The Costs and Benefits of Monitoring Providers: Evidence from Medicare Audits*

Maggie Shi †

November 5, 2021

Job Market Paper

Latest version: (link)

Abstract

This paper examines the extent to which government programs should monitor for wasteful expenditure when outsourcing to third parties, taking into account the costs and savings associated with monitoring. I use novel administrative data to study the largest Medicare monitoring program aimed at identifying and reclaiming payments for unnecessary inpatient admissions. I exploit plausibly exogenous variation across hospitals and across patients, and find that the majority of savings are due to the deterrence of future admissions. I do not find evidence that the marginal patient denied admission is harmed, suggesting that hospitals fine-tune their response to target unnecessary care. But in doing so, they incur compliance costs and adopt technology specifically aimed at assessing the medical necessity of care. For every $1,000 in savings to Medicare, hospitals incur $216 in compliance costs. My welfare calculations imply that despite the costs, increasing monitoring improves welfare.

*I am grateful to Wojciech Kopczuk, Adam Sacarny, Pietro Tebaldi, and Michael Best for their input and support with this project. 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, Geoff Clarke, Melinda Pitts, Bernard Salanie, and seminar participants at EEA, APPAM Student Research Series, ASHEcon, WEAI Graduate Student Workshop, EHEC, SHESG, Columbia Health Policy, and Columbia Applied Micro for helpful comments and feedback. I thank Mohan Ramanujan, Daniel Feenberg, Elizabeth Adams, Jean Roth, Adrienne Henderson, and Ashley Badami for their assistance in accessing and managing the data. I gratefully acknowledge fellowship support from the Agency for Healthcare Research and Quality (#R36HS027715-01). All errors are my own.

†Department of Economics, Columbia University. m.shi@columbia.edu

1 Introduction

Much of government expenditure is spent on outsourcing the provision of goods and services to third parties. In the U.S., contracted goods and services account for 40 percent of federal discretionary spending, 73 percent of defense spending, and nearly all of the spending within Medicare, the federal health insurance program that covers the elderly and disabled (U.S. Government Accountability Office, 2019; U.S. Department of Defense, 2021; Boards of Trustees for Medicare, 2021). Such arrangements can give rise to wasteful expenditure if third parties have incentive to provide more than what is deemed necessary by policymakers. A natural solution is to establish monitoring mechanisms that identify overprovision and deny the corresponding payments on a case-by-case basis (Nalebuff and Scharfstein, 1987; Laffont and Tirole, 1992). But the tradeoff is that these mechanisms can be costly – there are monitoring costs for the government, compliance costs for third parties, and downstream costs that may arise if the quality of goods and services changes. Understanding whether monitoring improves welfare thus requires estimates of both the costs and savings associated with it.

In this paper, I study the effect of monitoring in Medicare on government savings, provider compliance costs, and patient health outcomes. Using two identification strategies, one at the provider level and one at the patient level, I find that monitoring saves Medicare money by reducing unnecessary care, but also imposes considerable compliance costs on providers. In response to monitoring, providers scale back expenditure *without* harming patients, but in doing so their administration costs increase as they adopt IT to detect unnecessary care. For every \$1,000 in Medicare savings, providers incur \$216 in compliance costs. Taking into account the costs and savings, I find that monitoring is welfare-improving.

In particular, I study the largest Medicare monitoring program: the Recovery Audit Contractor (RAC) program. Through this program, private auditing firms (i.e., RACs) are paid a contingency fee to conduct manual post-payment reviews ("audits") of medical claims to identify and reclaim erroneous Medicare payments, such as payment for unnecessary care.[1] I focus on RAC audits of hospital admissions, Medicare's largest service expenditure category. In the first five years of the program, RACs manually audited four percent of all Medicare hospital admissions. I study hospital responses to RAC audits with an instrumented difference-in-difference across hospitals, and patient health effects with a difference-in-difference across patients who visit a hospital's emergency department (ED).

The central identification challenge is that auditing is endogenous, since RACs do not

[1] Medicare defines necessary care as services that are "proper and needed for diagnosis or treatment..., meet the standards of good medical practice in the local area, and aren't mainly for the convenience of you or your doctor" (Centers for Medicare and Medicaid Services, 2006).

audit randomly. In the hospital-level strategy, I exploit plausibly exogenous variation in audit intensity across different RACs, focusing on hospitals close to the border between different RACs' jurisdictions. To measure audit intensity, I use novel administrative data on the universe of RAC audits, which I then link to hospital data on Medicare admissions, administration costs, and IT adoption. I estimate an instrumented difference-in-difference specification that compares hospitals on the high-audit side of the border to their neighbors on the low-audit side, before and after a major expansion of audit scope in 2011.

I find that monitoring through the RAC program saves Medicare money: it not only reclaims payments from audited admissions, but, importantly, also deters admissions – the vast majority of savings are derived from this deterrence effect. A one percentage point increase (relative to an average of 2.2) in a hospital's 2011 audit rate leads to a two percent decrease in admissions in following years. Exposure to increased monitoring has a persistent effect, as the reductions in admissions continue even once the RAC program is significantly scaled back. RAC audits mostly deter admissions that Medicare considers most likely to be unnecessary – namely, short stay admissions with length of stay ≤ 2. However, I also find evidence that monitoring leads to a short-term uptick in hospital administration costs. One source of these costs is the technology that hospitals adopt in response to RAC audits. Hospitals subject to higher audit rates are more likely to install "medical necessity checking" software, which is used specifically to identify unnecessary care at risk of denial by a payer (3M, 2016).

Given that monitoring reduces admissions, a natural question arises of what effect it had on patients and their health. However, estimating the patient health effects at the *hospital* level is challenging: the changes in admission volume induce changes in the composition of admitted patients, and it is difficult to identify the counterfactual patients who were not admitted. To circumvent these issues, I switch to a patient-level empirical strategy. This second strategy focuses on patients who visit the ED, a context where I can observe who is admitted as inpatient or not. Specifically, I leverage the "Two Midnights rule," which barred audits in cases where the patient spent two or more *midnights* in the hospital, *including* time in the ED. The rule effectively increased audit likelihoods for ED visits in which the patient arrived to the hospital after midnight.

I use a difference-in-difference specification on hospital discharge data to compare ED visits associated with before- and after-midnight arrivals. I find that once the policy is introduced, hospitals cut back on inpatient admissions for after-midnight arrivals. However, I do *not* find evidence that patients who arrived after midnight were more likely to revisit a hospital within 30 days, which is a proxy for patient health that is observable in discharge data. This suggests that the marginal patient's health was unaffected, despite the patient being denied inpatient admission. Hospitals targeted patients in the middle of the severity distribution, some of whom faced up to a 25 percent reduction in admission likelihood. But even among these patients, I detect no increase in revisit rates.

Taken together, the empirical estimates suggest that the costs and benefits of the RAC program derive mostly from the Medicare savings and monetary costs associated with monitoring, rather than any effects on patient health. I then use these empirical estimates to calculate the welfare effect of a marginal increase in monitoring. I adapt Keen and Slemrod (2017)'s sufficient statistics welfare framework to the Medicare monitoring context. This framework accounts for the government's monitoring costs, which I calculate in my context using RACs' contingency fees. I also adapt the framework to include *private compliance costs*, motivated by my findings on hospital administration costs. Given the dynamics of the effects I find, in which compliance costs are mostly incurred upfront but Medicare savings accrue over several years as more admissions are deterred, I calculate the cumulative welfare effect of an increase in the 2011 audit rate across multiple subsequent years.

The welfare analysis shows that under the assumption of no effect on patient welfare, a marginal increase in the 2011 audit rate is welfare-improving in the long run. After five years, the societal value of savings from monitoring outweighs the hospital compliance costs incurred upfront. But absent these compliance costs, the welfare gain from increased monitoring would have been *immediate*, and the welfare effect after five years would be *nine times larger*. Thus, while the substantial Medicare savings and null patient health effect make monitoring worthwhile, the overall welfare improvement from monitoring is attenuated considerably by providers' compliance costs.

In studying the RAC program, I contribute to our understanding of policies to reduce the provision of unnecessary healthcare. I present, to my knowledge, the first quasi-experimental evidence on the effects of monitoring in Medicare. Despite the large fiscal impact of RAC auditing (as well as similar monitoring initiatives),[2] there is little academic work studying RACs outside of select hospitals (Sheehy et al., 2015, 2017). I highlight costs and savings beyond what was included in policymakers' cost-benefit analyses of the program (Centers for Medicare and Medicaid Services, 2011b). While policymakers only considered the total payments *reclaimed* by RACs, I find that this calculation is missing two important components: the savings from deterred admissions and the costs for providers to comply.

[2]The Medicare and Medicaid programs collectively spend $1.5 billion per year on monitoring (Department of Health and Human Services, 2021). For example, Medicare conducts stratified randomized audits through the Comprehensive Error Rate Testing program, continuing medical reviews and education through the Targeted Probe and Educate program, and also directs Medicare Administrative Contractors (MAC), Zone Program Integrity Contractors (ZPIC), and Supplemental Medical Review Contractors (SMRC) to conduct a variety of pre-payment and post-payment reviews on an as-needed basis (Centers for Medicare and Medicaid Services, 2016).

These findings also shed light on how healthcare providers respond to non-financial incentives. Administrative actions like monitoring reduce the *realized* price for care (i.e., the price after taking into account denials or billing costs), but ideally only in cases where the care has low clinical value. We know that healthcare providers respond to *contracted* prices, by changing the quantity and type of care provided (Cutler, 1995; Ellis and McGuire, 1996; Clemens and Gottlieb, 2014; Einav et al., 2018; Eliason et al., 2018; Alexander and Schnell, 2019; Gross et al., 2021; Gupta, 2021), or by changing how care is documented (Silverman and Skinner, 2004; Dafny, 2005; Sacarny, 2018; Gowrisankaran et al., 2019). In contrast, less is known about how providers respond to administrative mechanisms that change the realized price, despite the fact that such mechanisms are used widely by almost all payers.[3] By studying provider responses to post-payment reviews, I contribute to a nascent literature on non-financial incentives like billing complexity (Dunn et al., 2021), fraud detection (Leder-Luis, 2020; Nicholas et al., 2020; Howard and McCarthy, 2021), and prior authorization (Brot-Goldberg et al., 2021; Roberts et al., 2021).

More generally, this paper illustrates an example of a potential downside to well-intentioned public policy: high compliance costs for the third parties involved. Previous work on other programs has found that individuals and firms often face private costs when they interact with the government, often in instances where the individual or firm has something to gain – for example, when applying for benefits (Nichols and Zeckhauser, 1982; Currie, 2006; Deshpande and Li, 2019) or requesting tax refunds and credits (Kopczuk and Pop-Eleches, 2007; Zwick, 2021). I document an instance where third parties incurred substantial private costs to *save* money on behalf of the government, going so far as to install technology to identify wasteful expenditure. My findings lend credence to the notion that simply reducing wasteful government expenditure is not sufficient for a policy to be worth implementing – the costs must be considered as well.

The rest of the paper proceeds as follows. Section 2 describes the policy context of the RAC program and the data I use. Section 3.1 describes the hospital-level empirical strategy, and Section 3.2 describes the patient-level empirical strategy on ED visits. Section 4 presents the empirical results, and Section 5 incorporates and interprets these results in a welfare analysis framework. Section 6 concludes.

[3] Beyond Medicare, Medicaid also has its own RACs and State Medicaid Fraud Control Units. Almost every private insurer conducts some form of utilization review to monitor providers and assess quality and costs (Dranove and Satterthwaite, 2000). For example, see the discussion of auditing in the following insurer provider manuals: Humana (2020); UnitedHealthcare (2020); Empire Blue Cross Blue Shield (2020). In a study of remittance data on claim denials of outpatient visits, Gottlieb et al. (2018) find that 18 percent of Medicaid fee-for-service claims are challenged, while for Medicare fee-for-service it is 7 percent, and different private payers challenge between 2 and 10 percent of claims.

2 Policy Context and Data

2.1 Unnecessary Inpatient Stays and the Recovery Audit Contractor Program

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,[4] the payment rate depends on the patient's diagnosis, pre-existing health conditions, and procedures conducted during the stay – importantly, it does not generally depend on the admission's length of stay.

Over time, policymakers became increasingly concerned with one area of vulnerability: unnecessary inpatient stays, which they felt were particularly common among short 0-2 day stays (Centers for Medicare and Medicaid Services, 2011b). The Medicare Payment Advisory Commission (MedPAC), a non-partisan government agency, contended that hospitals were admitting patients for short inpatient stays because they were very profitable (Medicare Payment Advisory Commission, 2015). MedPAC estimated that the payment-to-cost ratio for short stays was over two times higher than that of longer stays. Appendix Section A.1 describes the Medicare inpatient prospective payment system and short stays in greater detail.

To address this issue, in 2011 Medicare directed contractor firms ("RACs") in the RAC program to begin monitoring and reclaiming payments for unnecessary inpatient admissions. RAC audits are carried out by four private firms,[5] each of which operates in its own geographic region and is in charge of conducting all RAC audits for Medicare claims in its region. The regions are illustrated in Figure 1a – they fall along state lines and, in the context of medical claims reviews, are unique to the RAC program.[6] RAC audits were introduced nationally in 2009 after a pilot program in select states, but RAC activity was fairly limited until 2011. In 2011, Medicare began allowing RACs to audit and correct payments for unnecessary inpatient stays. There was a 537 percent increase in the number of audits from 2010 to 2012 (Figure 1b).

95 percent of RAC audits for inpatient stays are conducted as follows: the RAC first

[4] 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 post-acute care, in case Medicare pays a per diem rate (Office of the Inspector General, 2019).

[5] In addition to working as RACs, the firms also conduct data analysis and recovery services for other clients in the U.S. and in other countries, working across a variety of different sectors, like healthcare, debt recovery, and tax collection.

[6] The RAC regions are also used by Durable Medical Equipment Medicare Administrative Contractors, who do not process medical claims.

runs a proprietary algorithm on Medicare claims data to flag individual claims for issues like missing documentation, incorrect coding, or – starting in 2011 – unnecessary care. A medical professional hired by the RAC, typically a nurse or a coder, then requests and manually reviews all documentation associated with the flagged claim. The medical professional makes a determination about whether Medicare made an overpayment. If the medical professional determines that there was an overpayment, then they can correct it by demanding a payment back from the provider.[7] There is no additional penalty associated with a corrected payment. The RAC firms are paid a negotiated contingency fee on the payments they correct: 9–12.5 percent, depending on the firm, of the reclaimed payment after appeals. Figure D1 illustrates the full claims auditing and appeals process, including the remaining 5 percent of inpatient stay audits that are automated reviews.

Figure 1b illustrates the total and reclaimed payments per hospital for inpatient stay audits, by year of audit. At the program's peak, RACs were reclaiming an average of $1 million per hospital annually (three percent relative to the average hospital's Medicare inpatient revenue of $32 million).[8] By 2020, 96 percent of hospitals had at least one inpatient stay that was audited. Hospital audit rates are correlated across years (Figure D2). RAC audits were scaled back significantly in 2014 and 2015, as Medicare paused the program to evaluate complaints by hospitals and industry stakeholders of overly aggressive auditing. Appendix Section A.2 describes the RAC regions, RAC firms, audit process, and timeline of the RAC program in greater detail.

Two years after expanding RACs' audit scope to include medical necessity, in August 2013 Medicare introduced a new rule to clarify which admissions were allowed to be audited for medical necessity. Under the so-called "Two Midnights" rule, Medicare counted the number of *midnights* a patient's entire time in the hospital crossed – this includes the time spent in the ED, in outpatient, and in inpatient.[9] If the patient's total time at the hospital crossed two midnights, then the stay was presumed to be necessary and RACs were barred from

[7]RACs can also identify underpayments, which are corrected by refunding the payment to the provider. In 2011, 6 percent of inpatient stay audits resulted in an underpayment determination.

[8]RAC audits of inpatient stays drop off significantly in 2014, when Medicare paused RAC operations to "review and refine" the program, in response to complaints from industry stakeholders and providers, who inundated the audit appeals system (Foster and McBride, 2014). Appendix Section A.2 covers the timeline of the RAC program in more detail.

[9]Midnight cutoffs are surprisingly common in hospital 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 how many midnights a patient's *entire stay crosses*, starting from the *ED arrival hour* (i.e., the hour the patient is recorded as first stepping foot into the hospital) if the patient entered through the ED. In contrast, the rules studied by these two papers focus on how many midnights a patient's *hospital admission* crossed, starting from the *hospital admission hour* (i.e., the hour that the patient is formally admitted as inpatient or, in the case of newborns, born).

auditing this stay for medical necessity. If the patient's stay did not cross two midnights, then RACs could audit it (Centers for Medicare and Medicaid Services, 2017). Among Medicare patients who enter a hospital through the ED,[10] the Two Midnights rule effectively increased audit likelihoods for patients who arrived at the ED after midnight, relative to those who arrived before.

2.2 Data

The hospital-level analysis uses four main datasets. First, I use novel, audit-level administrative data on the RAC program. The data spans 2010 to 2020 and includes claim-specific information 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 dataset covers 4.5 million audits of inpatient stays.

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

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

Fourth, I use data on IT adoption from the Healthcare Information and Management Systems Society (HIMSS) Analytics Database, which is a yearly survey of IT used by hospitals and other healthcare providers. HIMSS asks hospitals each year to report the types of IT they are installing or have already installed. In particular, I focus on the installation of medical necessity checking software, which hospitals use to identify potentially unnecessary care that could result in billing denials. Additionally, to study heterogeneity across hospital types, I also use hospital characteristics from the Medicare Provider of Services file and hospital merger data via Cooper et al. (2019).

Table 1 presents summary statistics by RAC region. Hospitals in regions B (Midwest) and C (South) have much lower audit rates than hospitals in regions A (Northeast) and D (West). Within each region, rural hospitals, small hospitals, non-profit hospitals, and hospitals with a higher share of short stay Medicare admissions are more likely to be audited (Figure

[10] 73 percent of Medicare inpatient admissions originate in the ED.

D3). Appendix Section A.3 discusses additional claim-level and hospital-level characteristics associated with auditing in further detail.

In the patient-level analysis of ED visits, I use the Florida State Emergency Department Database (SEDD) and State Inpatient Database (SID) between 2010 and 2015. I focus on Florida because it is the only state that reports *ED arrival hour* in the publicly available SID and SEDD datasets; Medicare's Inpatient and Outpatient files do not report this variable. The most granular unit of time for ED arrival in my data is 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 proxy for patient health after an ED visit by considering whether the patient revisits any hospital in Florida shortly after, either as an ED visit or an inpatient visit.[11] This choice is dictated by the fact that mortality is not observable in hospital discharge data like SID and SEDD.

Table 2 reports summary statistics for before- and after-midnight arrivals before the Two Midnights rule (in 2013Q2). Figure 2 plots the quarterly share of before- and after-midnight Medicare ED arrivals who are admitted as inpatient. Prior to the Two Midnights rule, after-midnight arrivals are more likely to be admitted as inpatient, but this gap in admission rate closes once the Two Midnights rule is implemented in 2013Q3.

3 Empirical Strategies

3.1 Effect of Monitoring on Hospital Outcomes: Hospital Admissions, Revenue, Costs, and IT Adoption

The aim of the hospital-level analysis is to understand how a marginal increase in a hospital's 2011 audit rate would affect its behavior in subsequent years. I leverage plausibly exogenous variation in 2011 audit rates driven by how aggressive a hospital's RAC is.

Border Hospital Sample: Figure 1a illustrates the variation in aggressiveness across RACs in 2011. Along the borders between RAC regions, there are sharp changes in audit rate from one side of the border to the other. The changes in audit rate across the RAC border are twice as large as the changes across state borders within each RAC region. The RAC border spans multiple states, so the differences at the border cannot be attributed to any individual state.

My research design compares subsets of hospitals close to the border, where I define "close" as being within 100 miles of it.[12] Since these border hospitals are geographically

[11]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.

[12]In robustness tests, I check that the results are not sensitive to the 100-mile sample definition.

close and serve overlapping patient pools, they should be relatively similar in terms of the characteristics of their Medicare admissions. Table E1 explores this by comparing hospitals with above- and below-median audit rates in the overall sample and in the border hospital sample. While the differences in audit rates are similar across the two samples, the differences in Medicare admission volume, payment per admission, and total Medicare inpatient revenue are smaller in the border sample than in the overall sample.

Neighbor Comparison Groups: In order to compare hospitals that are close *to each other*, rather than just hospitals that are close to the border, for each hospital I identify a unique set of neighbors and call this its "neighbor comparison group." I define a hospital's neighbor comparison group to be the hospitals on the *other* side of the border, within a 100-mile radius.[13] I then include a fixed effect for each group, interacted with a year indicator, to account for time-varying local trends.

Figure D5 illustrates an example of how I construct a 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%. Its neighbors in the neighbor comparison group are the hospitals on the other side of the border, within 100 miles – hospitals in Kansas (RAC Region D) who 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 these group-year fixed effects improves upon a specification with just border fixed effects for two reasons. Prior research has documented substantial geographic variation in Medicare utilization and spending (Skinner, 2011; Finkelstein et al., 2016). Each RAC border spans hundreds of miles. Therefore, comparing hospitals that are geographically far from each other, such as in a specification with just border (or border-year) fixed effects, risks confounding from local trends in utilization, spending, and patient health. Additionally, identifying a unique set of neighbors for each hospital allows for the inclusion of hospitals at the corner of intersections of borders, without having to arbitrarily assign hospitals to a single border.

Because a hospital can be in many other hospitals' neighbor comparison groups, the sample includes repeated hospital observations. Duplicate observations will have correlated errors. To account for this, I divide the border into smaller segments and cluster at the border segment level. Figure D6 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.

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

Specification: The event study specification for the hospital-level strategy is:

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

In Equation 1, Y_{ht} is an outcome for hospital h in year t, X_h^{2011} is the hospital's 2011 audit rate, $\phi_{g(h)t}$ is a neighbor comparison group-year fixed effect, and ψ_h is a hospital fixed effect. I estimate Equation 1 on the border hospital sample.

There is a β^{τ} for each year τ between 2007 and 2015, omitting 2010. β_{τ} can be interpreted as the effect of a one percentage point increase in *2011 audit rate* on a hospital outcome in year τ, relative to 2010.

