MONITORING FOR WASTE: EVIDENCE FROM MEDICARE AUDITS *

Maggie Shi†

September 9, 2023

Abstract

This paper examines the tradeoffs of monitoring for wasteful public spending. By penalizing unnecessary spending, monitoring improves the quality of public expenditure and incentivizes firms to invest in compliance technology. 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. I do not find evidence that the health of the marginal patient is harmed, indicating that monitoring primarily deters low-value care. Monitoring does increase provider administrative costs, but these costs are mostly incurred upfront and include investments in technology to assess the medical necessity of care.

JEL: H51, H83, I00, I13, I18, M42, M48

*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, Michael Ricks, Jeff Clemens, Katherine Meckel, Zarek Brot-Goldberg, Ashvin Gandhi, 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.

†University of Chicago Harris School of Public Policy and NBER. Email: m.shi@uchicago.edu. Mailing address: 1307 E. 60th Street, Chicago, IL 60637.

1 Introduction

Combating waste is a perennial problem for public programs. The Office of Management and Budget (2022) estimated that over seven percent of U.S. federal spending is wasted. 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). 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). However, policymakers may be wary of monitoring too aggressively because it is unclear whether it can successfully reduce waste or if it just introduces needless regulatory costs. But despite the importance of this question, there is little empirical evidence on the magnitude and nature of the tradeoffs associated with monitoring for waste in public spending.

This paper considers these tradeoffs in the context of Medicare, the federal health 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: estimates suggest that 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 health care becomes increasingly digitized, there has been significant progress in the development of technology to improve the efficiency of healthcare delivery (Hillestad et al., 2005). Policymakers have devoted billions of dollars in recent years to subsidize the adoption of this technology, with the hopes of reducing healthcare costs (Burde, 2011; Atasoy et al., 2019). Monitoring could therefore also serve as an additional policy lever to incentivize providers to seek out new ways to improve the cost-effectiveness of their spending.

On the other hand, the social costs of excessive oversight may be high here as well. Poorly targeted responses to monitoring could have dire implications for patient health. Pressuring providers

[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.

to cut back spending could deter necessary care, especially if providers are initially unsure about what is medically necessary for a patient (Doyle et al., 2015). Given the complexity of identifying unnecessary care, monitoring could also impose considerable compliance costs on providers. If these costs stem mostly from the "back and forth" of the monitoring process, this would simply add to providers' already-high administrative burden (Cutler and Ly, 2011; Dunn et al., 2020). But this is less of a concern if the costs stem from investments made to improve cost-effectiveness. Thus, the extent to which Medicare should monitor for wasteful spending depends on the balance between the savings from reducing unnecessary care and the nature of the costs imposed on patients and providers.

I study this question in the context of 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 lens for examining the effects 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 Medicare 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 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 annual hospital cost reports and surveys.

To motivate the empirical analysis, I consider a model of hospital behavior and Medicare audits to understand how monitoring affects admissions and technology adoption. Hospitals assess whether patients need to be admitted by observing a noisy signal of each patient's benefit from

[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).

admission. They set an admission threshold and admit patients whose signals are above the threshold. Thus the threshold determines how many patients the hospital expects to admit. Medicare reimburses hospitals for admissions but also conducts audits to uncover and penalize admissions with low true benefit. In setting the admission threshold, hospitals trade off the changes in patient benefit, which they value inherently because they are partially altruistic, with changes in reimbursement, treatment costs, and expected audit penalties. Prior to setting their threshold, hospitals can purchase technology that improves their ability to assess patient need by reducing the noise in their patient benefit signal. Adopting technology is costly but increases hospitals' payoff from admissions. Hospitals adopt only if the gains to doing so are greater than the fixed adoption cost. The model illustrates how auditing can shape hospital behavior both directly, by lowering the return to the marginal admission, but also indirectly, by increasing the return to investments in diagnostic ability. As a result, increasing the audit rate can change both the quantity and quality of hospital admissions.

I then examine the effects of monitoring on hospital behavior and patient outcomes in the data and arrive at three core empirical 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. Ninety percent of these savings stem from the deterrence of future spending, rather than the recovery of prior spending. Second, monitoring primarily deters low-value admissions. Hospitals are less likely to admit patients with higher audit risk, but 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. Most of the administrative costs hospitals incur can be attributed to such upfront costs rather than ongoing hassle costs. Taken together, the results show that monitoring providers reduces unnecessary care, and one way it does so is by incentivizing providers to adopt technology to improve their diagnostic ability.

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 that compares hospitals subject to differentially aggressive RACs, and another that compares patient cohorts who face exogenously different audit likelihoods.

To understand how hospitals respond 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 that are neighbors to each other but who are subject to different RACs, leveraging sharp differences in auditing between different RAC jurisdictions. Hospitals subject to a more-aggressive RAC reduce their admissions more -– 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. Hospitals scale back mostly on short stays and stays with diagnoses associated with high Medicare payment error rates. Among these high-error diagnoses, both emergent and non-emergent admissions decrease. 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 adopt technology in order to identify which patients to no longer admit. Hospitals subject to more audits are 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).[3] 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 at the onset of the program expansion in 2011. This suggests that provider compliance costs comprise mostly of the fixed costs from investments like technology adoption,

[3] Specifically, medical necessity checking software is a type of clinical decision support technology which "provides clinicians, staff, patients or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care... [including] diagnostic support, and contextually relevant reference information" (Office of the National Coordinator for Health Information Technology, 2018).

rather than the ongoing hassle costs of interacting with the RACs.

I then turn to the question of the effect on patient health – did the reductions stem mostly from unnecessary admissions? Because patient composition changes as hospital volume decreases, it is challenging to compare patient outcomes across hospitals. To address this, I identify a set of patients within a hospital who are likely to be marginal admissions: those arriving in the ED whose audit risk depends on an arbitrary threshold rule. In particular, I consider a rule which generated exogenous variation in audit risk across ED patients in the same hospital: the "Two Midnights rule." This rule was implemented in 2013 and 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. I 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. However, 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 admission 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 this software is employed.

Compared to the large literature studying tax enforcement on the revenue side, there is less work looking at monitoring for waste on the expenditure side.[4] This is in spite of the fact that governments conduct a considerable amount of this kind of monitoring. In the U.S., there are sev-

[4]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).

eral public entities solely devoted to uncovering waste in public spending, including the Offices of Inspector General and the Government Accountability Office. Monitoring for wasteful spending is likely understudied because what constitutes "waste" is often ambiguously defined and notoriously difficult to measure.[5] This paper fills this gap in the literature by using patient health outcomes as a measure of spending quality in the healthcare setting.

Given that policymakers only considered the recovered payments when assessing the cost-effectiveness of the RAC program, the large deterrence effect that I find effect is particularly striking (Centers for Medicare and Medicaid Services, 2012). Though deterrence plays a central role in economic theories of enforcement, in practice the evaluations of these policies often focus only on measuring the direct effects (Becker, 1968; Allingham and Sandmo, 1972). For example, the reports to Congress submitted by the Offices of Inspector General of various federal agencies only list the wasteful spending directly uncovered through their investigations.[6] This paper contributes to a broader empirical literature spanning criminal enforcement, tax compliance, and litigation which has also demonstrated sizeable deterrence effects.[7] Together these results underline the importance of incorporating measures of deterrence into cost-effectiveness evaluations.

The RAC program also serves as a useful context for studying how monitoring combats waste that arises in part due to unintentional errors. Rather than being the result of deliberate fraud, these may be errors that are simply less costly to ignore rather than to correct, like admitting a patient who ends up not needing it. Even if providers do not intend *ex ante* to deliver unnecessary care, assessing patient health needs is a complicated task and providers often make mistakes in assessing patient need (Chan and Gruber, 2020). At baseline, hospitals may not have sufficient incentive to root out low-value admissions if they are still reimbursed for them. I show that monitoring can incentivize investments to correct these errors. By penalizing low-value care, RAC audits motivate

[5]For example in Olken (2007), measuring wasteful spending in public infrastructure projects required assembling teams to take core samples from roads and then comparing the reported and actual amounts of construction material used.

[6]Link to reports: https://www.oversight.gov/reports (last accessed July 2023)

[7]Leder-Luis (2023) finds a 7-to-1 ratio of deterred spending to settlement funds in whistleblower lawsuits for fraudulent public spending, Kleven et al. (2011) finds a deterrence effect of 42 cents per dollar of adjustment for tax audits, and Di Tella and Schargrodsky (2004) find a deterrence effect of 20 cents per dollar spent on additional police presence via fewer car thefts.

hospitals to make costly improvements to their admissions process, such as installing medical necessity checking software.

A similar dynamic arises in other enforcement contexts as well. Given the complexity of the tax code, some under-reporting of tax liability may be the result of taxpayers making genuine mistakes rather than attempts to evade (Kopczuk, 2007). Increasing the threat of audit can incentivize taxpayers to purchase e-filing software or to hire an accountant to catch these mistakes. In the WIC and SNAP programs, transitioning retailers from paper vouchers to an electronic benefit system can increase program integrity by flagging price discrepancies and reducing the distribution of ineligible products, both of which may be unintentional (Meckel, 2020). Greater monitoring of retailers can incentivize them to adopt electronic cash registers, which mitigate these issues by recording transactions with product IDs and prices.[8] Broadly speaking, compliance technology can correct errors that an individual or firm may have previously turned a blind eye to. Thus if some of the private costs associated with monitoring stem from these kinds of investments, then the compliance costs of monitoring may not all be deadweight loss.

