ONLINE APPENDIX: FOR ONLINE PUBLICATION ONLY

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

Maggie Shi

University of Chicago Harris School of Public Policy and NBER

A Policy Context

A.1 Medicare Inpatient Prospective Payment System and Short Stays

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

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

From the patient's point of view, it is often difficult to differentiate between an observation stay and a short inpatient stay (Span, 2012). Thus, a hospital's costs for an observation stay are likely similar to the costs for a short inpatient stay. However, hospitals earn much more from Medicare for

admitting a patient for a short inpatient stay rather than for an outpatient observation stay: among DRGs common to both inpatient and observation stays, Medicare payments for inpatient stays were two to three times higher than payments for observation stays (Medicare Payment Advisory Commission, 2015).

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

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

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

A.2 RAC Program Details

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

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

RAC Audit Process RACs conduct postpayment reviews to identify and correct overpayments or underpayments for claims for inpatient care, outpatient care, long-term care, and durable medical equipment in the last three years. Figure H1 illustrates the claims auditing and appeals process, using 2011 inpatient audit rates as an example. Each RAC develops and runs its own proprietary algorithm on claims data to identify claims with potential payment errors. In 2011, RACs' auditing scope for inpatient claims included incorrect or incomplete coding, DRG validation, and medical necessity reviews. Five percent of audits were "automated reviews," which rely solely on claims data to make a determination based on clearly outlined Medicare policies. The rest of the audits were "complex reviews," in which a medical professional (for example, coder, nurse, or therapist) employed by the RAC submits a medical record request and manually reviews all documentation associated with an inpatient stay. It is up to the medical professional to determine

whether an overpayment or underpayment was made. If they find an error, then they can demand that the provider repays Medicare (or vice versa). Providers can appeal demands by first requesting a redetermination by the RAC and then escalating it to higher levels of appeals – for example, by requesting that a separate contractor reconsider the case, requesting a hearing by an administrative law judge, or escalating it to a review by the Medicare Appeals Council.

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

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

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

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

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

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

A.3 Characteristics of Audits and Audited Hospitals

Given Medicare policymakers' focus on short stays as the main source of unnecessary admissions, I examine audit frequency as a function of an admission's length of stay in Figure H13. Admissions with a length of stay of two or fewer days have much higher rates of auditing than longer admissions. The majority of audit recoveries of short stays result in the full payment being reclaimed. I also consider audit frequencies by base DRG group. Section 5 discusses how these groups were

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

ranked by CMS by severity of payemnt errors.

I next consider hospital-level characteristics and their correlation with audit rate in Figure H2. The RAC region a hospital is in is highly correlated with its audit rate. Within each region, rural hospitals, small hospitals, non-profit hospitals, and hospitals with a higher share of short stay Medicare admissions are more likely to be audited. Although almost every hospital was subject to an audit by 2020, in any given year there is a substantial portion of hospitals that do not face any audits. In 2011, 15 percent of hospitals had an audit rate of 0 percent. The share of hospitals with no audits varies across RAC regions from 2 to 23 percent.

B Model

B.1 Model Setup and Predictions

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

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

where:

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

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

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

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

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

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

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

A hospital will adopt technology that reduces signal noise from σ_H^2 to σ_L^2 if the cost to adopt is less than the difference between the expected payoffs with and without technology, denoted as K. If hospitals face a distribution of adoption costs, then as the difference between payoffs increases, more hospitals will adopt technology.

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

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

more technology adoption.

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

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

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

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

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

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

B.2 Medicare's Problem

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

Hospitals face identical patient populations and payoff functions with and without technol-

ogy, but vary uniformly in their fixed cost of technology adoption $K \in [0, \overline{K}]$. Medicare's payoff includes both patient benefit and expenditure. Medicare expenditure has three components: expenditure on hospital admissions, the cost of conducting audits, and the amount recouped as penalties. Medicare values population-level patient benefit, which is the sum of patient benefit across adopting and non-adopting hospitals. Let K^* be the threshold adoption cost as defined in Equation 12. Share $s = \frac{K^*}{\overline{K}}$ of hospitals will adopt technology, choose threshold τ^A and expect to admit q^A patients each. Share $1 - s$ of hospitals will not adopt technology, choose threshold τ^N, and expect to admit q^N patients each. Normalizing the total number of hospitals to 1, let the total admissions in the population be $Q = sq^A + (1-s)q^N$. α_M is the value Medicare places on patient benefit relative to expenditure, and $(1+\lambda)$ is the fiscal externality of raising a dollar of government revenue. Altogether, Medicare's payoff is:

