MONITORING FOR WASTE: EVIDENCE FROM MEDICARE AUDITS*

Maggie Shi

This article examines the trade-offs 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 health care 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 up-front and include investments in technology to assess the medical necessity of care. *JEL codes:* H51, H83, I00, I13, I18, M42, M48.

I. Introduction

Combating waste is a perennial problem for public programs. The Office of Management and Budget (2022) estimated that over 7% of U.S. federal spending is wasted. Economic theory prescribes a straightforward solution: more effort should be devoted to monitoring and penalizing wasteful spending

* This article 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 article. 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 Salanié, 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.

The Quarterly Journal of Economics (2023), 1–57. https://doi.org/10.1093/qje/qjad049.
Advance Access publication on September 28, 2023.

(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, policy makers 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. Despite the importance of this question, there is little empirical evidence on the magnitude and nature of the trade-offs associated with monitoring for waste in public spending.

This article considers these trade-offs 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 health care providers, who then have considerable latitude over spending decisions. Perhaps unsurprisingly, waste is widespread: estimates suggest that up to 13% 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 health care delivery (Hillestad et al. 2005). Policy makers have devoted billions of dollars in recent years to subsidize the adoption of this technology, with the hopes of reducing health care costs (Burde 2011; Atasoy, Greenwood, and McCullough 2019). Monitoring could therefore also serve as an additional policy lever to incentivize providers to seek 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 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,

1. Medicare expenditure accounts for 15% of federal spending (Cubanski, Neuman, and Freed 2019), so wasteful Medicare spending alone accounts for 2% of total federal spending.

this would simply add to providers' already high administrative burden (Cutler and Ly 2011; Dunn, Gottlieb, and Shapiro 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 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, 4% of all hospital admissions—Medicare's largest expenditure category—were audited, and 1% 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 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 and conducts audits to uncover and penalize admissions with low true benefit. In setting the admission threshold,

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

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 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 up-front 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. 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 use a difference-in-differences 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 1 percentage point (46%) increase in the share of admissions audited leads to a 2% drop in admissions. This effect persists even when auditing is scaled back in later years. Eighty-nine percent of the savings from the marginal audit stem from the deterrence of future admissions, and the remaining 11% are from the payments that 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 nonemergent admissions decrease. Extrapolating these effects to the overall hospital sample, I calculate that the RAC program led to upward 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 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; AccuReg 2022; Experian Health 2022).[3] Accordingly, hospital administrative costs rise: for every $1,000 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, rather than the ongoing hassle costs of interacting with the RACs.

I 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 in a hospital who are likely to be marginal admissions: those arriving in the ED whose

3. Specifically, medical necessity checking software is a type of clinical decision support technology that "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).

audit risk depends on an arbitrary threshold rule. In particular, I consider a rule that 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- versus after-midnight ED patients, pre and post the 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 30 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% 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 United States, there are several 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 article fills this gap in

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, Erard, and Feinstein (1998) and Slemrod and Yitzhaki (2002).

5. For example, in Olken (2007), measuring wasteful spending in public infrastructure projects required assembling teams to take core samples from roads

the literature by using patient health outcomes as a measure of spending quality in the health care setting.

Given that policy makers only considered the recovered payments when assessing the cost-effectiveness of the RAC program, the large deterrence effect that I find 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 article contributes to a broader empirical literature spanning criminal enforcement, tax compliance, and litigation, which has also demonstrated sizable deterrence effects.[7] Together these results underscore 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 hospitals to make costly improvements to their admissions process, such as installing medical necessity checking software.

A similar dynamic arises in other enforcement contexts. Given the complexity of the tax code, some underreporting of tax

---

and then comparing the reported and actual amounts of construction material used.

6. Link to reports: https://www.oversight.gov/reports.

7. Leder-Luis (forthcoming) finds a seven-to-one 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.

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 Special Supplemental Nutrition Program for Women, Infants, and Children and the Supplemental Nutrition Assistance Program, 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 disregarded. Thus if some of the private costs associated with monitoring stem from these kinds of investments, the compliance costs of monitoring may not all be deadweight loss.

This article also provides direct measures of the various social costs that 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 health care 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]

8. Interestingly, WIC/SNAP presents a case in which policy makers 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. (2024).

10. Examples of the former include Cutler (1995); Ellis and McGuire (1996); Clemens and Gottlieb (2014); Einav, Finkelstein, and Mahoney (2018); Eliason et al. (2018); Alexander and Schnell (2019); Gupta (2021); Gross et al. (forthcoming). Examples of the latter include Silverman and Skinner (2004); Dafny (2005); Sacarny (2018).

In contrast, less is known about how providers respond to nonfinancial incentives like monitoring, even though they are employed by both private and public insurers (Gottlieb, Shapiro, and Dunn 2018). This article contributes to a growing literature on how providers respond to various forms of nonfinancial incentives: prepayment denials (League 2023; Dunn et al. 2024), fraud enforcement (Nicholas et al. 2020; Howard and McCarthy 2021; Leder-Luis forthcoming), and prior authorization (Roberts et al. 2021; Brot-Goldberg et al. 2023).

The rest of the article proceeds as follows. Section II describes the policy context of the RAC program. Section III describes the model. Section IV.A describes the data for the empirical analysis, Section IV.B explains the hospital-level empirical strategy, and Section IV.C explains the patient-level empirical strategy. Section V presents the empirical results and compares the findings across the two empirical strategies. Section VI concludes.

II. Policy Context

Medicare spent $147 billion, or 19% 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, policy makers became increasingly concerned with one area of perceived waste: unnecessary short (0–2 day) stays (Centers for Medicare and Medicaid Services 2011; U.S. Department of Health and Human Services Office of Inspector General 2013). The Medicare Payment Advisory Commission (MedPAC), a nonpartisan 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

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

twice that of longer stays. Indeed, economists have long pointed this out as a potential vulnerability to a prospective payment system.[12] Online Appendix 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 in its geographic jurisdiction, or RAC region. Figure I, Panel A 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% from 2010 to 2012, which translated into a 1,211% increase in the value of payments reclaimed per hospital (Figure I, Panel B).[14]

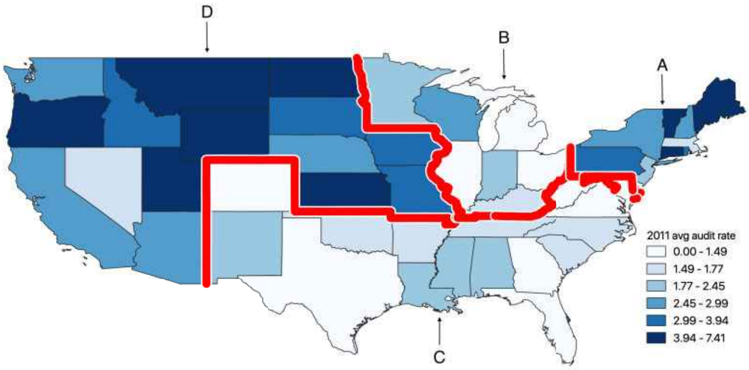
Ninety-five percent of inpatient stay RAC audits involve a manual review: the RAC first runs a proprietary algorithm on Medicare claims data to flag individual claims for issues such as missing documentation, incorrect coding, or—starting in 2011—unnecessary care. A medical professional hired by the 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 6% resulted in an underpayment determination.

12. As Ellis and McGuire (1986, 141–142) 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 (MACs), who do not process claims for medical care, but rather claims for equipment and supplies ordered by health care providers. This includes oxygen equipment, wheelchairs, and blood testing strips. The Part B MACs in charge of processing and denying medical claims use smaller regions, some of which share boundaries with RAC regions (League 2023). In Online Appendix 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.

(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

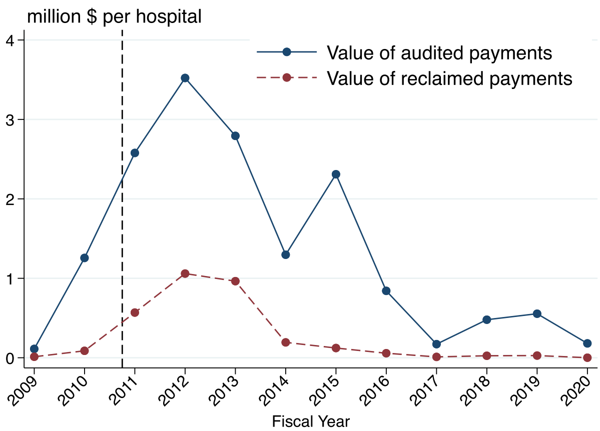


FIGURE I

RAC Audit Activity

Panel A plots the 2011 average state audit rates, where the 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 heavy 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.

Once the complex review is finished, RACs send a letter to providers that outlines whether a 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%, 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. Online Appendix Figure G1 illustrates the full process for claims auditing and appeals, including the remaining 5% of inpatient stay audits that do not involve a manual documentation review.

Figure I, Panel illustrates average per hospital RAC activity, by year of audit (which is often after the year the claim was originally paid). At the program's peak, RACs were reclaiming $1 million per hospital annually, or 3% of the average hospital's Medicare inpatient revenue of $32 million. By 2020, 96% 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). Online Appendix 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 health care 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 in the RAC program which generated differences in audit risk across patients in 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% 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.

III. 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

15. For example, the Experian product sheet (last accessed July 2023) offers to "prevent claim denials with access to timely and updated medical necessity content," "improve cash flow by proactively identifying 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).

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 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 because they are partially altruistic but also because they want to avoid audit penalties for ex post low-benefit stays. Whether 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.

III.A. 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 whom to admit—under threshold rule τ, it will admit all patients with $y \geqslant \tau$. Let the likelihood that a patient with benefit value x is admitted under threshold τ be $P(x; \tau) = 1 - \Phi(\frac{\tau - x}{\sigma})$. 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 II, Panel A 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

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 $-\frac{C_H}{2}(\beta q(\tau))^2$.

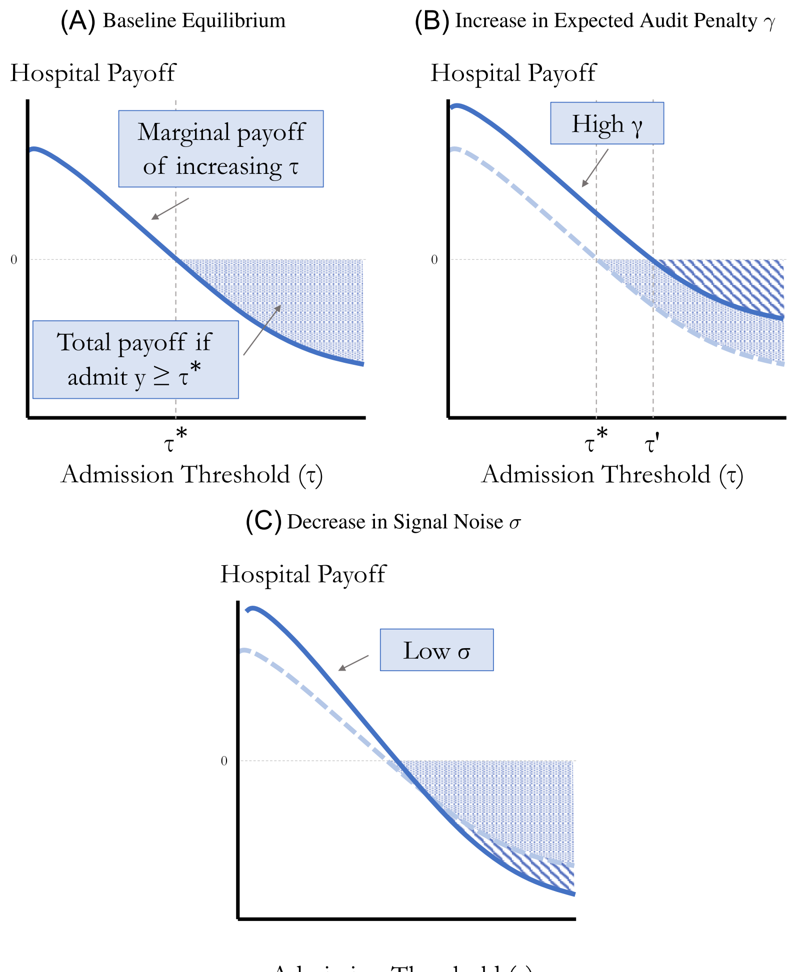


FIGURE II

Model Illustration

These figures illustrate the model as described in Section III and Online Appendix B. Panel A illustrates the baseline equilibrium. The x-axis is the admission threshold τ; the hospital admits when a patient has signal $y \geqslant \tau$. 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 $y \geqslant \tau^*$. 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

FIGURE II

(*Continued*) 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.

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

Figure II, Panel B 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 Online Appendix 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.

III.B. Signal Quality and Technology Adoption

Figure II, Panel C 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. Online Appendix 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 II, Panel C. 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.

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

In other words, hospitals will respond to audits by adopting technology if $\frac{dK}{d\gamma}$ is positive. As shown in Online Appendix 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.

Online Appendix 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.

IV. DATA AND IDENTIFICATION STRATEGIES

IV.A. 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%

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 data set 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% 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 Online Appendix 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 health care 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. In addition, to study heterogeneity across hospital types, I 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, nonprofit hospitals, and hospitals with a higher share of short-stay Medicare admissions are more likely to be audited (Online Appendix Figure G2). Online Appendix 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

TABLE I
HOSPITAL SUMMARY STATISTICS

	Sample		RAC region			
	Overall (1)	Border (2)	A (3)	B (4)	C (5)	D (6)
Panel 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 nonprofit	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 nonchain	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)

TABLE I
CONTINUED

	Sample		RAC region			
	Overall (1)	Border (2)	A (3)	B (4)	C (5)	D (6)
Panel B: Medicare inpatient characteristics						
Admissions	3,465.75	3,151.42	4,264.70	3,845.22	3,262.61	2,928.68
	(3,205.86)	(3,069.49)	(3,591.67)	(3,383.92)	(3,260.47)	(2,399.90)
Mean payment ($)	8,617.36	7,366.40	9,349.37	8,177.97	7,578.76	10,393.64
	(3,179.31)	(2,349.10)	(3,461.79)	(2,433.87)	(2,663.76)	(3,501.44)
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	2,960	510	489	571	1,237	663
N border hospitals	510	510	41	184	191	94

Notes. This table presents 2010 summary statistics of hospital characteristics and Medicare inpatient admissions by sample and RAC region. Standard deviations are 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 $\leqslant 2$. 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 2011). "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 100 miles of the RAC border with at least one hospital in their neighbor comparison group.

because it is the only state that reports ED arrival hour in the publicly available data for both the inpatient and emergency department data sets; 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 as an inpatient visit.[20] I use this proxy because mortality is not observable in hospital discharge data such as SID and SEDD. Online Appendix Table FI 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 three 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 III plots the quarterly share of before- and after-midnight Medicare ED arrivals who are admitted as inpatients. Prior to the two midnights rule, after-midnight arrivals are more likely to be admitted as inpatients, 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

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 mismeasurement (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.

TABLE II
PATIENT SUMMARY STATISTICS BY ED ARRIVAL HOUR, PRE AND POST THE TWO MIDNIGHTS RULE

	ED arrival hour			
	Prepolicy		Postpolicy	
	Before MN (1)	After MN (2)	Before MN (3)	After MN (4)
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 ($)	24,171	26,068	25,757	26,572
	(43,629)	(49,564)	(47,944)	(52,421)
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	31,419	17,690	32,420	17,637

Notes. This table presents summary statistics of characteristics of traditional Medicare patients in Florida who arrived in the ED within three hours of midnight in 2013Q2 (prepolicy) and in 2014Q2 (postpolicy). Standard deviations are in parentheses. "Share inpatient" is the share of ED patients admitted as inpatients (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 9:00 a.m. and 3:00 p.m. 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.

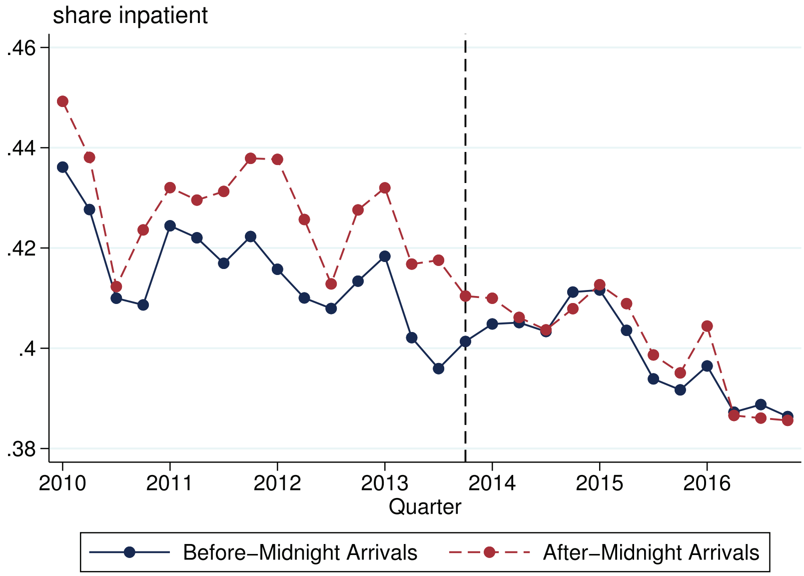


FIGURE III

Inpatient Admission Rates from ED, Before- versus After-Midnight ED Arrivals in Florida

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

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 prepolicy and postpolicy periods, which supports making a parallel-trends assumption about the before- and after-midnight arrivals.

IV.B. 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 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-differences 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.

1. *Border Hospital Sample.* Figure I, Panel A 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 100 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 nonprofit than the overall sample. Because these characteristics correlate with audit rate, border hospitals have a higher 2011 audit rate than the overall sample. In addition, a larger share of border hospitals come from RAC regions B and C.

2. *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. Online Appendix Table FII 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.

Online Appendix Figure G3 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%. The members of its neighbor comparison group are the hospitals on the other side of the border within 100 miles—in this case, that would be hospitals in Kansas (RAC Region D) that face a much higher average audit rate of 5.42%. Together, the Oklahoma hospital and its neighbors in Kansas form the neighbor comparison group for the Oklahoma hospital.

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

21. In identifying a unique set of neighbors for each hospital, I follow Dube, Lester, and Reich (2010), whose state border-county identification strategy allows individual counties to be paired with unique sets of adjacent comparison counties.

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. Online Appendix Figure G4 illustrates the border segments used for clustering, with each segment in a different color. Each border segment is 100 miles, except for segments that cross state lines, which are split at the state border.

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

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

In equation (3), Y_{ht} is an outcome for hospital h in year t, X_h^{2011} is the hospital's 2011 audit rate, ϕ_{gt} is a hospital's neighbor comparison group g-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 I, Panel B), 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 1 percentage point higher 2011 audit rate, where this behavior could be a response to the 2011 audit rate or to any subsequent auditing in later years.

4. *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:

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

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:

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

To interpret the coefficients as the effect of a 1 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:

$$Y_{ht} = \beta^{post} \mathbb{1}[t \geqslant 2011] \times X_h^{2011} + \phi_{gt} + \psi_h + \varepsilon_{ht}. \tag{6}$$

In this case, the reduced-form specification is:

$$Y_{ht} = \mathbb{1}[t \geqslant 2011] \times Z_h^{2011} \beta^{post} + \phi_{gt} + \psi_h + \varepsilon_{ht}. \tag{7}$$

5. *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 eight 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 I, Panel A 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 Online Appendix Table FII, 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. Although 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 that treatment effects are homogeneous. 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, whereas the RAC in Region C is a health care 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 a response to both 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 American Hospital Association's (AHA) 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 (de Chaisemartin and D'Haultfœuille 2018). One concern is that if hospitals on opposite sides of the border are very different at baseline, 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 (Online Appendix Table FII).