This paper also provides direct measures of the various social costs monitoring 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.[9] The hospital setting is a unique context where two forms of these costs – provider administrative costs and patient health outcomes – can be observed more readily.

Finally, these results shed further light on how healthcare providers respond to incentives. It has been well-documented that providers respond to financial incentives, either by changing what care they provide or how they document this care.[10] In contrast, less is known about how providers

[8]Interestingly, WIC/SNAP presents a case in which policymakers deemed it worthwhile to purchase the technology on behalf of retailers. As discussed in Meckel (2020), when Texas was transitioning to fully electronic dispensing of WIC/SNAP benefits, the state reimbursed retailers without electronic cash registers for the full installation and maintenance costs.

[9]Recent examples include Kopczuk and Pop-Eleches (2007); Deshpande and Li (2019); Finkelstein and Notowidigdo (2019); Meckel (2020); Zwick (2021); Dunn et al. (2023).

[10]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

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 (Dunn et al., 2023; League, 2022), fraud enforcement (Nicholas et al., 2020; Howard and McCarthy, 2021; Leder-Luis, 2023), and prior authorization (Roberts et al., 2021; Brot-Goldberg et al., 2023).

The rest of the paper proceeds as follows. Section 2 describes the policy context of the RAC program. Section 3 describes the model. 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,[11] the payment rate depends on 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

latter include Silverman and Skinner (2004); Dafny (2005); Sacarny (2018)

[11]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).

that of longer stays. Indeed, economists have long pointed this out as a potential vulnerability to a prospective payment system.[12] 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.[13] 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).[14]

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. The medical professional then reviews the documentation and determines whether Medicare made a payment error. Fifty-seven percent of manual reviews conducted in 2011 resulted in no finding, 37 resulted in an overpayment determination, and six percent resulted in an underpayment determination.

Once the complex review is finished, RACs send a letter to providers that outlines whether a

[12] As Ellis and McGuire (1986) write, "a prospective payment system employs incentives to reduce treatment, except at one critical point: the decision to admit a patient. Questionable or low-value admissions are, in fact, especially profitable for hospitals."

[13] The RAC regions are also used by Durable Medical Equipment Medicare Administrative Contractors (MAC), 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. The Part B Medicare Administrative Contractors in charge of processing and denying medical claims use smaller regions, some of which share boundaries with RAC regions (League, 2022). In Section C.1 I conduct placebo tests on MAC borders in the interior of each RAC region.

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

payment error was identified, the amount of overpayment demanded or underpayment refunded, and references supporting the decision. 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. Providers can appeal demands by requesting redetermination by the RAC and then escalating it to higher levels of 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 improve their admissions process to mitigate 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 spending. The top sources of spending were training and education programs and tracking software purchases (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. This software informs providers 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. Vendor marketing materials point to the ability to provide information in real time as a key feature of the software, suggesting that it is most relevant in emergent cases (3M, 2016; Experian Health, 2022).[15] Adoption of health IT like this is often touted as a way to reduce wasteful healthcare spending – in 2009, Congress passed the HITECH Act and devoted almost $30 billion to subsidizing health IT adoption with the explicit goal of improving cost-effectiveness (Burde, 2011; Dranove et al., 2014).

I also leverage an additional policy within the RAC program which generated differences in audit risk across patients within a hospital. 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.[16] 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 Model of Hospital Admissions and Technology Adoption

Consider a model of hospital behavior in which hospitals make two decisions: how high to set their threshold for admission and whether to invest in technology to better evaluate the medical necessity of each admission. For each patient, hospitals observe a noisy signal of their benefit

[15] For example, the Experian product sheet (link; last accessed July 2023) offers to "prevent claim denials with access to timely and updated medical necessity content," "improve cash flow by proactively identify procedures that may fail medical necessity," and "help protect [the provider] from regulatory fines by staying compliant with Medicare regulations and policies."

[16] 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).

from admission and admit them if this signal is above an established threshold. In choosing this threshold, hospitals trade off the gains from more admissions — the additional patient benefit and revenue – with the cost of these admissions — the treatment cost and audit penalties for low-benefit admissions. Hospitals can improve the quality of their patient benefit signal by adopting medical necessity checking technology at a fixed cost. Having a more precise signal allows them to screen better on patient benefit, which they value both because they are partially altruistic but also because they want to avoid audit penalties for *ex-post* low-benefit stays. Whether or not a hospital adopts comes down to whether the gains from adoption are larger than the fixed cost of investment.

The model delivers two predictions about how hospitals would respond to an increase in the audit rate. First, by raising the marginal cost of each admission, increased auditing leads hospitals to raise their admission threshold and thus reduce admissions. The deterred admissions are more likely to be low-benefit ones. Second, if the increase in audit penalties is smaller for hospitals who have adopted technology, then the value of adoption rises, leading more hospitals to adopt. I characterize the solution by backward induction, beginning with the admission threshold decision while holding technology fixed, and then moving on to the technology adoption decision.

Admission Threshold Patients are characterized by their true benefit from admission x, where x is drawn from a known distribution F. The hospital cannot directly observe x but instead observes a noisy signal: $y = x + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$. It uses a threshold rule to decide who to admit – under threshold rule τ, it will admit all patients with $y \geq \tau$. Let the likelihood that a patient with benefit value x is admitted under threshold τ be $P(x;\tau) = 1 - \Phi\left(\frac{\tau - x}{\sigma}\right)$. Then the hospital expects to admit $q(\tau) = \int_{-\infty}^{\infty} P(x;\tau)dF(x)$ patients, which is decreasing in τ: the lower the threshold, the more patients the hospital expects to admit. The hospital chooses τ to maximize its expected payoff, which is an additively separable function of the patient benefit and hospital profit from admissions:

$$\underbrace{E\left[U(\tau)\right]}_{\substack{\text{expected payoff}\\ \text{with threshold } \tau}} = \underbrace{\alpha B(\tau)}_{\substack{\text{value of total}\\ \text{patient benefit}}} + \underbrace{\underbrace{Rq(\tau)}_{\text{reimbursement}} - \underbrace{\frac{1}{2}Cq(\tau)^2}_{\substack{\text{treatment}\\ \text{costs}}} - \underbrace{\gamma\pi(\tau)}_{\substack{\text{audit penalties}\\ \text{for } x < \underline{h}}}}_{\text{hospital profit}} \tag{1}$$

Patient benefit enters the payoff because hospitals are partially altruistic. Integrating over the distribution of patient benefit, the expected total benefit is $B(\tau) = \int_{-\infty}^{\infty} xP(x;\tau)dF(x)$. α represents the hospital's marginal rate of substitution between patient benefit and profit – a hospital with higher α places relatively greater value on patient benefit (Ellis and McGuire, 1986).

Hospital profit has three components: reimbursement, treatment costs, and Medicare audit penalties. Medicare pays hospitals a constant reimbursement rate per admission and treatment costs are weakly convex in the number of admissions. If an admission is audited, Medicare observes the patient's true benefit and will penalize the hospital if the audit reveals that it was a low-benefit admission. Medicare defines a low-benefit admission as one with x below some threshold $\underline{h}$. The expected audit penalty for a low-benefit admission is $\gamma > 0$, which captures a combination of the audit rate and the penalty conditional on Medicare discovering it is low-benefit.[17] Thus the expected total audit penalty is just γ multiplied by the share of patients expected to have true benefit below $\underline{h}$: $\gamma\pi(\tau) = \gamma \int_{-\infty}^{\underline{h}} P(x;\tau)dF(x)$.[18]

The hospital chooses a threshold τ^* to maximize the expected payoff in Equation 1. Figure 2a illustrates the marginal and expected total payoff at τ^*. Because the hospital admits patients with a signal higher than τ^*, the expected payoff from these admissions is represented by the area to the right of τ^*, *above* the marginal payoff curve. This can be conceptualized as hospitals starting at a high threshold and then lowering it to admit more patients until the marginal benefit of lowering it further is equal to the marginal cost.

Model Prediction 1. *Holding fixed a hospital's technology decision, increased auditing reduces*

[17] For simplicity, I assume γ is a constant so all admissions are equally likely to be audited, and conditional on being penalized, receive the same penalty. This could be extended to allow for either the penalty or the audit likelihood to depend on the signal, the true benefit, or the difference between the two.

[18] Note that this specification assumes that undergoing audits is costless to the hospital. This could be extended by separating γ into an audit rate β and penalty ψ, and then adding a term into Equation 1 to capture the hassle cost associated with being audited, such as $-\frac{C_H}{2}(\beta q(\tau))^2$.

admissions and the decline will be more pronounced for low-benefit admissions.

Figure 2b depicts the effect of increasing the audit rate on the admission threshold. Under a higher audit rate, the hospital makes fewer, but higher-quality, admissions. As shown in Section B.1, the payoff of raising the admission threshold rises as the expected penalty γ increases. As the hospital raises its threshold, it admits fewer patients. However, the quality of the remaining admissions is higher – the reduction in admission likelihood is smaller for high-benefit admissions than it is for low-benefit ones.