Audit Rate Instrument: One reasonable concern with estimating Equation 1 is the endogeneity of a hospital's 2011 audit rate – i.e., that $E[\varepsilon_{ht}|X_h^{2011}] \neq 0$. To isolate variation driven by the RAC and not the hospital itself, I consider how aggressively the RAC audits *other hospitals*. 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 formally defined as:

$$Z_h^{2011} = \frac{1}{n_{s(h)} - 1} \sum_{h' \in \ s(h) \setminus h} X_{h'}^{2011} \ , \tag{2}$$

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

There are 8 instrumented variables in Equation 1, since X_h^{2011} is interacted with year indicators between 2007 and 2015 (omitting 2010). Thus I generate 8 instruments, each of which is an interaction of Z_h^{2011} with a year indicator.[14] For example, the first stage for the audit rate interacted with the 2012 year indicator is:

$$1[t = 2012] \times X_h^{2011} = 1[t = 2012] \times Z_h^{2011} \gamma^{2012} + \chi_{g(h)t} + \lambda_h + \nu_{ht} \ . \tag{3}$$

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

$$Y_{ht} = 1[t \geq 2011] \times X_h^{2011} \beta^{post} + \phi_{g(h)t} + \psi_h + \varepsilon_{ht} \ . \tag{4}$$

[14] Given the matching year interactions on the endogenous variable and the instrument, the first stage for each instrument is effectively a cross-sectional regression between X_h^{2011} and Z_h^{2011}, with controls. The second stage then estimates the coefficients of the fitted 2011 audit rate interacted with year indicators on Y_{ht}

In this case, I instrument for $1[t \geq 2011] \times X_h^{2011}$ by interacting Z_h^{2011} with a post-2011 indicator variable, and the first stage becomes:

$$1[t \geq 2011] \times X_h^{2011} = 1[t \geq 2011] \times Z_h^{2011} \gamma^{post} + \chi_{g(h)t} + \lambda_h + \nu_{ht} \ . \tag{5}$$

Identification Assumptions and Checks: The identification strategy relies on three assumptions: 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 third, that the leave-one-out audit rate affects hospital behavior only through its relationship to a hospital's audit rate (*exclusion restriction*).

Say that the pattern of sharp changes at the border in Figure 1a was driven *entirely* by hospitals or patients, not RACs. In that case, we would expect to see a similar pattern for hospital and patient characteristics as well. Figure D8a plots a hospital-level measure that is correlated with 2011 audit rates in the cross-section: the short stay share of 2010 Medicare admissions. Figure D8b plots the predicted 2011 audit rate, where the prediction depends on patient stay characteristics. Importantly, the prediction does *not* depend on the identity of the RAC. Neither of these measures displays sharp changes at the border, suggesting that the pattern in Figure 1a is indeed driven by RACs.

What drives these differences in audit intensity across RACs? One explanation could be spillovers from other hospitals in the same RAC region. This could be the case if a RAC combines data from across its region to train a single algorithm, rather than developing specific algorithms targeted to each hospital. It could also be the case if RACs set their strategies according to the average regional cost to audit, as opposed to the cost to audit each individual hospital. Another explanation could be that because each RAC comes from a different industry background (e.g., the RAC in Region A is a debt collection agency, while the RAC in Region C is a healthcare data analysis company), there are baseline productivity differences across RACs based on their prior experiences.

Identification also requires making the parallel trends assumption. With the inclusion of group-year fixed effects, we only need that hospitals on opposite sides of the border that are *geographically close to each other* do not differentially deviate from local trends. While this assumption is in principle untestable, a lack of preexisting differential trends in the event study would support making it. A potential violation of this assumption would be if the results are due to state policies rather than RAC audits. However, to generate the results below, these policies would have to be consistent across multiple states on one side of the border, and would all have to change in 2011. In robustness tests, I show that the results are robust to omitting individual states, meaning that the effect is not driven by a single state's policy changes.

Finally, we also need the exclusion restriction. This would be violated if the audit rates of leave-one-out hospitals depend on the hospital that is left out. This might happen if the hospital that is left out has a large market share within its local market. To address this concern, I run a robustness test that uses the average audit rate of hospitals in the same state but *in other markets* as an instrument, and find that the results are similar. I also show that the results are robust to using the state and RAC region audit rates as instruments.

3.2 Effect of Monitoring on Patient Outcomes: Admission Likelihood and Revisits

I next turn to the patient-level empirical strategy, which studies the effect of auditing on the likelihood of inpatient admission from the ED and subsequent patient health outcomes. I leverage the Two Midnights rule by splitting ED visits by whether the patient arrived before or after midnight, and then comparing them pre- and post-Two Midnights rule.

Specification: The event study specification is:

$$Y_v = \sum_{\tau=2010Q1}^{2016Q4} 1[q=\tau] \times 1[T \geq 00{:}00]\beta^{\tau} + \boldsymbol{W_v'\gamma} + \lambda_{hq} + \phi_{hT} + \varepsilon_v \ , \tag{6}$$

where ED visit v occurs in quarter q at hospital h, and the ED arrival hour of the visit is $T \in [21{:}00, 03{:}00)$ (i.e., between 9PM-3AM).[15] Y_v is the outcome of interest, such as an indicator for whether the visit resulted in an inpatient admission, or whether the patient revisited a hospital within 30 days. $1[q=\tau]$ is an indicator for whether the visit occurred in quarter τ, omitting 2013Q3. $1[T \geq 00{:}00]$ is an indicator for whether the patient arrived at the ED after midnight. λ_{hq} is a hospital-quarter fixed effect and ϕ_{hT} is a hospital-ED arrival hour fixed effect. W_v are controls for patient characteristics associated with the visit, including patient age, race, Hispanic, point of origin, indicator for whether last ED visit was within three days, number of chronic conditions, and average income in patient's zip code. β^{τ} is the coefficient of interest and can be interpreted as the effect of increased audit likelihood on after-midnight ED arrivals in quarter τ, relative to 2013Q3.

Equation 7 pools the event study into a single post-policy coefficient β:

$$Y_v = 1[q \geq 2013\text{Q3}] \times 1[T \geq 00{:}00]\beta + \boldsymbol{W_v'\gamma} + \lambda_{hq} + \phi_{hT} + \varepsilon_v \ , \tag{7}$$

[15] In robustness tests I check that the results are robust to using bandwidths ranging from one and five hours around midnight.

where $1[q \geq 2013Q3]$ is an indicator for whether the visit occurs after the Two Midnights rule is implemented in 2013Q3.

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 assumption, I check that there are no differential pre-trends between the two groups in the event study figures.

The second assumption is that there is no manipulation of the ED arrival hour. This would be violated if, for example, hospitals misreported after-midnight ED arrivals as arriving before midnight. If this were the case, we would expect to see bunching of ED arrivals right before midnight once the policy is implemented (i.e., an increase in the share of patients reported arriving between 11PM and midnight). Figure D9 plots the share of patients by ED arrival hour, pre- and post-policy. After the reform, there is no visual evidence of bunching. I test this empirically in Table E2 by considering whether there is a higher share of patients arriving in the hour before midnight (column 1), or a lower share of patients arriving after midnight (column 2), post-policy. Neither of these measures changes after the Two Midnights rule is implemented.

From a practical point of view, note that it may be challenging for hospitals to manipulate ED arrival hour in response to the Two Midnights rule. The arrival hour is recorded as soon as the patient walks in to the ED, which makes it more difficult to manipulate than a measure which is recorded later on. Additionally, to game the Two Midnights rule, hospitals would have to make after-midnight arrivals look like before-midnight ones. This would require them to actively move up a patient's ED arrival hour to an earlier time, rather than a more passive form of misreporting by "dragging their feet" to record a later arrival hour.

Another concern could be that hospitals respond to the Two Midnights rule by extending all stays to cross 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. In the SID and SEDD discharge data, I cannot directly observe how many midnights a patient's *entire time* in the hospital crossed. However, I do not find evidence that after-midnight patients have additional charges, diagnoses, or procedures after the rule is implemented (Table E3), suggesting that hospitals did not respond to the Two Midnights rule by extending stay duration.

4 Results

4.1 Hospital Admissions, Revenue, Costs, and IT Adoption

Results Figure 3 plots a binscatter of the cross-sectional relationship between the leave-one-out state audit rate and hospital audit rate in the border hospital sample. The leave-one-out audit rate explains 34 percent of the variation in the actual audit rate. There is a positive linear relationship between the two and it is not driven by outliers, which supports using a linear specification.

Figure 4 presents the first set of main results from Equation 1: the IV event study coefficients on hospital-level outcomes. Table 3 reports the yearly coefficients for 2011 to 2015 (for brevity, the pre-2011 coefficients are estimated but not reported in the table). Figures 4a and 4b plot the results for log Medicare admissions and log Medicare inpatient revenue, where inpatient revenue is defined as the sum of all Medicare inpatient payments. Hospitals with higher audit rates do not seem to be on differential pre-trends relative to their neighbors on the other side of the border, which supports making the parallel trends assumption. Starting in 2011, there is a decline and then a plateau in Medicare admissions and inpatient revenue among hospitals with higher audit rates. A one percentage point increase in 2011 audit rate results in a 1.1 percent decrease in admissions in 2011, which increases in magnitude to a 1.9 percent decrease by 2012 and 2013. Similarly, a one percentage point increase in 2011 audit rate results in a 1.0 percent decrease in inpatient revenue in 2011, and then a 1.7 percent decrease in 2012 and 2.8 percent decrease in 2013.

Next, I turn to the administrative burden RAC auditing had on hospitals. Figure 4 and Table 3 columns 5-6 present results on two dimensions of this burden: hospital administration costs and IT adoption. Figure 4c plots estimates of the effect on log administration costs, as reported in hospital cost reports. A one percentage point increase in RAC auditing in 2011 results in an immediate 1.5 percent uptick in administration costs, but this increase lasts for only about a year.

A potential source of these costs is any investment in technology to track audits or mitigate future ones. According to the AHA RACTrac survey, many hospitals reported installing tracking software in response to RAC audits (Figure D11). One particularly relevant type of technology is medical necessity checking software, which hospitals use to assess medical necessity, as defined by payer rules. Figure 4d presents the event study results for whether a hospital reported that it was installing this software in a given year. In response to a one percentage point increase in 2011 audit rate, hospitals were 2.2 percentage points more likely to report that they were installing or upgrading this software in 2012 (relative to the 59 percent of hospitals who had this software installed in 2010).

In Figure 5, I split admissions by their length of stay, given Medicare's concern over unnecessary short stays (US Department of Health and Human Services Office of Inspector General, 2013; Miller, 2015). The overall reduction in admissions is driven by a reduction in short stays, or admissions with length of stay ≤ 2. A one percentage point increase in audit rate results in a 4.4 percent decrease in short stay admissions and a 4.5 percent decrease in revenue from these stays in 2012 (Table 3). In contrast, there is a much smaller and statistically insignificant decrease in longer stay admissions.

Figure D12 plots the IV event study coefficients on the amount of payments demanded from audited claims. A one percentage point increase in audit rate in 2011 is associated with \$314,115 in demands in 2011 per hospital as well as additional demands in subsequent years, although the magnitude diminishes over time. In Figure D13 I consider whether hospitals substituted away from inpatient care to outpatient care – for example, to observation stays.[16] I find no evidence that, at the hospital level, hospitals increased outpatient spending and observation stays in response to audits of inpatient admissions.

Table E4 pools the post-2011 years of the main results into a single coefficient, as in Equation 4. Averaging across 2011 to 2015, there is a 1.5 percent reduction in overall admissions and a 2.2 percent reduction in short stay admissions relative to the pre-period. Table E5 considers heterogeneity across different hospital characteristics. Rural, for-profit, smaller, and non-chain hospitals are more responsive to audits. The increase in medical necessity checking software is driven by hospitals who do not have the software installed in 2010. Appendix Section B explores the robustness of the results to instrumenting for the share of claims that are *denied* rather than just audited, using varying bandwidths to define the hospital sample, excluding hospitals that are very close to the border, using alternative instruments for audit rate, removing individual states or neighbor comparison groups, and running a placebo test using state borders in the interior of each RAC region.

Discussion The results from the hospital-level analysis show that auditing saved money for Medicare by deterring unnecessary admissions, but the cost of identifying these admissions fell on hospitals. A back-of-the-envelope calculation comparing the total government savings to the compliance costs finds that for every \$1,000 in savings between 2011 and 2015, hospitals spent \$216 in compliance costs. Using the coefficients from Table 3 and Figure D12 for a one percentage point increase in audit rate, I calculate that the present discounted value

[16]Observation stays consist of short-term (often diagnostic) services provided at the hospital while a physician decides whether to admit a patient or send them home. Observation stays typically last less than 48 hours and are billed as an outpatient service, and are often cited as a more cost-effective alternative to a short inpatient stay (Medicare Payment Advisory Commission, 2015). Since observation stays occur in the hospital and can sometimes last more than one day, patients often cannot differentiate between an observation stay and an inpatient stay (Span, 2012).

of total government savings between 2011 and 2015 for the median hospital is $2.08 million in 2021 dollars (including savings from deterred admissions and reclaimed payments minus the contingency fee). The present discounted value of compliance costs associated with a one percentage point increase in audit rate is about $450k. Over 90% government savings from the RAC program are from deterred admissions, rather than reclaimed payments from prior admissions.[17]

The event studies in Figure 4 also illustrate the dynamics of hospitals' responses. Admissions and revenue decline steadily between 2011 and 2012, likely reflecting two factors: first, some of the 2011 admissions occurred before hospitals knew how aggressively they would be audited by RACs; and second, it may have taken time to implement practices or technology to reduce unnecessary admissions. But after 2012, admissions remained at their decreased levels – even in 2014 and 2015, when audit activity decreased significantly. In contrast, there was an immediate but short-lived increase in hospital administration costs in 2011. This timing is in line with hospitals making investments upfront to figure out how to comply with audits going forward. The installation of medical necessity checking software is an example of one such investment.

The results also suggest that prior to 2011, hospitals were not knowingly admitting unnecessary admissions (i.e., committing fraud). If they were, they would not need to install technology in order to stop. One might also expect that hospitals committing fraud would only reduce admissions while RACs are active, and ramp them back up once RAC activity decreases. Contrary to this, I provide evidence that experiencing high initial audit activity in 2011 had persistent effects on admission behavior, even absent high levels of contemporaneous auditing.

4.2 Patient Admission Likelihood and Revisit Likelihood

Figure 6 plots the event studies of the patient-level analysis of ED visits in Equation 6. Immediately after the Two Midnights rule is implemented, there is a drop in the share of after-midnight ED visits that result in an inpatient admission, relative to before-midnight visits. At the same time, there is an increase in the share of patients placed into observation (and never admitted). There is no clear trend in the pre-policy coefficients, which supports making the parallel trends assumption.

Table 4 reports the β coefficient from Equation 7. In columns 1 and 2, the coefficients on the inpatient indicator and observation indicator are symmetric in opposite directions.

[17]These numbers are calculated under the assumption that the hospital settled with CMS to return 68 percent of reclaimed payments. Under the assumption that a hospital does not settle, the total government savings are $2.6 million and deterred admissions account for 72% of the savings.

After the Two Midnights rule goes into effect, after-midnight arrivals are 0.7 percentage points (1.7 percent) less likely to be admitted as inpatient and 0.7 percentage points (14 percent) more likely to be placed in observation. There is no change in the share of patients who are sent home directly from the ED ("Not Admitted"). For ED patients on the margin of 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 negatively affected patient health outcomes. Panel 6d plots the event study results for an indicator of whether a patient revisited a hospital within 30 days of her ED visit, and column 4 in Table 4 reports the pooled coefficient. After-midnight patients were not more likely to revisit a hospital after the Two Midnights rule came into effect, despite their reduced inpatient admission rate. However, a null average effect may mask heterogeneity by patient severity, as only a subset of patients be should affected by auditing. Patients in the middle of the severity distribution should be more likely to be denied admission as a result of RAC audits, so one would also expect any effects on health outcomes to be concentrated among these patients as well.

To explore this, I predict a patient's severity based on information available at the outset of an ED visit. Using data on ED visits between 9AM and 3PM (i.e., outside of the time window 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 available during an ED visit.[18] 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 7, interacting β with an indicator for each decile.

Figure 7 plots the heterogeneity by severity results for inpatient status and for revisits within 30 days, and the coefficients are reported in Table E8. Inpatient admission status is unaffected by RAC audits for patients at the bottom and the top deciles. The reduction in admissions is concentrated in the middle of the severity distribution. There is a 5 percentage point decrease in admissions for patients in the fifth predicted decile, equal to a 25 percent reduction in admission likelihood. However, I do not see this pattern when the outcome is revisits within 30 days – the effect on revisits is statistically insignificant at all risk deciles. Thus, the overall null effect on revisits is not masking heterogeneity by patient severity – health outcomes are unaffected even for patients most likely to be denied admission due to the Two Midnights rule.

Table E6 reports heterogeneity of effect of the Two Midnights rule by hospital charac-

[18] This includes patient demographics like age-bin, sex, race, Hispanic indicator, point of origin indicator, and mean zip code income. It also includes hospital and quarter fixed effects, 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.

teristics – urban, teaching, for-profit, and smaller hospitals are more responsive, as well as hospitals with medical necessity checking software installed. Appendix Section B discusses robustness of the results to the bandwidth used to define before- and after-midnight ED arrivals, robustness to the time period used to measure hospital revisits, and a falsification test on non-Medicare patients, who should not be directly affected by the Two Midnights rule.

Discussion Similar to the hospital-level approach, in the analysis of ED visits at the patient level, I find that hospitals respond to audits by reducing inpatient admissions. Once the patient is already in the ED, it seems that hospitals change how they bill a patient's care (as observation or inpatient), but do not discharge them from the hospital or change the actual amount of care provided.[19] In contrast, in the across-hospital analysis I find a decrease in inpatient admissions *without* a symmetric increase in observation stays or outpatient care. The difference between the two sets of results could be driven by the subset of patients I focus on in the patient-level analysis: ED patients who have *already* arrived at the hospital. The reductions at the hospital level might reflect efforts to reduce admissions *before* patients even arrive at the hospital, like discouraging physician referrals and transfers, influencing ambulance referral patterns, or deciding to not expand ED capacity.

The results also speak to the usefulness of medical necessity checking software. The response to the Two Midnights rule is driven by the 67 percent of hospitals in Florida with this software installed in 2012. This software could be aiding providers in the decision between an inpatient or observation stay by notifying them of relatively obscure billing rules like the Two Midnights rule, which can depend on details irrelevant to a patient's actual medical necessity, like the patient's ED arrival hour.

Overall, I find that in response to auditing, hospitals reoptimized to reduce Medicare spending, and manage to do so without affecting patient health outcomes. This indicates that most of the welfare effect of RAC auditing comes through the savings to the government and the compliance costs incurred by hospitals, rather than through changes in the quality of care provided to patients.

5 Welfare Analysis

I next bring the empirical estimates together in a sufficient statistics framework to calculate the welfare effect of a marginal increase in 2011 audit rate. Because the empirical results are estimated from a major expansion of auditing scope in 2011, these welfare effects should be

[19] After-midnight patients have no additional charges, diagnoses, or procedures, and are not more likely to have an OR procedure (Table E3).

interpreted as the effect of a marginal increase in *2011 audit rate*, rather than as the effects of an increase in *contemporaneous* auditing.

5.1 Framework

Hospitals I assume that hospitals are altruistic in that they care about patient benefit as well as revenue (Chang and Jacobson, 2012). When audits began, RACs could audit prior admissions from the last three years, but hospitals could only change admissions going forward. To capture this distinction, I split admissions into the number of prior admissions n_P and the number of current admissions n. In total, $n_P + n$ admissions are at risk of audit (since RACs could also audit current admissions). Hospitals choose n to maximize their objective function:

Hospital's objective function:

$$\max_n \ \Pi\left(\underbrace{R(a, n_P + n) - k(a, n_P + n) - c(n)}_{\text{net hospital revenue}}, \quad \underbrace{b(n)}_{\substack{\text{patient benefit}\\ \text{or harm}}}\right). \tag{8}$$

The hospital faces audit rate a, which is the share of $n_P + n$ admissions that are audited. Net hospital revenue is comprised of the revenue from n current admissions minus the amount reclaimed from audits $R(a, n_P + n)$, net of the compliance costs $k(a, n_P + n)$ and the treatment cost $c(n)$. Because hospitals are altruistic, they also care about the patient benefit (or harm) from current admissions, $b(n)$.

Social Welfare I assume that the social welfare function is additively separable in its four components:

$$W = \underbrace{\left\{\max_n \Pi(n_P, a)\right\}}_{\text{(1) hospital objective function}} + \underbrace{V\left(G - R(a, n_P + n) - m(a, n_P + n)\right)}_{\text{(2) net government revenue}} + \underbrace{\Gamma\left(b(n)\right)}_{\substack{\text{(3) patient}\\ \text{benefit}\\ \text{or harm}}} - \underbrace{c(n)}_{\substack{\text{(4) treatment}\\ \text{cost}}}, \tag{9}$$

which are (1) the hospital's objective function, (2) the societal value of government revenue net of spending on inpatient stays and monitoring costs $m(a, n_P + n)$, (3) the societal value of the patient benefit from n admissions, and (4) the cost to treat n admissions.

Taking the derivative of the social welfare function with respect to audit rate a and applying the envelope theorem delivers the following first order condition at the optimal audit rate:

Social Welfare FOC:

$$\underbrace{(V' - \Pi')}_{\substack{\text{marginal value of public} \\ \text{funds vs. marginal value} \\ \text{of hospital revenue}}} \times \underbrace{(-R_a)}_{\substack{\text{marginal hospital} \\ \text{revenue}}} = \underbrace{\Pi' k_a}_{\substack{\text{marginal hospital} \\ \text{compliance cost}}} + \underbrace{V' m_a}_{\substack{\text{weighted marginal} \\ \text{gov't cost}}} + \underbrace{\Gamma' \frac{db}{dn}\frac{dn}{da}}_{\substack{\text{value of patient} \\ \text{benefit or harm}}} + \underbrace{\frac{dc}{dn}\frac{dn}{da}}_{\text{treatment cost}} \tag{10}$$

The marginal welfare effect of monitoring depends on the marginal effect on hospital revenue R_a, the marginal hospital compliance cost k_a, the marginal government monitoring cost m_a, the marginal effect on patients $\frac{db}{dn}\frac{dn}{da}$, and the marginal effect on treatment cost $\frac{dc}{dn}\frac{dn}{da}$. Audits facilitate a transfer from hospitals back to the government, and this transfer is only valuable if the marginal value of public funds (MVPF) is greater than the marginal value of hospital revenue ($V' > \Pi'$). I assume that V' and Π' are constants, and normalize Π' to 1 and assume V' to be 1.3 at baseline.[20]

The left-hand side of Equation 10 represents the value of the transfer of revenue from hospitals back to the government, and the right-hand side represents the costs of this transfer. At the optimal audit rate, the first order condition in Equation 10 holds. But if the left-hand side is greater than the right-hand side, then increased auditing is welfare-improving. Vice versa, if the right-hand side is greater than the left-hand side, then increased auditing is welfare-decreasing.