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-differences 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 Online Appendix C.1 that the results are not driven

24. Restricting the comparison to border hospitals allows me to make a weaker parallel-trends assumption than a comparison of all hospitals. Online Appendix Figure G8f 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.

by a particularly relevant confounder: the Medicare 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 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 (Online Appendix Figure G9). Finally, we need to make an assumption about monotonicity in audit intensity across RACs—that a particular hospital would be subject to more audits under a more aggressive RAC, and fewer audits under a less aggressive RAC (Imbens and Angrist 1994).

IV.C. Identifying the Effect of Monitoring on Patient Outcomes

I 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 postpolicy in a difference-in-differences specification.

1. *Specification.* The event study specification is:

$$(8) \quad Y_v = \sum_{\tau=2010}^{2016} \beta^{\tau} \mathbb{1}[y = \tau] \times \mathbb{1}[t \geqslant 00{:}00] + \boldsymbol{W}_v' \boldsymbol{\gamma} + \lambda_{hy} + \phi_{ht} + \varepsilon_v,$$

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 9 p.m. and 3 a.m.). 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 30 days. $\mathbb{1}[y = \tau]$ is an indicator for whether the visit occurred in

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

fiscal year τ, omitting 2013. $\mathbb{1}[t \geqslant 00{:}00]$ is an indicator for whether the patient arrived at the ED after midnight. λ_{hy} is a hospital-year fixed effect, and ϕ_{ht} is a hospital–ED-arrival-hour fixed effect. W_v are controls for patient characteristics, including patient age, race, Hispanic, point of origin, an indicator for whether their 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 postpolicy coefficient β:

$$(9) \quad Y_v = \beta \mathbb{1}[y \geqslant 2013Q3] \times \mathbb{1}[t \geqslant 00{:}00] + \boldsymbol{W}_{\boldsymbol{v}}'\boldsymbol{\gamma} + \lambda_{hq} + \phi_{ht} + \varepsilon_v.$$

Here $\mathbb{1}[y \geqslant 2013Q3]$ 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.

2. *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 pretrends 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 p.m. and midnight, once the policy is implemented. Online Appendix Figure G10 plots the share of patients by ED arrival hour, pre- and postpolicy—bunching before midnight does not appear postpolicy. I test this empirically in Online Appendix Table FIII 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)) postpolicy. 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. In addition, 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 (Online Appendix Table FIV), suggesting that hospitals did not respond to the two midnights rule by extending stay duration.

V. Results

V.A. *Hospital Outcomes: Admissions, Revenue, Costs, and IT Adoption*

Figure IV 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% 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 V 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 IV. Table III reports the yearly coefficients for 2011 to 2015.[26] Figure V, Panels A and B plot the results for log Medicare admissions and log Medicare inpatient revenue, where inpatient revenue is defined as the sum of all Medicare inpatient

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

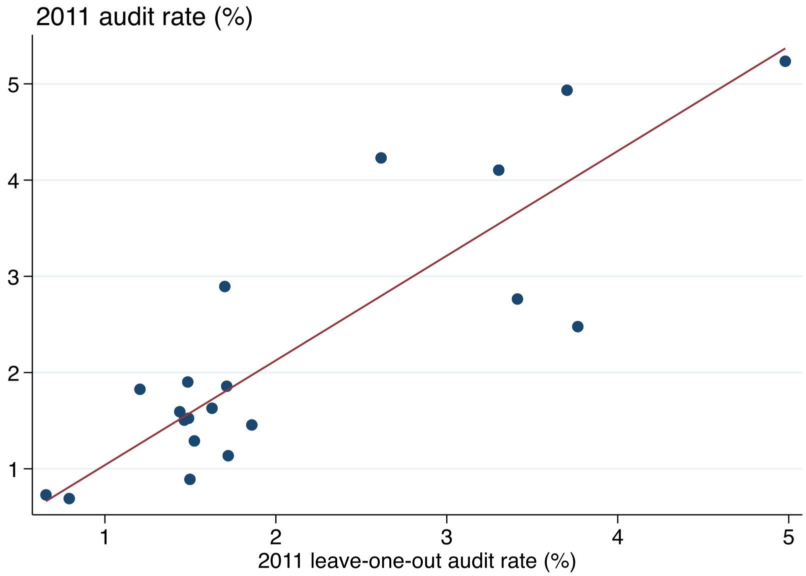


FIGURE IV

Binscatter of the 2011 Leave-One-Out State Audit Rate and the 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 100 miles of the RAC border with at least one hospital in their neighbor comparison group. Data: MEDPAR claims and CMS audit data.

payments. Before 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 1 percentage point increase in the 2011 audit rate results in a 1.1% decrease in admissions in 2011, which increases in magnitude to a 1.9% decrease by 2012 and 2013. Similarly, a 1 percentage point increase in the 2011 audit rate results in a 1.0% decrease in inpatient revenue in 2011, and then a 1.8% decrease by 2012 and a 2.8% decrease by 2013. Extrapolating to the overall hospital sample (albeit under fairly strong assumptions, as discussed in Online Appendix D) indicates

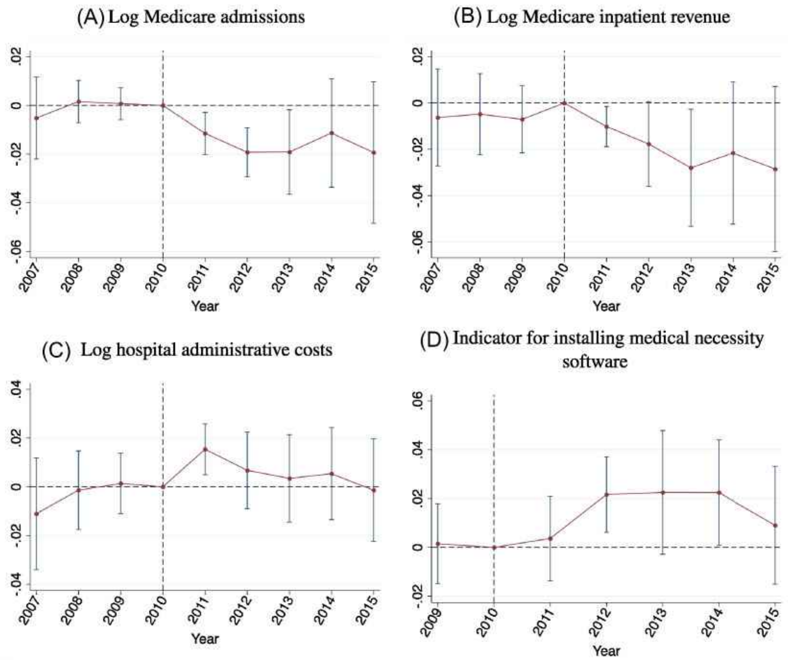


FIGURE V

Event Studies on the Effect of the 2011 Audit Rate on Hospital Outcomes

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 1 percentage point increase in the 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. The indicator for installing software is equal to one 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 100 miles of the RAC border with at least one hospital in their neighbor comparison group.

that RAC audits saved the Medicare program $9.28 billion between 2011 and 2015.

I turn to the administrative burden RAC audits impose on hospitals. Figure V and Table III, columns (5) and (6) present results on hospital administrative costs and IT adoption. Figure V, Panel C plots estimates of the effect on log administrative costs, as reported in hospital cost reports. A 1 percentage point increase in RAC auditing in 2011 results in an immediate 1.5% 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% 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 V, Panel D presents the event study results for whether a hospital reported installing medical necessity checking software in a given year. In response to a 1 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% increase relative to the 59% 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, Online Appendix Figure G12 plots the results for the payments directly reclaimed by RACs. A 1 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% 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

27. For a 1 percentage point increase in the 2011 audit rate, the government costs (the contingency fees paid to the RACs) by 2015 are $88,000 and the direct savings from reclaimed payments are $232,000. 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% 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 Online Appendix 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 $721,000, 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% of the savings.

TABLE III

EVENT STUDIES OF THE EFFECT OF THE 2011 AUDIT RATE ON HOSPITAL OUTCOMES, 2011–2015 COEFFICIENTS

	Overall		LOS ⩽ 2		Admin costs	Software installation
	Log adm. (1)	Log rev. (2)	Log adm. (3)	Log rev. (4)	Log costs (5)	Medical necc. (6)
2011 audit rate	−0.0115**	−0.0102**	−0.0145*	−0.0120***	0.0154***	0.0037
× 2011	(0.0044)	(0.0044)	(0.0074)	(0.0039)	(0.0053)	(0.0088)
2011 audit rate	−0.0192***	−0.0177*	−0.0457***	−0.0460***	0.0068	0.0217**
× 2012	(0.0051)	(0.0093)	(0.0111)	(0.0056)	(0.0080)	(0.0079)
2011 audit rate	−0.0191**	−0.0280**	−0.0282***	−0.0364***	0.0034	0.0225*
× 2013	(0.0089)	(0.0129)	(0.0082)	(0.0103)	(0.0092)	(0.0129)
2011 audit rate	−0.0113	−0.0216	−0.0241**	−0.0329**	0.0054	0.0225*
× 2014	(0.0114)	(0.0157)	(0.0092)	(0.0120)	(0.0096)	(0.0110)
2011 audit rate	−0.0193	−0.0285	−0.0208*	−0.0282**	−0.0014	0.0090
× 2015	(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 posp	510	510	510	510	510	506
Obs.	52,139	52,139	52,139	52,118	52,107	36,906
F	12.5	12.5	12.5	13.36	12.45	13.87

Notes. 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 V, 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 1 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. 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 one 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 100 miles of the RAC border with at least one hospital in their neighbor comparison group. * $p < .10$; ** $p < .05$; *** $p < .01$.

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 policy makers' concerns about short stays being the primary driver of unnecessary stays, Figure VI splits admissions by their length of stay. Online Appendix Figure G13a 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 made up 31% of stays on average in 2010. A 1 percentage point increase in the audit rate results in a 4.6% decrease in short-stay admissions and a 4.6% 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 VII 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 DRGs[29] by payment error calculated by the Comprehensive Error Rate Testing (CERT) Program in 2010, a Medicare program that randomly samples claims to calculate improper payment rates (Centers for Medicare and Medicaid Services 2011). 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. Online Appendix Figure G13b plots the audit rates for the top 20 highest error base DRGs. Figure VII, Panels A and B plot the event study results, which show larger and more sustained reductions in admissions for the top 20 base DRGs compared to DRGs

28. The value of compliance costs by 2015 is $455,000, compared to the total government savings of $2.08 million. Under the assumption that 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 $1,000 in Medicare savings.

29. DRGs can be categorized into groups of one to three 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 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).

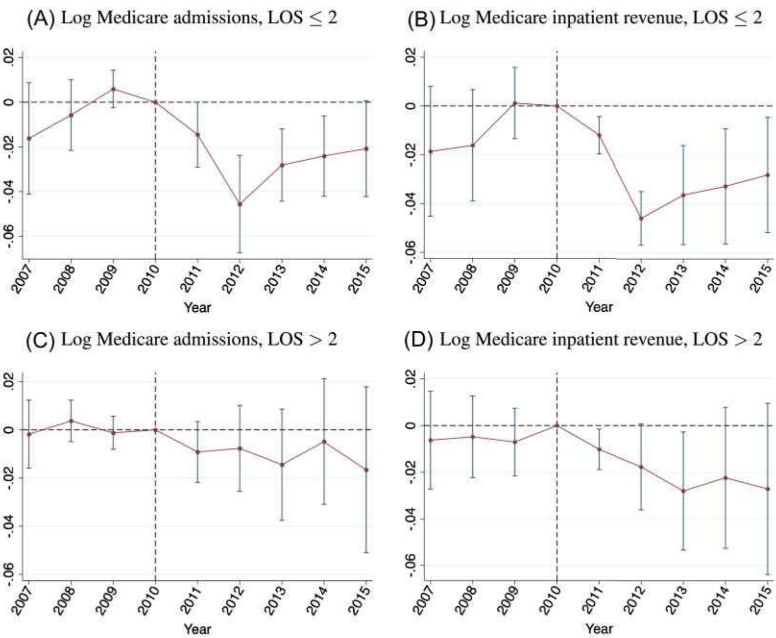


FIGURE VI

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

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 1 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 (LOS) 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 100 miles of the RAC border with at least one hospital in their neighbor comparison group.

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 policy makers 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).

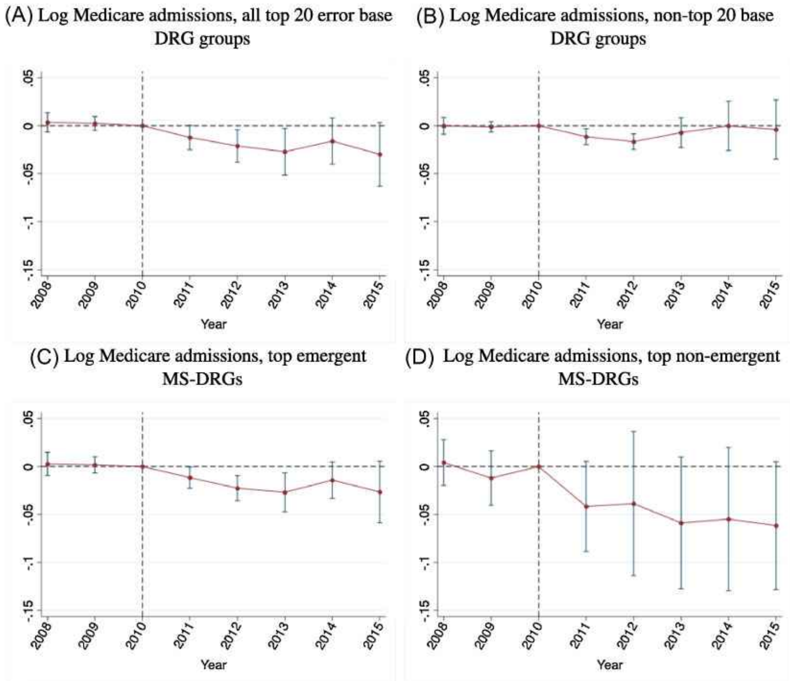


FIGURE VII

Event Studies on the Effect of the 2011 Audit Rate on Medicare Admissions, by Base Diagnosis Group Error Rates

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 1 percentage point increase in the 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 2011). 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 (ED rate 79%), chest pain (83%), GI hemorrhage (74%), respiratory infections (71%), esophagitis and miscellaneous digestive disorders (71%), kidney and UTI (69%), nutritional and metabolic (68%), renal failure (67%), syncope and collapse (78%), heart failure and shock (69%), cardiac arrhythmia (69%), pneumonia and pleurisy (65%), acute myocardial infarction (77%), chronic obstructive pulmonary disease (69%), hip and femur except major joint (82%), and intracranial hemorrhage or cerebral infarction (76%). Panel D plots admissions for the remaining four nonemergent MS-DRG base groups among the top 20: major joint replacement (ED rate 13%), permanent cardiac pacemaker (57%), drug-eluting stents (42%), and major bowel procedures (38%). The sample comprises hospitals within 100 miles of the RAC border with at least one hospital in their neighbor comparison group. Data: MEDPAR and CMS audit data.

The list of top 20 base DRGs includes both emergent (i.e., arising from an emergency) and nonemergent 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 nonemergent stays differ 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 Figure VII, Panels C and D, I split the top 20 base DRG groups into emergent and nonemergent diagnoses.[30] There are reductions in both emergent and nonemergent cases, with a larger effect (but noisier) for nonemergent stays of 5.1% after 2015 compared to a 2.1% decrease among emergent stays.

The fact that both emergent and nonemergent admissions decrease indicates that the overall reduction in admissions was not attained by adopting medical necessity checking software alone. The software is most useful for emergent cases, as its purpose is to relay information to on-the-ground providers as they make care decisions in real time. But the decision to reduce nonemergent 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 nonsurvey data.

The event studies in Figures V, VI, and VII 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

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

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 I, Panel B). 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 Online Appendix Figure G14, 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 in 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, Bubolz, and Skinner 2021; Leder-Luis forthcoming). This slower decline is consistent with hospitals needing time to implement improvements in their admissions processes, like incorporating newly installed software.

Online Appendix Table FV 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% reduction in overall admissions (although not statistically significant) and a 2.2% reduction in short-stay admissions relative to the preperiod. Online Appendix Table FVI considers heterogeneity in the effect by hospital characteristics. The results point to rural, for-profit, smaller, and nonchain 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 Online Appendix 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 Online Appendix C.2, I also consider the effect of RAC audits on coding. In addition to conducting audits for medical necessity, RACs could audit for coding errors such as upcoding.[32] 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, even though 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 Online 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.

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

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).

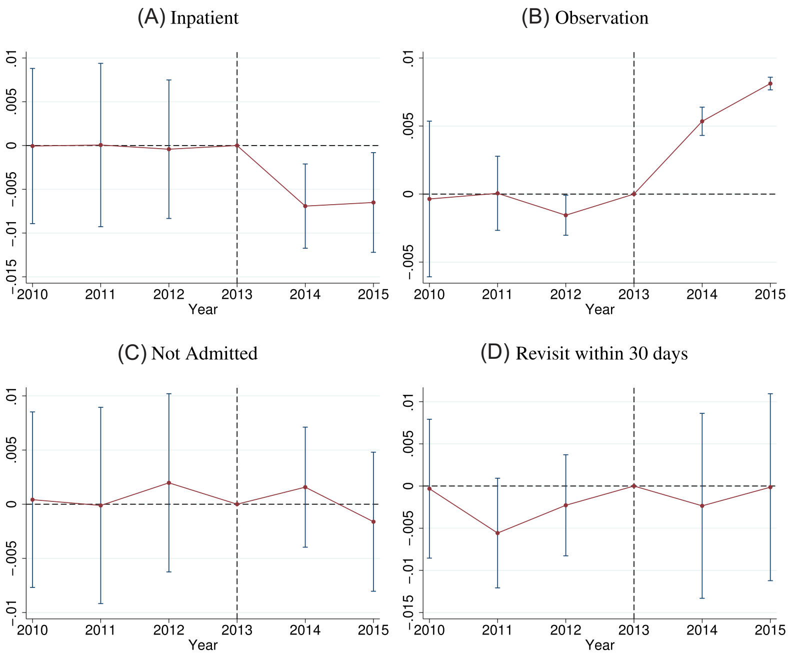


FIGURE VIII

Event Studies on the Effect of After-Midnight ED Arrival on Patient Status and Outcomes

This figure plots the coefficients and 95% confidence intervals for β^{τ} on $\mathbb{1}[y = \tau] \times \mathbb{1}[T_v \geqslant 00{:}00]$ 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 $\mathbb{1}[T_v \geqslant 00{:}00]$ 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 an 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. The sample consists of traditional Medicare patients who arrived in the ED within three hours of midnight in a Florida hospital. The 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.

V.B. *Patient Outcomes: Inpatient Admission Likelihood and Revisit Likelihood*

I turn to the results from the patient-level analysis. Figure VIII plots the event studies of the patient-level analysis of ED visits in equation (8). There is no clear trend in the prepolicy 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%) less likely to be admitted as inpatient and 0.7 percentage points (16.7%) 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. Figure VIII, Panel D plots the event study results for an indicator of whether a patient revisited a hospital within 30 days of her ED visit, and Table IV, column (4) 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 III 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 a.m. and 3:00 p.m. (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 30 days of the visit, based on information

TABLE IV

AFTER-MIDNIGHT ED ARRIVAL HOUR DIFFERENCE-IN-DIFFERENCES COEFFICIENTS ON PATIENT STATUS AND REVISITS

	Medicare				Non-Medicare
	Inpatient (1)	Observation (2)	Not admitted (3)	Revisit 30d (4)	Inpatient (5)
β	-0.007^{***}	0.007^{***}	0.000	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Prereform 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	1,254,857	1,254,857	1,254,857	1,254,857	7,428,583

Notes. Standard errors are in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[y \geqslant 2013Q3] \times \mathbb{1}[T_v \geqslant 00{:}00]$ of the specification in equation (9), where $\mathbb{1}[y \geqslant 2013Q3]$ is an indicator for whether the visit occurred after the two midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geqslant 00{:}00]$ 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 an 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. The sample for columns (1)–(4) consists of traditional Medicare patients who arrived in the ED within three hours of midnight in a Florida hospital. The sample for column (5) consists of all non-Medicare patients who arrived in the ED within three 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. * $p < .10$, ** $p < .05$, *** $p < .01$.