Signal Quality and Technology Adoption Figure 2c depicts the effect of reducing the variance of the signal on the hospital's marginal and total payoffs. As the variance σ^2 decreases, the slope of the marginal payoff curve steepens, making the hospital's payoff more elastic with respect to τ. With a more precise signal, the hospital's ability to screen based on its chosen threshold improves. Section B.1 shows that by Blackwell's informativeness theorem, reducing the noisiness of the benefit signal increases the hospital's expected payoff (Blackwell, 1951, 1953). However, note that the effect on total patient benefit is ambiguous. Since $B(\tau)$ is decreasing in threshold τ, overall patient benefit will only increase if reducing signal noise leads the hospital to lower its admission threshold.

Prior to setting the admission threshold, the hospital can choose whether to adopt technology to reduce the variance of its signal from σ_H^2 to σ_L^2. By altering the hospital's expected payoff curve, technology adoption can change both the quantity and quality of admissions, as illustrated in Figure 2c. If the technology was free, then all hospitals would choose to adopt because the expected payoff is greater under a more informative signal. But if technology is costly to adopt, then the adoption decision becomes a threshold rule where a hospital adopts only if the gains from adoption are greater than the cost of the investment.

Model Prediction 2. *If technology reduces audit penalties and hospitals face a distribution of adoption costs, then increasing the audit rate leads to more technology adoption.*

A hospital will adopt technology that reduces signal noise from σ_H^2 to σ_L^2 if the cost to adopt is less than the difference between the expected payoffs with and without technology, denoted below as K. If the difference between the payoffs increases with the audit rate, then increased auditing leads to more adoption.

$$\underbrace{K}_{\substack{\text{threshold}\\\text{adoption cost}}} = \underbrace{\max_{\tau} E[U(\tau;\sigma_L)]}_{\text{payoff with tech}} - \underbrace{\max_{\tau} E[U(\tau;\sigma_H)]}_{\text{payoff without tech}}. \tag{2}$$

In other words, hospitals will respond to audits by adopting technology if $\frac{dK}{d\gamma}$ is positive. As shown in Appendix Section B.1, the sign of $\frac{dK}{d\gamma}$ depends on the difference in audit penalties with and without technology. If hospitals with technology face lower penalties, then $\frac{dK}{d\gamma} > 0$ so adoption is increasing with the audit rate. A sufficient condition for this to be true would be if Medicare has a high audit penalty threshold $\underline{h}$ and the admission thresholds the hospital chooses with and without technology are relatively close to each other.

Appendix Section B.2 further extends the model to incorporate Medicare's problem of setting the audit rate. I also consider the conditions under which Medicare would choose to directly purchase the technology on behalf of hospitals, rather than indirectly encouraging adoption by conducting costly audits.

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 (and 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 the Medicare Outpatient claims and the Master Beneficiary file to assess ED visit outcomes in Section E.

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 H2). Appendix Section A.3 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.[19] The most granular unit of time for ED arrival in my data is the hour. SEDD includes 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.[20] 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), inpatient stays admitted from the ED in Florida (SID/SEDD), and patients admitted from a Florida ED who arrived at the ED within 3 hours of midnight (SID/SEDD). 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 3 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, are less likely to be white or female, and are sicker (i.e., more chronic conditions, 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.

[19] 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 using the Medicare claims will contain some degree of mis-measurement (ResDAC, 2022).

[20] 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.

4.2 Identifying the Effect of Monitoring on Hospital Outcomes

The aim of the hospital-level identification strategy is to understand how hospital behavior responds to audits. To understand the causal effect of auditing, I 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 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 remaining claims eligible for audit. 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.

To address concerns about spatially correlated patterns of hospital behavior, 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 consider 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, rather than that all hospitals in different regions 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."[21] 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 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. Within neighbor comparison groups, the 2011 audit rate is uncorrelated or weakly correlated with 2010 hospital characteristics for the border hospital sample. In contrast, these correlations are statistically significant and much larger in magnitude in the overall sample, further supporting the rationale to focus on border hospitals.

Figure H3 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 of 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 group-year fixed effects improves upon a specification with just border 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

[21] 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 comparison counties.

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 H4 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:

$$Y_{ht} = \sum_{\tau=2007}^{2015} \beta^{\tau} \mathbb{1}[t=\tau] \times X_h^{2011} + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \, . \tag{3}$$

In Equation 3, 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. To allow for dynamic responses, the main results are presented in the form of an event study with a β^{τ} for each year τ between 2007 and 2015, omitting 2010. Recall that RAC audits occurred not just in 2011 but throughout 2011-2015 (Figure 1b), so subsequent auditing may be endogenous to the initial audit rate. Thus the β^{τ} coefficients should be interpreted as capturing the behavior in year τ of hospitals subject to a one percentage point higher 2011 audit rate, where this behavior could be response to the 2011 audit rate or to any subsequent auditing in later years.

Audit Rate Instrument: One concern with estimating Equation 3 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 instance, if RACs target lower-quality hospitals and admissions at these 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. Specifically, 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:

$$Z_h^{2011} = \frac{1}{n_{s(h)} - 1} \sum_{h' \in \ s(h) \backslash h} X_{h'}^{2011} \ , \tag{4}$$

where $X_{h'}^{2011}$ is the 2011 audit rate for a 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:

$$Y_{ht} = \sum_{\tau=2007}^{2015} \gamma^{\tau} \mathbb{1}[t = \tau] \times Z_h^{2011} + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \ . \tag{5}$$

In order to interpret the coefficients as the effect of a one percentage point increase in the 2011 audit rate (as in Equation 3), I scale the γ^{τ} coefficients in Equation 5 by the correlation between X_h^{2011} and Z_h^{2011}, after accounting for hospital-group fixed effects.[22]

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

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

In this case, the reduced form specification is:

$$Y_{ht} = \mathbb{1}[t \geq 2011] \times Z_h^{2011} \beta^{post} + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \ . \tag{7}$$

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*).

[22] 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. This is implemented in a two-stage procedure to allow for clustering in the estimation of standard errors.

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.

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,[23] 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 a RAC 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 common 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. Finally, it could also be driven by differences in the contingency fee a RAC faces. While the structure of how each RAC was reimbursed was the same, each RAC faces a different contingency fee that they submit as part of their proposal bid, which is not publicly available. Thus the less-aggressive RACs could be the ones who negotiated a lower contingency fee and therefore face a lower return per audit.

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 are on 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

[23] 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.

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.[24]

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 (e.g., 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 about RAC behavior at the state-level – CMS did not release statistics about the size and scope of the program until much later. This is evidenced by the AHA's push to independently survey its members on their RAC experiences to gather information on the program.

Since a hospital's audit rate is continuous and therefore "fuzzy," I also need to assume that hospitals in the border sample have homogeneous treatment effects (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 within neighbor comparison groups, hospitals that are subject to different audit rates are still relatively similar by other measures (Table GII).

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. Non-time-varying confounders like existing state policies are absorbed by the hospital fixed effect in the difference-in-difference specification. To violate the exclusion restriction, time-varying confounders would have to be consistent across multiple states and occur simultaneously in 2011. I check in robustness tests in Section C.1 that the results are not driven by a particularly relevant confounder: the Medi-

[24]Restricting the comparison to border hospitals allows me to make a weaker parallel trends assumption than a comparison of all hospitals. Figure H8f shows the results from an alternate specification that includes all hospitals; there is evidence of differential pretrends when comparing across all hospitals in different RAC regions.

care Administrative Contractors (MACs), who primarily process and deny Medicare claims before payment. While the MACs operate in smaller regions than the RAC regions, some of the boundaries of the MAC and RAC regions overlap. The results are robust to excluding these overlapping borders.

The exclusion restriction could also be violated by reverse causality – if, say, the leave-one-out audit rate also 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 H9). Finally, 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. 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:

$$Y_v = \sum_{\tau=2010}^{2016} \beta^\tau \mathbb{1}[y = \tau] \times \mathbb{1}[t \geq 00{:}00] + \boldsymbol{W_v'\gamma} + \lambda_{hy} + \phi_{ht} + \varepsilon_v \ , \tag{8}$$

where ED visit v occurs in fiscal year y[25] 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. $\mathbb{1}[y = \tau]$ 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

[25] Fiscal year y goes from October in calendar year $y - 1$ to September in calendar y.

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 quartile of average income in the 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 9 pools the event study into a single post-policy coefficient β:

$$Y_v = \beta \mathbb{1}[y \geq 2013\text{Q3}] \times \mathbb{1}[t \geq 00{:}00] + \boldsymbol{W_v'}\boldsymbol{\gamma} + \lambda_{hq} + \phi_{ht} + \varepsilon_v \; . \tag{9}$$

Here $\mathbb{1}[y \geq 2013\text{Q3}]$ is an indicator for whether 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, or an increase in the share of patients reported arriving between 11:00 PM and midnight, once the policy is implemented. Figure H10 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 to game 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)).

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 4 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 5 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 4. Table III reports the yearly coefficients for 2011 to 2015.[26] Figures 5a and 5b plot the results for log Medicare admissions and log Medicare inpatient revenue, where inpatient

[26]For brevity, the pre-2011 coefficients are estimated but not reported in the table.

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. Extrapolating to the overall hospital sample (albeit under fairly strong assumptions, as discussed in Appendix Section D) indicates that RAC audits saved the Medicare program $9.28 billion between 2011 and 2015.

I next turn to the administrative burden RAC audits impose on hospitals. Figure 5 and Table III columns 5-6 present results on hospital administrative costs and IT adoption. Figure 5c 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 these higher administrative costs. Figure 5d 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).