Given the empirical results on the dynamics of hospital responses, the time horizon considered is important. If hospitals incur fixed costs like a large upfront investment in technology, then these costs should be compared to the present discounted value of savings over a multiyear horizon. To remain agnostic about the time horizon for calculating welfare, I calculate the cumulative savings and costs in each year between 2011 and 2018.

[20] 1.3 is a commonly-used MVPF in cost-benefit analyses (Finkelstein and Hendren, 2020). In subsequent analyses I explore how the results would change with varying values of MVPF.

5.2 Welfare Calculation and Results

I use the estimates derived from the IV event study in Figure 4 and Table 3 to inform the revenue effect R_a and the compliance cost effect k_a. To calculate the effect on government monitoring costs m_a, I multiply the reclaimed payments in Figure D12 by RACs' contingency fees. At baseline, I assume a contingency fee of 10.75% (the average of 9 and 12.5%).[21] Section C.1 describes this calculation in further detail.

For the marginal patient benefit $\frac{db}{dn}\frac{dn}{da}$, I assume in the baseline calculation that it is 0. This is motivated by the null result from the analysis on ED visits, which is also in line with other work which finds that the marginal hospitalization has no effect on patient health (Currie and Slusky, 2020). Patient health may not be the only component of patient welfare that is affected by audits – for example, patients could suffer psychological harm if they are denied admission when they believe it is necessary, but they could also be harmed by an unnecessary admission in terms of wasted time spent in the hospital. In Appendix Section C.2, I explore how the marginal welfare effect varies with different assumptions about the effects on patient welfare. At baseline assumptions, increasing monitoring is welfare-improving as long as the harm per patient denied admission is no more than \$190.

For the marginal treatment cost $\frac{dc}{dn}\frac{dn}{da}$, I assume that the cost incurred to treat the patient does not change. This is likely a lower bound on the treatment cost savings of increased monitoring, and assumes that hospitals substituted admissions with other forms of care that have the same cost. But if hospitals incurred lower treatment costs as a result of reducing admissions, then the savings from monitoring would be even larger. I relax this assumption with further calculations in Appendix Section C.2.

Figure 8 plots the cumulative difference between the marginal savings and marginal costs from a one-percentage point increase in 2011 audit rate – in other words, the difference between the left-hand side and right-hand sides of Equation 10. Increased auditing is welfare-improving if this value is positive, and welfare-reducing if this value is negative. Figure 8 plots this value in three cases that decompose the overall welfare effect: (1) audits deter admissions and increase compliance costs (baseline calculation); (2) there is no effect on compliance costs; and (3) there is no deterrence effect on admissions.

Increasing the 2011 audit rate is welfare-improving five years after 2011. The estimates imply that a one percentage point increase in 2011 audit rate results in a marginal welfare improvement of \$57,000 by 2015; across all 2,901 hospitals eligible for RAC audit, this is equivalent to a welfare improvement of \$165 million. Case (2) shows that absent compliance costs, a higher audit rate is always welfare-improving. Comparing cases (1) and (2), the

[21] Medicare does not report each individual RAC's contingency fee, just that the fees range from 9-12.5%.

marginal welfare effect by 2015 would be almost *9 times larger* ($512k per hospital) if hospitals did not face any compliance costs. The gap between the welfare effects in cases (1) and (2) diminishes over time as more savings accrue through deterred admissions. Comparing case (1) to case (3), we see that the key to the positive welfare effect is the deterrence of current and future admissions. If audits simply collect money back from prior admissions, a higher audit rate is always welfare-reducing, since the reclaimed payments would not cover the compliance costs.

In Appendix Section C.2, I explore additional calculations under varying assumptions about government monitoring costs, treatment costs, patient health, and the marginal value of public funds. I also calculate the marginal cost of funds (MCF), in the spirit of Slemrod and Yitzhaki (2001) and Hendren and Sprung-Keyser (2020). The MCF in the baseline calculation is 1.27 in 2015, which means that a policy which pairs RAC audits with expenditure with an MVPF over 1.27 would be welfare-improving. The MVPF of Medicare is estimated to be 1.63, so one welfare-improving policy would be to redirect the money saved from RAC monitoring back into Medicare (Finkelstein and McKnight, 2008; Hendren and Sprung-Keyser, 2020).

6 Conclusion

Governments often monitor the third parties they contract with to ensure the cost-effectiveness of public expenditure. The welfare effect of increased monitoring depends on the money it saves, the costs to conduct or comply with monitoring, as well as any changes in service quality it induces. I study these outcomes in the context of monitoring for unnecessary hospital admissions by Medicare. Monitoring causes hospitals to reduce admissions, particularly the ones most likely to be unnecessary. These reductions translate into savings for Medicare, in addition to the payments directly reclaimed from audits. At the patient level, hospitals are less likely to admit patients who, if admitted, have a greater probability of being audited. But despite being denied admission, these patients were not more likely to revisit the hospital at a later date, suggesting that their health outcomes did not worsen as a result.

While I do not find evidence that the reduction in admissions harmed patients, monitoring did come at a substantial private compliance cost to *hospitals*. In response to increased monitoring, hospitals increased their administration costs as they invested in technology to detect unnecessary care. The savings from monitoring accrued over several years, mostly driven by sustained reductions in unnecessary admissions. Given the high upfront compliance costs for providers and the fact that the savings to Medicare accrued over time, monitoring through the RAC program is welfare-improving only after five years.

The findings in this paper highlight an important unintended consequence of policy-making: that well-intentioned policies can be costly to implement or comply with. Reducing unnecessary government expenditure is not sufficient for a policy to be welfare-improving – it must also not be costly to implement or comply with. My findings suggest that there may be scope for policymakers to look for ways to reduce the administrative burden of complying with cost-saving measures like monitoring. This is especially pertinent within the healthcare context, where government programs are the largest payer and the administrative burden on providers is already relatively high (Cutler and Ly, 2011; Himmelstein et al., 2014; Papanicolas et al., 2018). I document an instance where the third parties contracting with the government – in this case hospitals – incurred private costs to save money *on behalf* of the government. The welfare gain from monitoring is much smaller once we take these compliance costs into account. Overall, the findings point to the importance of considering the tradeoff between *all* sources of benefits and costs, both public and private, in evaluating the welfare effects of policy.

References

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

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

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, 1st Quarter 2014," May 2014.

Boards of Trustees for Medicare, "2021 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds," Technical Report August 2021.

Brot-Goldberg, Zarek, Samantha Burn, Timothy Layton, and Boris Vabson, "Rationing Medicine Through Paperwork: Authorization Restrictions in Medicare," 2021.

Centers for Medicare and Medicaid Services, "CMS Glossary," May 2006.

_ , "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.

_ , "Guidance on Hospital Inpatient Admission Decisions," MLN Matters SE1037 July 2012.

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

_ , "Hospital Appeals Settlement Process," 2014.

_ , "Medicare Claim Review Programs," Technical Report September 2016.

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

Chang, Tom and Mireille Jacobson, "What do Nonprofit Hospitals Maximize? Evidence from California's Seismic Retrofit Mandate," in "in" August 2012, p. 61.

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.

Currie, Janet, "The Take-up of Social Benefits," in "Poverty, the Distribution of Income, and Public Policy, (New York: Russell Sage)" 2006, pp. 80–148.

_ **and David Slusky**, "Does the Marginal Hospitalization Save Lives? The Case of Respiratory Admissions for the Elderly," Technical Report w26618, National Bureau of Economic Research January 2020.

Cutler, David, "The Incidence of Adverse Medical Outcomes under Prospective Payment," *Econometrica*, 1995, *63* (1), 29–50. Publisher: Econometric Society.

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.

Department of Health and Human Services, "FY2022 CMS Congressional Justification of Estimates for Appropriations Committees," 2021.

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.

Dranove, David and Mark A. Satterthwaite, "Chapter 20 The industrial organization of health care markets," in "Handbook of Health Economics," Vol. 1, Elsevier, January 2000, pp. 1093–1139.

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. Publisher: The MIT Press.

Dunn, Abe, Joshua Gottlieb, Adam Shapiro, and Daniel Sonnenstuhl, "A Denial A Day Keeps the Doctor Away," in "in" April 2021.

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

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.

Empire Blue Cross Blue Shield, "Provider Manual," Technical Report July 2020.

Finkelstein, Amy and Nathaniel Hendren, "Welfare Analysis Meets Causal Inference," *Journal of Economic Perspectives*, November 2020, *34* (4), 146–167.

— **and Robin McKnight**, "What did Medicare do? The initial impact of Medicare on mortality and out of pocket medical spending," *Journal of Public Economics*, 2008, *92* (7), 1644–1668. Publisher: Elsevier.

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

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

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

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

Gross, Tal, Adam Sacarny, Maggie Shi, and David Silver, "Regulated Revenues and Firm Behavior: Evidence from a Medicare Overhaul," in "in" April 2021.

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

Hendren, Nathaniel and Ben Sprung-Keyser, "A Unified Welfare Analysis of Government Policies*," *The Quarterly Journal of Economics*, August 2020, *135* (3), 1209–1318.

Himmelstein, David U., Miraya Jun, Reinhard Busse, Karine Chevreul, Alexander Geissler, Patrick Jeurissen, Sarah Thomson, Marie-Amelie Vinet, and Steffie Woolhandler, "A Comparison Of Hospital Administrative Costs In Eight Nations: US Costs Exceed All Others By Far," *Health Affairs*, September 2014, *33* (9), 1586–1594. Publisher: Health Affairs.

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

Humana, "Provider Manual (2020)," Technical Report GHHKSLSEN IMO No. 3440 2020.

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

Keen, Michael and Joel Slemrod, "Optimal tax administration," *Journal of Public Economics*, August 2017, *152*, 133–142.

Kleven, Henrik Jacobsen and Claus Kreiner, "The marginal cost of public funds: Hours of work versus labor force participation," *Journal of Public Economics*, 2006, *90* (10-11), 1955–1973. Publisher: Elsevier.

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

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. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMsa1702321.

Laffont, Jean-Jacques and Jean Tirole, "Cost Padding, Auditing and Collusion," *Annales d'Économie et de Statistique*, 1992, (25/26), 205–226. Publisher: [GENES, ADRES].

Leder-Luis, Jetson, "Can Whistleblowers Root Out Public Expenditure Fraud? Evidence from Medicare," July 2020, p. 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.

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.

Nalebuff, Barry and David Scharfstein, "Testing in Models of Asymmetric Information," *The Review of Economic Studies*, April 1987, *54* (2), 265–277.

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.

Nichols, Albert L. and Richard J. Zeckhauser, "Targeting Transfers through Restrictions on Recipients," *The American Economic Review*, 1982, *72* (2), 372–377. Publisher: American Economic Association.

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 November 2019.

Papanicolas, Irene, Liana R. Woskie, and Ashish K. Jha, "Health Care Spending in the United States and Other High-Income Countries," *JAMA*, March 2018, *319* (10), 1024–1039.

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," in "in" 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. Publisher: The University of Chicago Press.

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

Sheehy, Ann M., Charles Locke, Jeannine Z. Engel, Daniel J. Weissburg, Stephanie Mackowiak, Bartho Caponi, Sreedevi Gangireddy, and Amy Deutschendorf, "Recovery Audit Contractor Audits and Appeals at Three Academic Medical Centers," *Journal of Hospital Medicine*, April 2015, *10* (4), 212–219.

_ , Jeannine Z. Engel, Charles F. S. Locke, Daniel J. Weissburg, Kevin Eldridge, Bartho Caponi, and Amy Deutschendorf, "Hospitalizations With Observation Services and the Medicare Part A Complex Appeals Process at Three Academic Medical Centers," *Journal of Hospital Medicine*, April 2017, *12* (4), 251–255. WOS:000399578000007.

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, "Integrating Expenditure and Tax Decisions: The Marginal Cost of Funds and the Marginal Benefit of Projects," *National Tax Journal*, 2001, *54* (2), 189–202. Publisher: National Tax Association.

Span, Paula, "In the Hospital, but Not Really a Patient," June 2012. Cad: 0.

UnitedHealthcare, "2020 UnitedHealthcare Care Provider Administrative Guide," Technical Report 2020.

U.S. Department of Defense, "Defense Spending by State Fiscal year 2019," Technical Report January 2021.

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.

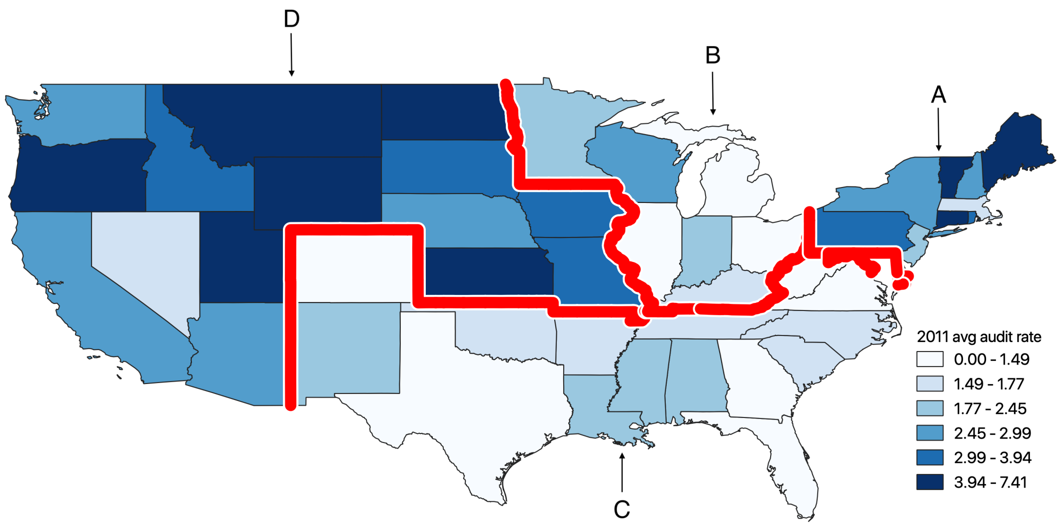
U.S. Government Accountability Office, "Federal Government Contracting for Fiscal Year 2018," May 2019.

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

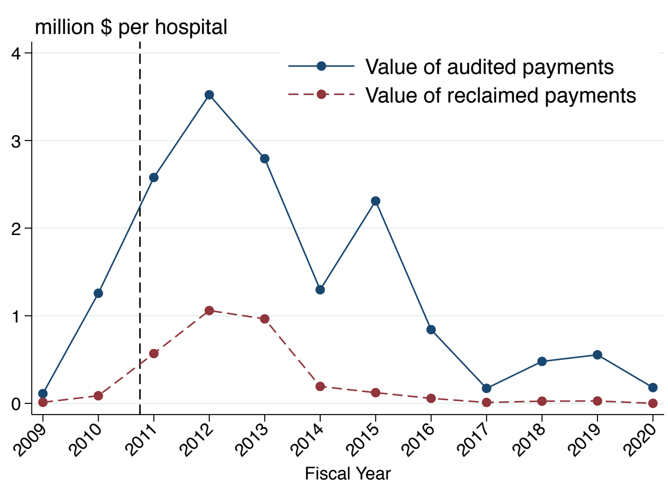
7 Figures

Figure 1. RAC Audit Activity

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

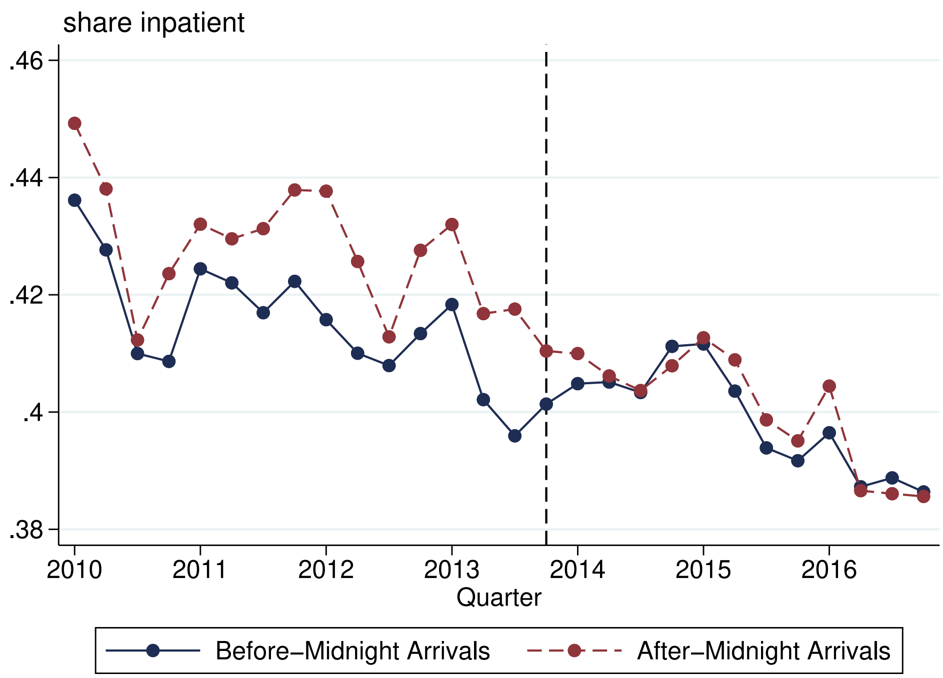


(b) Value of Audited Inpatient Payments and Net Reclaimed Payments per Hospital, by Year of Audit



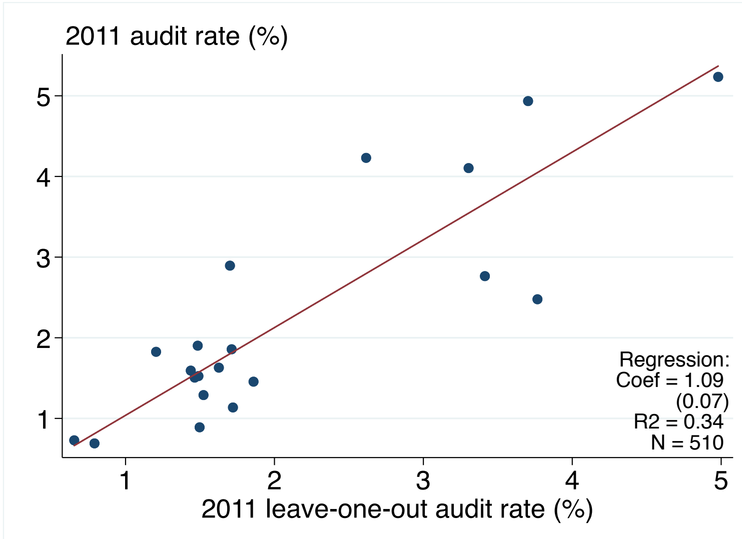
Panel (a) plots the 2011 average state audit rates, where audit rate is defined as the share of a hospital's 2008-2011 claims that were audited by RACs. The RAC regions are: Region A (Northeast), Region B (Midwest), Region C (South), and Region D (West). Darker shades denote higher audit rate. The red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

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



This figure plots the share of traditional Medicare patients admitted as inpatient from the emergency department, among Florida patients who arrived within 3 hours before midnight (9-11:59PM), in the blue solid line, and 3 hours after midnight (12:00-2:59AM), in the red dashed line. The dashed vertical line denotes 2013Q3, which is when the Two Midnights rule is implemented. Data: HCUP SID/SEDD.

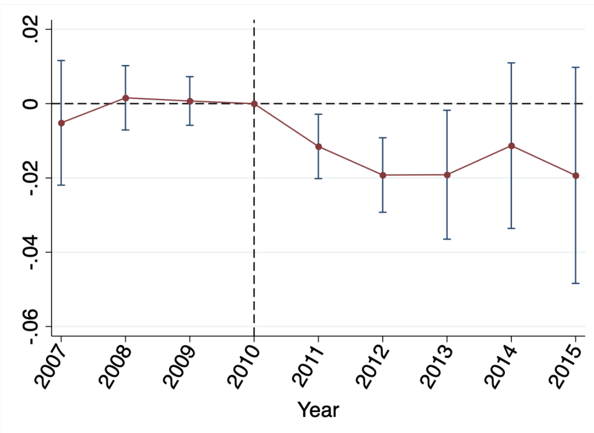
Figure 3. Binscatter of 2011 Leave-one-out State Audit Rate and 2011 Hospital Audit Rate, Border Hospital Sample



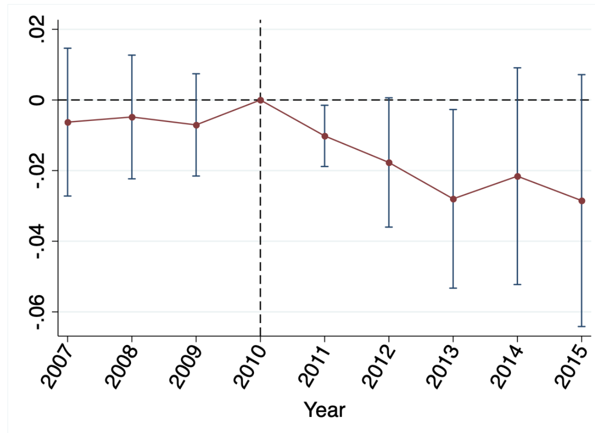
This figure plots a binscatter of 2011 hospital audit rate compared to the instrument, 2011 leave-one-out state audit rate. 2011 audit rate is defined as the share of 2008-2011 inpatient claims that were audited by RACs in 2011. Leave-one-out state audit rate is defined as the average audit rate of all other hospitals in the same state as a given hospital. The sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR claims and CMS audit data.