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.

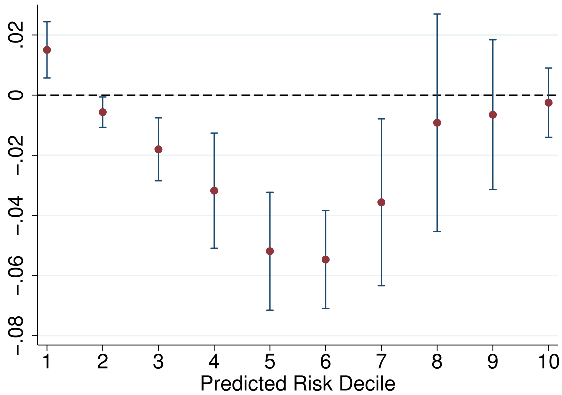
Figure IX plots the heterogeneity by severity results for inpatient status and for revisits within 30 days. The two midnights rule has no effect on admission rates for patients at the bottom and top severity deciles. Instead, the reduction in admissions stems primarily from the middle of the severity distribution. There is a 5 percentage point, or 25%, 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 robustness check, 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 Online Appendix Table FVII, and likewise find no effect on revisits on these subpopulations.

Online Appendix Table FVIII 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. Online Appendix 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.

V.C. Discussion

1. *Technology Adoption Mechanism.* Taken together, the hospital-level and patient-level results underscore the role that

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.



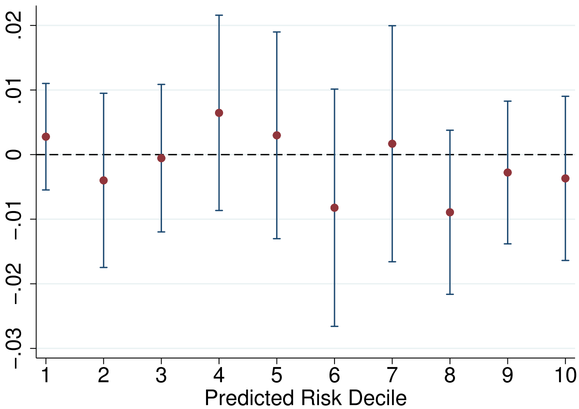


FIGURE IX

Heterogeneity of After-Midnight ED Arrival Coefficient by Patient Severity

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 $\mathbb{1}[y \geqslant 2013Q3] \times \mathbb{1}[T_v \geqslant 00{:}00]$, where $\mathbb{1}[y \geqslant 2013Q3]$ is an indicator for whether the visit occurred after 2013Q3, and $\mathbb{1}[T_v \geqslant 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. Panel A plots results for an indicator for whether the patient was admitted as inpatient from the ED, and Panel B 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 9:00 a.m. and 3:00 p.m. of an indicator for being admitted

FIGURE IX

(*Continued*) 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.

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 this software targets. They also show that the response to the two midnights rule was driven by hospitals with this software already installed.

Online Appendix Figure G11 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 when RACs focused on unnecessary admissions (Online Appendix Figure G11a). 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 in the same RAC region, those that were more heavily penalized by RACs were more likely to adopt the software in later years (Online Appendix Figure G11b). In 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 (Online Appendix Figure G11c). 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.

2. *Comparing the Two Approaches.* There are also important differences between the hospital-level and patient-level results that 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%) of Medicare admissions originate in the ED, and Online Appendix Table FI 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 nonemergent 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 nonemergent stays. Online Appendix Figure FI 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 (Online Appendix Figure G16).[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 III), possibly because

34. However, due to possible mismeasurement of emergency department visits in Medicare claims data, these results should be interpreted as suggestive. This is explained in further detail in Online Appendix E.

their audit rate decreases. We can use the model discussed in Section III 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 II, Panel C, 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 multiyear 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 focus on monitoring aggressively up-front 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, Gowrisankaran, and Langer 2020).

There is an additional distinction between the two policy environments that could explain this discrepancy: the level of confidence hospitals had in whether they could be retroactively punished in the future. With the two midnights rule, hospitals could be fairly confident that their admissions would be protected by the rule from future audits. However with the 2014 pause, hospitals could not be sure that auditing would not increase again in later years. RACs had a lookback period of three 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 to before. But it is unlikely that

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% prepay reviews.

hospitals could have anticipated this trajectory for the RAC program at the onset of the pause.

VI. Conclusion

In this article I consider the trade-offs 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 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% 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 nonemergent 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 that varied patients' audit rate depending on when a patient arrives at 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.

University of Chicago and National Bureau of Economic Research, United States

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

Data Availability

The data underlying this article are available in the Harvard Dataverse, https://doi.org/10.7910/DVN/WS9OAQ (Shi 2023).

References

3M, “3M APCfinderTM Software with Medical Necessity Validation,” 2016. https://multimedia.3m.com/mws/media/204326O/3m-apcfinder-with-mn-fact-sheet.pdf

AccuReg, “EngageCare Provider Medical Necessity,” 2022. https://accuregsoftware.com/wp-content/uploads/2021/04/EngageCare-Medical-Necessity.pdf

Alexander, Diane, and Molly Schnell, “The Impacts of Physician Payments on Patient Access, Use, and Health,” NBER Working Paper no. 26095, 2019. https://doi.org/10.3386/w26095

Allingham, Michael G., and Agnar Sandmo, “Income Tax Evasion: A Theoretical Analysis,” *Journal of Public Economics*, 1 (1972), 323–338. https://doi.org/10.1016/0047-2727(72)90010-2

Almond, Douglas, and Joseph J. Doyle, “After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays,” *American Economic Journal: Economic Policy*, 3 (2011), 1–34. https://doi.org/10.1257/pol.3.3.1

American Hospital Association, “Exploring the Impact of the RAC Program on Hospitals Nationwide: Results of AHA RACTrac Survey, 1st Quarter 2012,” (2012). https://www.aha.org/guidesreports/2012-02-15-ractrac-survey-exploring-impact-rac-program-hospitals-nationwide

———, “Exploring the Impact of the RAC Program on Hospitals Nationwide: Results of AHA RACTrac Survey, 4th Quarter of 2012,” 2013. https://www.aha.org/system/files/content/11/11Q4ractracresults.pdf

Andreoni, James, Brian Erard, and Jonathan Feinstein, “Tax Compliance,” *Journal of Economic Literature*, 36 (1998), 818–860. https://www.jstor.org/stable/2565123

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*, 40 (2019), 487–500. https://doi.org/10.1146/annurev-publhealth-040218-044206

Baron, David P., and David Besanko, “Regulation, Asymmetric Information, and Auditing,” *RAND Journal of Economics*, 15 (1984), 447–470. https://www.jstor.org/stable/2555518

Becker, Gary S., “Crime and Punishment: An Economic Approach,” *Journal of Political Economy*, 76 (1968), 169–217. https://doi.org/10.1086/259394

Blackwell, David, “Comparison of Experiments,” in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2 (Berkeley: University of California Press, 1951), 93–103. https://projecteuclid.org/proceedings/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Second-Berkeley-Symposium-on-Mathematical-Statistics-and/Chapter/Comparison-of-Experiments/bsmsp/1200500222

———, “Equivalent Comparisons of Experiments,” *Annals of Mathematical Statistics*, 24 (1953), 265–272. https://www.jstor.org/stable/2236332

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer, “Escalation of Scrutiny: The Gains from Dynamic Enforcement of Environmental Regulations,” *American Economic Review*, 110 (2020), 2558–2585. https://doi.org/10.1257/aer.20181012

Brot-Goldberg, Zarek C., Samantha Burn, Timothy Layton, and Boris Vabson, "Rationing Medicine Through Bureaucracy: Authorization Restrictions in Medicare," NBER Working Paper no. 30878, 2023. https://doi.org/10.3386/w30878

Burde, Howard, "Health Law. The HITECH Act—An Overview," *Virtual Mentor*, 13 (2011), 172–175. https://doi.org/10.1001/virtualmentor.2011.13.3.hlaw1-1103

Centers for Medicare and Medicaid Services, "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.

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

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

——— "Improper Payment Rates and Additional Data," Technical report, 2022.

Chan, David C., "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, 124 (2016), 734–770. https://doi.org/10.1086/685910

Chan, David C., and Jonathan Gruber, "Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions," *AEA Papers and Proceedings*, 110 (2020), 279–283. https://doi.org/10.1257/pandp.20201032

Clemens, Jeffrey, and Joshua D. Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 104 (2014), 1320–1349. https://doi.org/10.1257/aer.104.4.1320

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," *Quarterly Journal of Economics*, 134 (2019), 51–107. https://doi.org/10.1093/qje/qjy020

Cubanski, Juliette, Tricia Neuman, and Meredith Freed, "The Facts on Medicare Spending and Financing," Kaiser Family Foundation Issue Brief, 2019. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101754945-pdf

Cunningham, Susan, Mark McMillan, Sara Slayton O'Rourke, and Eric Schweikert, "Cracking Down on Government Fraud with Data Analytics," McKinsey Report, 2018. https://www.mckinsey.com/industries/public-sector/our-insights/cracking-down-on-government-fraud-with-data-analytics

Cutler, David, "The Incidence of Adverse Medical Outcomes under Prospective Payment," *Econometrica*, 63 (1995), 29–50. https://doi.org/10.2307/2951696

Cutler, David M., and Dan P. Ly, "The (Paper)Work of Medicine: Understanding International Medical Costs," *Journal of Economic Perspectives*, 25 (2011), 3–25. https://doi.org/10.1257/jep.25.2.3

Dafny, Leemore S., "How Do Hospitals Respond to Price Changes?," *American Economic Review*, 95 (2005), 1525–1547. https://doi.org/10.1257/000282805775014236

de Chaisemartin, Clément, and Xavier D'Haultfœuille, "Fuzzy Differences-in-Differences," *Review of Economic Studies*, 85 (2018), 999–1028. https://doi.org/10.1093/restud/rdx049

Deshpande, Manasi, and Yue Li, "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 11 (2019), 213–248. https://doi.org/10.1257/pol.20180076

Di Tella, Rafael, and Ernesto Schargrodsky, "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack," *American Economic Review*, 94 (2004), 115–133. https://doi.org/10.1257/000282804322970733

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*, 123 (2015), 170–214. https://doi.org/10.1086/677756
Dranove, David, Chris Forman, Avi Goldfarb, and Shane Greenstein, "The Trillion Dollar Conundrum: Complementarities and Health Information Technology," *American Economic Journal: Economic Policy*, 6 (2014), 239–270. https://doi.org/10.1257/pol.6.4.239
Dube, Arindrajit, T. William Lester, and Michael Reich, "Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties," *Review of Economics and Statistics*, 92 (2010), 945–964. https://doi.org/10.1162/REST_a_00039
Dunn, Abe, Joshua D. Gottlieb, Adam Hale Shapiro, Daniel J. Sonnenstuhl, and Pietro Tebaldi, "A Denial a Day Keeps the Doctor Away," *Quarterly Journal of Economics*, 139 (2024), https://doi.org/10.1093/qje/qjad035
Dunn, Abe, Joshua D. Gottlieb, and Adam Hale Shapiro, "Administration above Administrators: The Changing Technology of Health Care Management," *AEA Papers and Proceedings*, 110 (2020), 274–278. https://doi.org/10.1257/pandp.20201031
Einav, Liran, Amy Finkelstein, and Neale Mahoney, "Provider Incentives and Healthcare Costs: Evidence From Long–Term Care Hospitals," *Econometrica*, 86 (2018), 2161–2219. https://doi.org/10.3982/ECTA15022
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*, 108 (2018), 3232–3265. https://doi.org/10.1257/aer.20170092
Ellis, Randall P., and Thomas G. McGuire, "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply," *Journal of Health Economics*, 5 (1986), 129–151. https://doi.org/10.1016/0167-6296(86)90002-0
———, "Hospital Response to Prospective Payment: Moral Hazard, Selection, and Practice-Style Effects," *Journal of Health Economics*, 15 (1996), 257–277. https://doi.org/10.1016/0167-6296(96)00002-1
Experian Health, "Product Sheet: Medical Necessity," (2022). https://www.experian.com/content/dam/marketing/na/healthcare/brochures/medical-necessity-brochure.pdf
Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Sources of Geographic Variation in Health Care: Evidence from Patient Migration," *Quarterly Journal of Economics*, 131 (2016), 1681–1726. https://doi.org/10.1093/qje/qjw023
Finkelstein, Amy, and Matthew J. Notowidigdo, "Take-Up and Targeting: Experimental Evidence from SNAP," *Quarterly Journal of Economics*, 134 (2019), 1505–1556. https://doi.org/10.1093/qje/qjz013
Foster, Anne, and B. Scott McBride, "CMS Announces Changes to RAC Program; Temporary Pause in Document Requests," *JD Supra*, March 7, 2014. https://www.jdsupra.com/legalnews/cms-announces-changes-to-rac-program-te-07317/.
Gottlieb, Joshua D., Adam Hale Shapiro, and Abe Dunn, "The Complexity of Billing and Paying for Physician Care," *Health Affairs*, 37 (2018), 619–626. https://doi.org/10.1377/hlthaff.2017.1325
Gross, Tal, Adam Sacarny, Maggie Shi, and David Silver, "Regulated Revenues and Hospital Behavior: Evidence from a Medicare Overhaul," *Review of Economics and Statistics*, (forthcoming). https://doi.org/10.1162/rest_a_01254
Gupta, Atul, "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program," *American Economic Review*, 111 (2021), 1241–1283. https://doi.org/10.1257/aer.20171825
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?," RAND Corporation Technical report, 2005. https://doi.org/10.7249/RB9136

Howard, David H., and Ian McCarthy, “Deterrence Effects of Antifraud and Abuse Enforcement in Health Care,” *Journal of Health Economics*, 75 (2021), 102405. https://doi.org/10.1016/j.jhealeco.2020.102405

Imbens, Guido W., and Joshua D. Angrist, “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62 (1994), 467–475. https://doi.org/10.2307/2951620

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*, 79 (2011), 651–692. https://doi.org/10.3982/ECTA9113

Kopczuk, Wojciech, “Tax Simplification and Tax Compliance: An Economic Perspective,” Columbia University Working Paper, 2007.

Kopczuk, Wojciech, and Cristian Pop-Eleches, “Electronic Filing, Tax Preparers and Participation in the Earned Income Tax Credit,” *Journal of Public Economics*, 91 (2007), 1351–1367. https://doi.org/10.1016/j.jpubeco.2006.11.003

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*, 377 (2017), 1055–1064. https://doi.org/10.1056/NEJMsa1702321

Laffont, Jean-Jacques, and Jean Tirole, “Using Cost Observation to Regulate Firms,” *Journal of Political Economy*, 94 (1986), 614–641. https://doi.org/10.1086/261392

League, Riley, “Administrative Burden and Consolidation in Health Care: Evidence from Medicare Contractor Transitions,” Working Paper, 2023. https://rileyleague.github.io/files/MAC_transitions.pdf

Leder-Luis, Jetson, “Can Whistleblowers Root Out Public Expenditure Fraud? Evidence from Medicare,” *Review of Economics and Statistics* (forthcoming). https://doi.org/10.1162/rest_a_01339

Meckel, Katherine, “Is the Cure Worse than the Disease? Unintended Effects of Payment Reform in a Quantity-Based Transfer Program,” *American Economic Review*, 110 (2020), 1821–1865. https://doi.org/10.1257/aer.20160164

Medicare Payment Advisory Commission, “Hospital Short-Stay Policy Issues,” Technical report, 2015.

———, “July 2020 Data Book: Health Care Spending and the Medicare Program,” Technical report, 2020.

Nicholas, Lauren Hersch, Caroline Hanson, Jodi B. Segal, and Matthew D. Eisenberg, “Association between Treatment by Fraud and Abuse Perpetrators and Health Outcomes among Medicare Beneficiaries,” *JAMA Internal Medicine*, 180 (2020), 62–69. https://doi.org/10.1001/jamainternmed.2019.4771

Office of Management and Budget, “Annual Improper Payments Datasets,” Technical report, 2022.

Office of the Inspector General, “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, 2019.

———, “HHS-OIG Strategic Plan 2020–2025,” Technical report, 2020.

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

Olken, Benjamin A., “Monitoring Corruption: Evidence from a Field Experiment in Indonesia,” *Journal of Political Economy*, 115 (2007), 200–249. https://doi.org/10.1086/517935

O’Malley, A. James, Thomas A. Bubolz, and Jonathan S. Skinner, “The Diffusion of Health Care Fraud: A Bipartite Network Analysis,” *Social Science Medicine*, 327 (2023), 115927. https://doi.org/10.1016/j.socscimed.2023.115927

ResDAC, "Anomaly in MedPAR/Inpatient Claims When Counting Emergency Department Visits," correspondence with the author, 2022. https://mshi311.github.io/website/Shi_ResDAC_email.pdf
Roberts, James, Ryan McDevitt, Paul Eliason, Jetson Leder-Luis, and Riley League, "Enforcement and Deterrence of Medicare Fraud: The Case of Nonemergent Ambulance Rides," NBER Working Paper no. 29491, 2021.
Rose, Liam, "The Effects of Skilled Nursing Facility Care: Regression Discontinuity Evidence from Medicare," *American Journal of Health Economics*, 6 (2020), 39–71. https://doi.org/10.1086/706864
Sacarny, Adam, "Adoption and Learning across Hospitals: The Case of a Revenue-Generating Practice," *Journal of Health Economics*, 60 (2018), 142–164. https://doi.org/10.1016/j.jhealeco.2018.06.005
Shi, Maggie, "Replication Data for: 'Monitoring for Waste: Evidence from Medicare Audits'," (2023), Harvard Dataverse, https://doi.org/10.7910/DVN/WS9OAQ.
Silverman, Elaine, and Jonathan Skinner, "Medicare Upcoding and Hospital Ownership," *Journal of Health Economics*, 23 (2004), 369–389. https://doi.org/10.1016/j.jhealeco.2003.09.007
Skinner, Jonathan, "Causes and Consequences of Regional Variations in Health Care," in *Handbook of Health Economics*, Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds. (Amsterdam: Elsevier, 2011), 45–93. https://doi.org/10.1016/B978-0-444-53592-4.00002-5
Slemrod, Joel, and Shlomo Yitzhaki, "Tax Avoidance, Evasion, and Administration," in *Handbook of Public Economics*, vol. 3, A. J. Auerbach and Martin Feldstein, eds. (Amsterdam: Elsevier, 2002), 1423–1470. https://doi.org/10.1016/S1573-4420(02)80026-X
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 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*, 24 (2017), 422–431. https://doi.org/10.1111/acem.13140
Zwick, Eric, "The Costs of Corporate Tax Complexity," *American Economic Journal: Economic Policy*, 13 (2021), 467–500. https://doi.org/10.1257/pol.20180406

ONLINE APPENDIX: FOR ONLINE PUBLICATION ONLY

Appendix to: "Monitoring for Waste: Evidence from Medicare Audits"

Maggie Shi

University of Chicago Harris School of Public Policy and NBER

A Policy Context

A.1 Medicare Inpatient Prospective Payment System and Short Stays

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

The prospective payment system is in general perceived to work well at keeping inpatient hospital spending relatively low for the Medicare program [Lopez et al., 2020]. However, one persistent issue with IPPS that has been noted by policymakers is the high number of short stays. A CMS report found that "a large percentage of medically unnecessary [payment] errors are related to hospital stays of short duration... these services should have been rendered at a lower level of care" [Centers for Medicare and Medicaid Services, 2011b]. One less intensive alternative to an inpatient stay is an outpatient observation stay, which consists of short-term (often diagnostic) services provided at the hospital while a physician decides whether to formally admit a patient as inpatient or send them home. Observation stays typically last less than forty-eight hours and are billed as an outpatient service [Medicare Payment Advisory Commission, 2015].