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

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

Should Medicare Purchase the Technology for Hospitals? We can then use this framework to consider an additional choice for Medicare: whether to purchase the technology on behalf of all hospitals. Say that technology adoption increases total patient benefit, but voluntary technology adoption is low and it is expensive to conduct audits. In this case, it may be worthwhile for Medicare to purchase the technology on behalf of hospitals and then require that they use it. This

would capture policies like the HITECH Act, which directly subsidized the adoption of health IT (Burde, 2011).

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

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

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

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

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

C Additional Analyses

C.1 Robustness and Placebo Analysis

Hospital-Level Analysis As a robustness test, in Figure H5 I regress on a hospital's denial rate – the share of claims for which a denial is made after audit – rather than its audit rate. Equation 18

defines the relationship between denial rate and audit rate.

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

Since 41 percent of audits in 2011 resulted in a demand in the main sample and the denial rate is monotonically increasing in audit rate (Figure H6), one would expect that a hospital's response to a one-percentage point increase in the denial rate should be about twice the response to one percentage point increase in the audit rate. Indeed, this is what the results reflect; for example, hospitals reduced admissions by 2.5 percent in 2012 in response to a one-percentage point increase in the 2011 audit rate, and they reduced admissions by 5.7 percent in 2012 in response to a one-percentage point increase in the denial rate.

In Figure H7, I show that the results are robust to the inclusion of control variables. The main specification hospital, year, and neighbor comparison group-year fixed effects. Non-time-varying hospital characteristics are absorbed by the hospital fixed effects, and variables which vary over time nationally or within a local area are absorbed by the year and group-year fixed effects. Thus, the control variables I include are 2010 hospital characteristic-year fixed effects. These controls will capture trends over time across different types of hospitals – for example, if there are divergent trends between non-profit and for-profit hospitals. The control variables consist of the following variables, interacted with a year fixed effect: indicator for above-average 2010 beds, urban status, hospital profit type, teaching status, chain status in 2010, indicator for above-average 2010 short stay share, indicator for above-average 2010 administrative cost share, and indicator for above-average top 20 error MS-DRG share.

In Figure H8, I show that the results are robust to alternative sample definitions. Figure H8a reproduces the event study from the main specification for the outcome of log Medicare admissions, in which the sample is defined as all hospitals within 100 miles of the RAC border and the coefficient is scaled by the correlation between a hospital's audit rate and its leave-one-out state audit rate. This is robust to changing the sample to all hospitals within 50 miles (Figure H8b) or 150

miles (Figure H8c) of the border, although the results are noisier with a shorter distance. One concern with spatial identification strategies is the potential for spillovers to neighboring units. Here, the concern would be about spillovers from high-audit hospitals to low-audit hospitals across the border. On the one hand, if patients were redirected from a hospital near the border in a high-audit rate state to a nearby hospital in a low-audit rate state, then this would bias the coefficients to be larger in magnitude. On the other hand, if hospitals on the low-audit side internalize their high-audit neighbors' audit rates in making their admission decisions, this would bias the coefficients to be smaller in magnitude.

These spillovers should be less of a concern as the distance from the border increases or if the hospitals closest to the border are excluded. Figure H8d shows similar results when restricting the sample to hospitals that are at least 10 miles away from the border, demonstrating that the result is not driven by such spillovers. Finally, Figure H8e shows that the results are similar when restricting the sample to hospitals with audit rates greater than 0 percent, meaning that the results are driven by variation in auditing across hospitals on the intensive, rather than the extensive, margin.

Figure H9 shows that the results are robust to using alternative instruments to scale the reduced form effect. The main specification instruments for a hospital's audit rate using the leave-one-out state audit rate in order to capture the variation in audit intensity that is unrelated to the hospital's own behavior. Figure H9a plots the results of using the state audit rate (which includes the hospital) as an instrument. Figure H9c shows that the results using the leave-one-out RAC region audit rate, rather than the state audit rate, are similar.