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

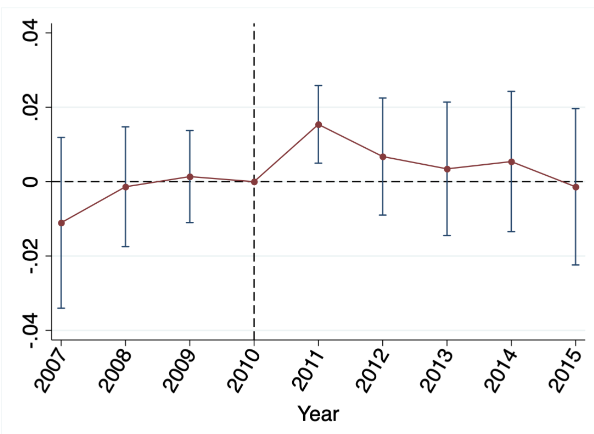
(a) Log Medicare admissions



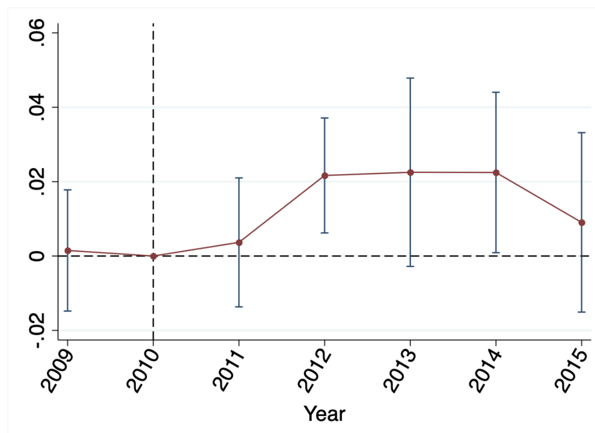
(b) Log Medicare inpatient revenue



(c) Log hospital administrative costs



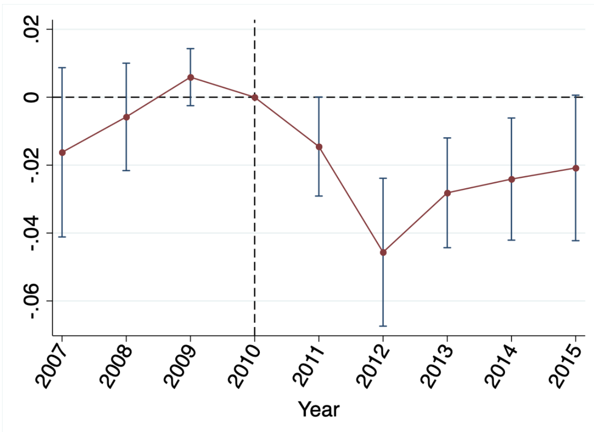
(d) Indicator for installing medical necessity software



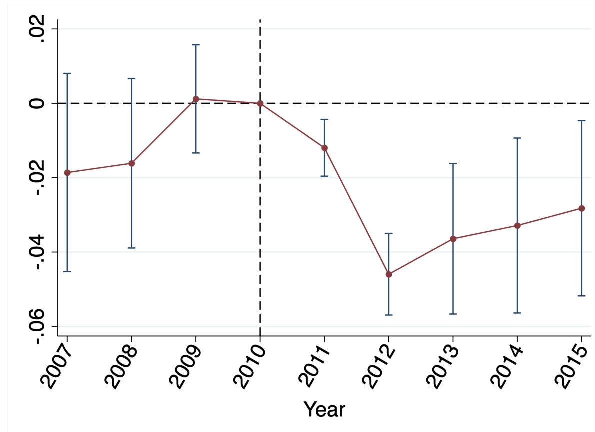
This figure plots event studies of the IV coefficients and 95% confidence intervals of the specification in Equation 1. The omitted year is 2010. Each coefficient is an estimate of the effect of a 1pp increase in 2011 audit rate on a hospital-level outcome. Medicare admissions and revenue are from MEDPAR. Inpatient revenue is the sum of all Medicare inpatient payments. Net 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. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their neighbor comparison group.

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

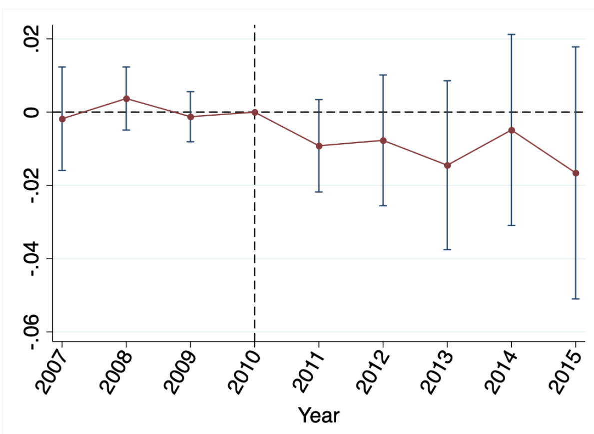
(a) Log Medicare admissions, LOS ≤ 2



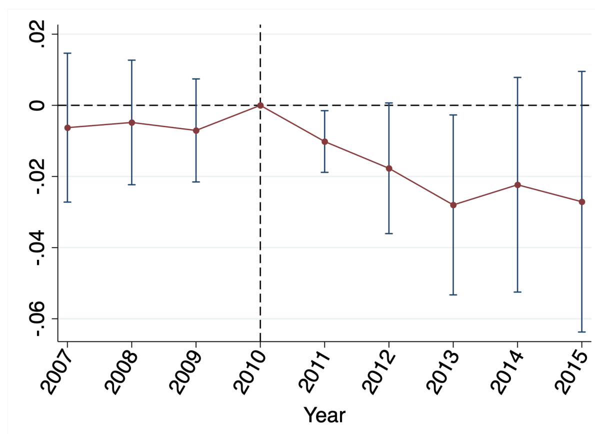
(b) Log Medicare inpatient revenue, LOS ≤ 2



(c) Log Medicare admissions, LOS > 2

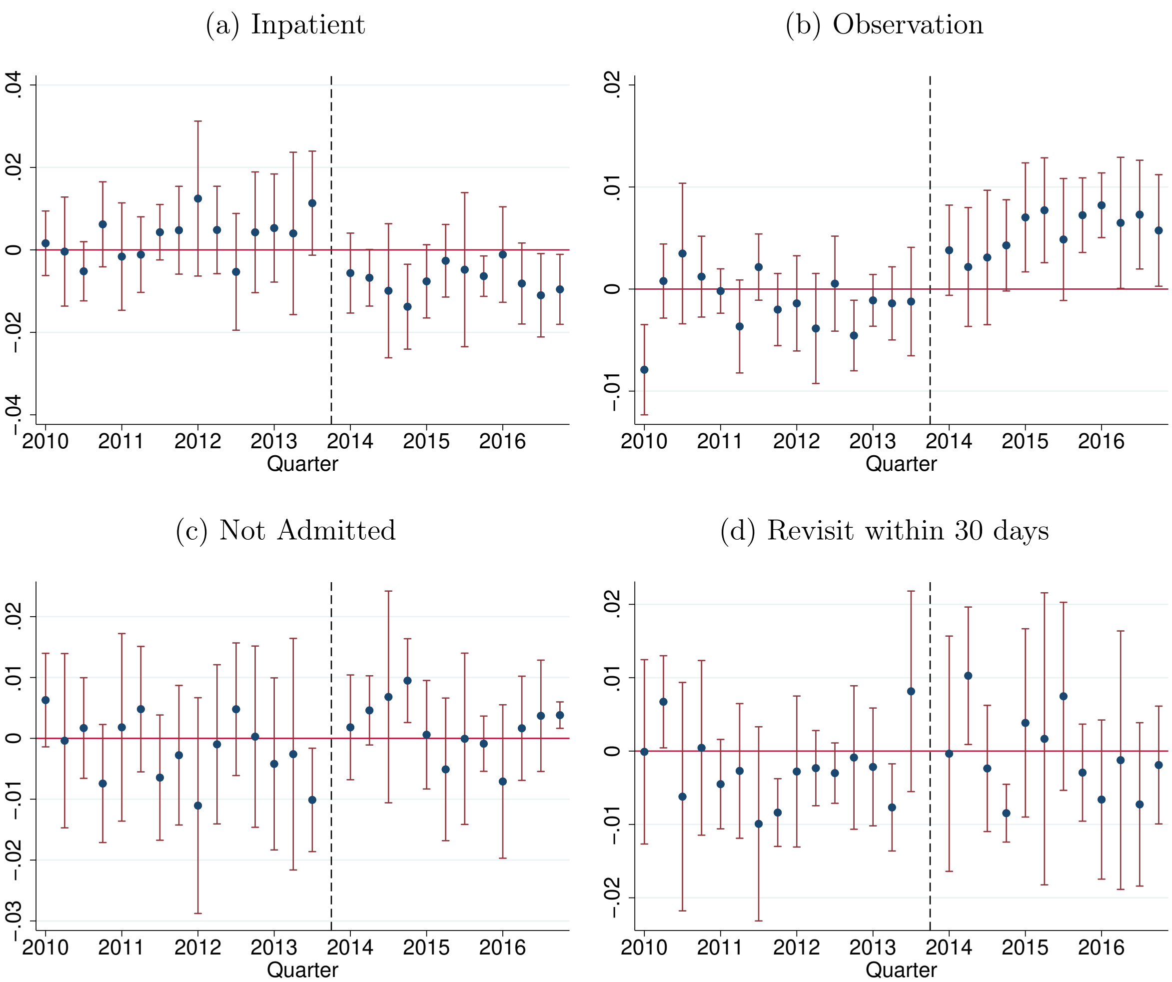


(d) Log Medicare inpatient revenue, LOS > 2



This figure plots event studies of the IV coefficients and 95% confidence intervals of the specification in Equation 1. The omitted year is 2010. Each coefficient is an estimate of the effect of a 1pp increase in 2011 audit rate on the volume and revenue of short stay admissions and longer admissions, from MEDPAR. Length of stay is counted as the difference in days between the admission and discharge date. Inpatient revenue is the sum of all Medicare inpatient payments. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their neighbor comparison group.

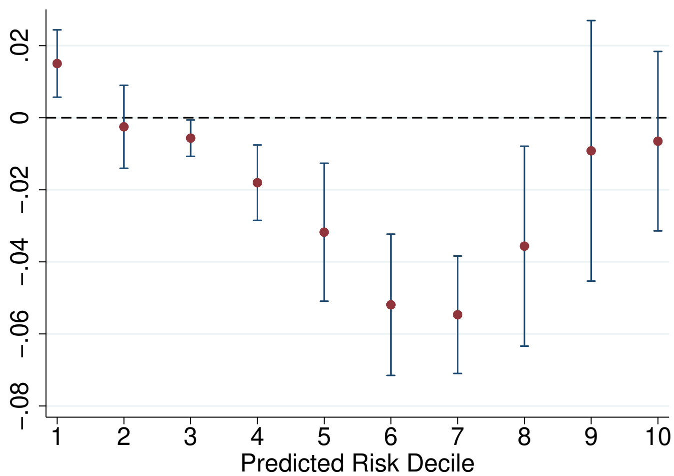
Figure 6. Event Studies on Effect of After-Midnight ED Arrival on Patient Status and Outcomes



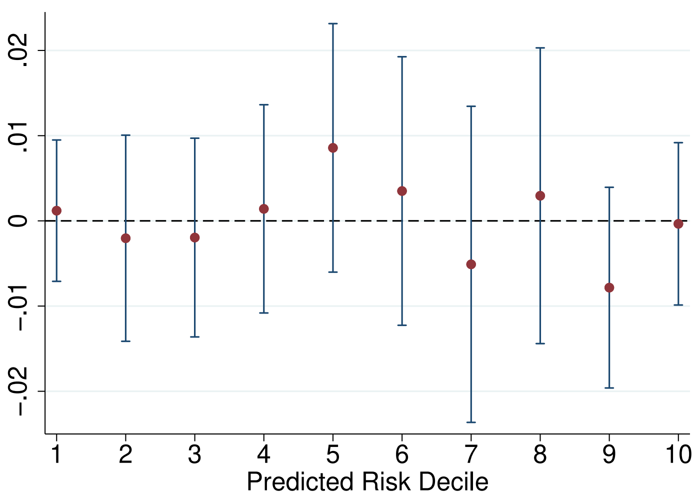
This figure plots the coefficients and 95% confidence intervals for β^{τ} on $1[q = \tau] \times 1[T_v \geq 00{:}00]$ of the specification in Equation 7, where $1[q = \tau]$ is an indicator for whether the visit occurred in quarter τ, and $1[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. The results are clustered at the ED arrival hour and quarter level. The omitted quarter is 2013Q3. “Inpatient” is an indicator for whether the patient was eventually admitted as inpatient from the ED. “Observation” is an indicator for whether the patient was placed in observation status and was never admitted. “Not Admitted” is an indicator equal to one when a patient is neither admitted nor placed in observation status. “Revisit within 30 days” is an indicator for whether the patient had another ED visit or inpatient stay within 30 days of the ED visit. Sample consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Data: HCUP SID/SEDD.

Figure 7. Heterogeneity of After-Midnight ED Arrival Coefficient by Patient Severity

(a) Inpatient

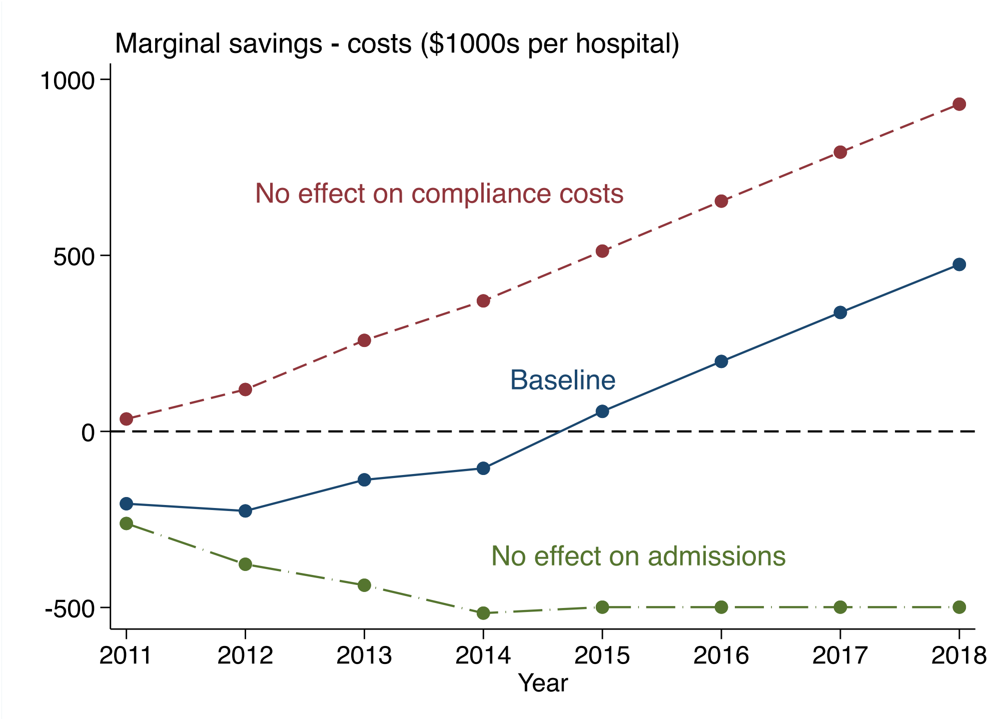


(b) Revisit in 30 days



This figure plots estimates and 95% confidence intervals of the β coefficient in Equation 7, interacted with an indicator for predicted severity decile. β is the coefficient on $1[q \geq 2013Q3] \times 1[T_v \geq 00{:}00]$, where $1[q \geq 2013Q3]$ is an indicator for whether the visit occurred after 2013Q3, and $1[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. The top panel plots results for an indicator for whether the patient was admitted as inpatient from the ED, and the bottom panel plots results for an indicator for whether the patient revisited any hospital in Florida within 30 days of the ED visit. The results are clustered at the ED arrival hour and quarter level. Patient risk is predicted by estimating a logit using ED visits between 9AM and 3PM of an indicator for being admitted within 30 days of an ED visit on patient demographics, current ED visit information, and information on any prior visits in the last 365 days. Demographics include age-bin, sex, race, Hispanic indicator, point of origin indicator, and mean zip code income. Information on current visit includes hospital and quarter. Information on previous visits includes the number of visits/inpatient stays/length of stay in the last month or last year, as well as any diagnoses and procedures recorded in stays within the last month or last year. Figure D19 plots the mean outcomes for each decile. Data: HCUP SID/SEDD.

Figure 8. Welfare Analysis Estimates



This figure plots the per-hospital welfare effect by year, or the difference between the savings and costs of auditing in Equation 10, of increasing the 2011 audit rate by one percentage point. Increasing audits is welfare-improving if this value is positive and welfare-reducing if this value is negative. This figure plots this value in three cases: (1) audits deter admissions and increase compliance costs (baseline); (2) audits deter admissions but have no compliance costs; and (3) audits do not deter admissions and have compliance costs. Table 5 lists the parameters and estimates used to calculate the welfare effects for each case.

8 Tables

Table 1. Hospital Summary Statistics by RAC Region

	(1)	(2)	(3)	(4)
	RAC Region			
	A	*B*	*C*	*D*
	A. Hospital Characteristics			
2011 audit rate	3.01	1.79	1.36	3.33
	(2.29)	(1.21)	(1.18)	(2.73)
Beds	239.63	200.76	196.86	195.93
	(197.77)	(174.67)	(190.88)	(147.14)
Share urban	0.83	0.70	0.64	0.82
Share non-profit	0.88	0.79	0.46	0.63
Share for-profit	0.05	0.09	0.29	0.19
Share government	0.07	0.12	0.24	0.18
Share non-chain	0.47	0.37	0.33	0.32
Total costs (million $)	276.36	212.53	156.68	222.61
	(345.14)	(274.69)	(210.96)	(229.22)
Net admin costs (million $)	35.50	33.66	23.39	35.55
	(42.63)	(44.20)	(34.65)	(38.03)
	B. Medicare Inpatient Admissions Characteristics			
Admissions	4291.56	3806.69	3239.88	2905.55
	(3666.92)	(3378.66)	(3297.37)	(2386.25)
Mean payment ($)	9413.40	8354.75	7784.83	10732.17
	(3442.28)	(2373.00)	(2558.62)	(3530.16)
Total payments (million $)	45.95	35.83	29.17	33.14
	(54.65)	(40.20)	(35.89)	(33.06)
Average short stay share	0.27	0.32	0.30	0.31
	(0.07)	(0.07)	(0.08)	(0.08)
Observations	489	571	1237	663
Obs w/in 100 miles of RAC border	164	224	282	118

This table presents 2011 summary statistics of hospital characteristics and Medicare inpatient admissions by RAC region. Standard deviation is in parentheses. Bed size, urban status, and profit type status come from the Medicare Provider of Services file. Non-chain status comes from hospital merge data via Cooper et al. (2019). Total and administration 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.

Table 2. Patient Summary Statistics by ED Arrival Hour

	(1)	(2)
	ED Arrival Hour	
	Before MN	*After MN*
Share inpatient	0.40	0.42
	(0.49)	(0.49)
Share observation	0.05	0.05
	(0.21)	(0.22)
Average charges ($)	23966.55	25881.27
	(43649.05)	(50655.54)
Average age	68.04	68.22
	(17.33)	(17.28)
Share white	0.78	0.77
	(0.41)	(0.42)
Share hispanic	0.12	0.11
	(0.32)	(0.31)
Share female	0.57	0.54
	(0.50)	(0.50)
Average n of chronic conditions	3.95	4.17
	(3.57)	(3.64)
Share inpatient in last 30 days	0.13	0.14
	(0.33)	(0.34)
Share hospital visit in last 30 days	0.28	0.30
	(0.45)	(0.46)
Share hospital visit in next 30 days	0.27	0.29
	(0.45)	(0.45)
Share hospital visit in next 60 days	0.38	0.39
	(0.48)	(0.49)
Share hospital visit in next 90 days	0.44	0.45
	(0.50)	(0.50)
Observations	32793	18467

This table presents summary statistics of characteristics of traditional Medicare patients in Florida who arrived in the ED within 3 hours of midnight in 2013Q2. Standard deviation is in parentheses. "Share inpatient" is the share of ED patients admitted to inpatient (this includes patients who could have initially been placed in observation and eventually admitted). "Share observation" is the share of patients who are placed in outpatient observation only. Data: HCUP SID/SEDD.

Table 3. Event Studies of Effect of 2011 Audit Rate on Hospital Outcomes, 2011-2015 Coefficients

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

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the state and border segment level. This table reports the IV event study coefficients for 2011-2015 of the specification in Equation 1, or the effect of a 1pp increase in 2011 audit rate on a hospital-level outcome in a given year, relative to 2010. For brevity, the pre-2011 coefficients are estimated but not reported in the table. Omitted year is 2010. Columns 1 and 2 report the effect on the log number of Medicare inpatient admissions and log Medicare inpatient revenue from the MEDPAR data, and columns 3 and 4 report the effect on short stay admissions and revenue. Column 5 reports the effect on log net administration costs from HCRIS data. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Column 6 reports the effect on an indicator for installing medical necessity software application, which is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in the HIMSS data. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their "neighbor comparison group."

Table 4. After-Midnight ED Arrival Hour Difference-in-Difference Coefficients on Patient Status and Revisits

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

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient on $1[q \geq 2013Q3] \times 1[T_v \geq 00{:}00]$ of the specification in Equation 7, where $1[q \geq 2013Q3]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $1[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "Inpatient" is an indicator for whether the patient was eventually admitted as inpatient from the ED. "Observation" is an indicator for whether the patient was placed in observation status and was never admitted. "Not Admitted" is an indicator equal to one when a patient is neither admitted nor placed in observation status. "Revisit within 30 days" is an indicator for whether the patient had another ED visit or inpatient stay within 30 days of the ED visit. Sample for columns 1-4 consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. The sample for column 5 consists of all non-Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Data: HCUP SID/SIDD.

Table 5. Welfare Analysis Parameters

	(1)	(2)	(3)
	Model Assumptions		
	Baseline	*No Compliance Costs*	*No Deterrence*
	A. Estimates		
Effect on admissions	2011-2015: estimates after 2015: 2015 estimate	2011-2015: estimates after 2015: 2015 estimate	all years: 0
Effect on compliance costs	2011-2015: estimates after 2015: 0	all years: 0	all years: 0
Payments demanded	2011-2015: estimates after 2015: 0	2011-2015: estimates after 2015: 0	2011-2015: estimates after 2015: 0
Avg 2010 inpatient revenue	$15,029,306	$15,029,306	$15,029,306
Avg 2010 compliance cost	$12,822,887	$12,822,887	$12,822,887
	B. Parameters		
RAC contingency fee	10.75%	10.75%	10.75%
Marginal value of public funds	1.3	1.3	1.3
Discount rate	2%	2%	2%
Share of demanded pmts refunded	68%	68%	68%

This table lists the parameters and assumptions for the three welfare calculations depicted in Figure 8 and described in Section C.1: (1) audits deter admissions and increase compliance costs (baseline); (2) audits deter admissions but have no compliance costs; and (3) audits do not deter admissions and have compliance costs.. Effect on admissions and compliance costs are from Table 3. Payments demanded are from Figure D12. The 2010 hospital revenue and hospital compliance costs are the median values for hospitals in the border hospital sample.

A Additional Policy Context

A.1 Medicare Inpatient Prospective Payment System and Short Stays

Medicare pays for inpatient hospital admissions through the inpatient prospective payment system (IPPS), in which Medicare pays a fixed amount per inpatient stay within broad categories of diagnoses called Medicare Severity Diagnoses 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 incentive to keep costs low. The payment rate for each DRG reflects the national average cost of treating a patient across all cases, and is revised each year based on claims data in the last two years. The per-stay payment is adjusted based on a patient's preexisting chronic conditions in order to account for the patient's severity. It is also adjusted by hospital-specific factors like wage index, teaching status, share of low-income patients, and number of unusually costly outlier cases. The prospective payment system generally works well to keep inpatient hospital spending relatively low for the Medicare program (Lopez et al., 2020).