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

Policymakers considered various alternative solutions to address unnecessary short stays before settling on RAC audits. They were wary of reducing the payment rate for short stays or penalizing high rates of short stays, due to concerns that hospitals would simply keep patients for longer to evade these policies [Medicare Payment Advisory Commission, 2015].[36] There is evidence that length of stay is highly malleable in the face of reimbursement "jumps" in other contexts [Einav et al., 2018; Eliason et al., 2018]. Additionally, short stays constitute almost a third of inpatient stays. Their prevalence suggests that not all short stays are unnecessary, and cutting payments for short stays across the board would reduce payments for some necessary stays.

Aside from the RAC program, Medicare enacted other monitoring and education programs to measure and mitigate unnecessary inpatient stays. They measured payment errors across different discharge and service types through the Comprehensive Error Rate Testing Program (CERT) in 2010, which randomly samples Medicare claims to calculate improper payment rates [Centers for Medicare and Medicaid Services, 2011b]. The CERT reports then informed provider education programs, like the "Targeted Probe and Educate" program, which involves claim reviews and one-on-one education sessions for providers, as well as the PEPPER and Comparative Billing Reports (CBR) programs which distributes provider-specific reports on which of their discharge types and services were most vulnerable to improper payments. See the CMS websites for the TPE program (link; last accessed September 2023) and the PEPPER and CBR programs (link; last accessed September 2023).

[36] In testimony to Congress, MedPAC's executive director stated concerns that this policy "would create a financial incentive to extend an inpatient stay from one to two days" [Miller, 2015].

A.2 RAC Program Details

RAC Regions In the context of medical claims processing and reviews, the jurisdictions used for RAC regions are unique, though they do share some overlapping boundaries with Part B Medicare Administrative Contractors (MACs), who primarily process Medicare claims and can deny claims before payment. The RAC regions do align exactly with the regions of Durable Medical Equipment MACs. However, the DME MACs only process payments for durable medical equipment like prosthetics, orthotics, and other devices, and they do not process claims for medical services [Medicare Contractor Management Group, 2017]. Medicare posts a separate contract solicitation for each region, and firms submit separate bids.

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

RAC Audit Process RACs conduct postpayment reviews to identify and correct overpayments or underpayments for claims for inpatient care, outpatient care, long-term care, and durable medical equipment in the last three years. Figure G1 illustrates the claims auditing and appeals process, using 2011 inpatient audit rates as an example. Each RAC develops and runs its own proprietary algorithm on claims data to identify claims with potential payment errors. In 2011, RACs' auditing scope for inpatient claims included incorrect or incomplete coding, DRG validation, and medical necessity reviews. Five percent of audits were "automated reviews," which rely solely on claims data to make a determination based on clearly outlined Medicare policies. The rest of the audits were "complex reviews," in which a medical professional (for example, coder, nurse, or therapist) employed by the RAC submits a medical record request and manually reviews all documentation associated with an inpatient stay. It is up to the medical professional to determine whether an overpayment or underpayment was made. If they find an error, then they can demand that the provider repays Medicare (or vice versa). Providers can appeal demands by first requesting a redetermination by the RAC and then escalating it to higher levels of appeals – for example, by requesting that a separate contractor reconsider the case, requesting a hearing by an administrative law judge, or escalating it to a review by the Medicare Appeals Council.

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

In the face of a sudden rise in auditing and overpayment demands, hospitals began mounting a campaign to fight back. Hospitals started appealing high volumes of RAC determinations, and some hospital systems worked with the American Hospital Association (AHA) to file lawsuits and complaints against Medicare over RAC audits.[37] Between 2011 and 2013, the number of appeals that reached the administrative-judge level of the appeals process increased by 500 percent, and by mid-2014 there was a backlog of 800,000 appeals at that level [Medicare Payment Advisory Commission, 2015]. The AHA also began tracking the effect of RAC activity on its own through the quarterly RACTrac Survey of hospitals. Many hospitals reported that RAC audits imposed significant administrative burdens on them; for example, 11 percent of hospitals reported costs associated with managing the RAC program of over $100,000 [American Hospital Association, 2014].

Hospitals and industry stakeholders filed several complaints with Medicare stating that RAC

[37] See the AHA website for a list of all past and ongoing litigation: (link; last accessed September 2023).

audits were overly aggressive. As a result, in 2014 Medicare paused almost all RAC audits by significantly limiting their scope [Foster and McBride, 2014]. Other Medicare contractors such as MACs picked up additional review responsibilities after the RAC audits were paused.[38] Medicare maintained that the pause on RAC audits was temporary and would resume at previous levels, but it is clear from Figure 1b that RAC auditing never returned to its peak level after the pause. The pause began at the end of 2014Q1 and was originally meant to end in 2014Q3. After several quarters of delayed resumption, inpatient RAC audits finally resumed in 2015Q4, although RACs were much more constrained in how many audits they could conduct compared to before.

The pause occurred because CMS came under intense pressure to scale back the RAC program as the AHA began to organize its members together to file lawsuits and lodge complaints about RAC audits. Hospitals also coordinated a "DDOS attack"-style campaign to overwhelm the RAC appeals process [Bagley, 2014]. Between 2011 and 2013, the number of appeals that reached the administrative-judge level of the appeals process increased by 500 percent, and by mid-2014 there was a backlog of 800,000 appeals at that level [Medicare Payment Advisory Commission, 2015]. In response, CMS announced a one-time option to settle appeals by offering hospitals 68 percent of each appealed denied inpatient claim, in exchange for hospitals dropping all of their appeals rather than settling them one by one. As a result, hospitals dropped almost 350,000 appeals in exchange for $1.5 billion in settled denials [Centers for Medicare and Medicaid Services, 2014].

A.3 Characteristics of Audits and Audited Hospitals

Given Medicare policymakers' focus on short stays as the main source of unnecessary admissions, I examine audit frequency as a function of an admission's length of stay in Figure G13. Admissions with a length of stay of two or fewer days have much higher rates of auditing than longer admissions. The majority of audit recoveries of short stays result in the full payment being reclaimed. I also consider audit frequencies by base DRG group. Section 5 discusses how these groups were

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

ranked by CMS by severity of payemnt errors.

I next consider hospital-level characteristics and their correlation with audit rate in Figure G2. The RAC region a hospital is in is highly correlated with its audit rate. 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. Although almost every hospital was subject to an audit by 2020, in any given year there is a substantial portion of hospitals that do not face any audits. In 2011, 15 percent of hospitals had an audit rate of 0 percent. The share of hospitals with no audits varies across RAC regions from 2 to 23 percent.

B Model

B.1 Model Setup and Predictions

The hospital chooses its admission threshold τ^* to maximize its expected payoff:

$$\underbrace{E\left[U(\tau)\right]}_{\substack{\text{expected payoff} \\ \text{with threshold } \tau}} = \underbrace{\alpha B(\tau)}_{\substack{\text{value of total} \\ \text{patient benefit}}} + \underbrace{\underbrace{Rq(\tau)}_{\text{reimbursement}} - \underbrace{\frac{1}{2}Cq(\tau)^2}_{\substack{\text{treatment} \\ \text{costs}}} - \underbrace{\gamma\pi(\tau)}_{\substack{\text{audit penalties} \\ \text{for } x < \underline{h}}}}_{\text{hospital profit}} \tag{10}$$

where:

- Admission likelihood: $P(x;\tau) = 1 - \Phi\left(\frac{\tau - x}{\sigma}\right)$
- Total patient benefit: $B(\tau) = \int_{-\infty}^{\infty} xP(x;\tau)dF(x)$
- Number of admissions: $q(\tau) = \int_{-\infty}^{\infty} P(x;\tau)dF(x)$
- Number of admissions penalized by an audit: $\pi(\tau) = \int_{-\infty}^{\underline{h}} P(x;\tau)dF(x)$

In order to ensure that τ^* is a local maximum, $U(\tau)$ must be globally convex so that $U''(\tau^*) \leq 0$. The first order condition at the equilibrium τ^* is:

$$\underbrace{\alpha B'(\tau^*) + Rq'(\tau^*)}_{\text{benefit of increasing admission threshold}} = \underbrace{Cq(\tau^*)q'(\tau^*) + \gamma\pi'(\tau^*)}_{\text{cost of increasing admission threshold}} \tag{11}$$

Model Prediction 1. *Holding fixed a hospital's technology decision, increased auditing reduces admissions and the decline will be more pronounced for low-benefit admissions.*

Increased auditing reduces admissions: As the expected penalty γ increases, the cost of increasing the admission threshold (and thus admitting fewer patients) decreases because $\pi'(\tau^*) < 0$. So as the audit rate increases, the threshold increases and admissions decrease. This can be shown by applying the implicit function theorem to Equation 11, which shows that $\frac{d\tau^*}{d\gamma} = -\frac{dU_\tau/d\gamma}{dU_\tau/d\tau} = \frac{\pi_\tau}{U_{\tau\tau}} > 0$. The denominator $U_{\tau\tau}$ is negative since τ^* is at a local maximum and $U(\tau)$ is globally convex, and the numerator is also negative because $\pi_\tau < 0$. Since $q(\tau)$ is decreasing in τ, then as γ increases, admissions decrease.

The decline will be more pronounced for low-benefit admissions: This can be shown by taking the cross derivative of the admission likelihood with respect to τ and x: $\frac{d^2P(x;\tau)}{d\tau dx} = -(\frac{\tau-x}{\sigma^3})\phi(\frac{\tau-x}{\sigma})$. This is negative when $x < \tau$ and positive when $x > \tau$. Since $\frac{dP}{d\tau} < 0$, this implies that the reduction in admission likelihood from an increase in τ is larger for low-benefit admissions (i.e., more negative) than for high-benefit ones. □

As σ^2 decreases, the hospital's expected payoff increases by Blackwell's informativeness theorem [Blackwell, 1951, 1953]. This theorem implies that a Bayesian agent using a signal of the state of the world to make a decision under uncertainty will have strictly greater payoff using a signal s over using a mean-preserving spread s', where s' has the same distribution as $s + \varepsilon$ and $E(\varepsilon|s) = 0$ [Bergin, 2005]. In other words, the agent will always prefer the less noisy signal.

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 as K. If hospitals face a distribution of adoption costs, then as the difference between payoffs increases, more hospitals will adopt technology.

$$\underbrace{K}_{\substack{\text{threshold} \\ \text{adoption cost}}} = \underbrace{\max_\tau E[U(\tau;\sigma_L)]}_{\text{payoff with tech}} - \underbrace{\max_\tau E[U(\tau;\sigma_H)]}_{\text{payoff without tech}}. \tag{12}$$

Model Prediction 2. *If technology reduces audit penalties, then increasing the audit rate leads to* *more technology adoption.*

This can be shown by applying the envelope theorem to Equation 12:

$$\begin{aligned}\frac{dK}{d\gamma} &= \underbrace{\frac{dU(\tau;\sigma_L)}{d\tau}\frac{d\tau}{d\gamma}}_{=0} + \frac{dU(\tau;\sigma_L)}{d\gamma} - \underbrace{\frac{dU(\tau;\sigma_H)}{d\tau}\frac{d\tau}{d\gamma}}_{=0} - \frac{dU(\tau;\sigma_H)}{d\gamma} \\ &= \frac{dU(\tau;\sigma_L)}{d\gamma} - \frac{dU(\tau;\sigma_H)}{d\gamma} \\ &= \pi(\tau_H;\sigma_H) - \pi(\tau_L;\sigma_L),\end{aligned} \tag{13}$$

where τ_H is the optimal threshold without technology and τ_L is the optimal threshold with technology. This is proportional to the difference in expected audit penalties with and without technology:

$$\gamma\frac{dK}{d\gamma} = \underbrace{\gamma\pi(\tau_H;\sigma_H)}_{\substack{\text{audit penalty}\\ \text{without tech}}} - \underbrace{\gamma\pi(\tau_L;\sigma_L)}_{\substack{\text{audit penalty}\\ \text{with tech}}}. \tag{14}$$

Thus the effect of increasing γ on adoption depends on the sign of the difference in audit penalties with and without technology. That is, if technology adoption reduces the audit penalty, then $\frac{dK}{d\gamma} > 0$ and increasing the audit rate leads to more technology adoption. □

Note that while $\gamma\pi(\tau_H;\sigma_H)$ and $\gamma\pi(\tau_L;\sigma_L)$ are each individually positive, the sign of their difference is theoretically ambiguous. A sufficient condition for $\frac{dK}{d\gamma} > 0$ would be if τ is relatively inelastic with respect to σ and $\underline{h}$ is relatively high, in which case the expected penalty rises with σ.

B.2 Medicare's Problem

The model discussed in Section 3 can be extended to capture Medicare's problem of setting the audit rate. Let the audit intensity parameter γ be defined as $\gamma = \beta\psi$, where β is the audit rate and ψ is the penalty when an audit uncovers a low-benefit admission. I assume that ψ is fixed, and Medicare chooses β to maximize its payoff. All admissions face the same audit likelihood regardless of x. I also assume that Medicare takes the reimbursement rate R as fixed.

Hospitals face identical patient populations and payoff functions with and without technology, but vary uniformly in their fixed cost of technology adoption $K \in [0, \overline{K}]$. Medicare's payoff includes both patient benefit and expenditure. Medicare expenditure has three components: expenditure on hospital admissions, the cost of conducting audits, and the amount recouped as penalties. Medicare values population-level patient benefit, which is the sum of patient benefit across adopting and non-adopting hospitals. Let K^* be the threshold adoption cost as defined in Equation 12. Share $s = \frac{K^*}{\overline{K}}$ of hospitals will adopt technology, choose threshold τ^A and expect to admit q^A patients each. Share $1 - s$ of hospitals will not adopt technology, choose threshold τ^N, and expect to admit q^N patients each. Normalizing the total number of hospitals to 1, let the total admissions in the population be $Q = sq^A + (1 - s)q^N$. α_M is the value Medicare places on patient benefit relative to expenditure, and $(1 + \lambda)$ is the fiscal externality of raising a dollar of government revenue. Altogether, Medicare's payoff is:

$$\mathbb{U}(\beta) = \underbrace{\alpha_M[sB(\tau^A) + (1-s)B(\tau^N)]}_{\text{Medicare value of patient benefit}} - (1+\lambda)\underbrace{\Big[\underbrace{RQ}_{\text{reimb.}} + \underbrace{\frac{1}{2}C_{aud}(\beta Q)^2}_{\text{audit costs}} - \underbrace{\beta\psi(s\pi(\tau^A) + (1-s)\pi(\tau^N))}_{\text{audit penalties}}\Big]}_{\text{Medicare expenditure}} \tag{15}$$

In choosing audit rate β, Medicare trades off changes in population-level patient benefit with the net effect on government spending from changes in hospital reimbursement, audit penalties, and audit costs. As β increases, Medicare's payoff changes in three ways. First, increasing β changes the technology adoption threshold, leading marginal hospitals to adopt technology and change from producing $B(\tau^N)$ patient benefit and q^N admissions to $B(\tau^A)$ patient benefit and q^A admissions. Second, admissions q^A and q^N decrease among all inframarginal adopting or non-adopting hospitals. Third, the number of audits and amount recouped via penalties change as the number and composition of admissions change.

Should Medicare Purchase the Technology for Hospitals? We can then use this framework to consider an additional choice for Medicare: whether to purchase the technology on behalf of all hospitals. Say that technology adoption increases total patient benefit, but voluntary technology adoption is low and it is expensive to conduct audits. In this case, it may be worthwhile for Medicare to purchase the technology on behalf of hospitals and then require that they use it. This would capture policies like the HITECH Act, which directly subsidized the adoption of health IT [Burde, 2011].

Whether it is worthwhile for Medicare to directly purchase the technology will depend again on a threshold cost rule. If G is the cost to purchase the technology for all hospitals, then Medicare's payoff from fully subsidizing the technology purchase is $\mathbb{W}(\beta) - (1+\lambda)G$, where $\mathbb{W}(\beta)$ is Medicare's payoff when all hospitals adopt:

$$\mathbb{W}(\beta) = \underbrace{\alpha_M B(\tau^A)}_{\substack{\text{value of patient benefit} \\ \text{when all adopt}}} - (1+\lambda)\underbrace{\left[Rq^A + \frac{1}{2}C_{aud}(\beta q^A)^2 - \beta\psi\pi(\tau^A)\right]}_{\text{Medicare expenditure when all adopt}}, \tag{16}$$

where Medicare has set β optimally. There is a threshold cost below which Medicare will choose to purchase the technology on hospitals' behalf and then require that they use it. Specifically, Medicare will do so if the purchase cost is less than threshold cost G^*:

$$\begin{aligned}(1+\lambda)G^* &= \mathbb{W}(\beta) - \mathbb{U}(\beta) \\ &= (1-s)\alpha_M[B(\tau^A) - B(\tau^N)] \\ &\quad - (1+\lambda)\left[R(q^A - Q) + \frac{\beta^2}{2}C_{aud}((q^A)^2 - Q^2) - \beta\psi(1-s)(\pi(\tau^A) - \pi(\tau^N))\right]\end{aligned} \tag{17}$$

As this threshold increases, Medicare is more willing to make this purchase. The threshold G^* is higher if technology adoption improves total patient benefit ($B(\tau^A) > B(\tau^N)$), if technology adoption reduces the number of admissions ($q^A < Q$), or if audit costs C_{aud} are high.

C Additional Analyses

C.1 Robustness and Placebo Analysis

Hospital-Level Analysis As a robustness test, in Figure G5 I regress on a hospital's denial rate – the share of claims for which a denial is made after audit – rather than its audit rate. Equation 18 defines the relationship between denial rate and audit rate.

$$Denial\ Rate_{ht} = \underbrace{P(Audit)}_{Audit\ Rate_{ht}} \times \underbrace{P(Demand|Audit)}_{Demand\ Rate_{ht}} \tag{18}$$

Since 41 percent of audits in 2011 resulted in a demand in the main sample and the denial rate is monotonically increasing in audit rate (Figure G6), one would expect that a hospital's response to a one-percentage point increase in the denial rate should be about twice the response to one percentage point increase in the audit rate. Indeed, this is what the results reflect; for example, hospitals reduced admissions by 2.5 percent in 2012 in response to a one-percentage point increase in the 2011 audit rate, and they reduced admissions by 5.7 percent in 2012 in response to a one-percentage point increase in the denial rate.

In Figure G7, I show that the results are robust to the inclusion of control variables. The main specification hospital, year, and neighbor comparison group-year fixed effects. Non-time-varying hospital characteristics are absorbed by the hospital fixed effects, and variables which vary over time nationally or within a local area are absorbed by the year and group-year fixed effects. Thus, the control variables I include are 2010 hospital characteristic-year fixed effects. These controls will capture trends over time across different types of hospitals – for example, if there are divergent trends between non-profit and for-profit hospitals. The control variables consist of the following variables, interacted with a year fixed effect: indicator for above-average 2010 beds, urban status, hospital profit type, teaching status, chain status in 2010, indicator for above-average 2010 short stay share, indicator for above-average 2010 administrative cost share, and indicator for above-average top 20 error MS-DRG share.