While using the leave-one-out audit rate strips away the direct effects of a hospital's own behavior, it still includes other hospitals surrounding a given hospital, whose audit rates may still reflect that hospital's behavior. This can be the case if, for example, a given hospital has a large market share. To address this, in Figures H9b and H9d I consider using the audit rate of other hospitals in the same state or RAC region in *other* markets, which I define using hospital referral regions. This instrument leverages hospitals whose behavior is less likely to be affected by a given hospital's behavior since they are in different markets. Similarly, one might be concerned that a

hospital's audit rate is correlated with the behavior and audit rates of other hospitals in the same hospital system. Figure H9e uses the audit rate of hospitals in the same state but different hospital systems in 2010. The results are robust to using these hospitals to instrument for a hospital's audit rate.

Because neighbor comparison groups can overlap, they can potentially span multiple border segments. Thus, clustering at the border segment-level may not capture the correlated errors across border segments, which would bias the standard errors. Given how the neighbor comparison groups are defined, there is no way to set border segments that eliminates this problem. However, it should be less of a concern for longer border segments, as hospitals in the same neighbor comparison group would now be less likely to point to different border segments. Figure H17 plots the event studies from using 50- and 150-mile border segments. While the the standard errors do increase as the segments become longer, the coefficients remain statistically significant.

To confirm that the results are not driven by a single state or hospital comparison group, Figure H18 plots the distribution of coefficients when one state or one hospital comparison group is removed from the sample at a time. The coefficients are always negative and the distribution is centered around the main effect.

Finally, I consider a falsification test using state borders in the *interior* of each RAC region. In the interior of each region, there is no change in RAC identity at state borders, so comparing hospitals across these interior borders does not capture exogenous variation driven by different audit strategies across RACs. Figure H19a illustrates the interior borders and the sample of hospitals within one hundred miles of the interior border (excluding hospitals that are within one hundred miles of the RAC border). I also restrict the falsification test just to interior borders between different *MAC* regions. As noted in Section 4.2, the fact that some Part B MAC region borders overlap with RAC region borders could make changes in MAC activity a time-varying confounder. Both of the falsification tests show no effect on admissions on the "high-audit side" of the interior borders (Figures H19b and H19c).

Patient-Level Analysis In Table GIX, I show that the Two Midnight rule difference-in-

difference results are robust to varying the sample to include patients who arrive between one and five hours of midnight. Table GIV shows that, in addition to a null effect on revisits within thirty days, there is no effect on revisits within sixty or ninety days.

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

C.2 Coding Response

To assess the effects of auditing on upcoding, I run the main specification on the log number of diagnoses per claim (ICD-9/ICD-10 diagnosis codes) in Figure H20. The number of reported diagnoses decreases at hospitals subject to more audits. The main results indicate that the remaining admissions tended to have longer lengths of stay, and are presumably sicker, than deterred admissions. Thus if hospitals didn't change how they coded, we would expect them to have more diagnoses per patient. These results suggest that in addition to changing utilization patterns, RAC audits led to less upcoding.

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

C.3 Rural Hospital Closures

The main results show that RAC audits decrease hospital revenue and increase their costs. This raises the concern that RAC auditing may have driven hospitals into financial distress and, given the prevalence of hospital closures in recent years, led them to close. Hospital closures are associated

with decreases in access to care and increases in patient mortality (Carroll, 2019; Gujral and Basu, 2019). To study whether RAC auditing led to hospital closures, I use data from the Sheps Center for Health Services Research on rural hospital closures between 2005 and 2022.[39] I adapt the main specification for the hospital-level analysis to study rural hospital closures. In the border hospital sample, no hospitals closed before 2012 – this is by definition, since the hospital had to be open in 2011 to be audited. Therefore there is no variation in the pre-2010 period to use a difference-in-differences framework. I run the following specification separately for each year Y in the post period:

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

which regresses a dummy for whether a rural hospital has closed in year Y, $Close_h^Y$, on its (instrumented) audit rate X_h^{2011}, after taking into account the hospital's neighbor comparison group. Figure H21 plots the β^Y coefficients for years where there is variation in closures among rural hospitals in the border sample (i.e., excluding 2012, 2017, and 2021). The results indicate that higher RAC auditing did not cause hospital closures.