One persistent issue with IPPS that was noted by policymakers, however, is the high number of short, unnecessary stays. Medicare states 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 would be 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 48 hours and are billed as an outpatient service. The use of observation stays among Medicare beneficiaries has grown over time (Medicare Payment Advisory Commission, 2015). An outpatient observation stay often precludes Medicare coverage for postacute care services at a skilled nursing facility (SNF), because Medicare requires at least a three-day inpatient stay in order to cover a SNF stay.

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 2-3 times higher than payments for observation stays (Medicare Payment Advisory Commission, 2015).

Policymakers considered various alternative policy solutions to address unnecessary inpatient stays. Medicare was wary of setting overly stringent admission requirements – Medicare's admission guidelines give a lot of deference to physicians in the admission decision. Medicare recognized the admission decision as a complex one, noting that providers must take into account many factors, including the "medical predictability of something adverse happening to the patient, the severity of the patient's condition, the need for and availability of diagnostics, the types of facilities available, hospital by-laws and admissions policies, and the relative appropriateness of treatment in each setting" (Centers for Medicare and Medicaid Services, 2012a). Medicare was also wary of reducing the payment rate for short stays or penalizing high rates of short stays, as policymakers were concerned hospitals would simply keep patients for longer to evade the policy (Medicare Payment Advisory Commission, 2015). There is evidence that hospitals delay discharging patients if there is incentive to do so (Jin et al., 2018). Finally, short stays (0-2 day stays) comprise almost a third of inpatient stays. the prevalence of short stays suggests that not *all* short stays are unnecessary, and cutting payments for short stays across the board would reduce payments for some *necessary* stays.

A.2 RAC Program Details

A.2.1 RAC Regions

In the context of medical claims processing and reviews, RAC regions are unique. Medicare Administrative Contractors (MACs) are contractors who process medical claims for Medicare; MACs operate in smaller regions that fall within RAC regions. The RAC regions do align with the regions of Durable Medical Equipment (DME) MACs, who process payments for durable medical equipment, prosthetics, orthotics, and supplies (Medicare Contractor Management Group, 2017). Medicare posts a separate contract solicitation for each region, and firms submit separate bids.

A.2.2 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 have a focus on healthcare (e.g., Health Data Insight, Cotiviti), while others serve other government agencies and corporations as well (e.g., 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.

A.2.3 RAC Audit Process

RACs conducted postpayment reviews to identify and correct overpayments or underpayments for inpatient, outpatient, long-term care, and durable medical equipment claims in the last three years. Figure D1 illustrates the claims auditing and appeal process, using 2011 inpatient audits as an example. Each RAC develops and runs its own proprietary algorithm on claims data to identify claims with potential payment errors. In 2011, RACs' auditing scope for inpatient claims included incorrect/incomplete coding, DRG validation, and medical necessity reviews, where the latter was added in 2011. Five percent of audits were "automated reviews," which rely solely on claims data to make a determination based on clearly outlined Medicare policies. 95 percent of audits were "complex reviews" in which a medical professional (e.g., 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's judgment to determine whether an overpayment or underpayment was made. Once the complex review is finished, RACs send a demand letter to providers which outlines whether a payment error was identified, the amount of overpayment or underpayment demanded, and references supporting the decision. 57 percent of complex reviews in 2011 result in no finding, 37 percent result in an overpayment demand (in which providers must return payment back to Medicare), and six percent result in an underpayment demand (in which Medicare returns payment to the provider). Hospitals can appeal demands by first requesting a redetermination by the RAC, and then escalating it to higher levels of appeals like requesting a reconsideration by a separate contractor or taking the appeal to court.

A.2.4 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 2005-2008 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, 2012b). As shown in Figure 1b, RAC audit activity peaked in 2011–2013, but 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[22]. 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, eleven 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 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 like MACs picked up auditing after the RACs were paused.[23] 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 resumed to its peak level after the pause in 2014. The pause began at the end of 2014Q1 and was originally meant to end in 2014Q3. After several quarters of delayed resumption, inpatient RAC audits finally resumed in 2015Q4, although they were subject to limitations to reduce provider burden. In August 2014, Medicare also announced a one-time option to settle appeals by offering hospitals 68 percent of the 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, in Figure D10 plots audit rates as a function of an admission's length of stay. Admissions with a length of stay of two or fewer days have much higher audit rates than longer admissions. Admissions with a length of stay less than 2 have an average audit rate of 4.2 percent, while admissions with a length of stay over 2 have an average audit rate of 0.7 percent. The majority of audits of short stays result in the full payment being reclaimed

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

[23]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 steps.

(Figure D4).

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 (Figure D3).

Although almost every hospital is 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. The share of hospitals with no audits varies across RAC regions from 2 to 23 percent (Figure D7).

B Robustness and Placebo Tests

Hospital-Level Analysis I run a specification that instruments for the denial rate, or the share of all admissions at a hospital that are audited and a payment is demanded from. I also consider heterogeneity by how "successful" audits are (from the RAC's point of view) by comparing hospitals with above- and below-average demand rates, or the share of audits that result in a payment demand. As a robustness test, in Figure D14 I instrument for a hospital's denial rate, or the share of claims that where a denial is made after audit, rather than its audit rate. Equation 11 defines the relationship between denial rate, audit rate, and demand rate.

$$Denial\ Rate_{ht} = \underbrace{P(Audit)}_{Audit\ Rate}{}_{ht} \times \underbrace{P(Demand|Audit)}_{Demand\ Rate}{}_{ht} \tag{11}$$

Since 41 percent of audits in 2011 resulted in a demand in the main sample, one would expect that the hospital response to a one-percentage point increase in denial rate should be about twice the response to one percentage point increase in audit rate. Indeed, this is what the results reflect – for example, hospitals reduced admissions by 2.5 percent in 2012 in response to a 1pp increase in 2011 audit rate, and they reduced admissions by 5.7 percent in 2012 in response to a 1pp increase in denial rate. The denial rate results track with the audit rate results, and combined with the heterogeneity by demand rate results, demonstrate that hospitals are mostly responding to the share of claims that are audited, rather than how successful the audits are. Hospitals with relatively low demand rates still reduce admissions in response to audits, even though a lower share of their audited claims are denied. They may still reduce admissions because simply going through the auditing process is costly – even if a claim is not eventually denied. So as hospitals learn about what RACs are targeting, they reduce admissions to deter future audits, not just future denials. Hospitals may want to avoid audits if the audit process itself is costly, which is consistent with other work that has shown that the “back-and-forth” interactions between providers and payers in the claim denial process is costly for providers, even when it doesn’t result in a denial (Dunn et al., 2021).

In Figure D15, I show that the results are robust to alternative sample definitions. Figure D15a reproduces the event study from the main specification for the outcome log Medicare admissions, where the sample is defined as all hospitals within 100 miles of the RAC border and the instrument for a hospital’s audit rate is its leave-one-out state audit rate. This is robust to changing the sample to be all hospitals within 50 miles (Figure D15b) or 150 miles (Figure D15c) of the border, although the results are noisier with a shorter distance. One concern with boundary discontinuity identification strategies is the potential for spillovers among hospitals very close to the border. For example, 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 upwards. Figure D15d shows similar results when restricting to hospitals that are at least 10 miles away from the border, demonstrating that the result is not driven by such spillovers. Finally, Figure D15e shows that the results are similar when restricting to hospitals with audit rates greater than 0, meaning that the results are driven by variation in auditing across hospitals on the intensive, rather than the extensive, margin.

Figure D16 shows that the results are robust to using alternative instruments for audit rate. 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 D16a plots the results of using the state audit rate (which includes the hospital) as an instrument. Figure D16c shows that the results using the leave-one-out *RAC region* audit rate, rather than state audit rate, are similar. While using the leave-one-out audit rate strips away the direct effects of a hospital’s own behavior, the leave-one-out audit rate 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 in its region. To address this, I consider using the audit rate of other hospitals in the same state in *other* markets, which I define using hospital referral regions. This instrument leverages hospitals whose behavior should not be affected by a given hospital’s behavior, since they are much further away in different markets. Figure D16b shows that the results are robust to using these hospitals to instrument for a hospital’s audit rate. To confirm that the results are not driven by a single state or hospital comparison group, Figure D17 plots the distribution of coefficients when one state or one hospital comparison group is removed from the sample. The coefficients are always negative and distributed 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 D18a illustrates the interior borders and the sample of hospitals within 100 miles of the interior border (excluding hospitals that are within 100 miles of the RAC border). The falsification test shows no effect on admissions on the high-audit side of the interior border (Figure D18b), in contrast to the main results which show a drop in admissions on the high-audit side of the RAC border.

Patient-Level Analysis In Table E7, I show that the Two Midnight difference-in-difference results are robust to varying the sample to include patients who arrive between 1 to 5 hours of midnight. Table E3 shows that, in addition to a null effect on revisits within 30 days, there is no effect on revisits within 60 or 90 days.

In column 5 of Table 4, 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 Two Midnights rule onto populations not covered by the rule.

C Welfare Analysis

C.1 Welfare Analysis Calculations

I next lay out the estimates required to calculate the marginal welfare effect in each year. Define θ_t and D_{at} as the estimates on log inpatient revenue and amount demanded in Table 3 and γ_t as the estimates on hospital administration costs in Table 3. Let I_{2010} be a hospital's inpatient revenue in 2010 (Table 5). Define I_{a_T} to be the present-discounted value of the marginal reduction in log inpatient revenue between 2011 and year T due to an exogenous increase in audit rate in 2011, relative to 2010. If θ_t is the estimated percent reduction in revenue in year t relative to 2010 (i.e., Table 3 column 2) and δ is the discount rate, then:[24]

$$I_{a_T} = \sum_{t=2011}^{T} \frac{\theta_t R_{2010}}{(1+\delta)^{t-2010}} \tag{12}$$

The total effect on government revenue also includes the money demanded back from audits D_{a_t} (i.e., Figure D12) less the contingency fee f paid to RACs, which ranges from 9 to 12.5 percent of the amount demanded, and scaled by the share s of demanded payments

[24] I_{a_T} is a negative number because θ_t is negative, and the marginal effect of increased auditing on hospital inpatient revenue is negative.

that was refunded to hospitals in later lawsuits.[25] For the main calculations I assume f to be the midway point between 9 and 12.5, 10.75 percent. If RACs are perfectly competitive and make zero profit, then f is equal to the monitoring costs to society; otherwise f is an upper bound. Define $-R_{a_T}$ to be the present-discounted value of the marginal savings to the government of increasing the 2011 audit rate.[26]

$$-R_{a_T} = -I_{a_T} + (1-s)\sum_{t=2011}^{T} \frac{(1-f)D_{a_t}}{(1+\delta)^{t-2010}} \tag{13}$$

For provider compliance costs, let k_{2010} be a hospital's 2010 compliance costs (Table 5, k_{a_T} be the present-discounted value of the marginal increase in compliance costs between 2011 and year T, and γ_t be the estimated percent increase in compliance costs in year t relative to 2010 (i.e., Table 3 column 5). Then:

$$k_{a_T} = \sum_{t=2011}^{T} \frac{\gamma_t k_{2010}}{(1+\delta)^{t-2010}} \tag{14}$$

The marginal effect on government monitoring costs m_{a_t} is defined as the contingency fee f multiplied by the money demanded back from audits each year.

$$m_{a_T} = \sum_{t=2011}^{T} \frac{f D_{a_t}}{(1+\delta)^{t-2010}} \tag{15}$$

C.2 Welfare Results Under Alternative Assumptions

In Figure D20 I compare the "most conservative" case to the "least conservative" one, where the most conservative case corresponds to the highest costs and the lowest savings. In the most conservative case, RACs charge the highest contingency fee of 12.5 percent, the deterrence effect on admissions is 0 after 2015, and Medicare has to refund 68 percent of demanded payments. In the least conservative case, RACs demand the lowest contingency fee of 9 percent, the effect on admissions is permanently negative after 2015, and Medicare keeps all the demanded payments. Even in the most conservative case, increasing audits is welfare-improving by 2015. The parameters used for this robustness test are reported in Table E9.

Figure D21 compares cases that relax the assumption of no change in treatment cost. In particular, I assume that the admission price is a markup above treatment cost, so that the change in total treatment cost in response to monitoring is a fraction of the change in

[25] Appendix Section A.2.4 discusses later lawsuits that refunded money to hospitals.

[26] $-R_{a_T}$ is a positive number.

inpatient revenue. Case (2) calculates the marginal welfare when treatment cost is 20% of the Medicare price, case (3) calculates it when treatment cost is 60% of the Medicare price, and case (4) calculates it when treatment cost is equal to the Medicare price. The baseline calculation, where there is *no* change in treatment cost in response to monitoring, is a lower bound on the welfare effect.

The welfare calculations hinge on two key assumptions: first, that the marginal effect on patient health of being denied admission is zero, and second, that the MVPF is 1.3, relative to a marginal value of hospital revenue of 1. To explore how these assumptions affect the findings, Figure D22 plots the relationship between marginal welfare per hospital in 2015, the marginal effect on patient health, and the MVPF. At a MVPF of 1.3, increasing the 2011 audit rate is still welfare-improving by 2015 as long as the harm per patient denied admission is less than \$190.

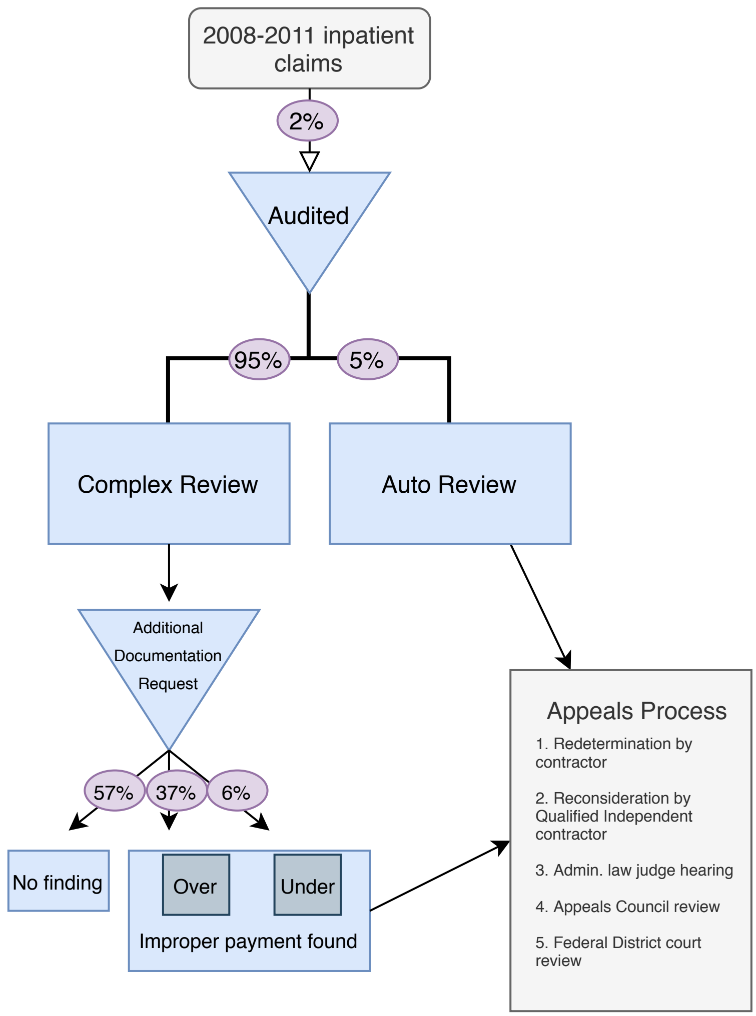
Finally, Figure D23 calculates the marginal cost of funds (MCF), which is the ratio between the net effect on patients and hospitals and the net effect on the government's budget (Slemrod and Yitzhaki, 2001; Kleven and Kreiner, 2006):

$$\text{marginal cost of funds} = \frac{k_a + \gamma' \frac{db}{dn_F}\frac{dn_F}{da} - \frac{dc}{dn_F}\frac{dn_F}{da} + R_a}{R_a - m_a}\ . \tag{16}$$

Comparing the MCF of a revenue-raising policy to the MVPF of government expenditure tells us whether a combined policy that raises revenue and spends it in this manner improves welfare. If the MCF is smaller than the MVPF, then the combined policy would be welfare-improving. Figure D23 compares the calculated marginal cost of funds at different points in time to a MVPF of 1.63, which is the estimated MVPF of Medicare (Finkelstein and McKnight, 2008; Hendren and Sprung-Keyser, 2020), and a MVPF of 1.3. The MCF crosses 1.63 by 2013, while it crosses 1.3 by 2015.

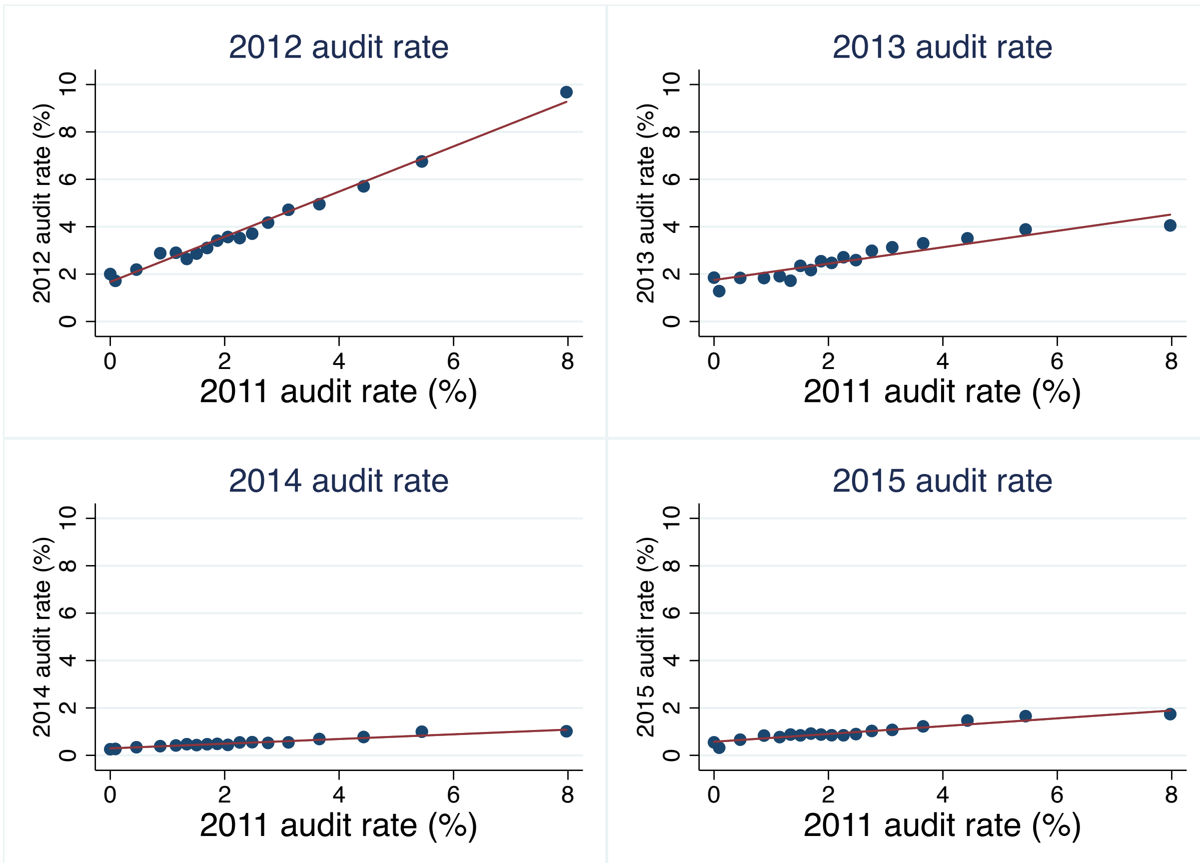
D Appendix Figures

Figure D1. 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 that occurred in 2011 of inpatient claims between 2008 and 2011. Data: CMS audit data.

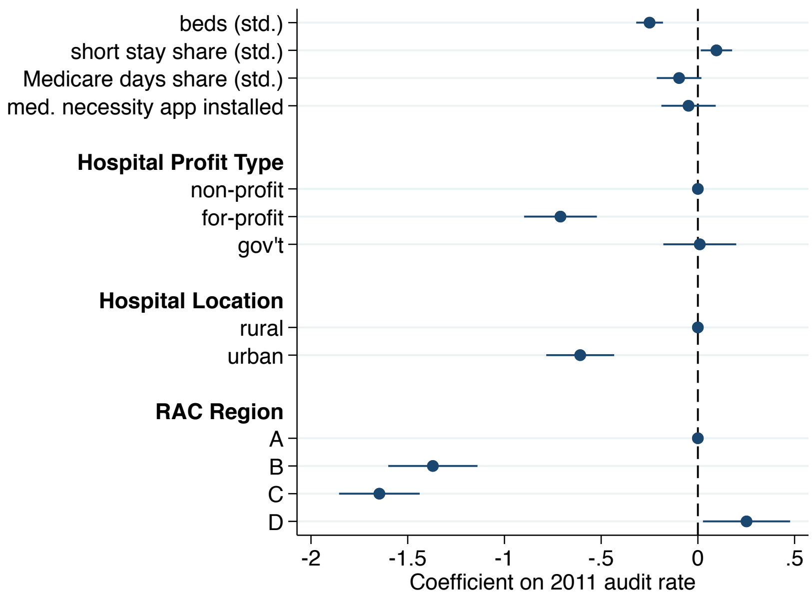
Figure D2. Correlation of 2011 Audit Rate with Later Year Audit Rates



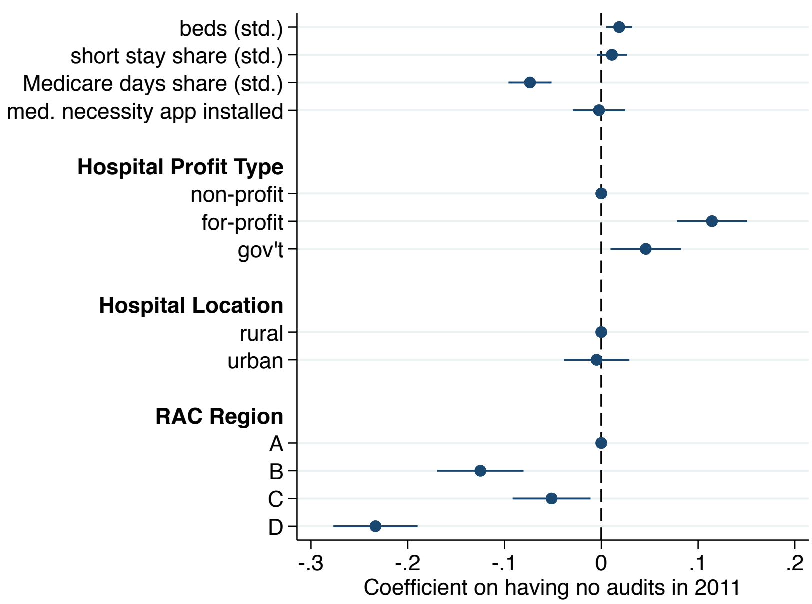
This figure plots binscatters of the correlation between hospital audit rates in 2011 and audit rates in subsequent years. Data: MEDPAR and CMS audit data.