To estimate the total savings from RAC audits, Figure H12 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.[27] 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 \$178–218 in compliance costs.[28]

Next, I explore the effects on different types of admissions to understand what stays are being deterred. Given policymakers' concerns about short stays being the primary driver of unnecessary stays, Figure 6 splits admissions by their length of stay. Figure H13a plots the audit rates by length of stay. The deterrence 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 7 then explores differences across diagnoses with different propensities for payment errors. Specifically, I categorize diagnoses by the payment error rate associated with each Medicare Severity Diagnosis Related Group (MS-DRGs, also referred to as DRGs). I use the ranking of base

[27]For a one percentage point increase in 2011 audit rate, the government costs (the contingency fees paid to the RACs) by 2015 are \$88k and the direct savings from reclaimed payments are \$232k. Including deterred admissions the total Medicare savings are \$2.08 million, so Medicare has a return of \$24. 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 A.2 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. The government cost remains the same since the contingency fees were paid before the payments were returned in the August 2014 settlement. Thus in this case, RAC audits save \$29 per dollar of monitoring costs, and deterred admissions account for 72 percent of the savings.

[28]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 \$178 in hospital compliance costs per \$1000 in Medicare savings.

DRGs[29] by payment error calculated by the Comprehensive Error Rate Testing (CERT) Program in 2010, a Medicare program which randomly samples 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 claim types, and RACs did not participate in this program. Figure H13b plots the audit rates for the top 20 highest error base DRGs. Figures 7a and 7b plot the event study results, which show larger and more sustained reductions in admissions for the top 20 base DRGs compared to DRGs outside of the top 20. This is consistent with hospitals focusing on reducing the types of diagnoses that Medicare signaled it was most concerned about. 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 mostly 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 7c and 7d I then split the top 20 base DRG groups into emergent and non-emergent diagnoses.[30] 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

[29] DRGs can be categorized 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).

[30] The event studies begin in 2008 to avoid capturing a 2007 reform to how DRGs are categorized (Gross et al., 2023).

on-the-ground providers as they make care decisions in *real time*. But the decision to reduce non-emergent admissions can be made at a higher level – say, if a hospital changes its policy on inpatient stays after elective procedures. AHA survey evidence shows that hospitals reported hiring utilization management consultants and undergoing training programs, which could reflect efforts to inform administrators how to set these policies (American Hospital Association, 2012). But in contrast to software adoption, these activities are not easily observed in non-survey data.

The event studies in Figures 5, 6, and 7 also 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 adopt 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 the administrative cost 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 in later years (Figure 1b). Instead, the one-time spike in administrative costs is consistent with hospitals making upfront investments like adopting technology, hiring consultants, or participating in training programs.

The dynamic effects should be interpreted as capturing hospitals' responses to a combination of the exogenous 2011 audit rate and all the (possibly endogenous) audit rates they faced in subsequent years. As shown in Figure H14, the high-audit regions' audit rates decrease over time relative to their highest point in 2012, while low-audit regions' audit rates continue to increase. Thus these estimates may understate what we would see if RAC audit rates persisted within region 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 as many 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 entirely due to hospitals knowingly admitting them. The event studies reveal that the full effect on admissions took several 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; O'Malley et al., 2021; Leder-Luis, 2023). This slower decline is consistent with hospitals needing time to implement improvements in their admissions processes, like incorporating newly installed software.

Table GV pools the post-2011 years of the main results into a single post-2011 coefficient, as in Equation 7. 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 GVI 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.[31] 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.1, I check that these results are robust to instrumenting for the share of claims that are denied rather than just audited, including controls for hospital characteristic-year time trends, 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 the state borders and the MAC borders in the interior of each RAC region.

In Appendix Section C.2, I also consider the effect of RAC audits on coding. In addition to conducting audits for medical necessity, RACs could also audit for coding errors such as upcoding.[32]

[31] 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).

[32] "Upcoding" refers to the practice of reporting additional (potentially unsubstantiated) diagnoses on a claim to maximize health insurance reimbursement. In the context of the Medicare inpatient DRG system, hospitals can bill for a more lucrative DRG by adding diagnoses to indicate a patient has more comorbidities or complications Silverman and Skinner (2004); Dafny (2005).

Five percent of audits reclaimed a partial payment, which could arise from coding corrections. In contrast, medical necessity corrections should lead to the full payment being reclaimed. I measure coding intensity as the number of diagnosis codes reported per Medicare admission and find that auditing reduces reported diagnoses, despite the fact that the patients still admitted are presumably sicker. This suggests that auditing may have also reduced upcoding, implying that the savings from deterred admissions may be an underestimate of the overall savings from the RAC program.

Finally in Appendix C.3, 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 welfare beyond the deterred admissions. I find that border hospitals subject to more auditing were no more likely to close in subsequent years, mitigating concerns about this channel.

5.2 Patient Outcomes: Inpatient Admission Likelihood and Revisit Likelihood

I next turn to the results from the patient-level analysis. Figure 8 plots the event studies of the patient-level analysis of ED visits in Equation 8. 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 9. 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 8d 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 indicates that the marginal admission deterred by auditing is a relatively low-value one.

However, because only a small subset of patients should be on the admission margin, this null average effect may be masking heterogeneity across patients. The model discussed in Section 3 predicts that the deterrence effect should be concentrated among relatively lower benefit admissions. The highest benefit patients will still be admitted and the lowest benefit patients were never admitted to begin with. Therefore, it should be patients in the middle of the benefit distribution who are most likely to be denied admission. 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.[33] 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 9, interacting β with an indicator for each severity decile.

Figure 9 plots the heterogeneity by severity results for inpatient status and for revisits within thirty days. The Two Midnights rule has no effect on admission rates for patients at the bottom and top severity deciles. Instead, the reduction in admissions stems primarily from the middle of the severity distribution. There is a 5 percentage point, or 25 percent, decrease in admission likelihood 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. As a robustnss check,

[33]This includes patient demographics such as age-bin, sex, race, a Hispanic indicator, a point-of-origin indicator, and quartile of 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.

I also conduct a subsample analysis that restricts to particularly vulnerable patient populations as defined by age, number of chronic conditions, race, and income in Table GVII, and likewise find no effect on revisits on these subpopulations.

Table GVIII 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 mostly driven by hospitals with the medical necessity checking software prior to the Two Midnights rule. Appendix Section C.1 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-Medicare patients, who should not be directly affected by the Two Midnights rule.

5.3 Discussion

Technology Adoption Mechanism Taken together, the hospital-level and patient-level results underscore the role that medical necessity checking software plays in helping hospitals identify unnecessary admissions, especially for emergent stays. The hospital-level results show that hospitals responded to RAC audits by installing this software. The patient-level results then demonstrate that the deterrence effect was concentrated among admissions with relatively lower patient health benefit, which is precisely the type of care that this software targets. They also show that the response to the Two Midnights rule was driven by hospitals with this software already installed.

Figure H11 provides three pieces of cross-sectional evidence that lend further support to this mechanism. First, hospitals with the software already installed in 2010 had lower denial rates, especially in the years where RACs focused on unnecessary admissions (Figure H11a). This indicates that having the software reduces audit penalties – the model predicts that increased auditing will lead to greater adoption only if hospitals with the technology have lower penalties. Second, among hospitals within the same RAC region, those that were more heavily penalized by RACs were more likely to adopt the software in later years (Figure H11b). Within each region, RACs focused more attention on hospitals that were making more unnecessary admissions. Therefore these hospitals should have been the ones with the most to gain from adopting medical necessity checking software, as their penalties without adopting are relatively high. Finally, hospitals that adopted software in 2011-2015 saw the largest decreases in high-error emergent stays, suggesting that hospitals use medical necessity software to target emergent stays in particular (Figure H11c). This is consistent with how vendors marketed the software, as being able to provide timely medical necessity information, which should be most relevant for emergent cases.

Comparing the Two Approaches There are also important differences between the hospital-level and patient-level results which warrant further investigation. The first is the difference in the patient population considered in each approach. The hospital-level results capture all Medicare inpatient stays, regardless of admission source. In contrast, the patient-level approach focuses on a much more narrow sample: patients who enter the ED in Florida around midnight. A large majority (73 percent) of Medicare admissions originate in the ED, and Table GI shows that the patient characteristics across the two samples are similar. But there is still the key difference that the patient-level sample consists only of emergent cases. We may therefore be concerned about the external validity of extrapolating the patient health results from the ED patient-level sample to the overall hospital-level sample, where we also see reductions in non-emergent stays.

The external validity of the patient-level health results is supported by the fact that patients in emergent stays tend to be in worse health compared to those in non-emergent stays. Figure H15 shows that 30-day mortality is higher among DRGs with a larger share of stays originating in the ED. Because emergent cases are most at risk of harm, we would expect that any negative effect on patient health should be more likely to appear in the ED sample. But I do not detect a negative health effect in this sample of higher-risk patients, even for the subset of these patients who face the largest reduction in admission rates. Another way to approach this is to extend the hospital-level specification to ED visits by incorporating the Medicare Outpatient file into the analysis. Consistent with the patient-level results, I find evidence at the hospital-level of increased observation stay usage, as well as a null effect on 30-day revisits and mortality among ED visits

(Figure H16).[34]

The second difference between the two approaches is what happens when audit rates decrease. The hospital-level results show that once the RAC program is scaled back in 2014, admissions do not rebound. In contrast, in the patient-level results, admissions appear to increase for before-midnight arrivals once the Two Midnights rule is in place (Figure 3), possibly because their audit rate decreases. We can use the model discussed in Section 3 to reconcile these two findings. In the hospital-level results, I find that auditing leads some hospitals to install medical necessity software. As illustrated in the example in Figure 2c, this may have fundamentally changed the payoff curve these hospitals faced, resulting in them choosing a different admission threshold. These hospitals may not have uninstalled the software once auditing is scaled back, either because uninstalling is costly or because they signed multi-year contracts with software vendors. This could explain why admissions do not rebound substantially in later years, despite the low audit rates in those years.