D Extrapolation to Overall Hospital Sample

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

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

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

Under these assumptions, I can calculate the extrapolated savings by multiplying the 2011-2015 event study coefficients on Medicare inpatient revenue (Figure 5b) and payments demanded (Figure H12) by each hospital's 2011 audit rate. Since the estimates are based on the logarithm of inpatient revenue and represent a percent change relative to the baseline in 2010, I multiple these coefficients by the hospital's 2010 inpatient revenue. Figure H22 plots the extrapolated savings for each hospital-year, compared to the actual changes in Medicare inpatient revenue and actual payments demanded. For both types of savings, the extrapolated and actual savings are positively correlated. This indicates that in the overall sample, hospitals subject to higher audit rates reduced their Medicare inpatient revenue more and were subject to more audit demands in subsequent years. Summing up the extrapolated savings across all hospitals from 2011 to 2015 implies that the RAC program saved the Medicare program $9.28 billion between 2011 and 2015, compared to the actual $11.74 billion reduction in inpatient spending and savings from audit recoveries in this period. Note, however, the relatively low R^2 from the regression between extrapolated and actual savings, indicating that much of the variation in savings is not explained by variation in 2011 audit rate.

E Hospital-level Emergency Department Visit Analysis

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

potential reliability concerns with these measures and then discuss the results.

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

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

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

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

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

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

Figure H16 shows the event studies from Equation 3 on the ED visit outcomes. The results from this analysis are largely consistent with the patient-level results. Among hospitals with a higher 2011 audit rate, the share of ED visits with outpatient observation services increases after 2011. However, ED visits at these hospitals do not seem to result in greater revisits or mortality. There is a small and statistically insignificant increase in revisit rates after 2012, but its magnitude (0.25%) is very small relative to the 2010 mean, 15%.

F Appendix Tables

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

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

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

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

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

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

Table GIII. ED Arrival Hour Manipulation Tests

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

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

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

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

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

Table GV. Across-Hospital Post-2011 Coefficient

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

* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors in parentheses and are clustered at the state and border segment level. This table reports the coefficients of the reduced form event study in Equation 7, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents the effect of a one percentage point increase in 2011 audit rate on a hospital-level outcome in 2011-2015. Columns 1-2 report two stage least squares outcomes for the number of and revenue from Medicare admissions overall, columns 3-4 report outcomes for the number of and revenue from Medicare admissions with length of stay ≤ 2, column 5 reports the outcomes for log net administration costs, and column 6 reports the outcomes for an indicator for installation of medical necessity software. Length of stay is counted as the difference in days between the admission and discharge date. Inpatient revenue is the sum of all Medicare inpatient payments. Net administration costs are salary and other costs in the "Administrative and General" category in HCRIS, net of reclassifications and adjustments. Indicator for installing software is equal to 1 if a hospital reports the status of a medical necessity software as "contracted/not yet installed," "installation in process," and "to be replaced" in HIMSS. The sample comprises hospitals within a hundred miles of the RAC border with at least 1 hospital in their neighbor comparison group. Data: MEDPAR, CMS audit data, HCRIS, and HIMSS.

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

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

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

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

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

* $p < .10$, ** $p < .05$,*** $p < .01$. Standard errors are in parentheses and are clustered at the ED arrival hour and quarter level. This table reports the β coefficient for different patient subsets on $\mathbb{1}[y \geq \text{2013Q3}] \times \mathbb{1}[T_v \geq \text{00:00}]$ of the specification in Equation 9, where $\mathbb{1}[y \geq \text{2013Q3}]$ is an indicator for whether the visit occurred after the Two Midnights rule was implemented in 2013Q3, and $\mathbb{1}[T_v \geq \text{00:00}]$ is an indicator for whether the ED arrival hour for the visit was after midnight. "Inpatient" is an indicator for whether the patient was eventually admitted as inpatient from the ED. "Revisit within 30 days" is an indicator for whether the patient had another ED visit or inpatient stay within 30 days of the ED visit. The sample consists of Medicare patients who arrived in the ED within 3 hours of midnight in a Florida hospital. Column 1 comprises the subset of patients in the top quartile of age, column 2 comprises patients in the top quartile of numbers of chronic conditions, column 3 comprises non-white patients, and column 4 comprises patients living in zip codes with the lowest quartile income. Regression includes hospital, hospital-quarter, hospital-hour fixed effects, and controls for age-sex bin, race, Hispanic indicator, point of origin indicator, last ED visit within 30 days indicator, number of chronic conditions, and quartile of mean zip code income. Data: HCUP SID/SIDD.