Figure D3. Correlation between Hospital Characteristics on 2011 Audit Rate and No Audit

(a) Outcome: 2011 hospital audit rate

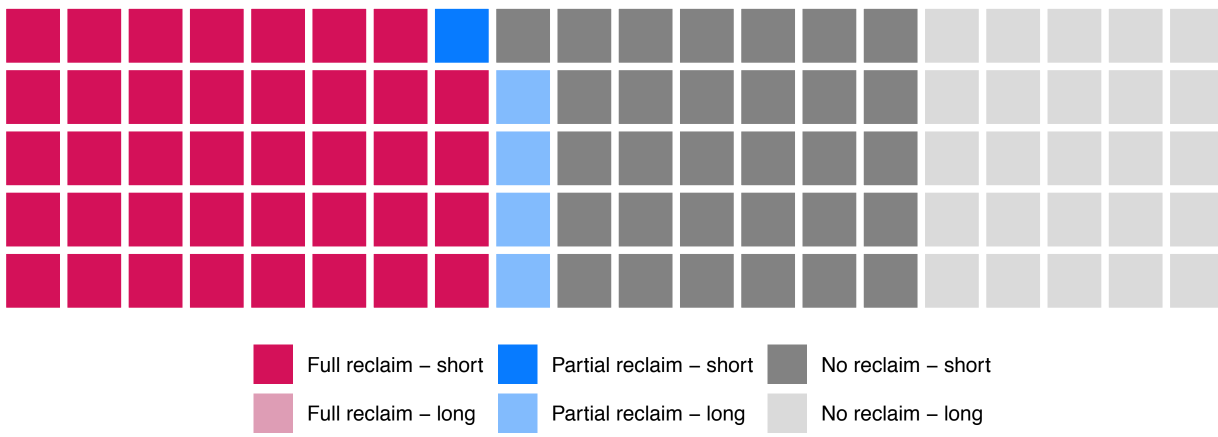


(b) Outcome: no audits at hospital in 2011



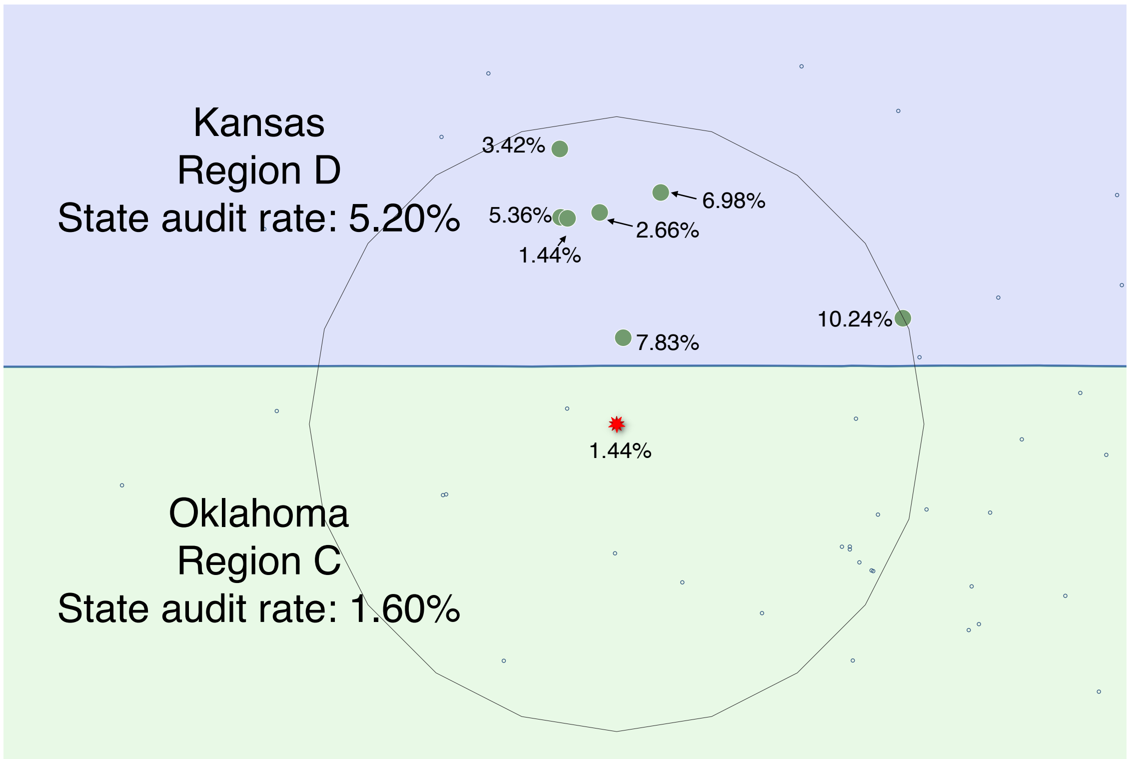
These figures plot coefficients from a regression of (a) a hospitals 2011 audit rate and (b) an indicator variable for whether a hospital was not audited in 2011 on 2010 hospital characteristics. Short stay share is the share of 2010 Medicare admissions with lengths of stay 0-2. Medicare days share is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

Figure D4. 2011 Audit and Denial Characteristics



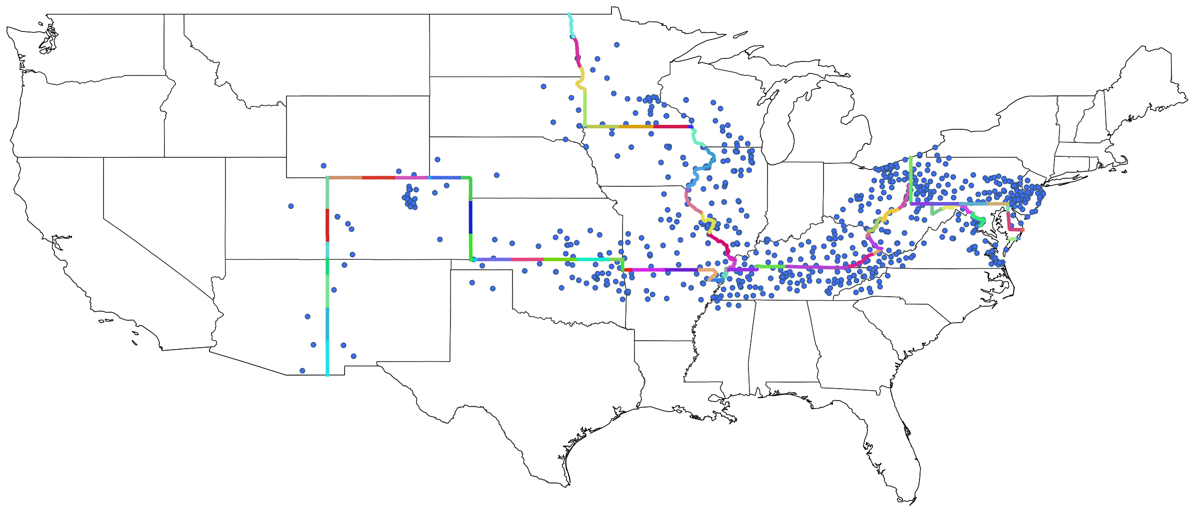
This figure is a waffle plot of 2011 audits of inpatient stays in 2008-2011, where each box represents one percent of total audits. The dark shaded boxes of each color denote audits of inpatient stays. The red and blue colored boxes denote audits that result in the full payment being reclaimed or a partial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

Figure D5. Example of Border Hospital and Neighbor Comparison Group Definition



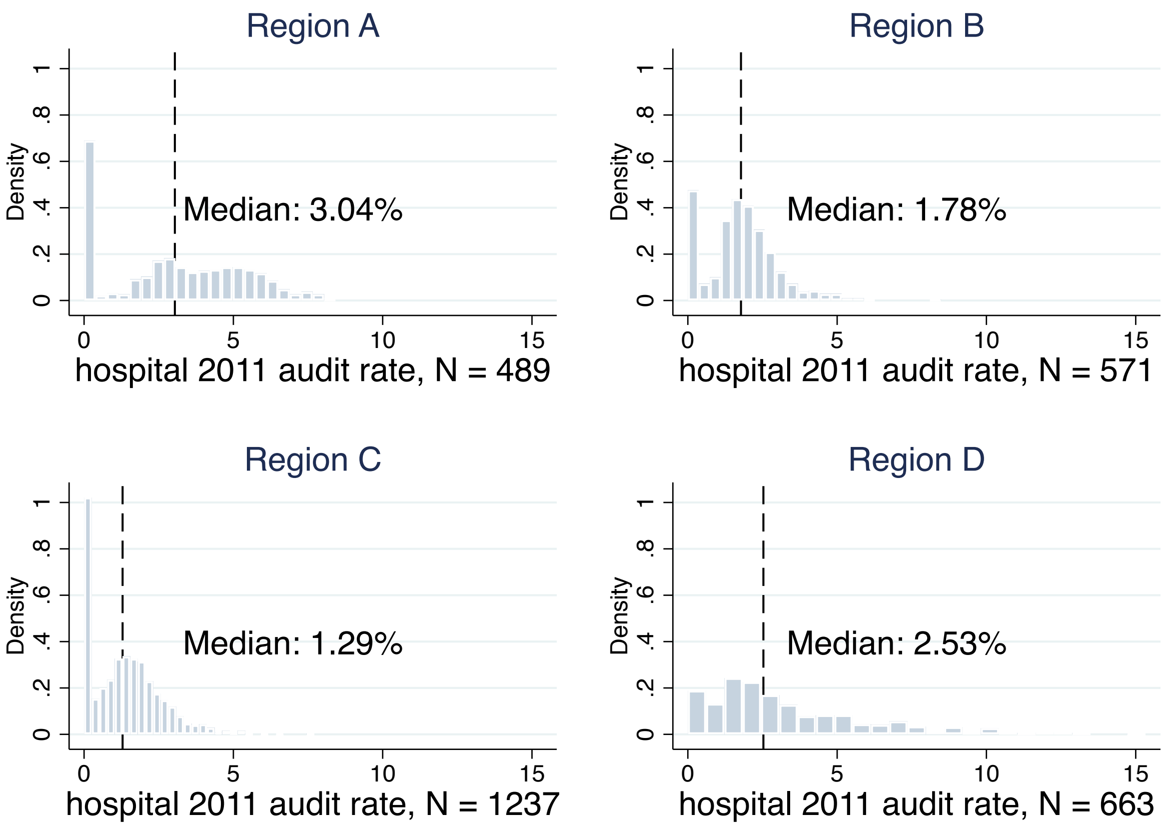
This figure illustrates how a "neighbor comparison group" is identified for each border hospital in the across-hospital empirical strategy. Neighboring hospitals are all hospitals within a 100 mile radius of a hospital, on the opposite side of the RAC border. In this example, the green circle hospitals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

Figure D6. RAC Border Segments and Hospitals Within 100 Miles



This figure shows how the RAC border is divided into 100 mile segments that do not cross state borders, and all hospitals within 100 miles of the RAC border.

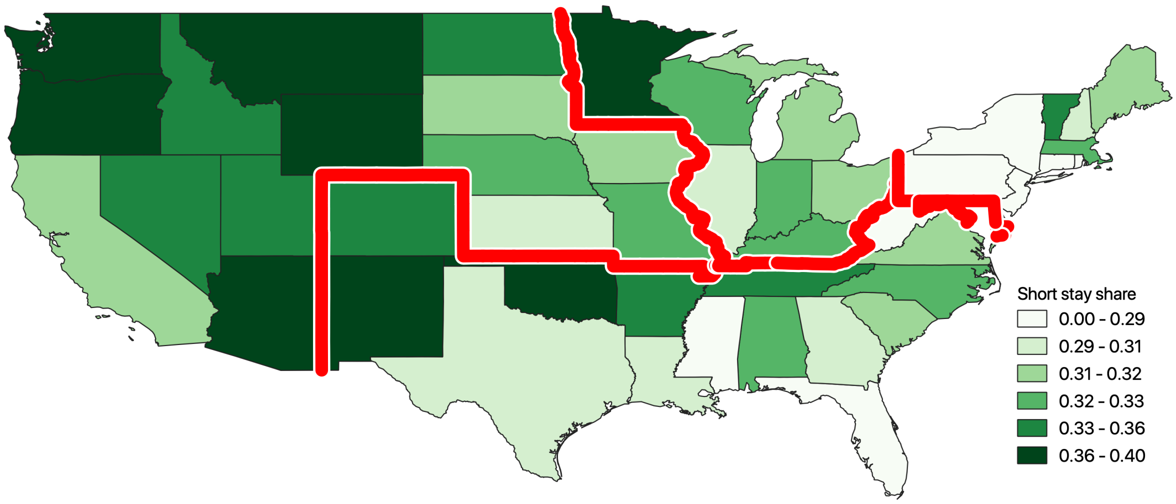
Figure D7. Histogram of 2011 Hospital Audit Rates by RAC Region



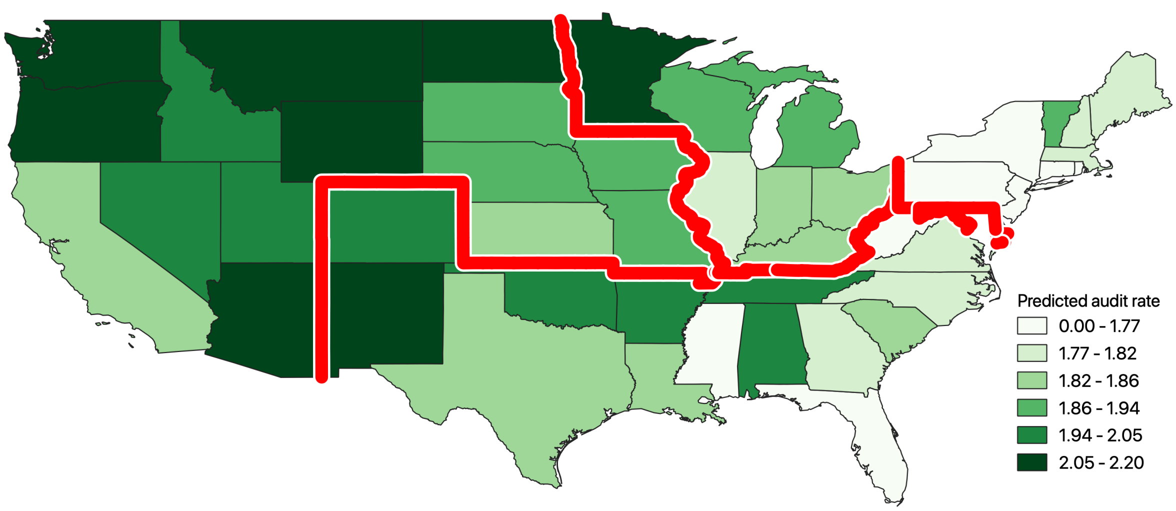
This figure plots the histogram of 2011 hospital audit rates by RAC region, where audit rate is defined as the percent of a hospital's 2008-2011 claims that were audited by RACs. Data: MEDPAR and CMS audit data.

Figure D8. 2010 Average Short Stay Share of Medicare Admissions and Predicted 2011 Audit Rate by HRR

(a) 2010 Average Short Stay Share of Medicare Admissions by State

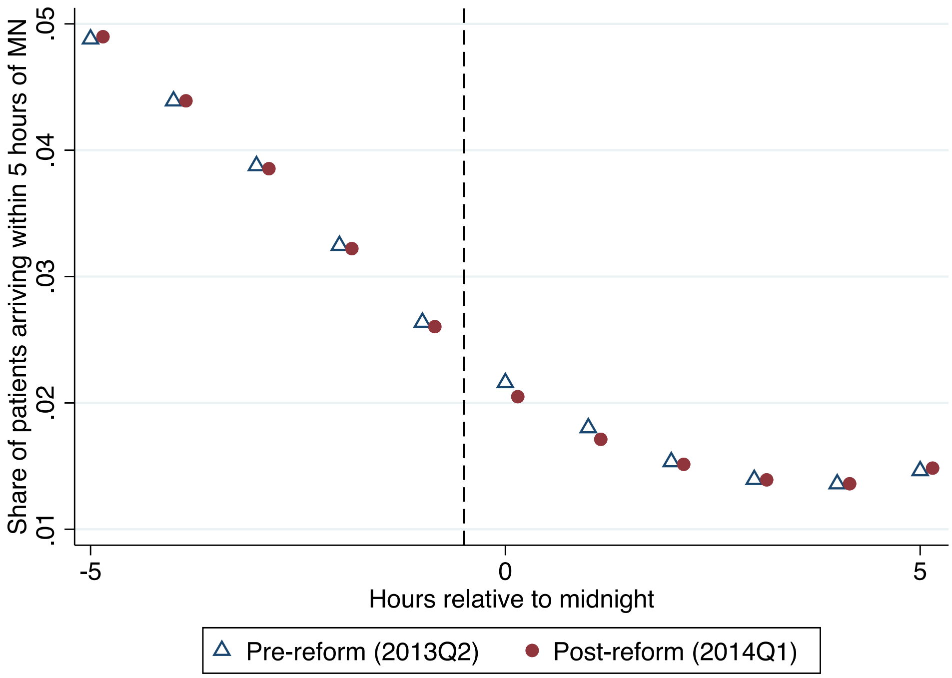


(b) Predicted 2011 Audit Rate by State



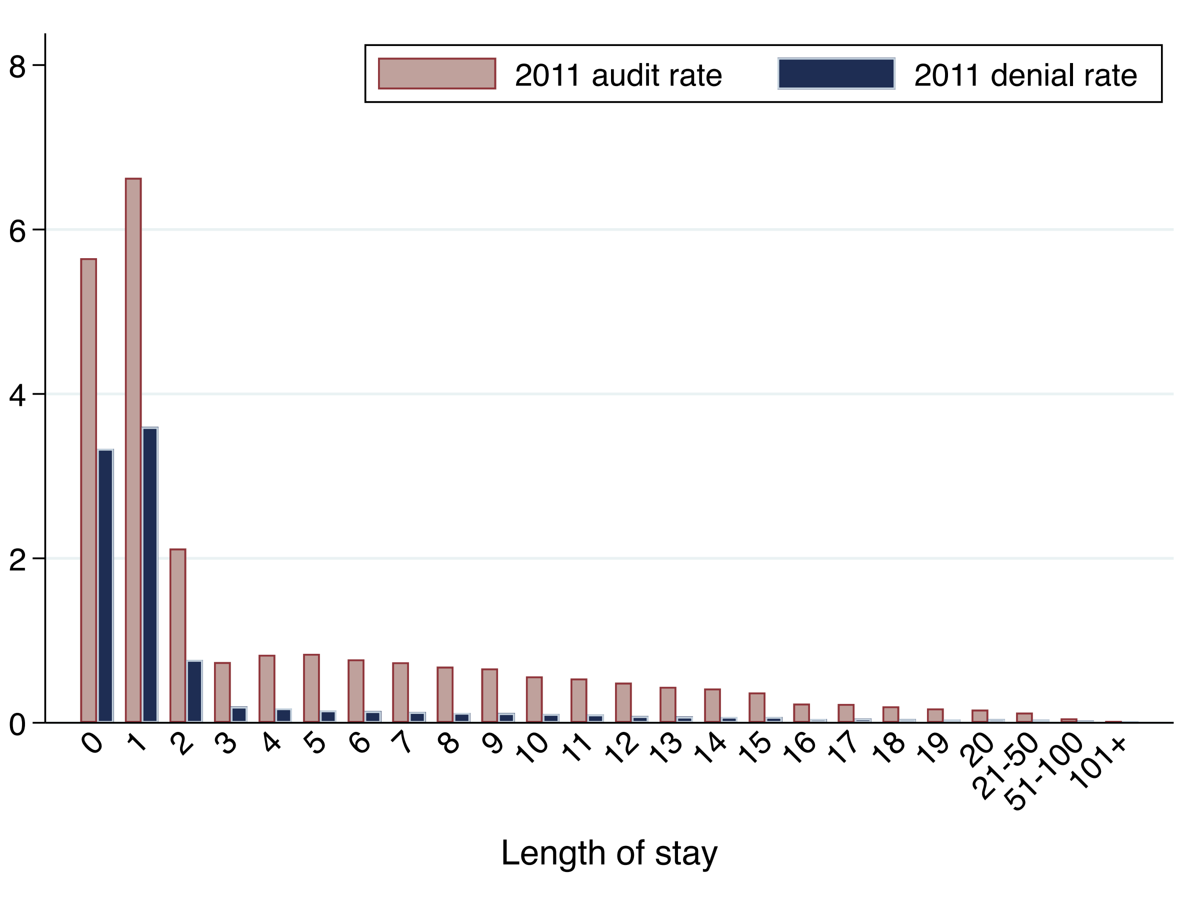
These figures plot state averages of hospital-level characteristics. The top panel plots the average share of Medicare admissions with a length of stay of 0-2 in 2010, and a darker shade is associated with a higher share. The bottom panel plots the predicted 2011 audit rate using characteristics of 2007-2009 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 of each hospital's 2007-2009 claims. The red line demarcates RAC regions, which are: Region A (Northeast), Region B (Midwest), Region C (South), and Region D (West). Darker shades denote higher audit rate. The red line demarcates RAC regions. Maryland was not audited under the RAC program as it uses a unique all-payer rate-setting system for hospital services. Data: MEDPAR and CMS audit data.