However in the patient-level sample, hospitals could only respond to the Two Midnights rule by moving along their existing payoff curve. Compared to the changes in admissions resulting from technology adoption, these changes may be more easily reversed when audit rates decrease. An interesting implication of the persistent hospital-level response is that Medicare may not need to continually monitor hospitals, but can rather focus on monitoring aggressively upfront to induce investment. This echos dynamic enforcement strategies employed by other CMS monitoring programs,[35] as well as other regulatory agencies like the Environmental Protection Agency (Blundell et al., 2020).

There is an additional distinction between the two policy environments that could also 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. However with the 2014 pause, hos-

[34]However, due to possible mis-measurement of emergency department visits in Medicare claims data, these results should be interpreted as suggestive. This is explained in further detail in Appendix Section E.

[35]Specifically, CMS's "Targeted Probe and Educate" program subjects providers with high denial rates to intensive claim reviews and one-on-one education. If providers do not improve within three rounds of these reviews, they will face even greater scrutiny like 100 percent prepay reviews (link, last accessed July 2023).

pitals could not be sure that auditing wouldn't increase again in later years. RACs had a lookback period of 3 years, so admissions in 2014-2015 could be audited as late as 2018. When it paused audits in 2014, Medicare emphasized that it was only a temporary pause. After multiple announced and subsequently delayed resumption dates over several quarters, inpatient RAC audits finally resumed in 2015Q4, although RACs were much more constrained compared than before. But it is unlikely that hospitals could have anticipated this trajectory for the RAC program at the onset of the pause.

6 Conclusion

In this paper, I consider the tradeoffs of monitoring for wasteful public spending by studying a large Medicare program that audited for unnecessary hospital admissions. I consider a model of hospital admissions and technology adoption to understand how monitoring interacts with hospital behavior. The model predicts that hospitals respond to increased audits by reducing low-value admissions, and it may also spur them to adopt technology to improve their diagnostic ability. In the empirical analysis, I first compare hospitals subject to differentially aggressive auditors and find that auditing has a large deterrence effect on hospital admissions – I estimate a \$24-29 return per dollar spent on monitoring. Almost 90 percent of the savings from audits come from the deterrence effect, rather than the actual savings recouped in the audits. There are decreases among admissions with both emergent and non-emergent diagnoses, and most of the reductions were concentrated among short stays. While hospital administrative costs do increase, these costs are short-lived and can be attributed in part to the adoption of software to improve compliance with medical necessity rules. I then look to patient health outcomes to assess whether these savings stemmed from reductions in low-value care. Drilling down to the patient-level, I leverage a policy which varied patients' audit rate depending on when a patient arrives to the ED. Here, I also find that hospitals respond to increased audit risk by decreasing admissions. I do not detect evidence of patient harm, as measured by hospital revisit rates, suggesting that the marginal admission deterred was an unnecessary one. Taken together, these results show that monitoring can be a highly effective tool to combat waste in public spending and improve compliance with policy goals.

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.

Atasoy, Hilal, Brad N. Greenwood, and Jeffrey Scott McCullough, "The Digitization of Patient Care: A Review of the Effects of Electronic Health Records on Health Care Quality and Utilization," *Annual Review of Public Health*, April 2019, *40*, 487–500.

Bagley, Nicholas, "Crashing Medicare," September 2014.

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.

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217. Publisher: University of Chicago Press.

Bergin, James, "Information," in James Bergin, ed., *Microeconomic Theory: A Concise Course*, Oxford University Press, May 2005, p. 0.

Blackwell, David, "Comparison of Experiments," in "Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability," Vol. 2, University of California Press, January 1951, pp. 93–103.

— , “Equivalent Comparisons of Experiments,” *The Annals of Mathematical Statistics*, 1953, *24* (2), 265–272. Publisher: Institute of Mathematical Statistics.

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer, “Escalation of Scrutiny: The Gains from Dynamic Enforcement of Environmental Regulations,” *American Economic Review*, August 2020, *110* (8), 2558–2585.

Brot-Goldberg, Zarek, Samantha Burn, Timothy Layton, and Boris Vabson, “Rationing Medicine Through Paperwork: Authorization Restrictions in Medicare,” Working Paper w30878, National Bureau of Economic Research January 2023.

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.

Chaisemartin, Clement and Xavier D’Haultfœuille, “Fuzzy Differences-in-Differences,” *The Review of Economic Studies*, April 2018, *85* (2), 999–1028.

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

— **and Jonathan Gruber**, “Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions,” *AEA Papers and Proceedings*, May 2020, *110*, 279–283.

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.

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.

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.

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.

_, Joshua Gottlieb, Adam Shapiro, and Daniel Sonnenstuhl, "A Denial A Day Keeps the Doctor Away," Working Paper w29010, National Bureau of Economic Research January 2023.

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.

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.

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.

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.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez, "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark," *Econometrica*, 2011, *79* (3), 651–692. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9113.

Kopczuk, W., "Tax simplification and tax compliance: An economic perspective," in "in" 2007.

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.

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.

Miller, Mark, "Hospital Short-Stay Policy Issues," May 2015.

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.

Office of the National Coordinator for Health Information Technology, "Clinical Decision Support | HealthIT.gov," April 2018.

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

O'Malley, A. James, Thomas A. Bubolz, and Jonathan S. Skinner, "The Diffusion of Health Care Fraud: A Network Analysis," Working Paper w28560, National Bureau of Economic Research March 2021.

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.

Tella, Rafael Di and Ernesto Schargrodsky, "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack," *American Economic Review*, March 2004, *94* (1), 115–133.

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.

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
Share non-chain	0.38	0.39	0.49	0.39	0.34	0.33
Share teaching	0.33	0.32	0.50	0.37	0.25	0.31
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, teaching status, and profit type status come from the Medicare Provider of Services file. Chain status comes from Cooper et al. (2019) merger data. 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 without inpatient admission. “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 × 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 × 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 × 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 × 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 × 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 5 and plotted in Figure 5, 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.

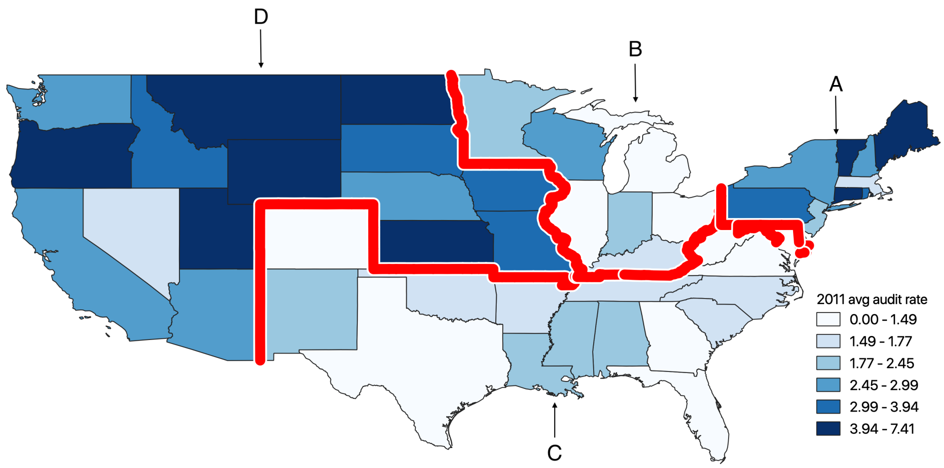
Table IV. After-Midnight ED Arrival Hour Difference-in-Difference Coefficients on Patient Status and Revisits

	(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

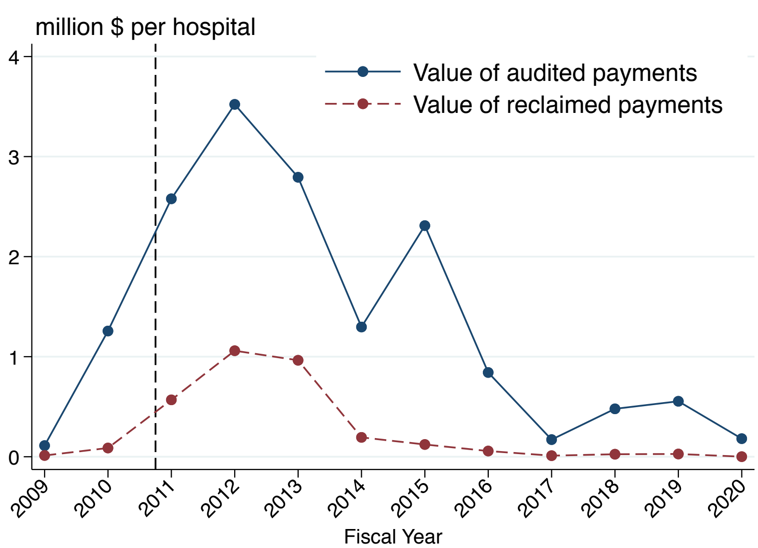
* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors are in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[y \geq \text{2013Q3}] \times \mathbb{1}[T_v \geq \text{00:00}]$ of the specification in Equation 9, where $\mathbb{1}[y \geq \text{2013Q3}]$ 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. "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 in a Florida hospital 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.