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

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

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

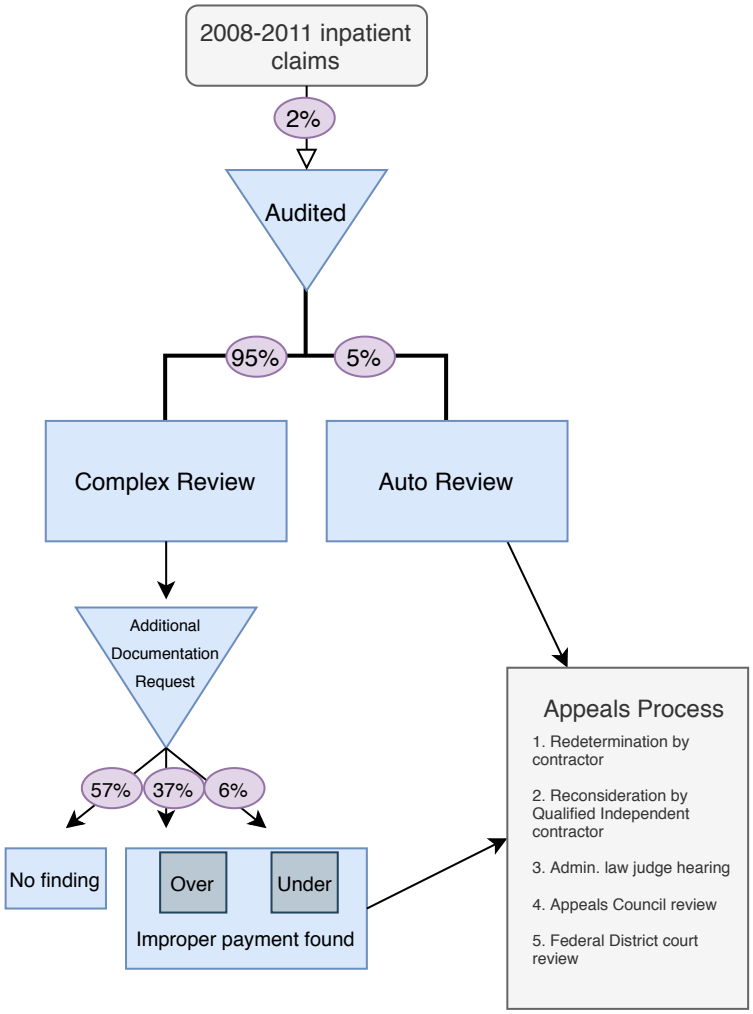
Table GIX. Robustness Test: Sample of Patients by ED Arrival Relative to Midnight