Figure D9. Share of Medicare ED Patients By Hour of ED Arrival



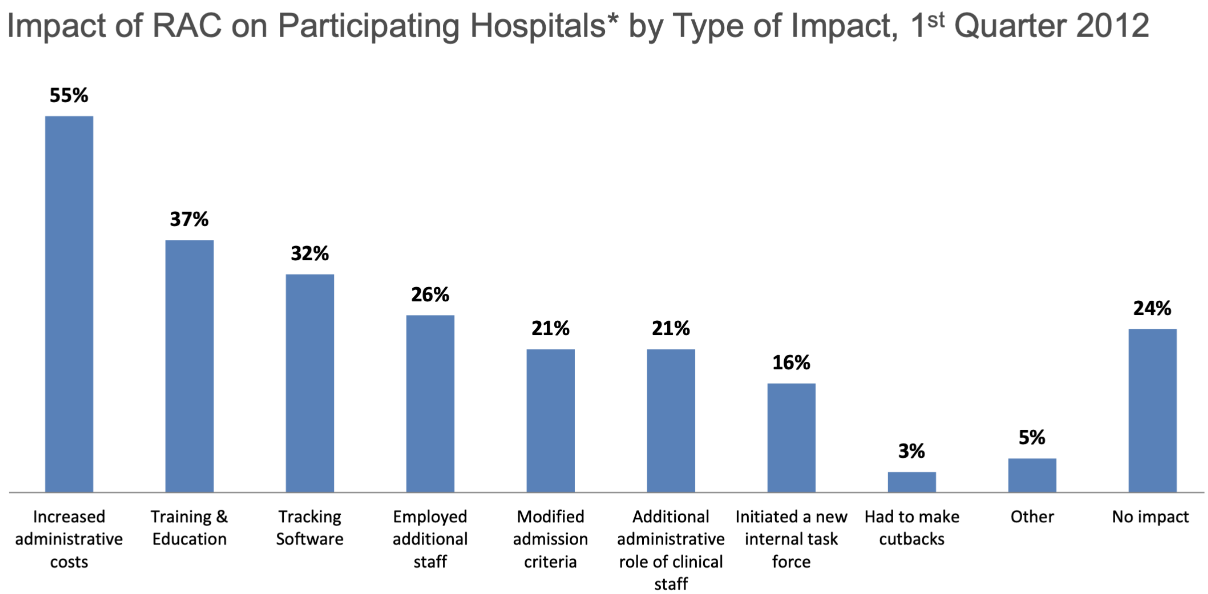
This figure plots the share of Medicare patients that arrive at the ED at each hour (relative to midnight) pre- and post-reform, among traditional Medicare patients who arrived in the ED within 5 hours of midnight in Florida. Data: HCUP SID/SEDD.

Figure D10. 2011 Audit and Denial Rates by Admission Length of Stay



This figure plots average 2011 audit rates and denial rates by an admission's length of stay. Audit rate is defined as the share of eligible admissions that were audited, and denial rate is the share of all eligible admissions that are audited and a payment is demanded from. Data: MEDPAR and CMS audit data.

Figure D11. RACTrac Survey on Hospital Administration Spending, 2012 Quarter 1



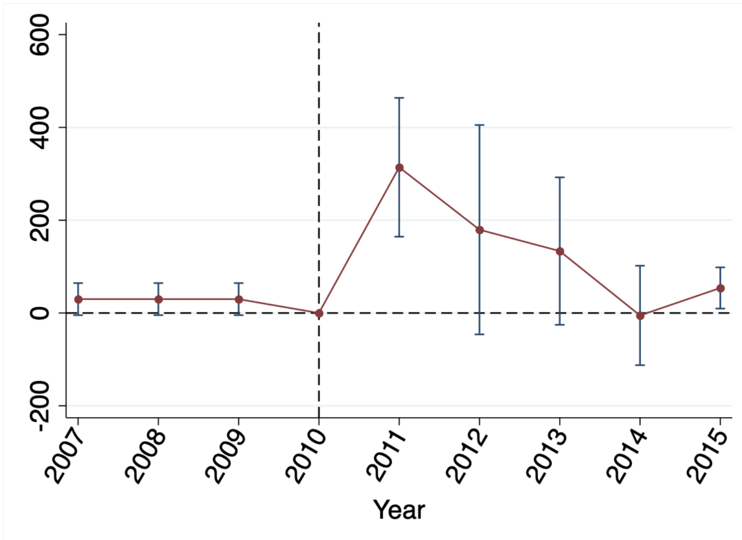
** Includes participating hospitals with and without RAC activity*

Source: AHA. (May 2012). *RACTrac* Survey

AHA analysis of survey data collected from 2,220 hospitals: 1,854 reporting activity, 366 reporting no activity through March 2012. Data were collected from general medical/surgical acute care hospitals (including critical access hospitals and cancer hospitals), long-term acute care hospitals, inpatient rehabilitation hospitals and inpatient psychiatric hospitals.

This figure is from a report published by the American Hospital Association on the RACTrac Survey, titled "Exploring the Impact of the RAC Program on Hospitals Nationwide" (American Hospital Association, 2012).

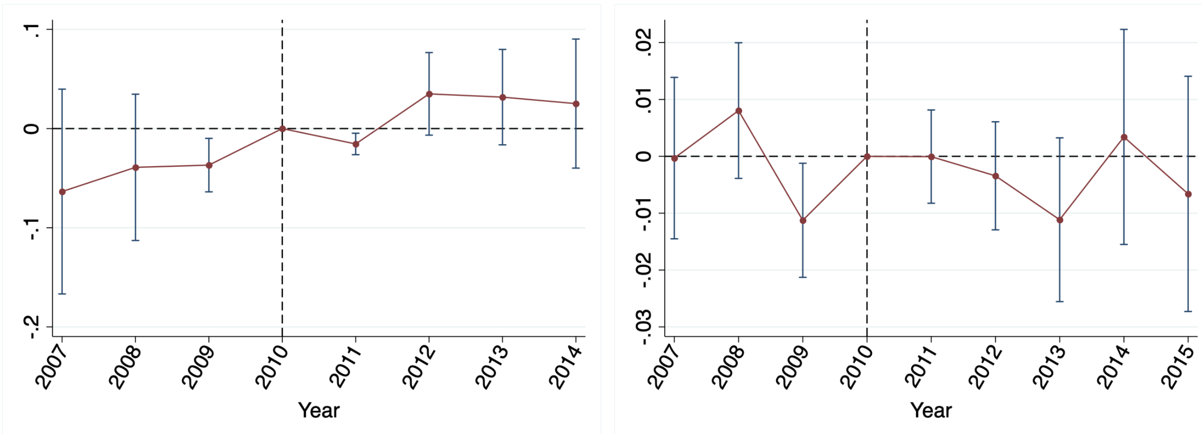
Figure D12. Event Study on Effect of 2011 Audit Rate on Amount Demanded ($1000s) from RAC Audits



This figure plots the event study of the IV coefficient and 95% confidence interval of the specification in Equation 1, where the outcome variable is the amount demanded ($) from audits of inpatient claims per hospital. Data: CMS audit data.

Figure D13. Event Studies on Effect of 2011 Audit Rate on Hospital Outpatient Revenue and Observation Stays

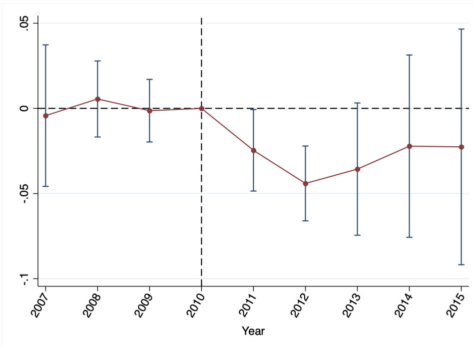
(a) Log observation stays (b) Log Medicare outpatient revenue



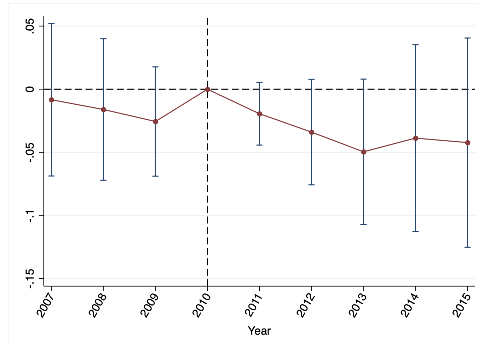
This figure plots event studies of the IV coefficients and 95% confidence intervals of the specification in Equation 1. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on a hospital-level outcome in a given year. Observation stays are defined as outpatient claims associated with revenue center "0760" or "0762," or the HCPCS procedure codes "G0378" or "G0379." Outpatient revenue is the sum of all Medicare outpatient payments. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their "neighbor comparison group." Data: Medicare outpatient claims.

Figure D14. Event Studies on Effect of 2011 Denial Rate on Medicare Admissions and Revenue, and Administrative Burden

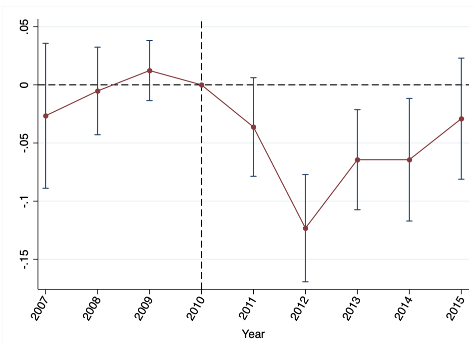
(a) Log Medicare admissions



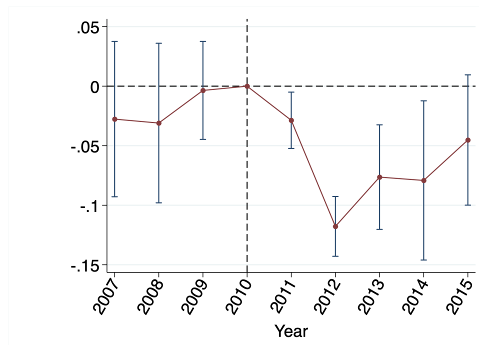
(b) Log Medicare inpatient revenue



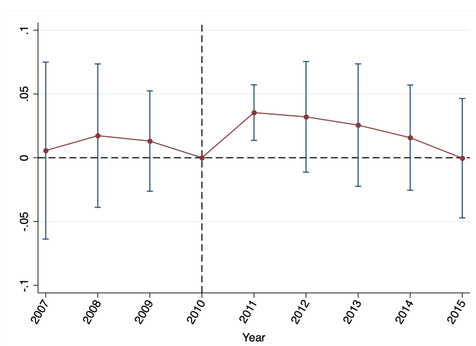
(c) Log Medicare admissions, LOS ≤ 2



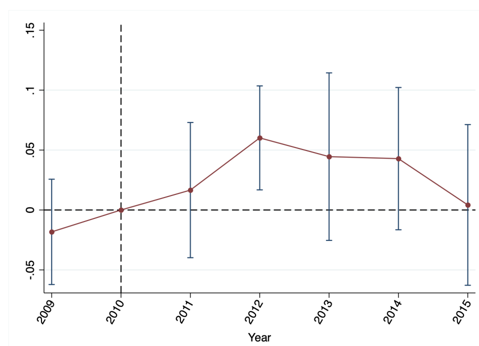
(d) Log Medicare inpatient revenue, LOS ≤ 2



(e) Log net administration costs



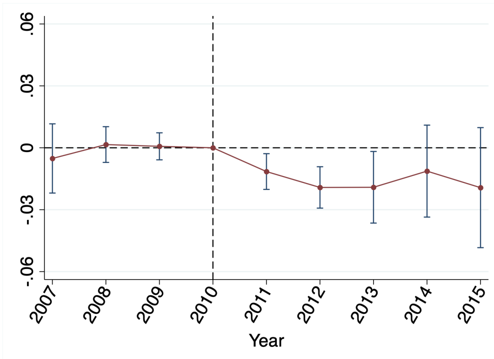
(f) Indicator for installing medical necessity software



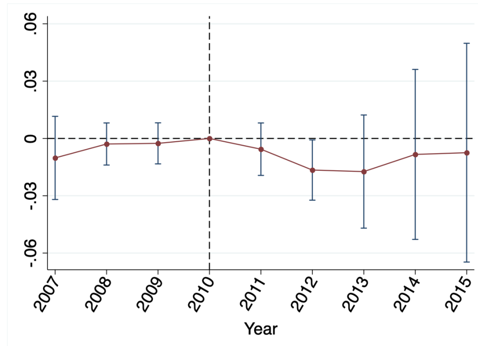
This figure plots event studies of the IV coefficients and 95% confidence intervals of the specification in Equation 1. The results are clustered at the state and border segment level. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 denial rate (share of claims that are audited and result in an overpayment/underpayment demand) on log Medicare admissions and revenue 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. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their "neighbor comparison group."

Figure D15. Robustness to Sample Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

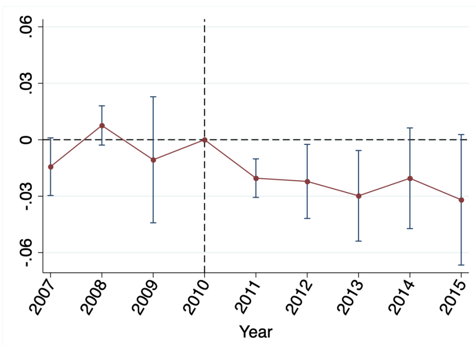
(a) Sample: within 100 miles (main)



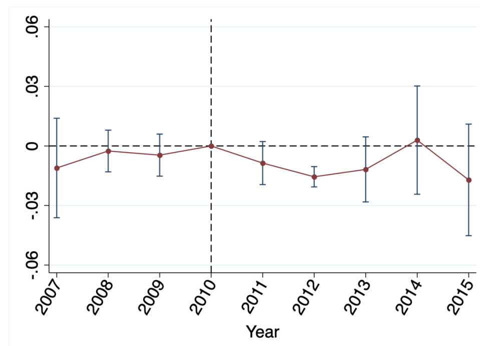
(b) Sample: within 50 miles



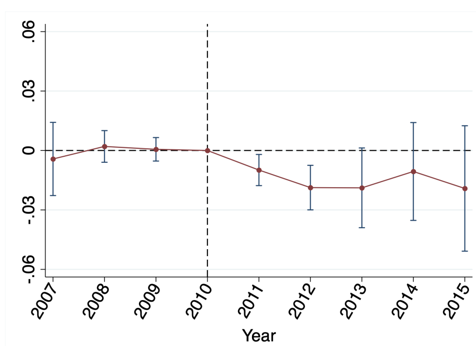
(c) Sample: within 150 miles



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



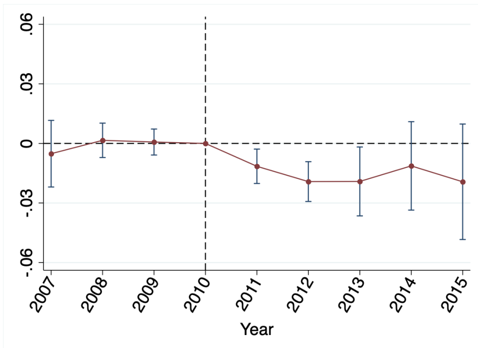
(e) Sample: within 100 miles, audit rate > 0



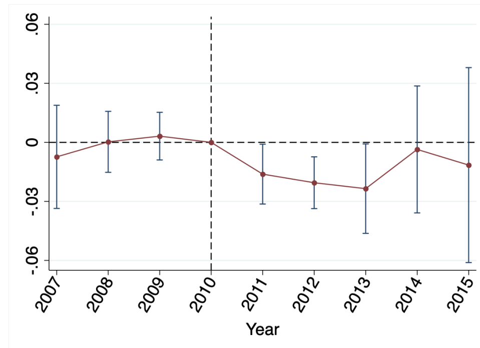
This figure plots robustness analysis event studies of the IV coefficients and 95% confidence intervals of the specification in Equation 1. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different 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. Data: MEDPAR.

Figure D16. Robustness to Instrument Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

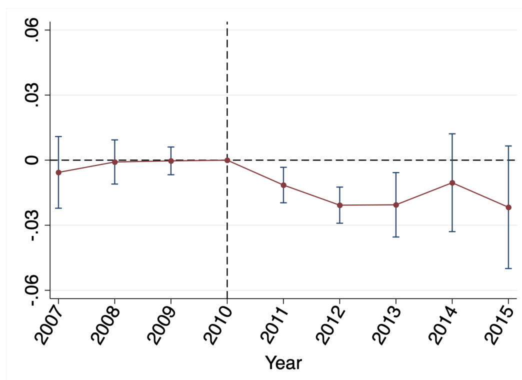
(a) Instrument: state audit rate



(b) Instrument: state rate, outside HRR



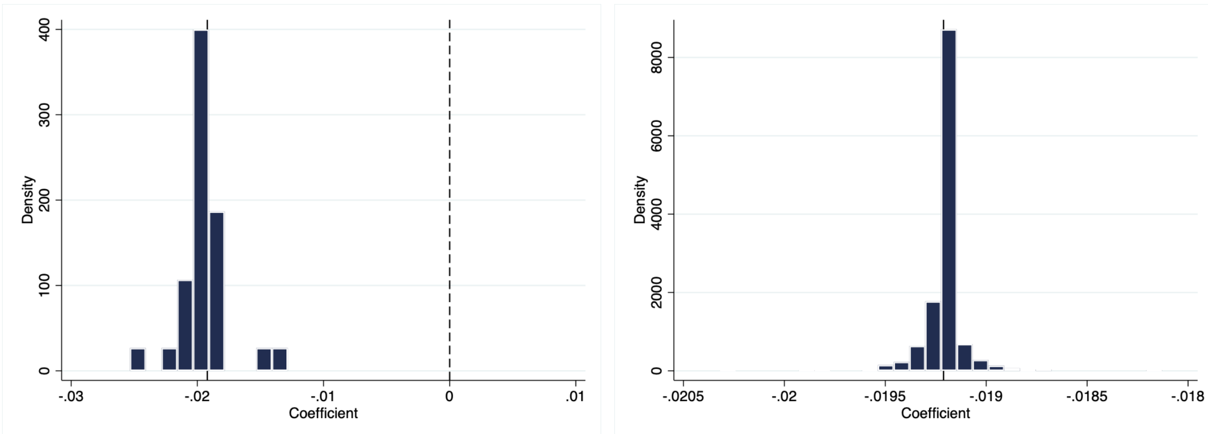
(c) Instrument: leave-one-out RAC region audit rate



This figure plots robustness analysis event studies of the IV coefficients and 95% confidence intervals of the specification in Equation 1. The omitted year is 2010. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on log Medicare admissions. The figures plot the results using different instruments for a hospital's 2011 audit rate. Panel (a) uses 2011 state audit rate and panel, (b) uses 2011 audit rate among hospitals in the same state but in different hospital referral regions (HRR) as the hospital, and (c) uses the 2011 audit rate of other hospitals in the same RAC region. Data: MEDPAR.

Figure D17. Robustness Test: Leave-one-out Coefficients of 2012 Effect of 2011 Audit Rate on Log Medicare Admissions

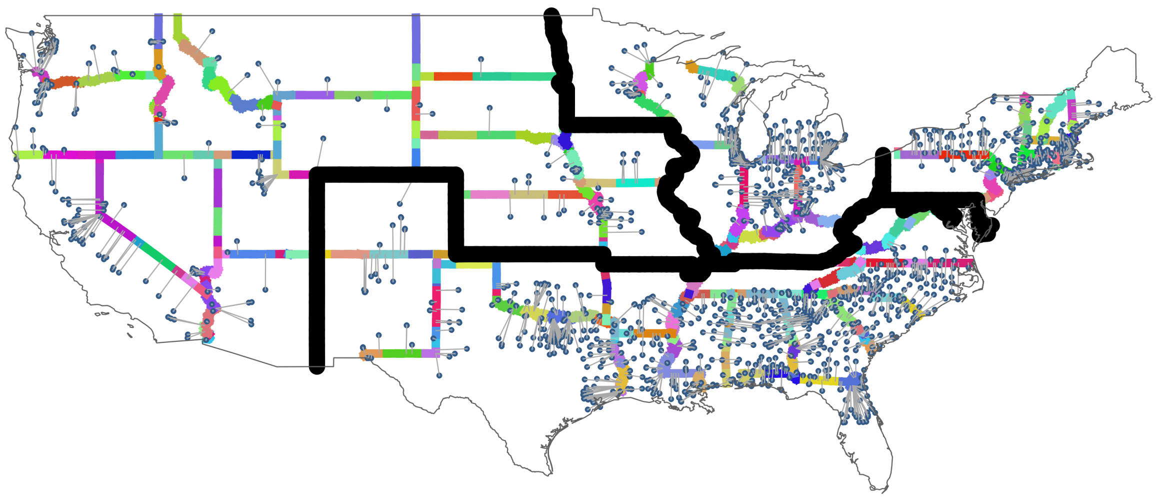
(a) Leaving 1 state out at a time (b) Leaving 1 comparison group out at a time



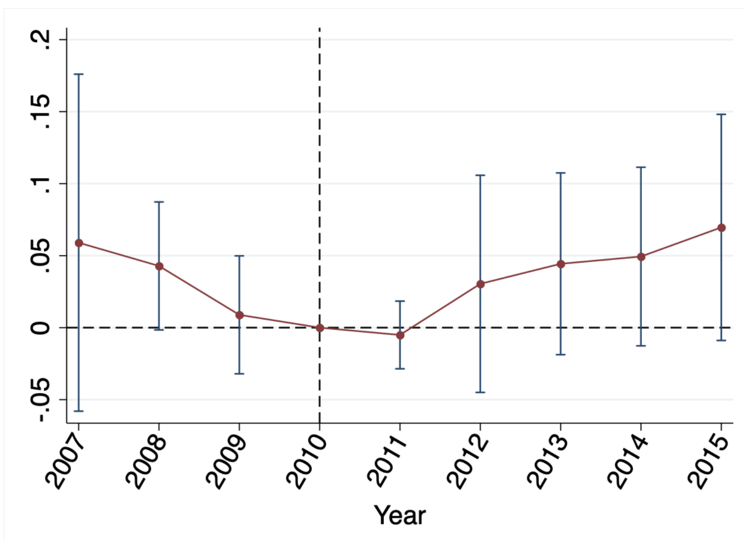
This figure plots distributions of the 2012 IV coefficient (β^{2012} from the specification in Equation 1), where the outcome is log Medicare admissions. Panel (a) plots the distribution of the coefficient when leaving one state out at a time, and panel (b) plots the distribution of the coefficient when leaving 1 hospital neighbor comparison group out at a time.