Figure 1. RAC Audit Activity

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



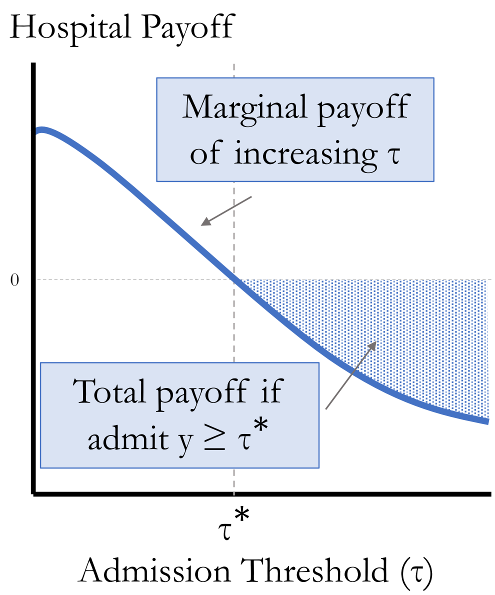
(b) Value of Audited Inpatient Payments and Net Reclaimed Payments per Hospital, by Year of Audit



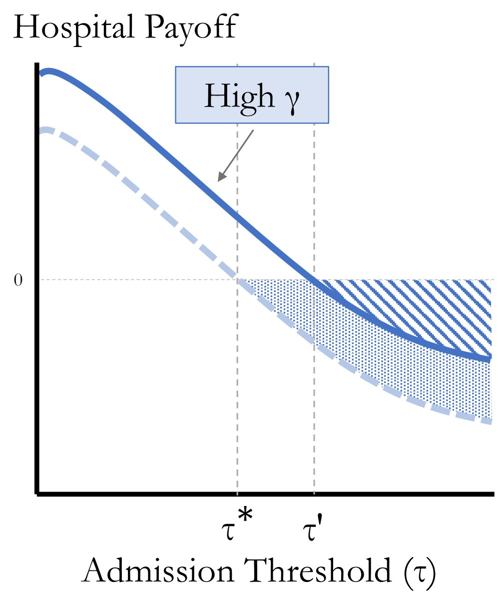
Panel (a) plots the 2011 average state audit rates, where audit rate is defined as the share of a hospital's 2008-2011 claims that were audited by RACs. The RAC regions are Region A (Northeast), Region B (Midwest), Region C (South), and Region D (West). Darker shades denote a higher audit rate. The red line demarcates RAC regions. Panel (b) plots the average per-hospital value of inpatient payments audited by RACs and the reclaimed payments, by year of audit. 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. Model Illustration

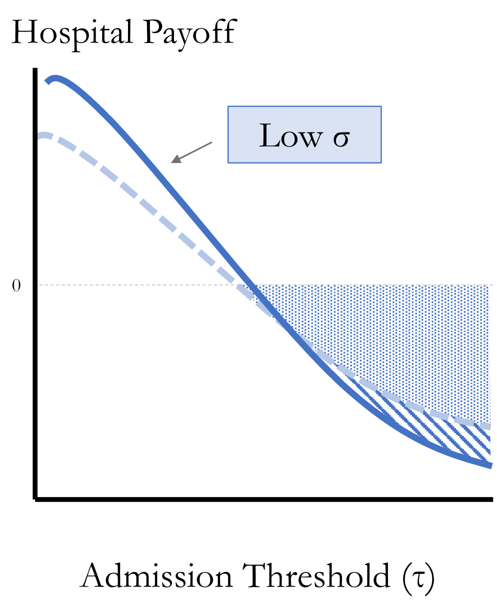
(a) Baseline Equilibrium



(b) Increase in Expected Audit Penalty γ

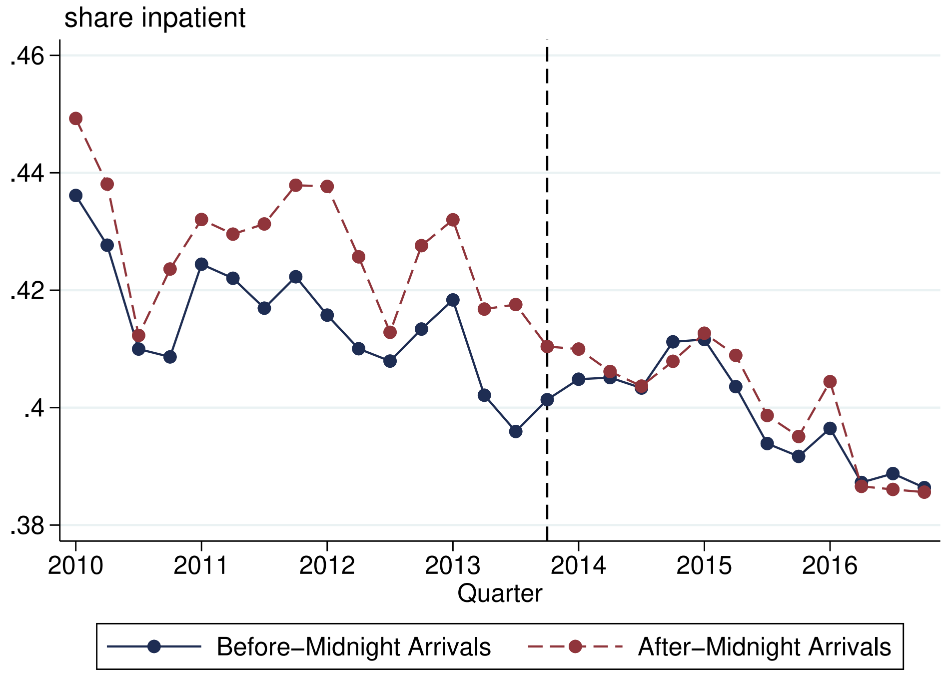


(c) Decrease in Signal Noise σ



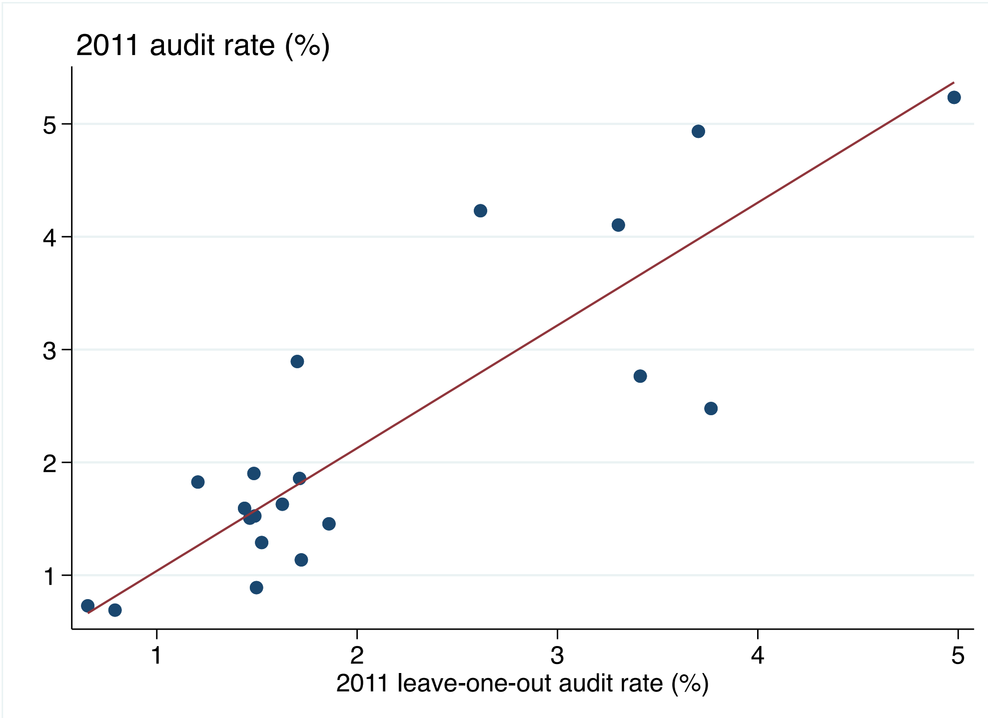
These figures illustrate the model as described in Section 3 and Appendix Section B. Panel (a) illustrates the baseline equilibrium. The x-axis is the admission threshold τ; the hospital admits when a patient has signal $y \geq \tau$. The y-axis is the marginal payoff of increasing τ, and the area to the right of τ^* and above the marginal payoff is the total expected payoff for admitting patients with $y \geq \tau^*$. The threshold is inversely related to $q(\tau)$, the expected number of admissions. Panel (b) illustrates the effect of increasing the expected audit penalty for low-benefit admissions, γ. The marginal payoff shifts up as the returns to increasing the threshold (i.e., reducing admissions) increase. As a result the equilibrium threshold increases from τ^* to τ', and the number of admissions decreases. Panel (c) illustrates the effect of reducing signal noise σ. The marginal payoff curve steepens as the payoff becomes more elastic with respect to τ. By Blackwell's informativeness theorem (Blackwell, 1951, 1953) the expected total payoff increases when σ decreases.

Figure 3. 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 was implemented. Data: HCUP SID/SEDD.