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

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

G Appendix Figures

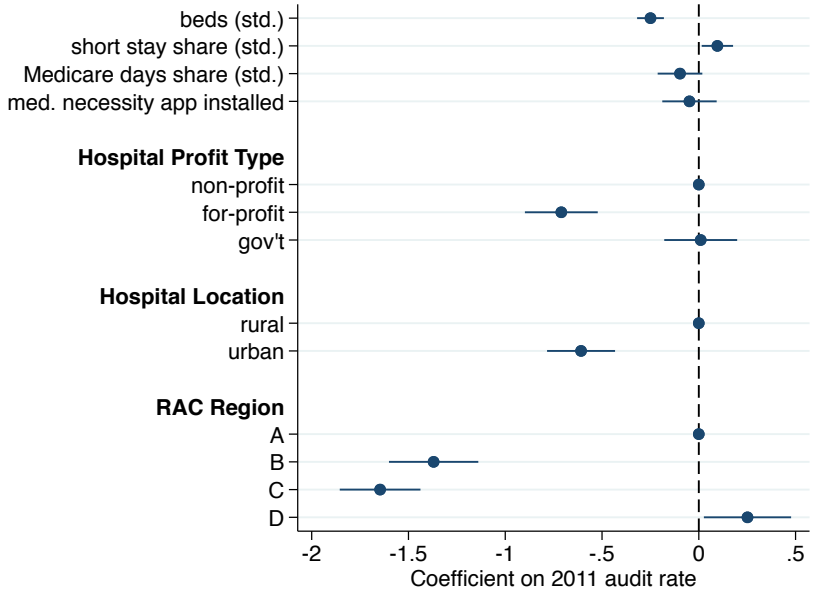
Figure H1. RAC Inpatient Claims Auditing and Appeals Process, 2011 Audits



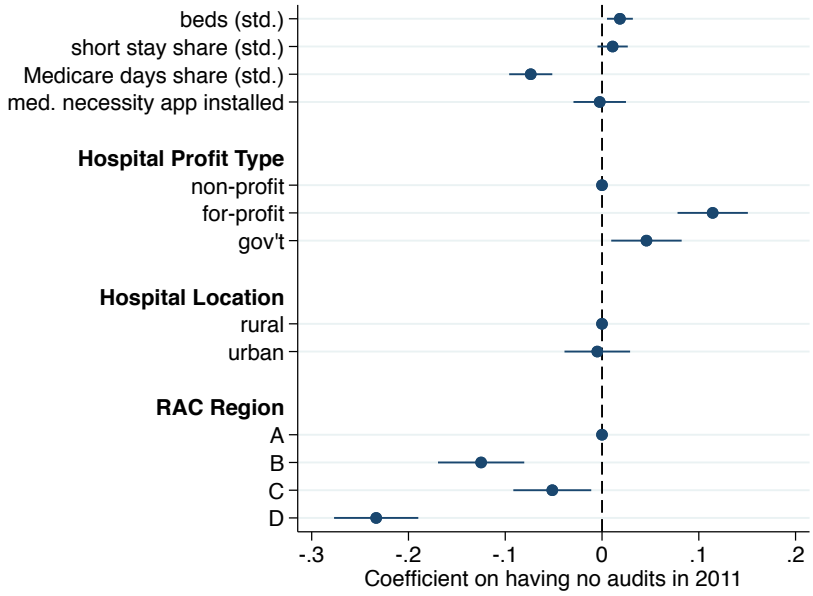
This figure illustrates the stages of the claims auditing and appeals process. The percentages in ovals denote the percent of claims that, conditional on reaching a given stage in the process, reach the next stage. The percentages are calculated based on audits in 2011 of inpatient claims between 2008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

Figure H2. Correlation between Hospital Characteristics on 2011 Audit Rate and No Audit