In Figure G8, I show that the results are robust to alternative sample definitions. Figure G8a reproduces the event study from the main specification for the outcome of log Medicare admissions, in which the sample is defined as all hospitals within 100 miles of the RAC border and the coefficient is scaled by the correlation between a hospital's audit rate and its leave-one-out state audit rate. This is robust to changing the sample to all hospitals within 50 miles (Figure G8b) or 150 miles (Figure G8c) of the border, although the results are noisier with a shorter distance. One concern with spatial identification strategies is the potential for spillovers to neighboring units. Here, the concern would be about spillovers from high-audit hospitals to low-audit hospitals across the border. On the one hand, if patients were redirected from a hospital near the border in a high-audit rate state to a nearby hospital in a low-audit rate state, then this would bias the coefficients to be larger in magnitude. On the other hand, if hospitals on the low-audit side internalize their high-audit neighbors' audit rates in making their admission decisions, this would bias the coefficients to be smaller in magnitude.

These spillovers should be less of a concern as the distance from the border increases or if the hospitals closest to the border are excluded. Figure G8d shows similar results when restricting the sample to hospitals that are at least 10 miles away from the border, demonstrating that the result is not driven by such spillovers. Finally, Figure G8e shows that the results are similar when restricting the sample to hospitals with audit rates greater than 0 percent, meaning that the results are driven by variation in auditing across hospitals on the intensive, rather than the extensive, margin.

Figure G9 shows that the results are robust to using alternative instruments to scale the reduced form effect. The main specification instruments for a hospital's audit rate using the leave-one-out state audit rate in order to capture the variation in audit intensity that is unrelated to the hospital's own behavior. Figure G9a plots the results of using the state audit rate (which includes the hospital) as an instrument. Figure G9c shows that the results using the leave-one-out RAC region audit rate, rather than the state audit rate, are similar.

While using the leave-one-out audit rate strips away the direct effects of a hospital's own behavior, it still includes other hospitals surrounding a given hospital, whose audit rates may still reflect that hospital's behavior. This can be the case if, for example, a given hospital has a large market share. To address this, in Figures G9b and G9d I consider using the audit rate of other hospitals in the same state or RAC region in *other* markets, which I define using hospital referral regions. This instrument leverages hospitals whose behavior is less likely to be affected by a given hospital's behavior since they are in different markets. Similarly, one might be concerned that a hospital's audit rate is correlated with the behavior and audit rates of other hospitals in the same hospital system. Figure G9e uses the audit rate of hospitals in the same state but different hospital systems in 2010. The results are robust to using these hospitals to instrument for a hospital's audit rate.

Because neighbor comparison groups can overlap, they can potentially span multiple border segments. Thus, clustering at the border segment-level may not capture the correlated errors across border segments, which would bias the standard errors. Given how the neighbor comparison groups are defined, there is no way to set border segments that eliminates this problem. However, it should be less of a concern for longer border segments, as hospitals in the same neighbor comparison group would now be less likely to point to different border segments. Figure G17 plots the event studies from using 50- and 150-mile border segments. While the the standard errors do increase as the segments become longer, the coefficients remain statistically significant.

To confirm that the results are not driven by a single state or hospital comparison group, Figure G18 plots the distribution of coefficients when one state or one hospital comparison group is removed from the sample at a time. The coefficients are always negative and the distribution is centered around the main effect.

Finally, I consider a falsification test using state borders in the *interior* of each RAC region. In the interior of each region, there is no change in RAC identity at state borders, so comparing hospitals across these interior borders does not capture exogenous variation driven by different audit strategies across RACs. Figure G19a illustrates the interior borders and the sample of hospitals within one hundred miles of the interior border (excluding hospitals that are within one hundred miles of the RAC border). I also restrict the falsification test just to interior borders between different *MAC* regions. As noted in Section 4.2, the fact that some Part B MAC region borders overlap with RAC region borders could make changes in MAC activity a time-varying confounder. Both of the falsification tests show no effect on admissions on the "high-audit side" of the interior borders (Figures G19b and G19c).

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

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

C.2 Coding Response

To assess the effects of auditing on upcoding, I run the main specification on the log number of diagnoses per claim (ICD-9/ICD-10 diagnosis codes) in Figure G20. The number of reported diagnoses decreases at hospitals subject to more audits. The main results indicate that the remaining admissions tended to have longer lengths of stay, and are presumably sicker, than deterred admissions. Thus if hospitals didn't change how they coded, we would expect them to have more diagnoses per patient. These results suggest that in addition to changing utilization patterns, RAC audits led to less upcoding.

One caveat to note in this analysis is the presence of two coding-related reforms in the study period. First, from 2007-2008, CMS transitioned from the DRG system to the MS-DRG system and recategorized many DRGs [Gross et al., 2023]. Second, the maximum number of diagnoses allowed on a Medicare inpatient claim increased from 9 to 25 in 2010. Both of these reforms applied to all hospitals, so a key assumption for identification is that they should not have differential effects across border hospitals subject to different RACs.

C.3 Rural Hospital Closures

The main results show that RAC audits decrease hospital revenue and increase their costs. This raises the concern that RAC auditing may have driven hospitals into financial distress and, given the prevalence of hospital closures in recent years, led them to close. Hospital closures are associated with decreases in access to care and increases in patient mortality [Carroll, 2019; Gujral and Basu, 2019]. To study whether RAC auditing led to hospital closures, I use data from the Sheps Center for Health Services Research on rural hospital closures between 2005 and 2022.[39] I adapt the main specification for the hospital-level analysis to study rural hospital closures. In the border hospital sample, no hospitals closed before 2012 – this is by definition, since the hospital had to be open in 2011 to be audited. Therefore there is no variation in the pre-2010 period to use a difference-in-differences framework. I run the following specification separately for each year Y in the post period:

$$Close_h^Y = X_h^{2011}\beta^Y + \phi_{g(h)} + \varepsilon_h \tag{19}$$

which regresses a dummy for whether a rural hospital has closed in year Y, $Close_h^Y$, on its (instrumented) audit rate X_h^{2011}, after taking into account the hospital's neighbor comparison group. Figure G21 plots the β^Y coefficients for years where there is variation in closures among rural hospitals in the border sample (i.e., excluding 2012, 2017, and 2021). The results indicate that higher RAC auditing did not cause hospital closures.

D Extrapolation to Overall Hospital Sample

This section describes the calculation to extrapolate the savings estimates from the border hospital sample to the overall RAC program. This calculation rests on fairly strong assumptions, but nonetheless may be of interest for gauging the magnitude of overall savings from the RAC program. First, we must assume that the savings scale linearly with audit rate, so that the effects estimated from a marginal increase in audit rate can be extrapolated beyond the support to a wide range of audit rates. Second, we must assume homogeneous treatment effects across hospitals in the border sample and overall. Note that while hospitals on opposite sides of the border are similar to each other (Table FII), the border hospital sample differs from the overall sample. Hospitals in

[39] Data available at https://www.shepscenter.unc.edu/programs-projects/rural-health/rural-hospital-closures/. Last accessed July 2023.

the border sample are smaller, more rural, more likely to be non-profit and disproportionately from the Midwest RAC region, Region B, (Table I). Third, this calculation assumes that even at high levels of auditing, there is still no effect on other outcomes that may affect welfare, like patient health or hospital closures.

Under these assumptions, I can calculate the extrapolated savings by multiplying the 2011-2015 event study coefficients on Medicare inpatient revenue (Figure 5b) and payments demanded (Figure G12) by each hospital's 2011 audit rate. Since the estimates are based on the logarithm of inpatient revenue and represent a percent change relative to the baseline in 2010, I multiple these coefficients by the hospital's 2010 inpatient revenue. Figure G22 plots the extrapolated savings for each hospital-year, compared to the actual changes in Medicare inpatient revenue and actual payments demanded. For both types of savings, the extrapolated and actual savings are positively correlated. This indicates that in the overall sample, hospitals subject to higher audit rates reduced their Medicare inpatient revenue more and were subject to more audit demands in subsequent years. Summing up the extrapolated savings across all hospitals from 2011 to 2015 implies that the RAC program saved the Medicare program \$9.28 billion between 2011 and 2015, compared to the actual \$11.74 billion reduction in inpatient spending and savings from audit recoveries in this period. Note, however, the relatively low R^2 from the regression between extrapolated and actual savings, indicating that much of the variation in savings is not explained by variation in 2011 audit rate.

E Hospital-level Emergency Department Visit Analysis

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

I use the MEDPAR (Inpatient) and Outpatient files to identify all ED visits by Medicare beneficiaries at the hospitals in my sample. Note that the ED outcomes I consider are *shares*, rather than counts. I use shares because of concerns about inconsistencies in the reporting of ED visit count across different data sources, different providers, and different time periods. Venkatesh et al. [2017] counts ED visits in Medicare claims in one year using four different definitions, and finds differences up to 17 percent across the different measures. Additionally, in attempting to construct a measure of ED visits across multiple years, I also found data anomalies across states that CMS's Research Data Assistance Center (ResDAC) confirmed were likely due to reporting errors.[40]

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

It is challenging to determine whether an ED visit resulted in an inpatient stay using the MEDPAR (Inpatient) and Outpatient files. This is because the inpatient stays with ED charges only capture a portion of all inpatient stays associated with an ED visit. A portion of ED visits that result in an inpatient stay are located in the MEDPAR file, while the rest are in the Outpatient file with no direct linkage to the associated inpatient stay. ResDAC cautions that "although one

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

can assume ER patients found in the inpatient data were admitted to the hospital, one *cannot* assume ER patients found in the outpatient data were not admitted to the hospital...some patients are transferred to a different hospital for admission and some hospitals bill ER and inpatient services separately" (emphasis in original) [Barosso, 2015]. A substantive share of Medicare patients undergo inter-hospital transfer, especially for diagnoses that are prevalent in the ED – for example, up to 50% of patients with heart attacks are transferred [Iwashyna et al., 2010]. Thus, it is difficult to discern in the Medicare claims which ED visits resulted in an inpatient stay.

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

Figure G16 shows the event studies from Equation 3 on the ED visit outcomes. The results from this analysis are largely consistent with the patient-level results. Among hospitals with a higher 2011 audit rate, the share of ED visits with outpatient observation services increases after 2011. However, ED visits at these hospitals do not seem to result in greater revisits or mortality. There is a small and statistically insignificant increase in revisit rates after 2012, but its magnitude (0.25%) is very small relative to the 2010 mean, 15%.

F Appendix Tables

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

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

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

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

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

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

Table FIII. ED Arrival Hour Manipulation Tests

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

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

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

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

* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[y \geq 2013Q3] \times \mathbb{1}[T_v \geq 00{:}00]$ of the specification in Equation 9, where $\mathbb{1}[y \geq 2013Q3]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "OR procedure" is an indicator for whether a patient received an OR procedure during their stay. "Revisit within 60/90 days" is an indicator for whether the patient had another ED visit or inpatient stay within 60/90 days of the ED visit. Sample comprises traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and quartile of mean zip code income. Data: HCUP SID/SEDD.

Table FV. Across-Hospital Post-2011 Coefficient

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		LOS ≤ 2		Admin Costs	Software Installation
	Log Adm.	*Log Rev.*	*Log Adm.*	*Log Rev.*	*Log Costs*	*Medical Necc.*
2011 audit rate	-0.0154	-0.0166	-0.0227**	-0.0227***	0.0087	0.0153*
$\times$ post-2011	(0.0092)	(0.0136)	(0.0096)	(0.0067)	(0.0100)	(0.0081)
Hosp FE	X	X	X	X	X	X
Nbr group FE	X	X	X	X	X	X
Hosp	510	510	510	510	510	506
N	52139	52139	52139	52118	52107	36906
F	104.98	104.98	104.98	103.89	104.68	84.15

* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors in parentheses and are clustered at the state and border segment level. This table reports the coefficients of the reduced form event study in Equation 7, 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 in 2011-2015. Columns 1-2 report two stage least squares outcomes for the number of and revenue from Medicare admissions overall, columns 3-4 report outcomes for the number of and revenue from Medicare admissions with length of stay ≤ 2, column 5 reports the outcomes for log net administration costs, and column 6 reports the outcomes for an indicator for installation of medical necessity software. Length of stay is counted as the difference in days between the admission and discharge date. Inpatient revenue is the sum of all Medicare inpatient payments. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in HIMSS. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, CMS audit data, HCRIS, and HIMSS.

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

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

* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors are in parentheses and are clustered at the state and border segment level. This table reports the coefficients of the reduced form event study in Equation 7, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome after 2011. Columns 1-2 report two stage least squares outcomes for the number of and revenue from Medicare admissions overall, columns 3-4 report outcomes for the number of and revenue from Medicare admissions with length of stay ≤ 2, column 5 reports the outcomes for log net administration costs, and column 6 reports the outcomes for an indicator for installation of medical necessity software. Length of stay is counted as the difference in days between the admission and discharge date. Inpatient revenue is the sum of all Medicare inpatient payments. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted," "installation in process," and "to be replaced" in the HIMSS data in 2012. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Omitted year is 2010. Data: MEDPAR, CMS audit data, HCRIS, HIMSS, Medicare Provider of Services, and Cooper et al. [2019] merger data.

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

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

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

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

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

* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[y \geq \text{2013Q3}] \times \mathbb{1}[T_v \geq \text{00:00}]$ of the specification in Equation 9, interacted with hospital characteristics. $\mathbb{1}[y \geq \text{2013Q3}]$ 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 variable for whether the patient was eventually admitted as inpatient from the ED (HCUP SID/SEDD). The sample consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and quartile of mean zip code income. Urban/rural, teaching/non-teaching, for-profit/government/non-profit, and bed size come form the Medicare Provider of Services file. Non-chain status come from Cooper et al. [2019]. Medical necessity application is an indicator which is equal to one if medical necessity checking application is listed as "live and operational," "contracted" "installation in process," or "to be replaced" in the HIMSS data in 2012. Data: MEDPAR, CMS audit data, HCRIS, HIMSS.

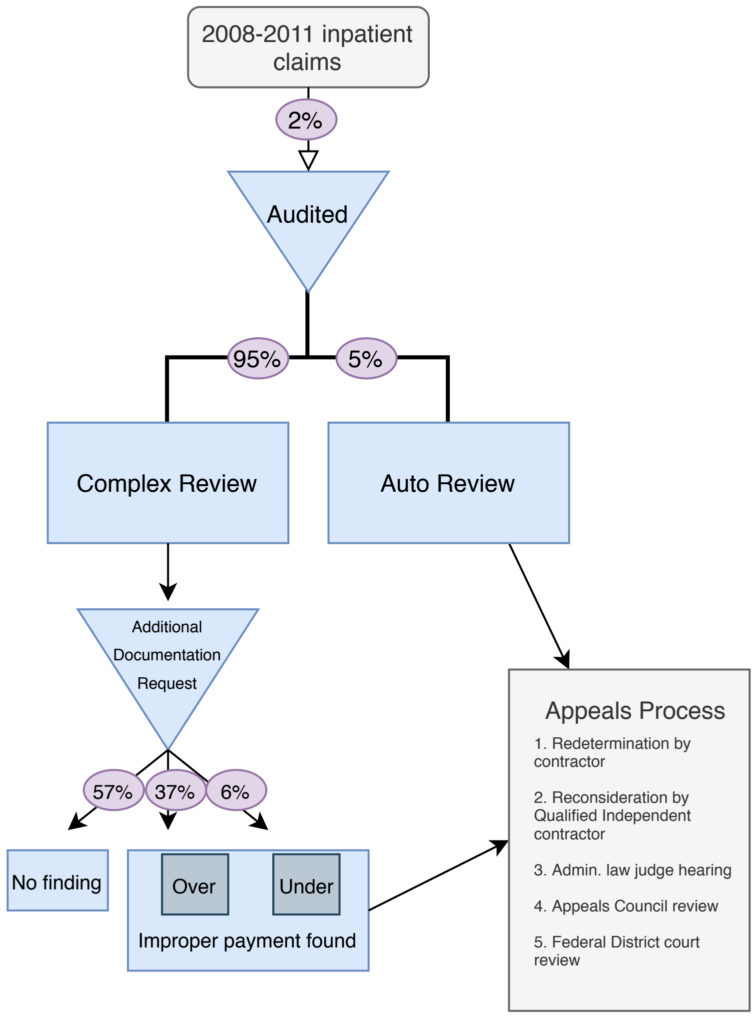
Table FIX. Robustness Test: Sample of Patients by ED Arrival Relative to Midnight

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

* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $\mathbb{1}[y \geq \text{2013Q3}] \times \mathbb{1}[T_v \geq \text{00:00}]$ 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. 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 quartile of zip code income. The samples comprise of traditional Medicare patients who arrive at the ED in a Florida hospital within 1 hour of midnight (11PM-12:59AM; column 1), within 2 hours of midnight (10PM-1:59AM; column 2); within 3 hours of midnight (9PM-2:59AM; column 3); within 4 hours of midnight (8PM-3:59AM; column 4); and within 5 hours of midnight (7PM-4:59AM; column 5). Data: HCUP SID/SEDD.