Figure D18. Falsification Test: Interior State Borders

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



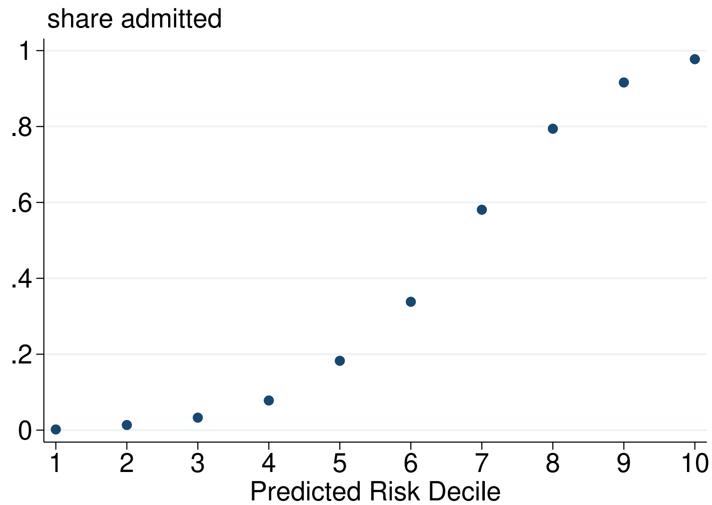
(b) Event Study on Effect of 2011 Audit Rate on Log Medicare Admissions



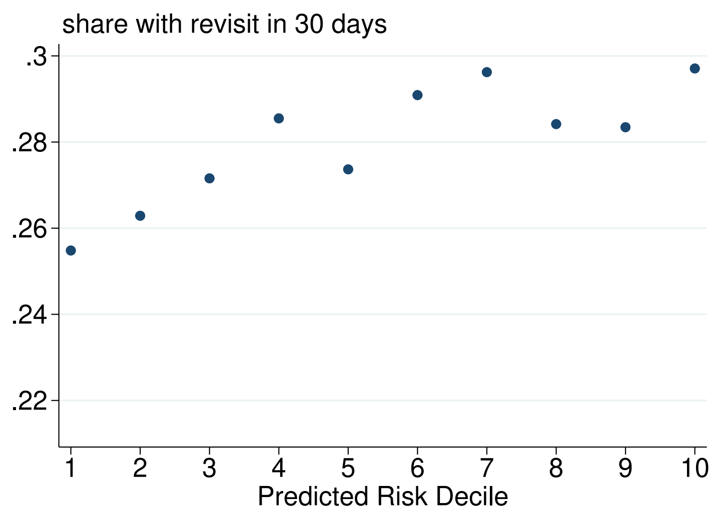
The top panel of this figure plots a map of state borders on the interior of RAC regions, divided into 100-mile segments that do not cross state borders. The RAC border is the thick black line. Each dot represents a hospital within 100 miles of the interior state borders, excluding hospitals that are in the main sample (within 100 miles of the RAC border). The line between the hospital and the interior state border denotes the closest interior state border to that hospital. The bottom panel of this figure plots the event study of the IV coefficient and 95% confidence interval of the specification in Equation 1, where the outcome variable is log Medicare admissions (MEDPAR). Sample is comprised of hospitals within 100 miles of the state interior border with at least 1 hospital in their "neighboring hospital comparison group" and are clustered at the state and border segment level.

Figure D19. Average Outcomes by Patient Severity

(a) Inpatient

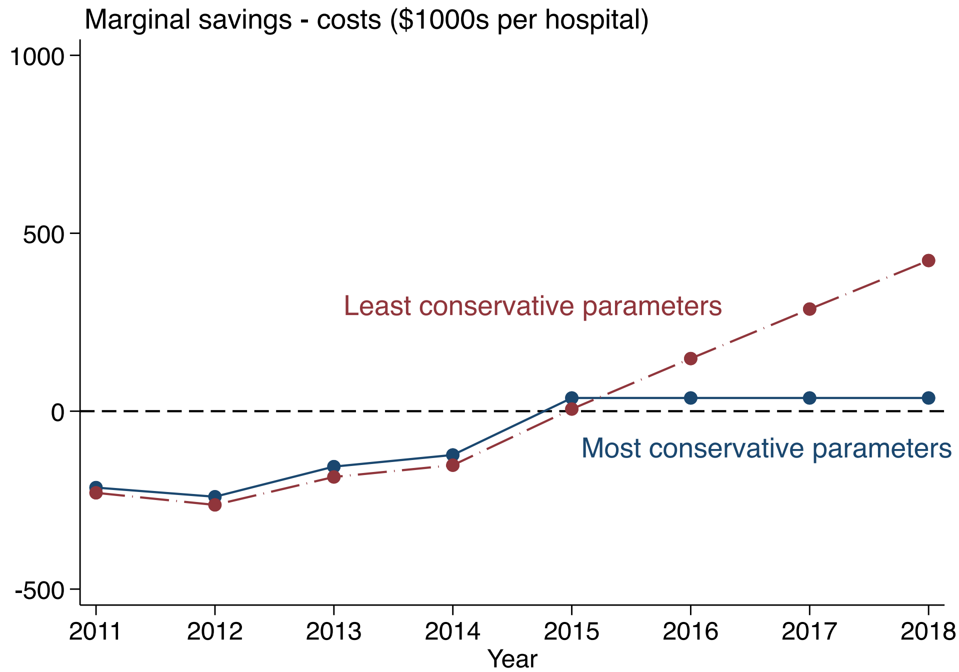


(b) Revisit in 30 days



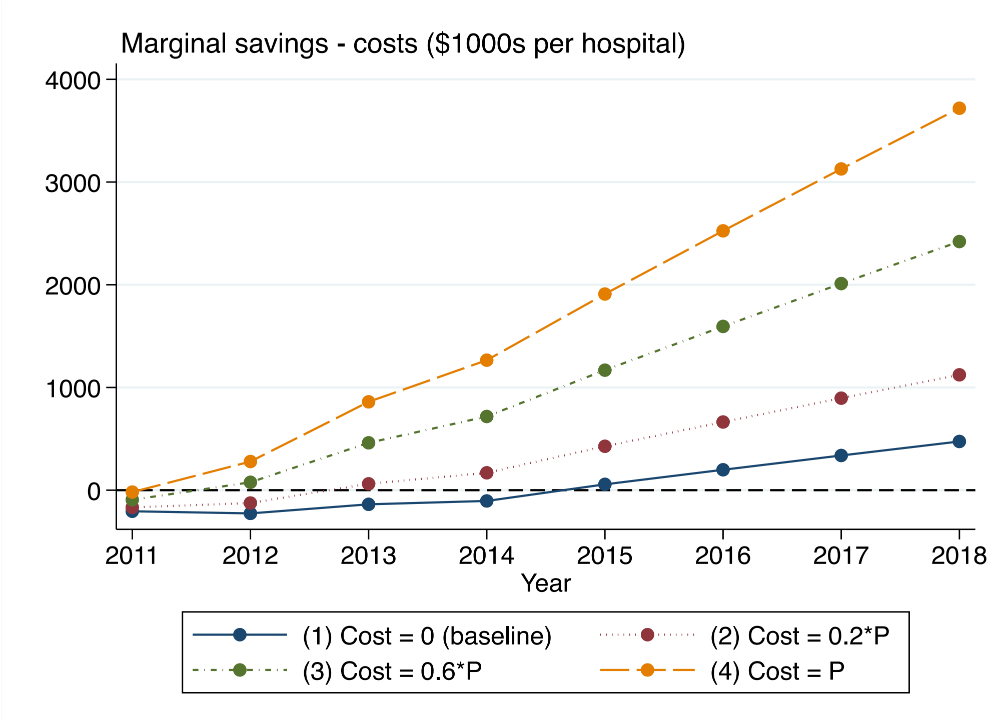
This figure plots (a) the share of patients admitted as inpatient from the ED and (b) the share of patients with a revisit within 30 days by predicted severity decile, in 2013Q2. Patient risk is predicted by estimating a logit using ED visits between 9AM and 3PM of an indicator for being admitted within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

Figure D20. Welfare Analysis Estimates, Most vs. Least Conservative



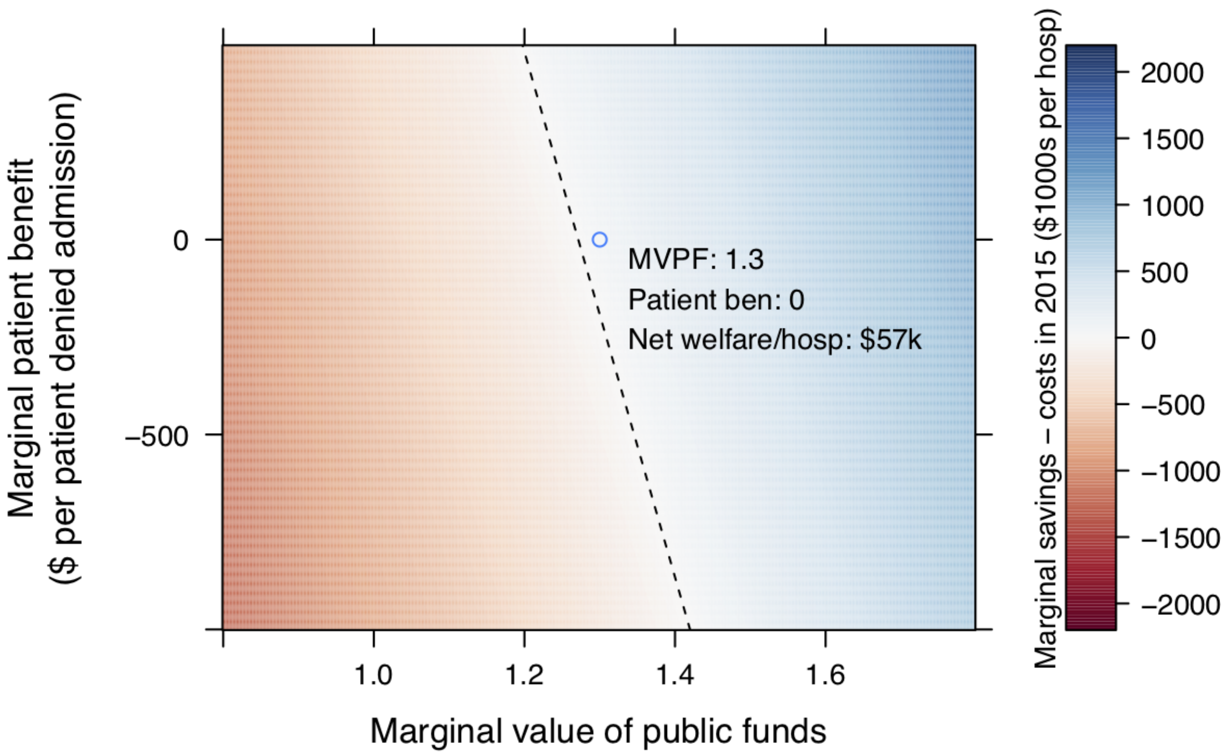
This figure plots the per-hospital welfare effect, or the difference between the savings and compliance/administrative costs of auditing, of increasing audits in 2011 by a given year. Increasing audits is welfare-improving if this value is positive and welfare-reducing if this value is negative. The figure plots the estimates from the most conservative case to the least conservative case. In the most conservative case, the RACs charge the highest contingency fee of 12.5 percent, the effect on admissions is 0 after 2015, and CMS has to refund 68 percent of demanded payments. In the least conservative case, RACs demand the lowest contingency fee of 9 percent, the effect on admissions is permanent after 2015, and CMS keeps all the demanded payments. Table D20 reports the parameters used to calculate each case.

Figure D21. Welfare Analysis Estimates, by Treatment Cost Assumptions



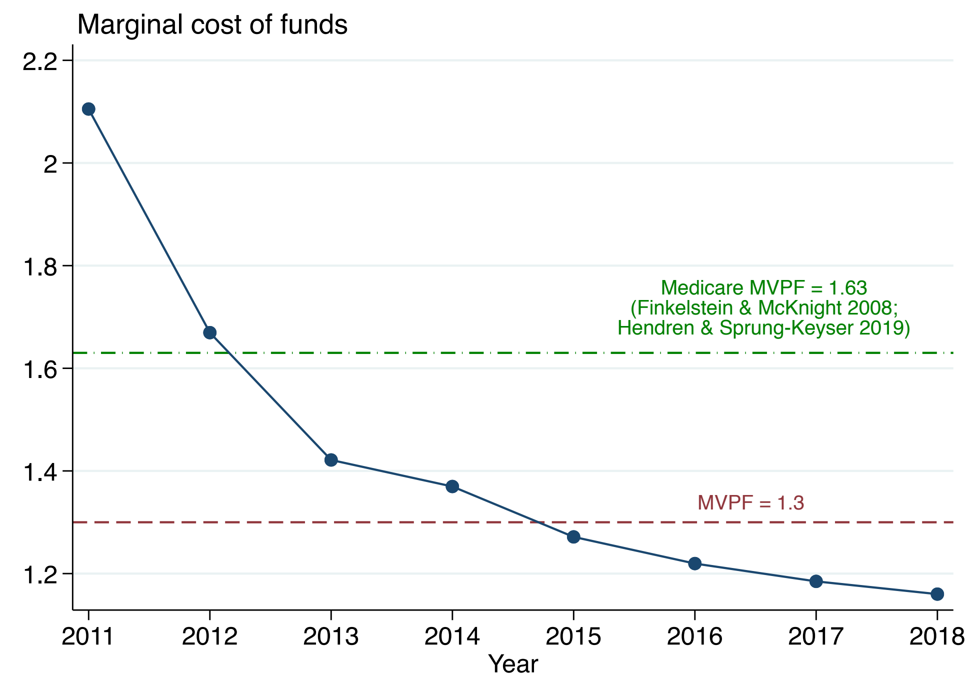
This figure plots the per-hospital welfare effect, or the difference between the savings and compliance/administrative costs of auditing, of increasing 2011 audits under varying assumptions about the change in treatment cost. Increasing audits is welfare-improving if this value is positive and welfare-reducing if this value is negative. This figure plots the following assumptions: (1) no change in treatment cost from reduced admissions; (2) a change in treatment cost per admission equal to 0.2× price of reduced admission; (3) a change in treatment cost per admission equal to 0.6× price of reduced admission; (4) a change in treatment cost per admission equal to the price of each reduced admission. Table 5 lists the rest of the parameters and estimates used to calculate the welfare effects.

Figure D22. Marginal Welfare Effect in 2015 by Patient Benefit and Marginal Value of Public Funds



This figure plots the per-hospital marginal welfare effect of increasing 2011 audits, with varying assumptions about the marginal value of public funds and the marginal patient benefit ($ per patient denied admission) in 2015. Increasing audits is welfare-improving if this value is positive (blue) and welfare-reducing if this value is negative (red). The blue point represents the baseline specification, which assumes a MVPF of 1.3 and no patient health effects from reduced admissions. The dotted line denotes the set of combinations of marginal patient benefit and marginal value of public funds where the marginal welfare effect is 0. Table 5 lists the other parameters and estimates used to calculate the welfare effects.

Figure D23. Marginal Cost of Funds



This figure plots the marginal cost of funds, taking into account cumulative savings and cumulative costs since 2011 (Slemrod and Yitzhaki, 2001; Kleven and Kreiner, 2006). The top horizontal dashed line represents a marginal value of public funds (MVPF) of 1.63, which is the estimated MVPF of Medicare (Finkelstein and McKnight, 2008; Hendren and Sprung-Keyser, 2020). The bottom horizontal dashed line represents a MVPF of 1.3, which is the assumed MVPF in the baseline welfare calculation. Table 5 lists the parameters and estimates used to calculate the welfare effects.

E Appendix Tables

Table E1. Summary Statistics of 2010 Hospital Characteristics by 2011 Hospital Audit Rate, Overall Sample vs. Border Sample

	(1)		(2)		(3)		(4)	
	Overall Hospitals				Border Hospitals			
	Above Median		Below Median		Above Median		Below Median	
	A. RAC Program Characteristics							
2011 audit rate	3.60	(1.89)	0.73	(0.65)	3.62	(2.09)	0.84	(0.63)
Share in Region A	0.23		0.11		0.15		0.02	
Share in Region B	0.19		0.20		0.31		0.42	
Share in Region C	0.28		0.54		0.27		0.48	
Share in Region D	0.30		0.16		0.28		0.09	
	B. Overall Characteristics							
Beds	182.04	(164.09)	228.76	(195.51)	176.66	(194.80)	181.73	(149.42)
Share urban	0.68		0.76		0.49		0.60	
Share non-profit	0.68		0.58		0.72		0.66	
Share for-profit	0.12		0.25		0.12		0.21	
Share government	0.20		0.16		0.16		0.14	
Share non-chain	0.42		0.31		0.44		0.35	
Total cost (million \$)	193.78	(248.46)	215.21	(269.58)	164.60	(294.09)	163.17	(222.23)
Net admin costs (million \$)	28.84	(39.11)	32.12	(39.59)	24.83	(44.68)	26.18	(47.14)
Share with medical necessity app.	0.67		0.68		0.73		0.68	
	C. Medicare Inpatient Admission Characteristics							
Admissions	3056.70	(3057.97)	3931.26	(3351.82)	3007.99	(3332.92)	3225.81	(2833.25)
Mean payment (\$)	8788.95	(3134.69)	9001.31	(3104.10)	7539.54	(2268.09)	7618.78	(2231.30)
Total payments (million \$)	30.28	(38.07)	39.03	(42.22)	26.66	(40.01)	28.13	(31.07)
Mean share stays, LOS = 0-2	0.31	(0.07)	0.30	(0.07)	0.31	(0.07)	0.31	(0.06)
N neighboring hospitals					16.29	(11.29)	17.13	(11.21)
Observations	1474		1430		255		255	

This table presents 2010 summary statistics for hospitals above and below the median 2011 audit rate for two samples: all hospitals ("Overall Hospitals") and hospitals within 100 miles of the RAC border that have at least 1 hospital their "neighbor comparison group" ("Border Hospitals"). Standard deviation is in parentheses. The median audit rate for the overall sample in 2011 was 1.78%. The median audit rate for border hospitals in 2011 was 1.60%. Bed size, urban status, and profit type status come from the Medicare Provider of Services file. Non-chain status comes from hospital merge 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).

Table E2. ED Arrival Hour Manipulation Tests

	(1) $[23{:}00 \leq T_v \leq 23{:}59]$	(2) $1[T_v \geq 00{:}00]$
$1[q \geq 2013Q3]$	-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 $1[q \geq 2013Q3]$], 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 11PM and midnight, and 0 otherwise. $1[T_v \geq 00{:}00]$ 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 E3. After-Midnight ED Arrival Coefficient on Stay Characteristics and Patient Outcomes

	(1) Total Charges ($)	(2) N Diagnoses	(3) N Procedures	(4) OR Procedure	(5) Revisit 60d	(6) Revisit 90d
β	42.707	-0.003	-0.005	-0.001	0.002	0.000
	(254.406)	(0.013)	(0.009)	(0.001)	(0.002)	(0.002)
Observations	1252735	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 $1[q \geq 2013Q3] \times 1[T_v \geq 00{:}00]$ of the specification in Equation 7, where $1[q \geq 2013Q3]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $1[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "OR procedure" is an indicator for whether a patient received an OR procedure during their stay. "Revisit within 60/90 days" is an indicator for whether the patient had another ED visit or inpatient stay within 60/90 days of the ED visit. Sample comprises traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Data: HCUP SID/SEDD.

Table E4. Across-Hospital Post-2011 IV Coefficient

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

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered at the state and border segment level. This table reports IV coefficients in Equation 4. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on a hospital-level outcome after 2011. For brevity, the pre-2010 coefficients are estimated but not reported in the table. Omitted year is 2010. 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. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their "neighbor comparison group."

Table E5. Heterogeneity of Across-Hospital Post-2011 IV 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. App.	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 in parentheses and are clustered at the state and border segment level. This table reports IV coefficients in Equation 4, interacted with 2010 hospital characteristics. Each coefficient estimates the effect of a 1pp increase in 2011 audit rate on a hospital-level outcome after 2011 (Post). 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. Sample is comprised of hospitals within 100 miles of the RAC border with at least 1 hospital in their "neighbor comparison group." Omitted year is 2010.

Table E6. 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 $1[q \geq 2013Q3] \times 1[T_v \geq 00{:}00]$ of the specification in Equation 7, interacted with hospital characteristics. $1[q \geq 2013Q3]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $1[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "Inpatient" is an indicator variable for whether the patient was eventually admitted as inpatient from the ED (HCUP SID/SEDD). The sample consists of traditional Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. Urban/rural, teaching/non-teaching, for-profit/government/non-profit, and bed size come form the Medicare Provider of Services file. Non-chain status come from Cooper et al. (2019). Medical necessity application is an indicator which is equal to one if medical necessity checking application is listed as "live and operational," "contracted/not yet installed," "installation in process," or "to be replaced" in the HIMSS data.

Table E7. 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 $1[q \geq 2013Q3] \times 1[T_v \geq 00{:}00]$ of the specification in Equation 7, where $1[q \geq 2013Q3]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $1[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and zip code income. The samples comprise of traditional Medicare patients who arrive at the ED in a Florida hospital within 1 hour of midnight (11PM-12:59AM; column 1), within 2 hours of midnight (10PM-1:59AM; column 2); within 3 hours of midnight (9PM-2:59AM; column 3); within 4 hours of midnight (8PM-3:59AM; column 4); and within 5 hours of midnight (7PM-4:59AM; column 5).

Table E8. After-Midnight ED Arrival Difference-in-Difference Coefficient, Heterogeneity by Patient Severity

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

* $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses and are clustered by the ED arrival hour and quarter. This table reports the $\beta \times (\text{Risk Decile 1})_v$ coefficient on $1[q \geq 2013Q3] \times 1[T_v \geq 00{:}00]$ of the specification in Equation 7, interacted with an indicator for the predicted risk decile of visit v. $1[q \geq 2013Q3]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $1[T_v \geq 00{:}00]$ is an indicator for whether the ED arrival hour for the visit was after midnight. Patient risk is predicted by estimating a logit using ED visits between 9AM and 3PM of an indicator for being admitted within 30 days of an ED visit on patient demographics, current ED visit information, and information on any prior visits in the last 365 days. Demographics include age-bin, sex, race, Hispanic indicator, point of origin indicator, and mean zip code income. Information on current visit includes hospital and quarter. Information on previous visits includes the number of visits/inpatient stays/length of stay in the last month or last year, as well as any diagnoses and procedures recorded in stays within the last month or last year.

Table E9. Robustness Analysis: Welfare Analysis Parameters

	(1)	(2)
	Model Assumptions	
	Most Conservative	*Least Conservative*
	A. Estimates	
Effect on admissions	2011-2015: est. coeffs after 2015: 0	2011-2015: est. coeffs after 2015: 2015 estimate
Effect on compliance costs	2011-2015: est. coeffs after 2015: 0	2011-2015: est. coeffs after 2015: 0
Payments demanded	2011-2015: est. coeffs after 2015: 0	2011-2015: est. coeffs after 2015: 0
2010 hospital revenue	\$15,029,306	\$15,029,306
2010 hospital compliance costs	\$12,822,887	\$12,822,887
	B. Parameters	
RAC contingency fee	12.5%	9%
Value of public funds	1.3	1.3
Discount rate	2%	2%
Share of demanded pmts refunded	68%	0%

This table lists the parameters and assumptions for "most conservative" and "least conservative" calculations, depicted in Figure D20. Effect on admissions and compliance costs are from Table 3 column 4. Payments demanded are from Figure 10. The 2010 hospital revenue and hospital compliance costs are the median values for hospitals in the sample for Table 3.