(a) Outcome: 2011 hospital audit rate

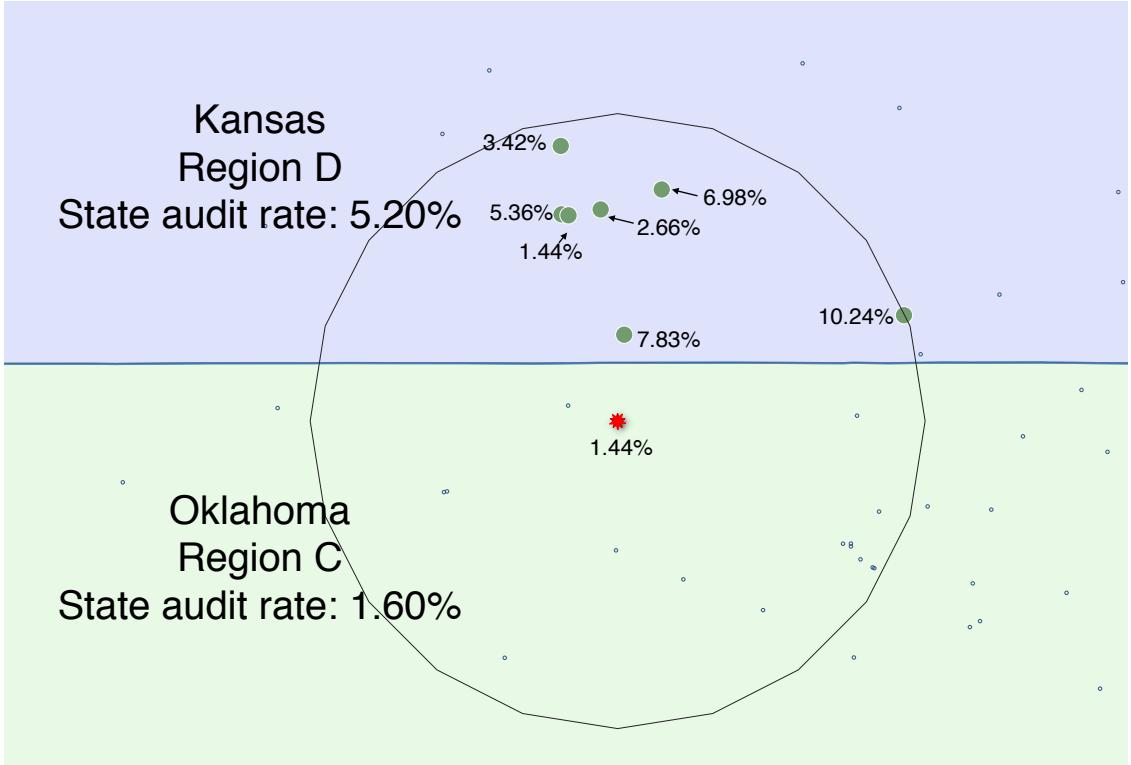


(b) Outcome: no audits at hospital in 2011



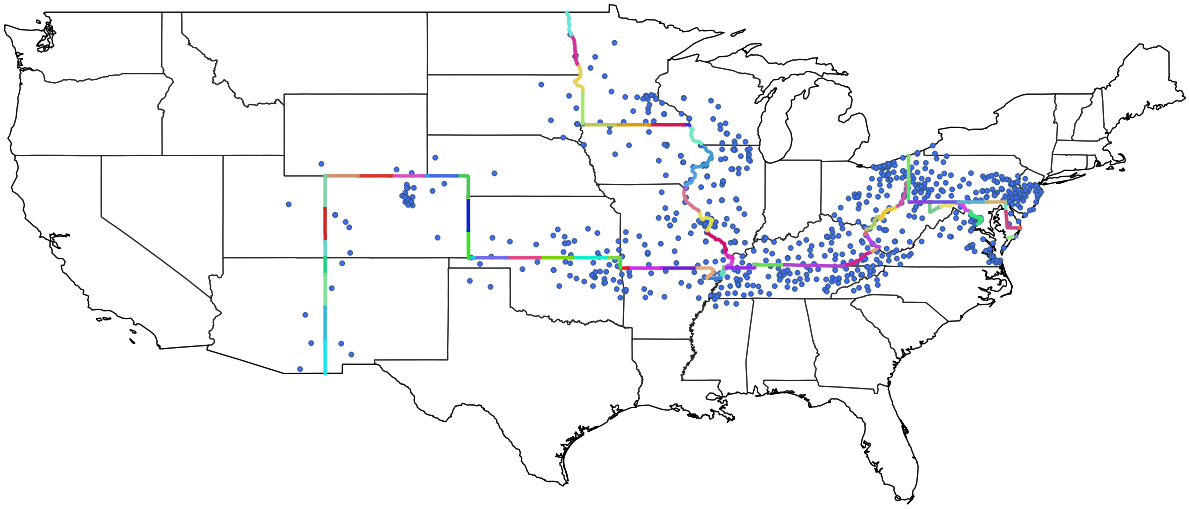
These figures plot coefficients from a regression of (a) a hospitals 2011 audit rate and (b) an indicator variable for whether a hospital was not audited in 2011 on 2010 hospital characteristics. Short stay share is the share of 2010 Medicare admissions with lengths of stay 0-2. Medicare days share is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