G Appendix Figures

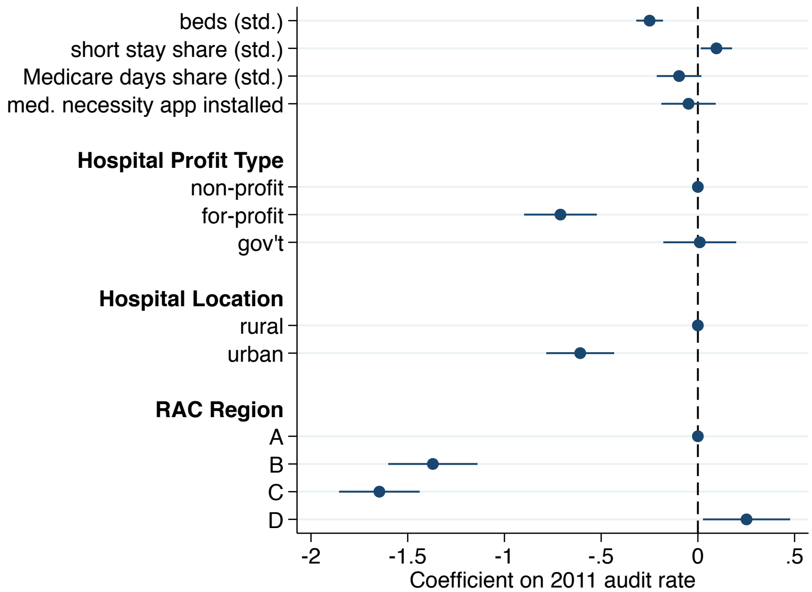
Figure G1. RAC Inpatient Claims Auditing and Appeals Process, 2011 Audits



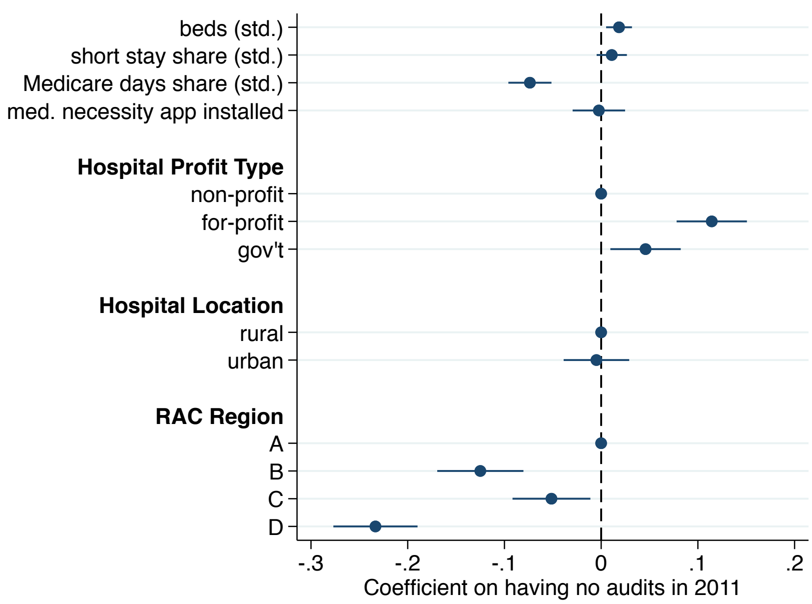
This figure illustrates the stages of the claims auditing and appeals process. The percentages in ovals denote the percent of claims that, conditional on reaching a given stage in the process, reach the next stage. The percentages are calculated based on audits in 2011 of inpatient claims between 2008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

Figure G2. Correlation between Hospital Characteristics on 2011 Audit Rate and No Audit

(a) Outcome: 2011 hospital audit rate

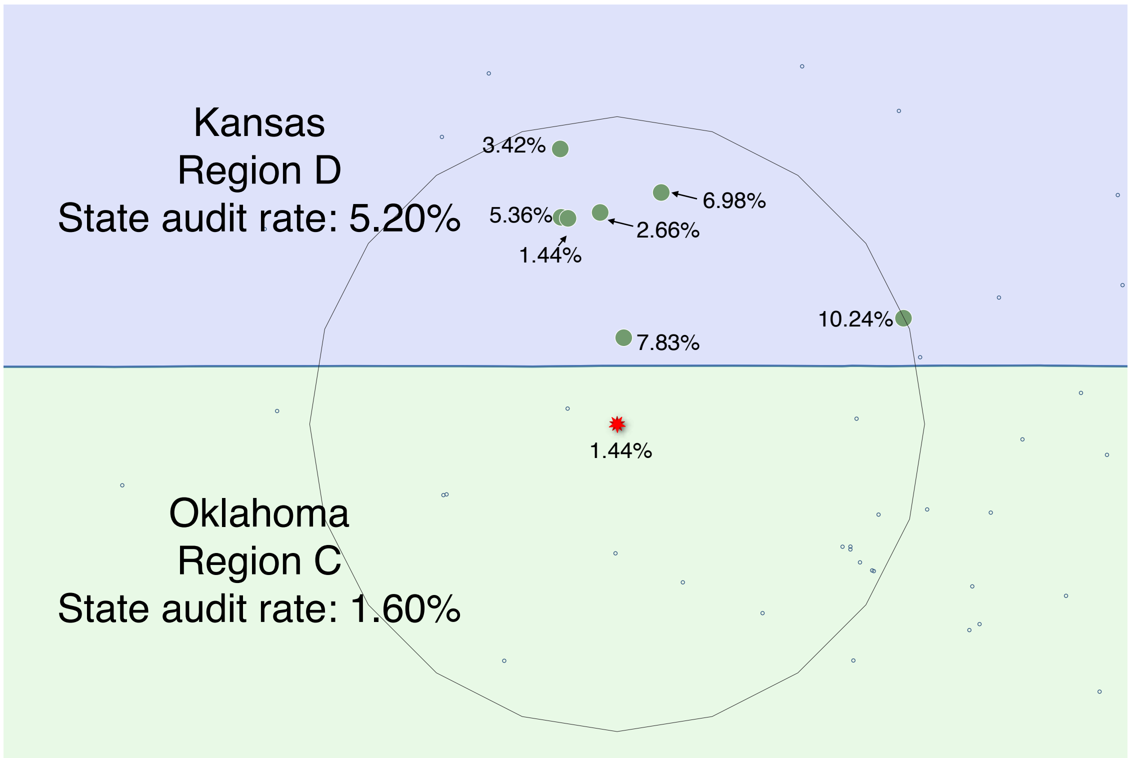


(b) Outcome: no audits at hospital in 2011



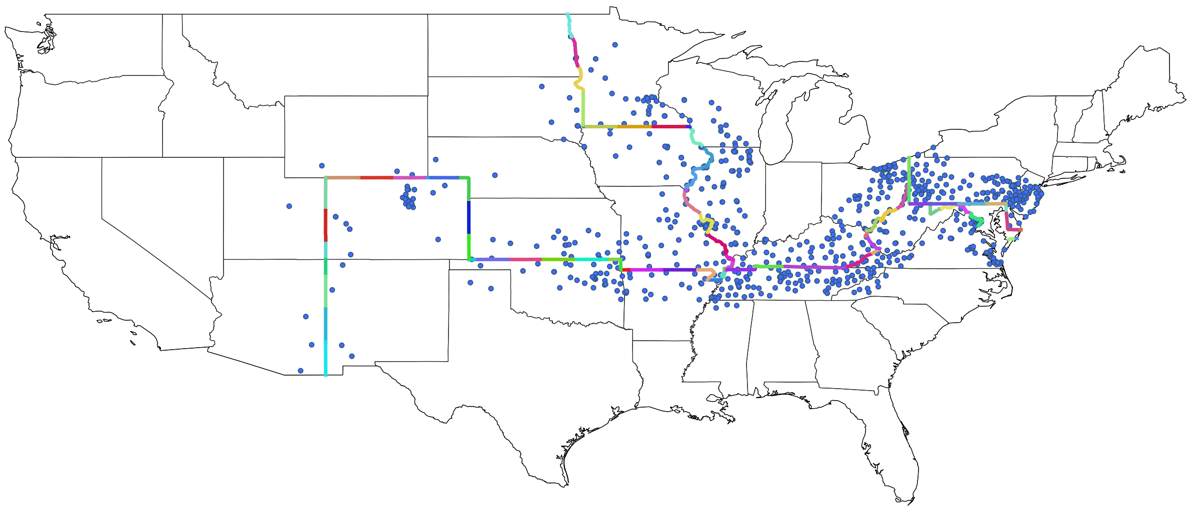
These figures plot coefficients from a regression of (a) a hospitals 2011 audit rate and (b) an indicator variable for whether a hospital was not audited in 2011 on 2010 hospital characteristics. Short stay share is the share of 2010 Medicare admissions with lengths of stay 0-2. Medicare days share is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

Figure G3. Example of Border Hospital and Neighbor Comparison Group Definition



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

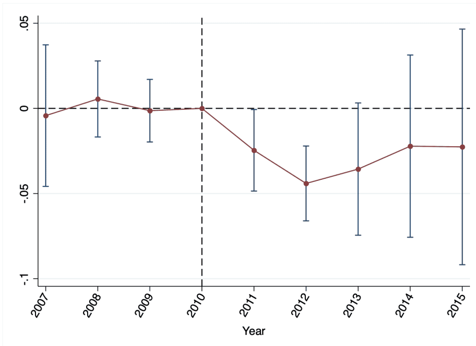
Figure G4. RAC Border Segments and Hospitals Within 100 Miles



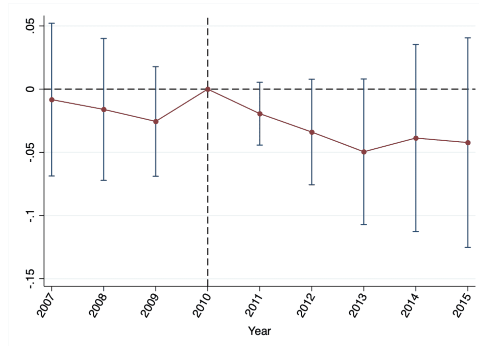
This figure shows how the RAC border is divided into one hundred mile segments that do not cross state borders, and all hospitals within one hundred miles of the RAC border. These border segments are used for clustering in Equation 3.

Figure G5. Event Studies on Effect of 2011 Denial Rate on Medicare Admissions and Revenue, and Administrative Burden

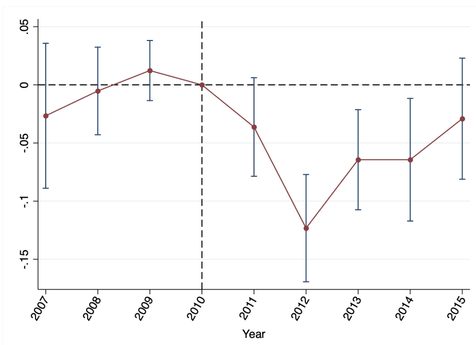
(a) Log Medicare admissions



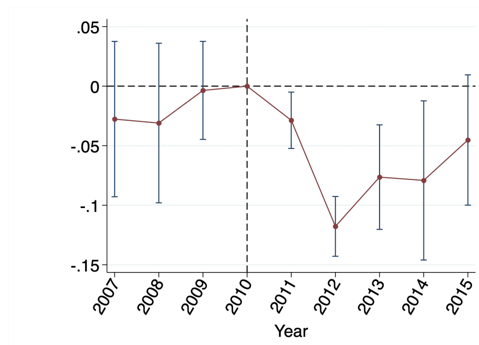
(b) Log Medicare inpatient revenue



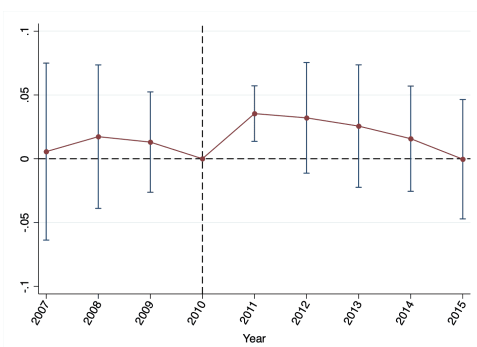
(c) Log Medicare admissions, LOS ≤ 2



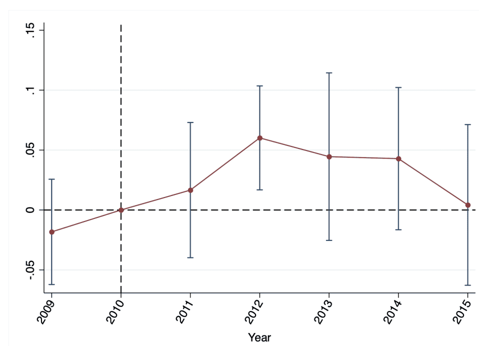
(d) Log Medicare inpatient revenue, LOS ≤ 2



(e) Log net administration costs

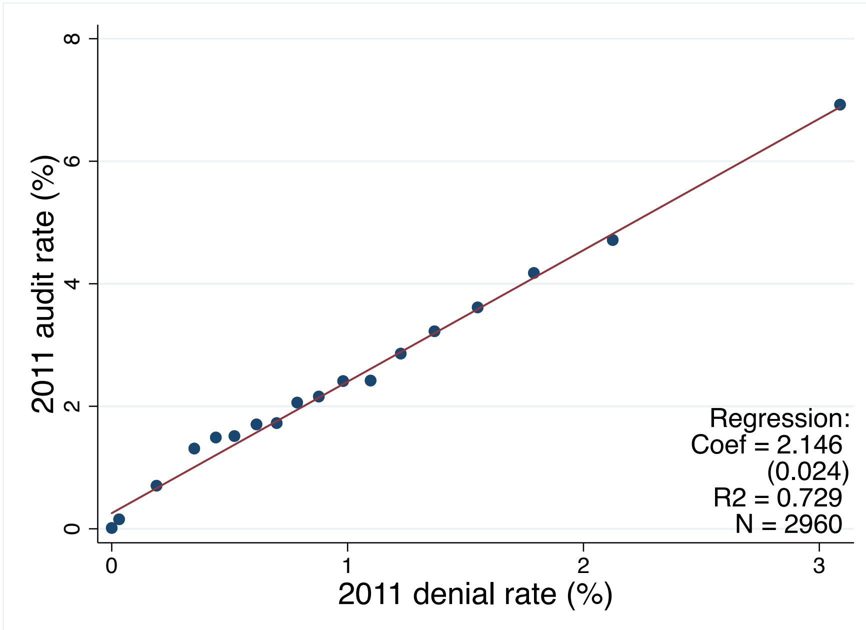


(f) Indicator for installing medical necessity software



This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5 using the denial rate rather than the audit rate, scaled by the correlation between the leave-one-out 2011 denial rate and the actual 2011 denial rate in the weighted border hospital sample. Denial rate is the share of claims that are audited and result in an overpayment demand or repayment for an underpayment. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 denial rate rate on a hospital-level outcome. Medicare admissions and revenue are from MEDPAR. Inpatient revenue is the sum of all Medicare inpatient payments. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in HIMSS. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, CMS audit data, HCRIS, HIMSS.

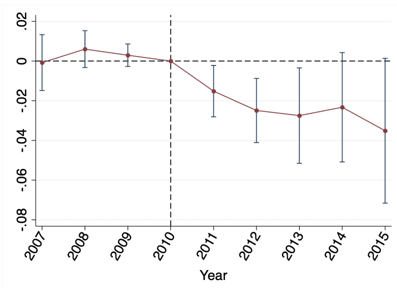
Figure G6. 2011 Audit Rate vs. 2011 Denial Rate



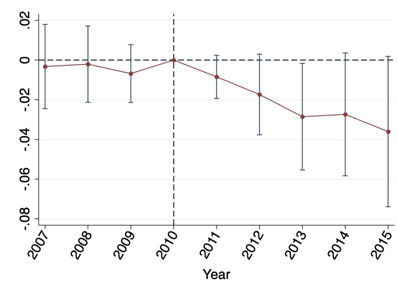
This figure plots a binscatter of 2011 hospital audit rate (share of claims subject to an audit) against the 2011 hospital denial rate (share of claims with reclaimed payment because of audit). Data: CMS audit data and MEDPAR.

Figure G7. Event Studies on Effect of 2011 Audit Rate on Hospital Outcomes, including Controls

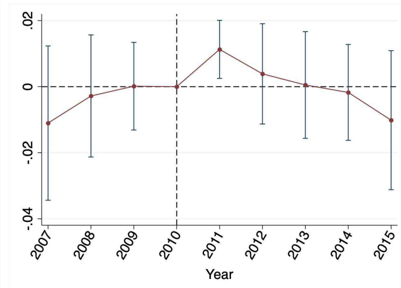
(a) Log Medicare admissions



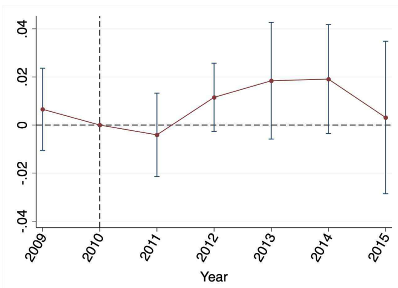
(b) Log Medicare inpatient revenue



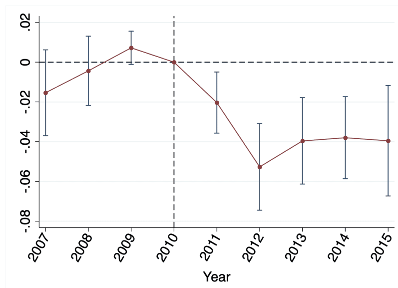
(c) Log hospital administrative costs



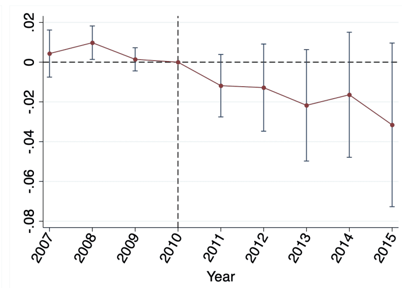
(d) Indicator for installing medical necessity software



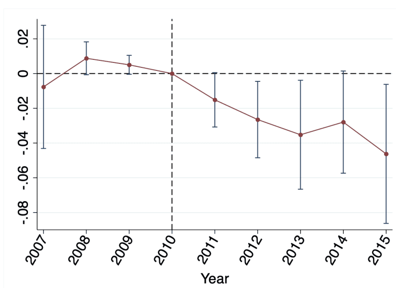
(e) Log Medicare admissions, LOS ≤ 2



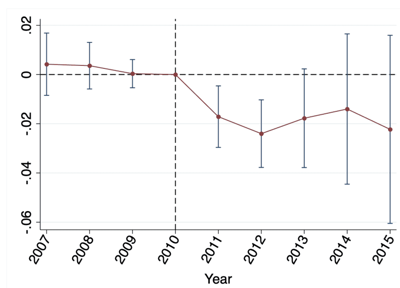
(f) Log Medicare admissions, LOS > 2



(g) Log Medicare admissions, top 20 error rate MS-DRGs



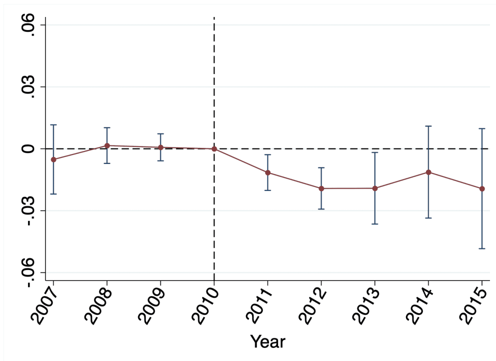
(h) Log Medicare admissions, non-top 20 error rate MS-DRGs



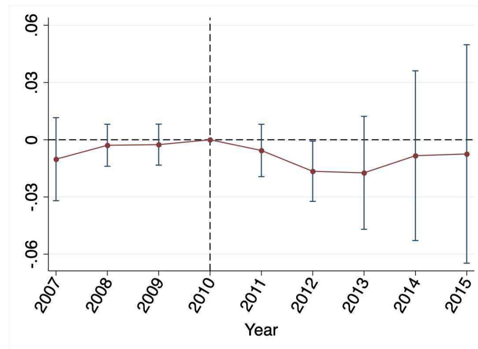
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5 including control variables, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. The control variables consist of the following variables interacted with a year fixed effect: indicator for above-average 2010 beds, urban, hospital profit type, teaching status, chain status in 2010, indicator for above-average 2010 administrative share, indicator for above-average 2010 short stay share, and indicator for above-average top 20 error MS-DRG share. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, CMS audit data, HCRIS, HIMSS, Medicare Provider of Services, and Cooper et al. [2019] merger data.