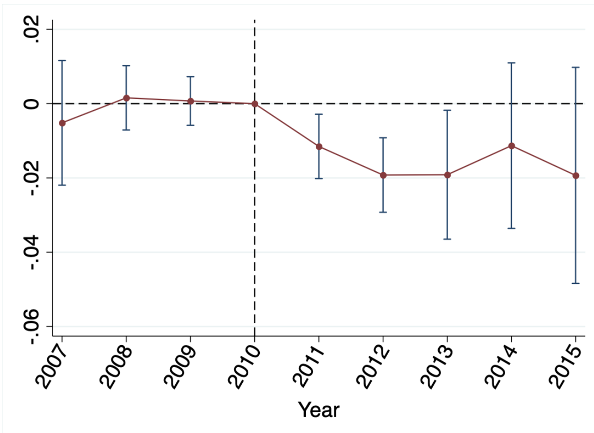
Figure 4. Binscatter of 2011 Leave-One-Out State Audit Rate and 2011 Hospital Audit Rate, Border Hospital Sample



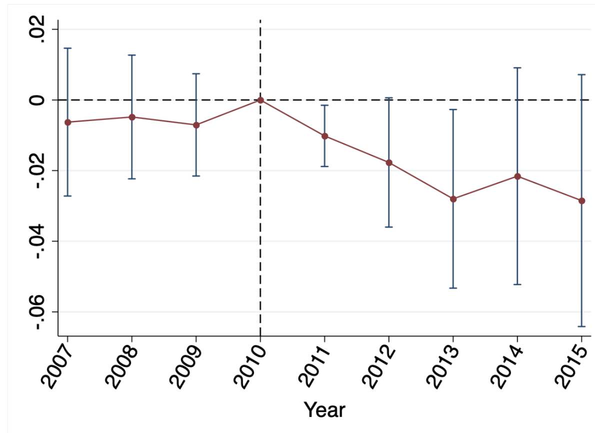
This figure plots a binscatter of the 2011 hospital audit rate compared to the 2011 leave-one-out state audit rate. The 2011 audit rate is defined as the share of 2008-2011 inpatient claims that were audited by RACs in 2011. The leave-one-out state audit rate is defined as the average audit rate of all 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 5. Event Studies on Effect of 2011 Audit Rate on Hospital Outcomes

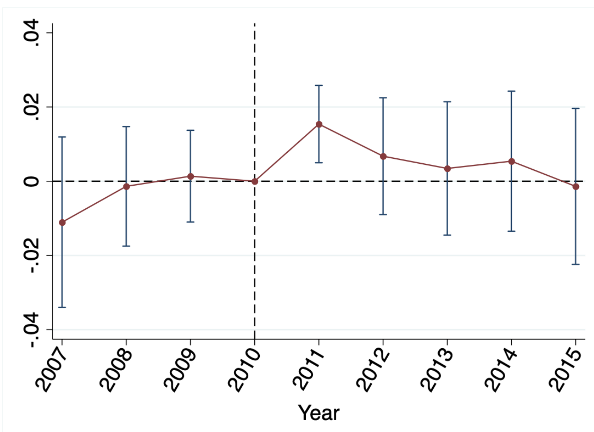
(a) Log Medicare admissions



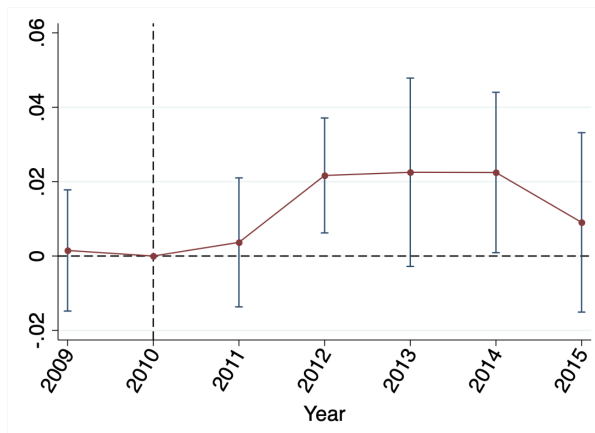
(b) Log Medicare inpatient revenue



(c) Log hospital administrative costs



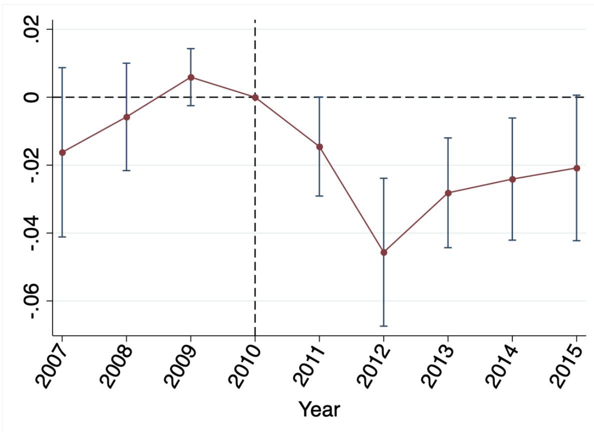
(d) Indicator for installing medical necessity software



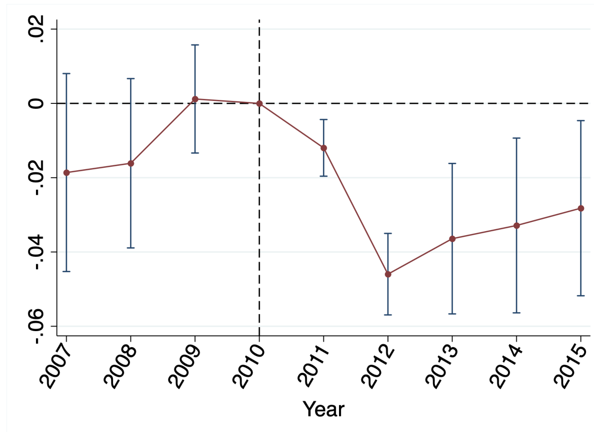
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents 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 6. Event Studies on Effect of 2011 Audit Rate on Medicare Admissions and Revenue, by Length of Stay

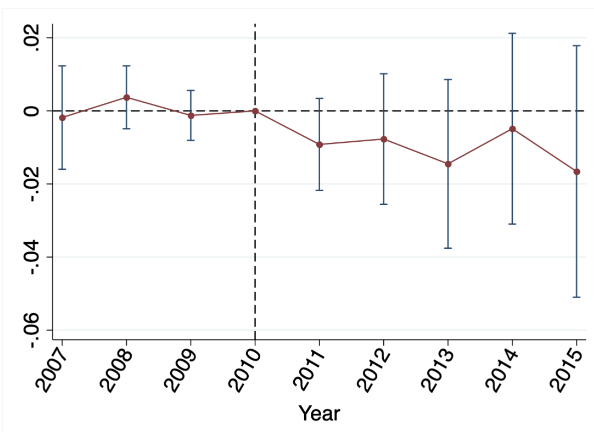
(a) Log Medicare admissions, LOS ≤ 2



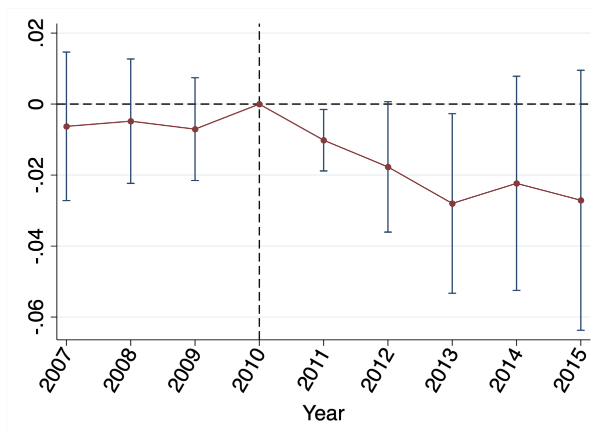
(b) Log Medicare inpatient revenue, LOS ≤ 2



(c) Log Medicare admissions, LOS > 2



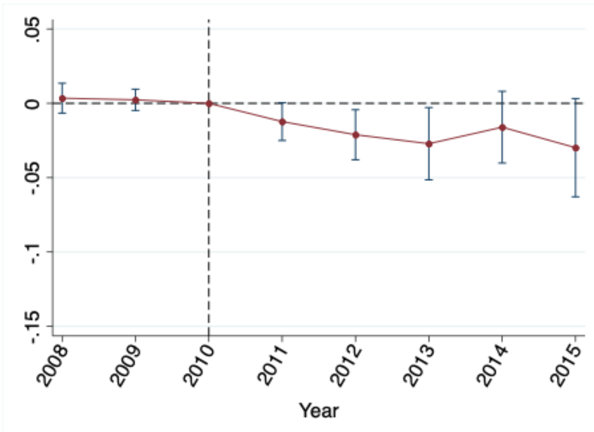
(d) Log Medicare inpatient revenue, LOS > 2



