue are calculated by subtracting a hospital's revenue in a given year (between 2011 and 2015) from its 2010 revenue. Actual audit demands are calculated using the RAC audit data, and adjusted for refunds to hospitals due to the lawsuit over appeals described in Section A.2. Each observation is a hospital-year. Section E describes in further detail how the extrapolated changes in Medicare inpatient revenue and audit demands are calculated. The sample is winsorized at the 99th percentile of actual changes in Medicare inpatient revenue.