Figure G8. Robustness to Sample Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

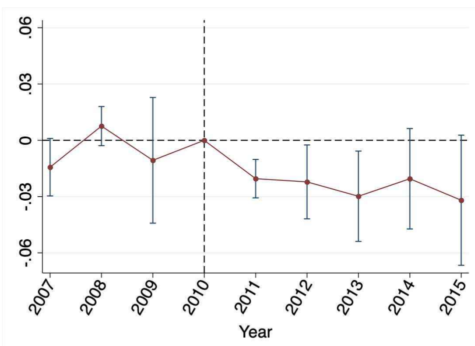
(a) Sample: within 100 miles (main)



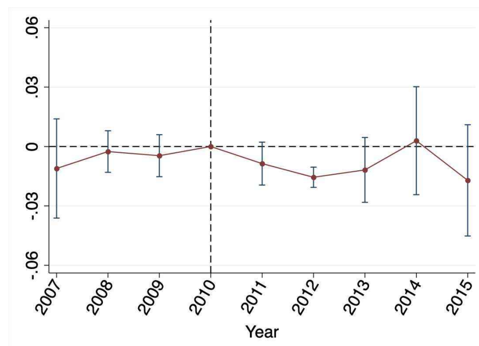
(b) Sample: within 50 miles



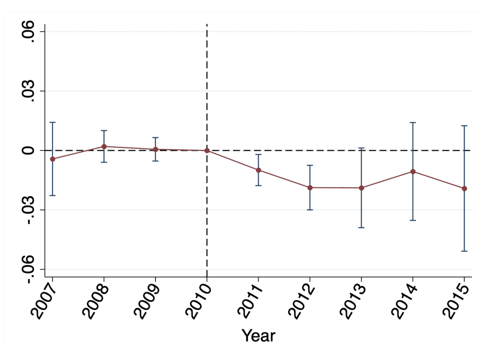
(c) Sample: within 150 miles



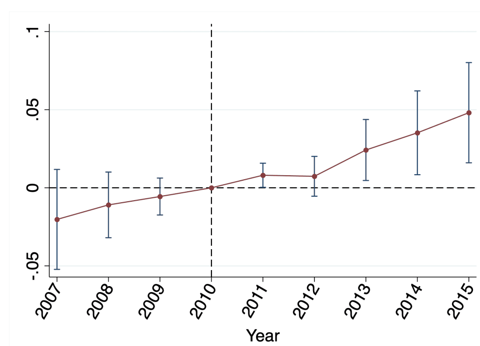
(d) Sample: w/in 100 miles, outside 10



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



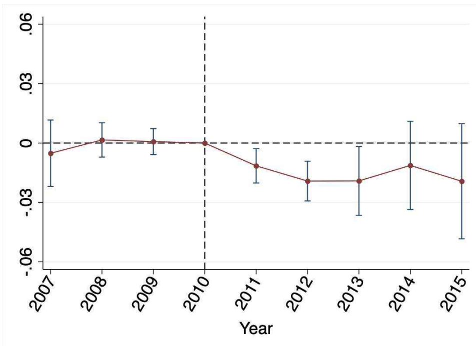
(f) Sample: non-border, all hospitals



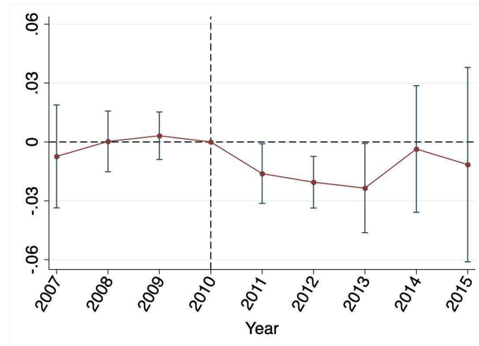
This figure plots robustness analysis event studies of the scaled reduced form coefficients and 95% confidence intervals of the specification in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient estimates the effect of a one percentage point increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different definitions of the border sample: (a) reproduces the main result and defines the border sample to be all hospitals within 100 miles of the RAC border; (b) defines the border sample to be all hospitals within 50 miles of the RAC border, (c) defines the border sample to be all hospitals within 150 miles of the RAC border, (d) defines the border sample to be all hospitals within 100 miles of the RAC border, excluding hospitals within 10 miles of the border, and (e) uses the 100 mile border sample and restricts to hospitals with 2011 audit rate greater than 0. Panel (f) plots the results for all hospitals (N=3014), in a specification where the hospitals audit rate is instrumented using the leave-one-out RAC region rate and includes hospital and year fixed effects. Data: MEDPAR and CMS audit data.

Figure G9. Robustness to Instrument Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

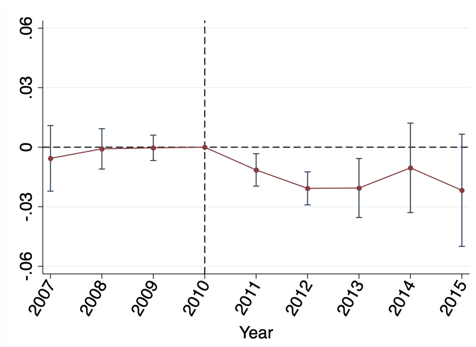
(a) Instrument: state audit rate



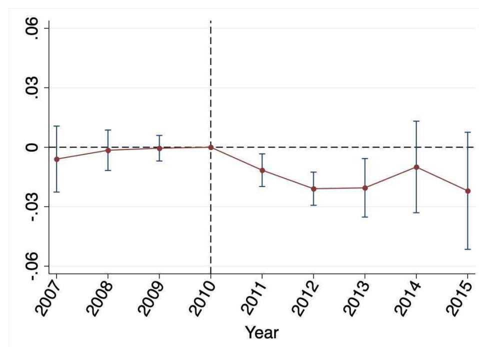
(b) Instrument: state rate, outside HRR



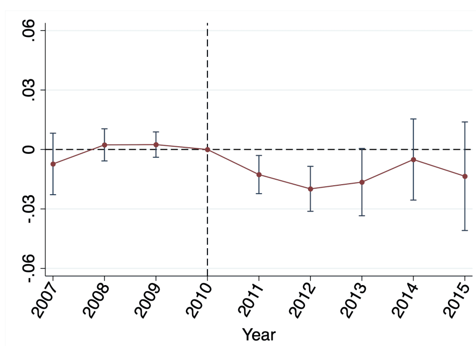
(c) Instrument: leave-one-out RAC region rate



(d) Instrument: RAC rate, outside HRR

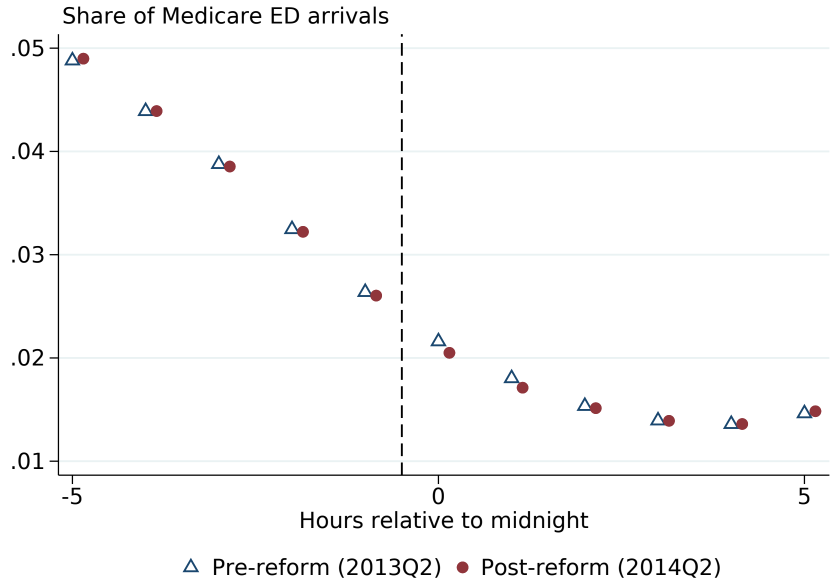


(e) Instrument: state rate, outside system



This figure plots robustness analysis event studies of the reduced form coefficients and 95% confidence intervals of the specification in Equation 5, scaled by the correlations between the instruments and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient estimates the effect of a one percentage point increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different instruments for a hospital's 2011 audit rate. Panel (a) uses 2011 state audit rate and panel, (b) uses 2011 audit rate among hospitals in the same state but in different hospital referral regions (HRR) as the hospital, (c) uses the 2011 audit rate of other hospitals in the same RAC region, (d) uses the 2011 audit rate of other hospitals in the same RAC region but in different HRRs, and (e) uses the 2010 audit rate of other hospitals in the same state but in different hospital systems in 2010. Data: MEDPAR, CMS audit data, and hospital systems from Cooper et al. [2019].

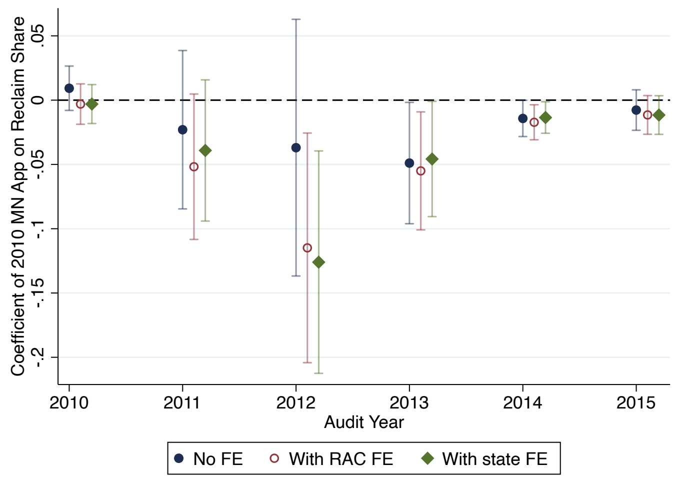
Figure G10. Share of Medicare ED Patients By Hour of ED Arrival



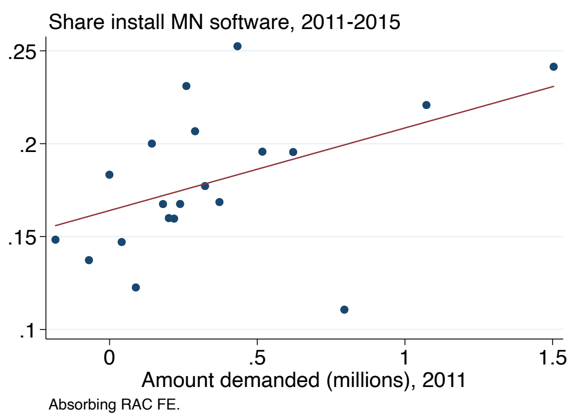
This figure plots the share of Medicare patients that arrive at the ED at each hour (relative to midnight) pre- and post-reform, among traditional Medicare patients who arrived in the ED within 5 hours of midnight in Florida. Data: HCUP SID/SEDD.

Figure G11. Cross-sectional Correlations with Medical Necessity Checking Software

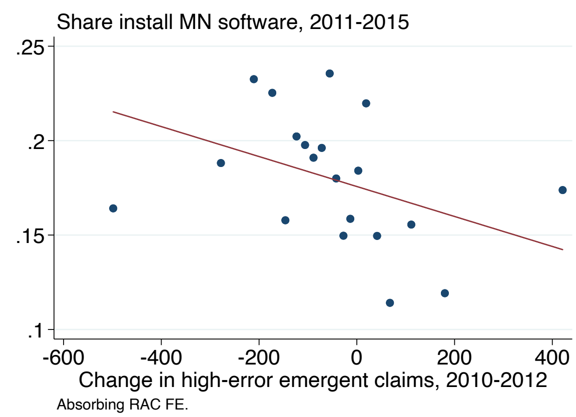
(a) Denial rate and MN software already installed in 2010



(b) Binscatter of 2011 demanded amount & share install MN software, 2011-15

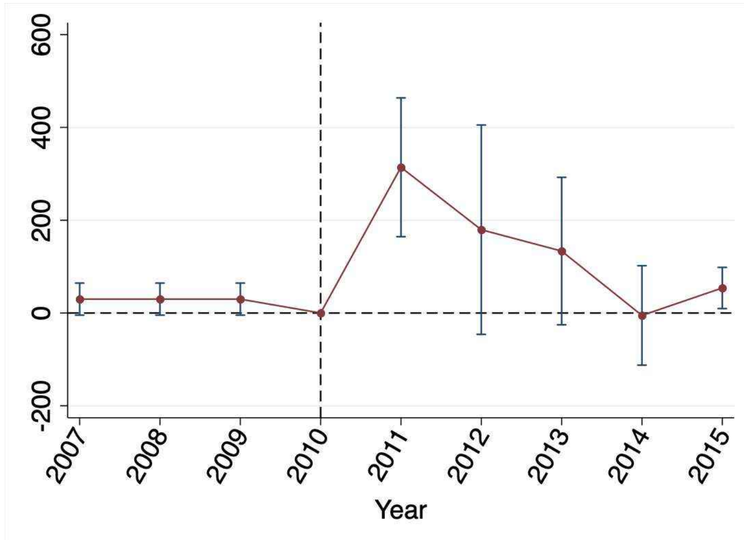


(c) Binscatter of 2011-12 emergent stay change & share install MN software, 2011-15



Panel (a) plots the coefficients of a regression between a dummy variable for whether a hospital has medical necessity checking software installed in 2010 and RAC denial rates in 2010 to 2015. The first specification has no fixed effects, the second specification has RAC region fixed effects, and the third specification has state fixed effects. Panel (b) plots a binscatter of 2011 demanded amount (million \$, winsorized at 95%) and a dummy variable for whether a hospital installs medical necessity checking software in 2011-2015, absorbing RAC fixed effects. The correlation has coefficient 0.04 (p-value: 0.035). Panel (c) plots a binscatter of the change in high-error emergent claims from 2011-2012 and a dummy variable for whether a hospital installs medical necessity checking software in 2011-2015, absorbing RAC fixed effects. The correlation has coefficient -0.00007 (p-value: 0.04). The correlation for the change in high-error non-emergent claims from 2010-2012 has coefficient -0.0001 (p-value: 0.43). Data: MEDPAR, HIMSS, and CMS audit data.

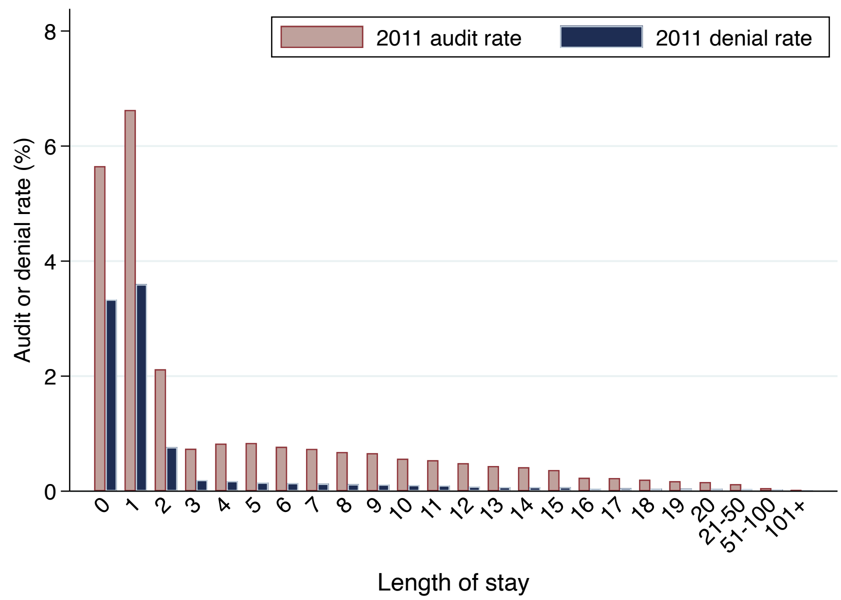
Figure G12. Event Study on Effect of 2011 Audit Rate on Payment Demanded ($1000s) from RAC Audits



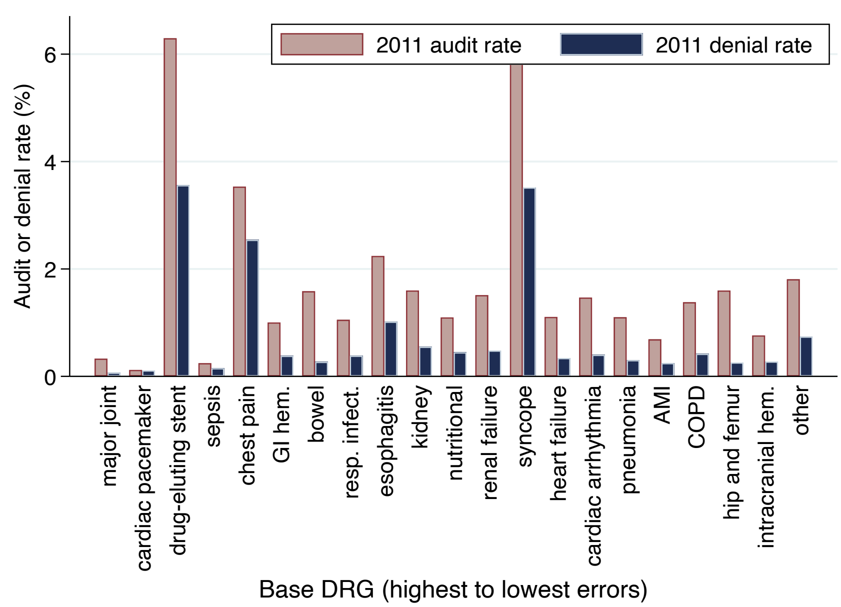
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. The outcome is the amount of payment demanded initially from RAC audits of inpatient stays, by year of audit. Data: CMS audit data.

Figure G13. 2011 Audit and Denial Rates by Stay Characteristics

(a) By Length of Stay



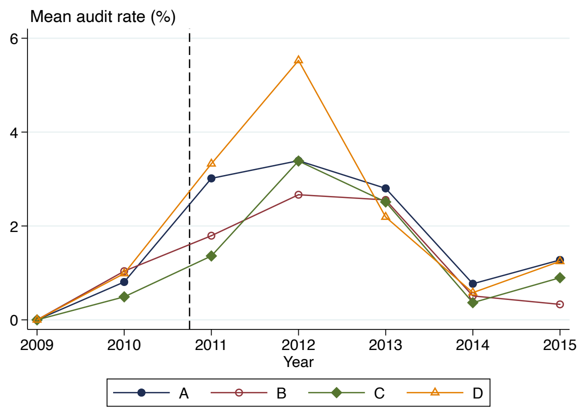
(b) By Base Diagnosis Related Group



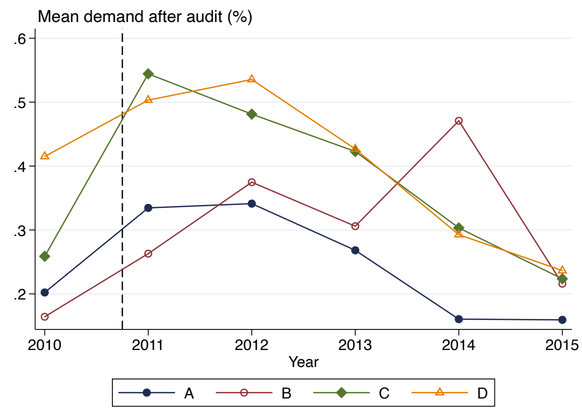
This figure plots the count of 2011 audit and denial rates by (a) an admission's length of stay and (b) its base DRG. Panel (b) shows the top 20 base DRGs with highest improper payments identified in the 2010 CERT report, in descending order of estimated improper payments [Centers for Medicare and Medicaid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

Figure G14. Audit Outcomes and Admissions by RAC Region and Year

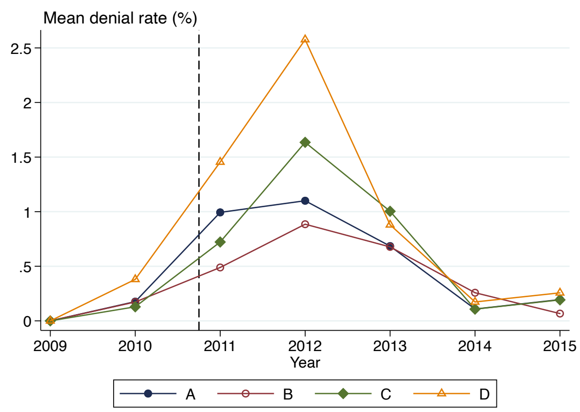
(a) Mean Audit Rate



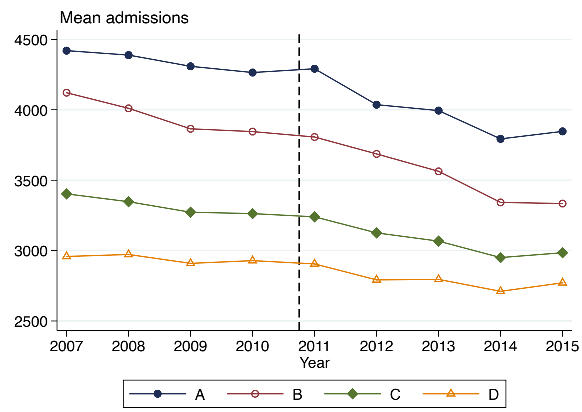
(b) Mean Demand After Audit Rate



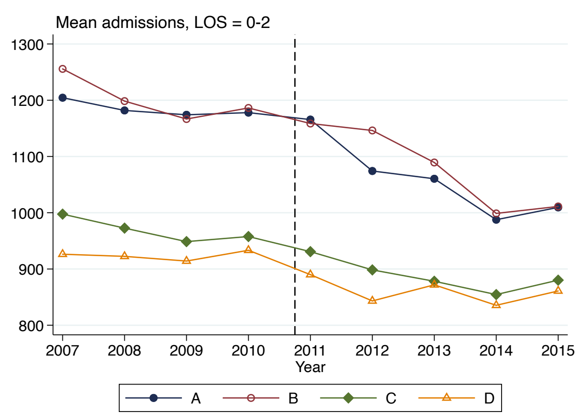
(c) Mean Denial Rate



(d) Mean Admissions

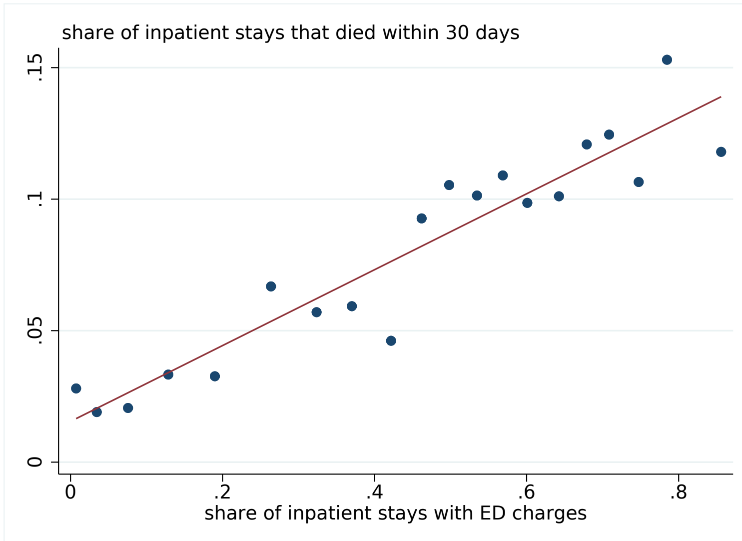


(e) Mean Admissions (LOS = 0-2)