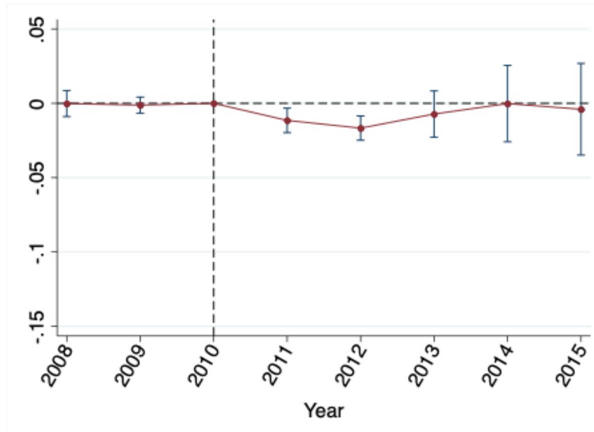
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents 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 7. Event Studies on Effect of 2011 Audit Rate on Medicare Admissions, By Base Diagnosis Group Error Rates

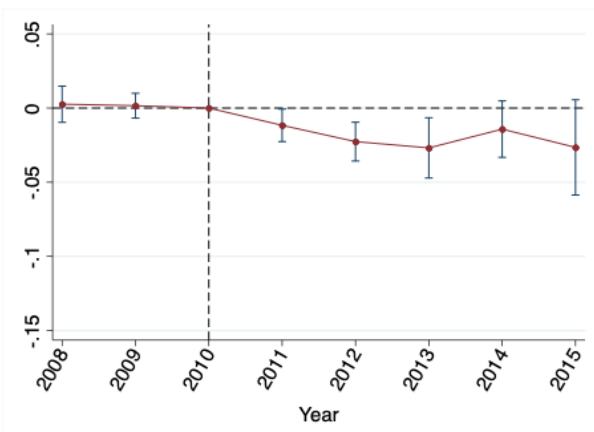
(a) Log Medicare admissions, all top 20 error base DRG groups



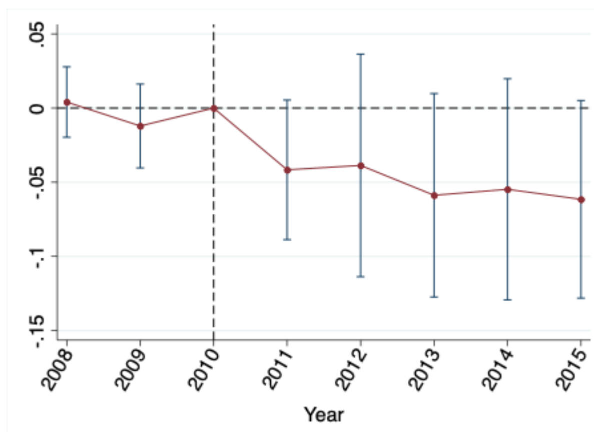
(b) Log Medicare admissions, non-top 20 base DRG groups



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

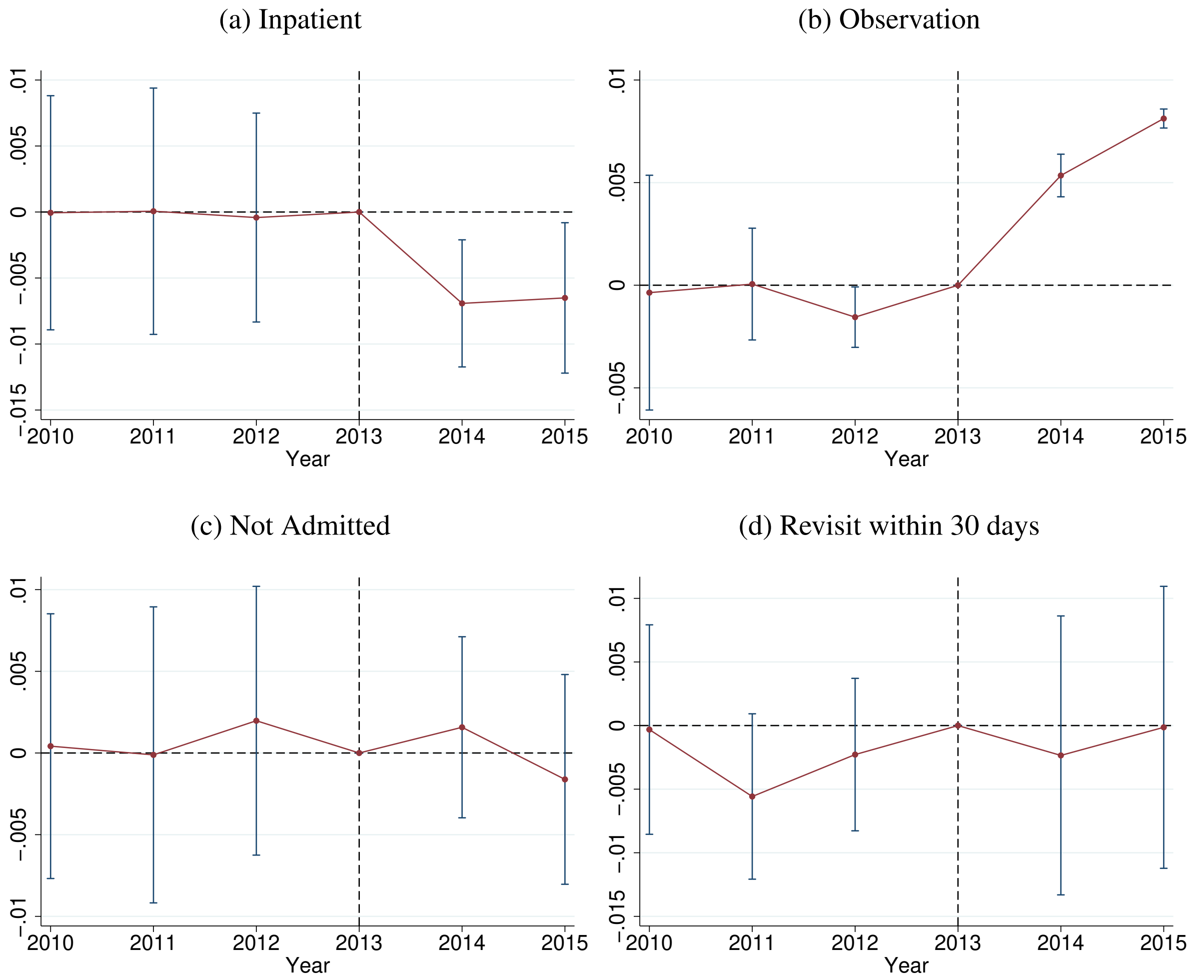


(d) Log Medicare admissions, top non-emergent MS-DRGs



This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5, 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 16 emergent MS-DRG base groups 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 4 non-emergent MS-DRG base groups 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%), and 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.

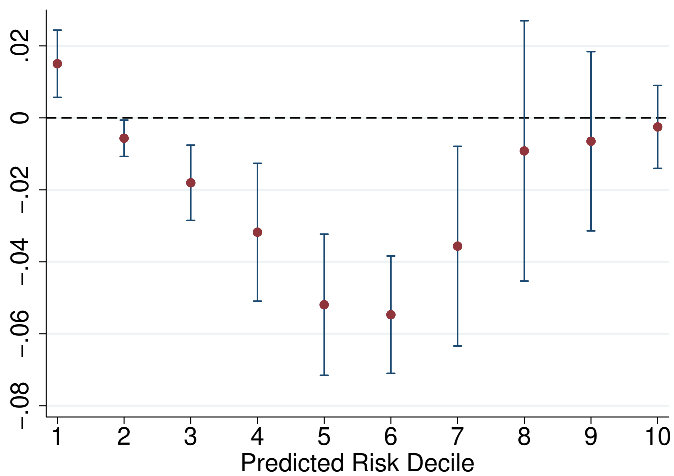
Figure 8. Event Studies on Effect of After-Midnight ED Arrival on Patient Status and Outcomes



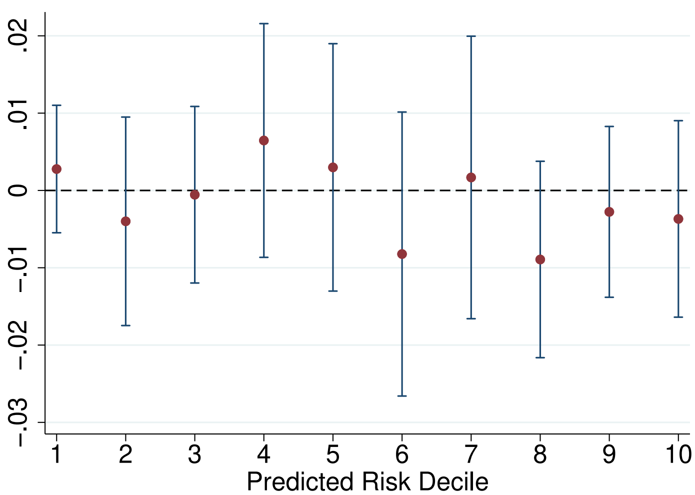
This figure plots the coefficients and 95% confidence intervals for β^{τ} on $\mathbb{1}[y = \tau] \times \mathbb{1}[T_v \geq 00{:}00]$ of the specification in Equation 9, where $\mathbb{1}[y = \tau]$ is an indicator for whether the visit occurred in fiscal year τ (i.e., October year $\tau - 1$ through 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 quartile of mean zip code income. Data: HCUP SID/SEDD.

Figure 9. 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 9, interacted with an indicator for predicted risk decile. β is the coefficient on $\mathbb{1}[y \geq 2013Q3] \times \mathbb{1}[T_v \geq 00{:}00]$, where $\mathbb{1}[y \geq 2013Q3]$ 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 quartile of 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. Data: HCUP SID/SEDD.