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

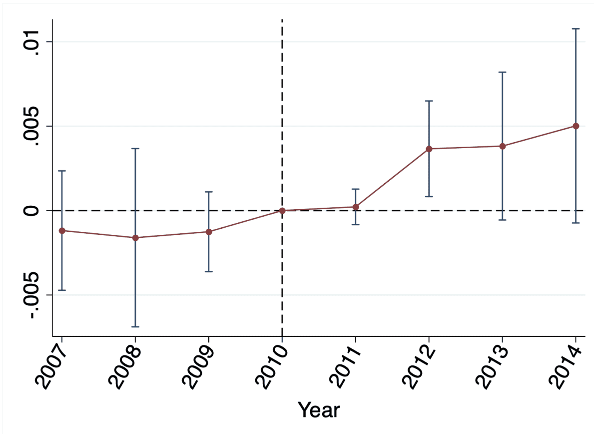
Figure G15. DRG ED Visit Share vs. Died Share



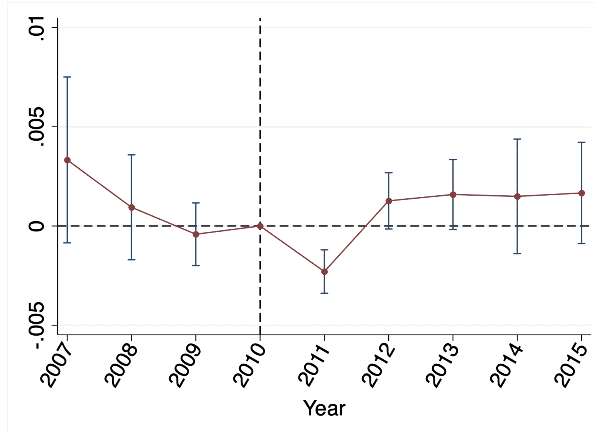
This figure plots a binscatter between the share of a DRG's admissions with ED charges and the share with a death within 30 days, among 2010 Medicare inpatient stays. Data: MEDPAR.

Figure G16. Event Studies on Effect of 2011 Audit Rate on Medicare ED Visits

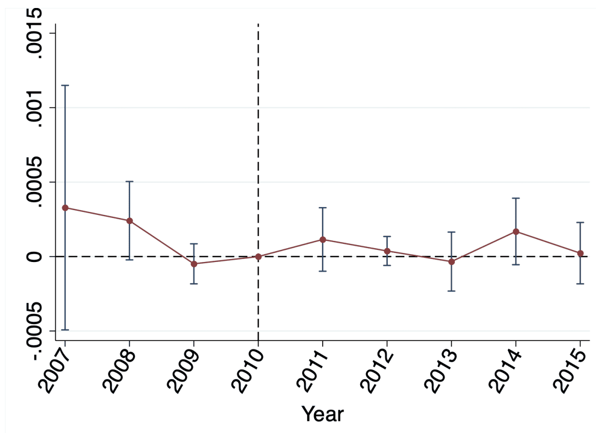
(a) Share with Outpatient Obs or "Suspected" Obs



(b) Share Revisit in 30 Days



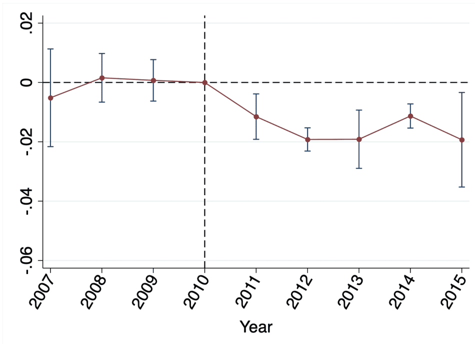
(c) Share Died in 30 Days



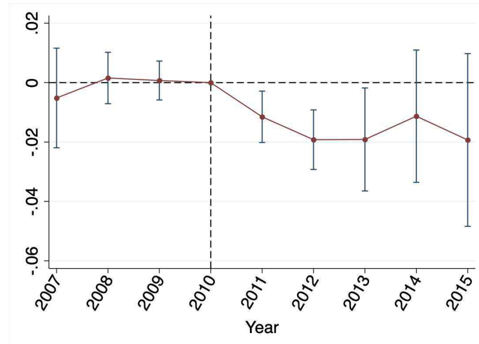
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome. Panel (a) shows the share of Medicare ED visits that report outpatient observation payment or where the outpatient stay spans two days ("suspected" observation stay). The 2010 mean was 12%. Panel (b) shows the share of Medicare ED visits with a revisit to the ED within 30 days (2010 mean: 15%). Panel (c) shows the share of visits with a beneficiary death within 30 days of visit (2010 mean: 1.4%). A Medicare ED is defined as an inpatient or outpatient claim with ED charges. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, Outpatient file, Master Beneficiary Summary file, and CMS audit data.

Figure G17. Robustness to Border Segment Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

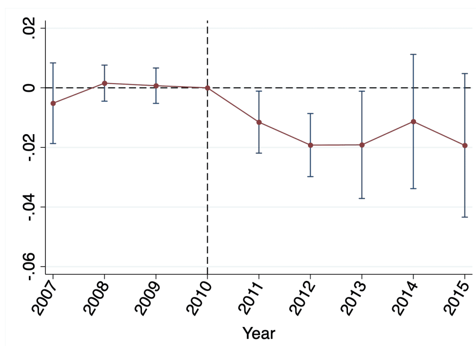
(a) 50-mile Border Segments



(b) 100-mile Border Segments (main)



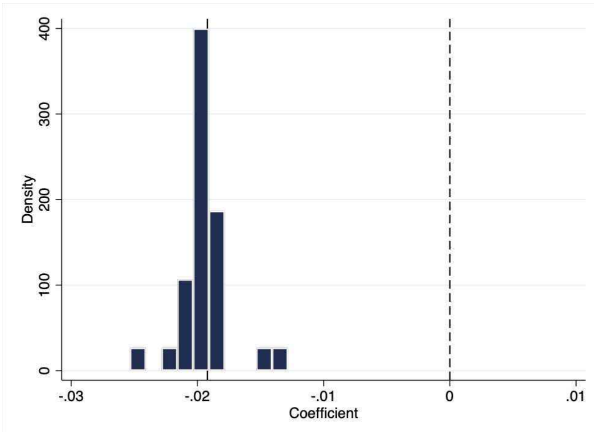
(c) 150-mile Border Segments



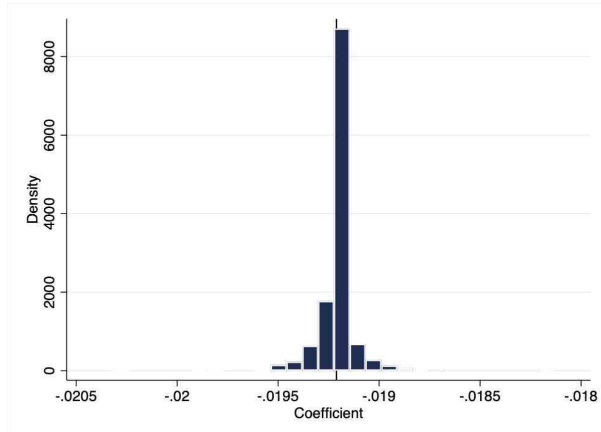
This figure plots event studies of the reduced form coefficients and 95% confidence intervals of the specification in Equation 5, scaled by the correlations between the instruments and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different border segment lengths used for clustering. The segments are defined such that they do not cross state lines. Panel (a) shows the results for 50-mile segments, (b) shows the main results for 100-mile segments, and (c) shows the main results for 150-mile segments. Data: MEDPAR and CMS audit data.

Figure G18. Robustness Test: Leave-One-Out Coefficients of 2012 Effect of 2011 Audit Rate on Log Medicare Admissions

(a) Leaving 1 state out at a time



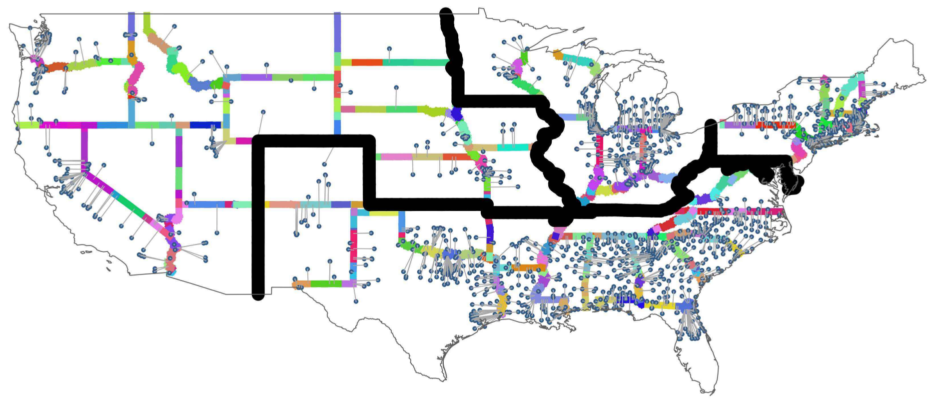
(b) Leaving 1 comparison group out at a time



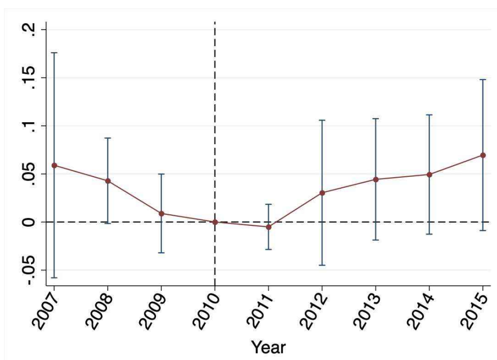
This figure plots distributions of the 2012 coefficient of the reduced form event study specification in Equation 5 on log Medicare admissions, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample the outcome. Panel (a) plots the distribution of the coefficient when leaving one state out at a time, and panel (b) plots the distribution of the coefficient when leaving one hospital neighbor comparison group out at a time. Data: MEDPAR and CMS audit data.

Figure G19. Falsification Test: Interior State Borders

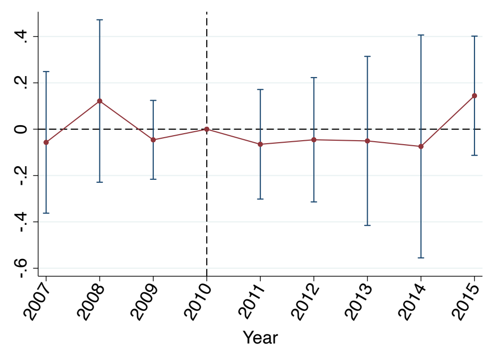
(a) Falsification Test Border Segments and Hospitals Within 100 Miles



(b) Event Study on All Interior Borders



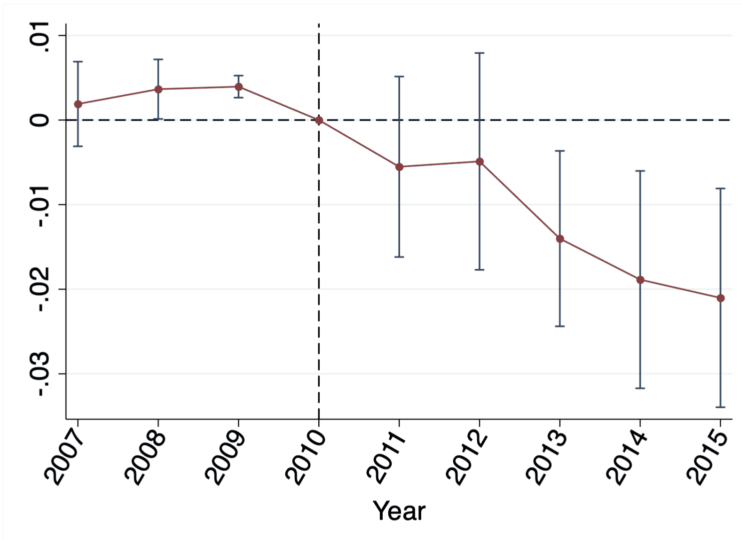
(c) Event Study on MAC Interior Borders



Panel (a) of this figure plots a map of state borders on the interior of RAC regions, divided into 100-mile segments that do not cross state borders. The RAC border is the thick black line. Each dot represents a hospital within 100 miles of the interior state borders, excluding hospitals that are in the main sample (within 100 miles of the RAC border). The line between the hospital and the interior state border denotes the closest interior state border to that hospital. Panel (b) plots the reduced form coefficient and 95% confidence interval of the specification in Equation 5 (scaled by correlation between 2011 audit rate and 2011 leave-one-out audit rate in the interior border hospital sample), where the outcome variable is log Medicare admissions (MEDPAR). Panel (c) plots the event study, restricted to the interior MAC borders. The interior MAC border sample consists of hospitals along the border between MAC Regions E and F (OR/ID/NV/UT/AZ/CA), F and G (ND/SD/MN/NE/IA), G and I (IN/IL/KY), M and J (SC/GA/TN/NC/VA), J and N (AL/GA/FL), and L and K (PA/NY), as defined in the MAC jurisdiction map in 2010 (link to 2010 MAC map; last accessed June 2023). Sample is comprised of hospitals within 100 miles of the state interior border with at least 1 hospital in their "neighbor hospital comparison group" and are clustered at the state and border segment level. Data: MEDPAR and CMS audit data.

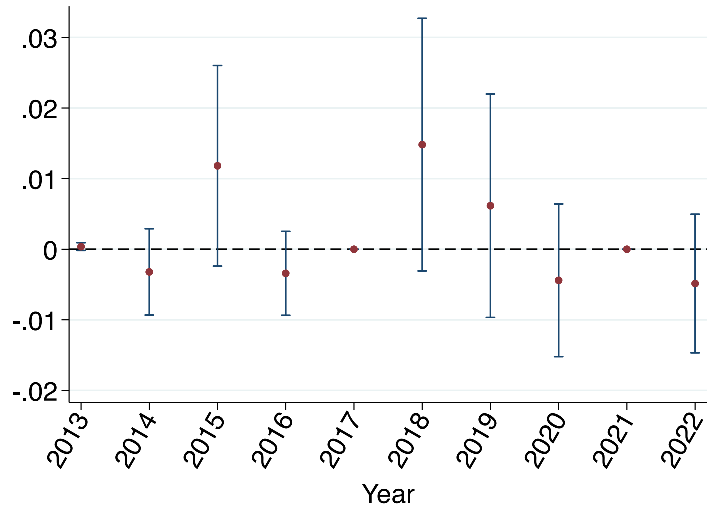
Figure G20. Event Studies of Effect of 2011 Audit Rate on Coding

(a) Log Mean Diagnoses per Admission



This figure plots event studies of the IV coefficients and 95% confidence intervals of the specification in Equation 3. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on a hospital-level outcome in a given year. The outcome is the log number of ICD-9 or ICD-10 diagnoses per claim. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their "neighboring hospital comparison group." Data: MEDPAR and CMS audit data.

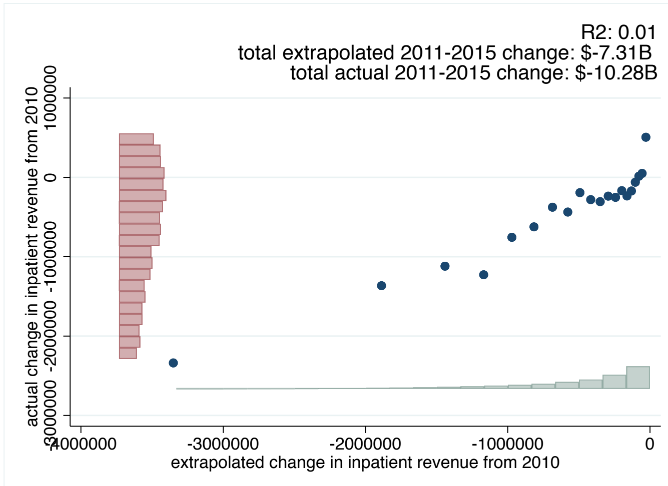
Figure G21. Coefficients of Effect of 2011 Audit Rate on Rural Hospital Closure in a Given Year



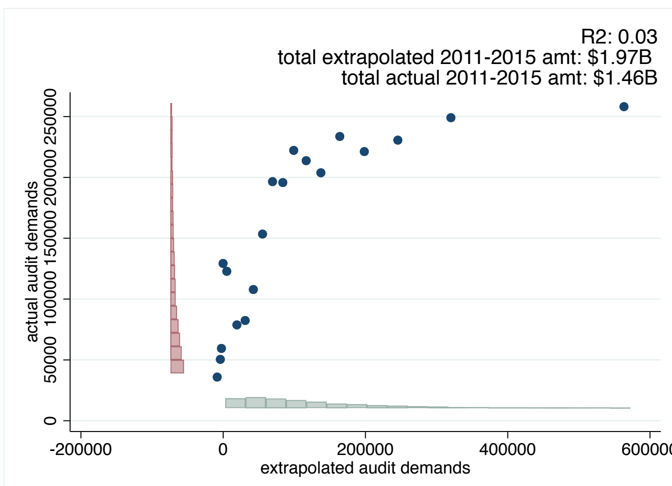
This figure plots the coefficients from individual regressions of the instrumented 2011 audit rate on a dummy for whether a hospital closed in a given year, for rural hospitals in the border sample. There are no closures prior to 2013 and no closures in 2017 and 2021 in the border hospital sample. Data: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

Figure G22. Extrapolation Exercise: Actual vs. Extrapolated Savings

(a) Savings from changes in Medicare inpatient spending



(b) Savings from audit demands



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

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, Wojciech, "Tax simplification and tax compliance: An economic perspective," 2007. Columbia University.

_ **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.

Shi, Maggie, "Replication Data for: 'Monitoring for Waste: Evidence from Medicare Audits'," 2023. Harvard Dataverse, https://doi.org/10.7910/DVN/WS9OAQ.

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.