Figure H3. Example of Border Hospital and Neighbor Comparison Group Definition



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

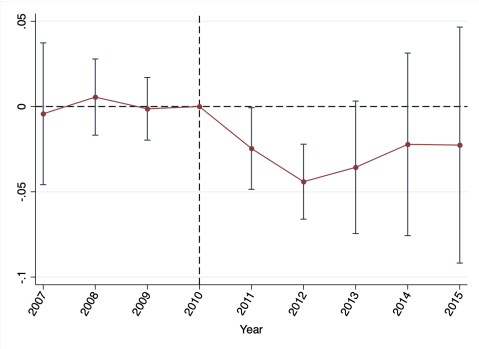
Figure H4. RAC Border Segments and Hospitals Within 100 Miles



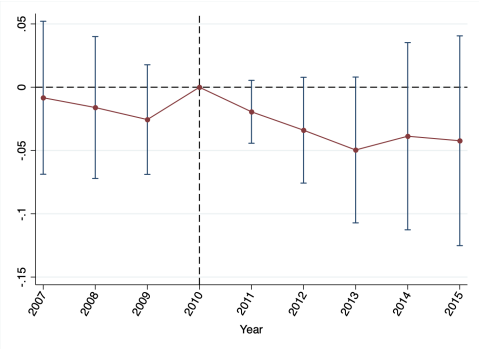
This figure shows how the RAC border is divided into one hundred mile segments that do not cross state borders, and all hospitals within one hundred miles of the RAC border. These border segments are used for clustering in Equation 3.

Figure H5. Event Studies on Effect of 2011 Denial Rate on Medicare Admissions and Revenue, and Administrative Burden

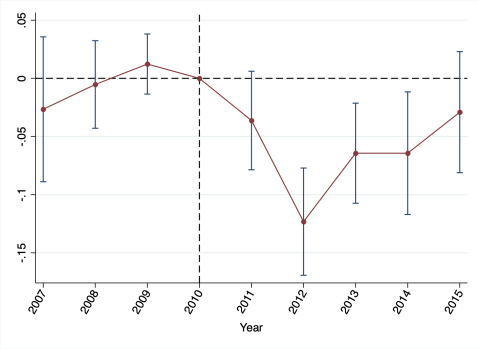
(a) Log Medicare admissions



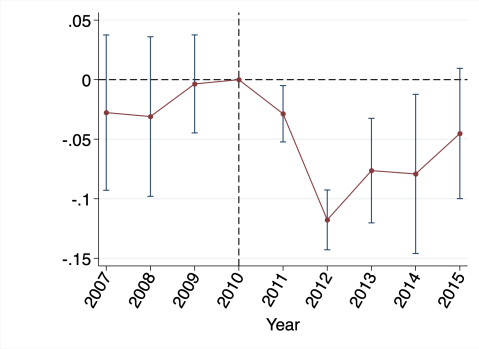
(b) Log Medicare inpatient revenue



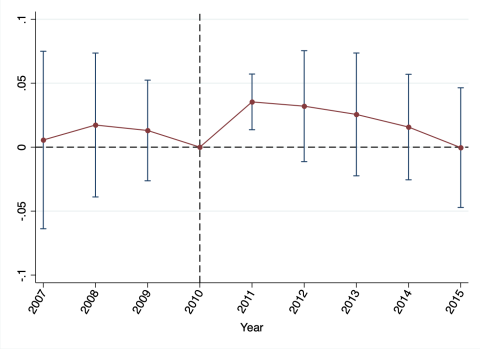
(c) Log Medicare admissions, LOS ≤ 2



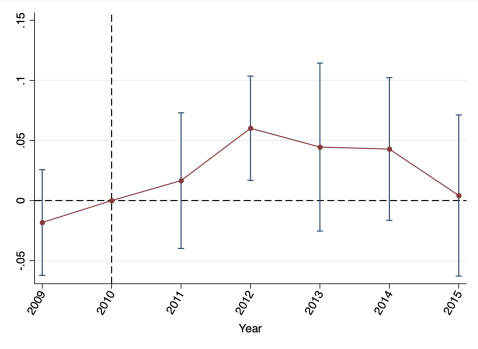
(d) Log Medicare inpatient revenue, LOS ≤ 2



(e) Log net administration costs

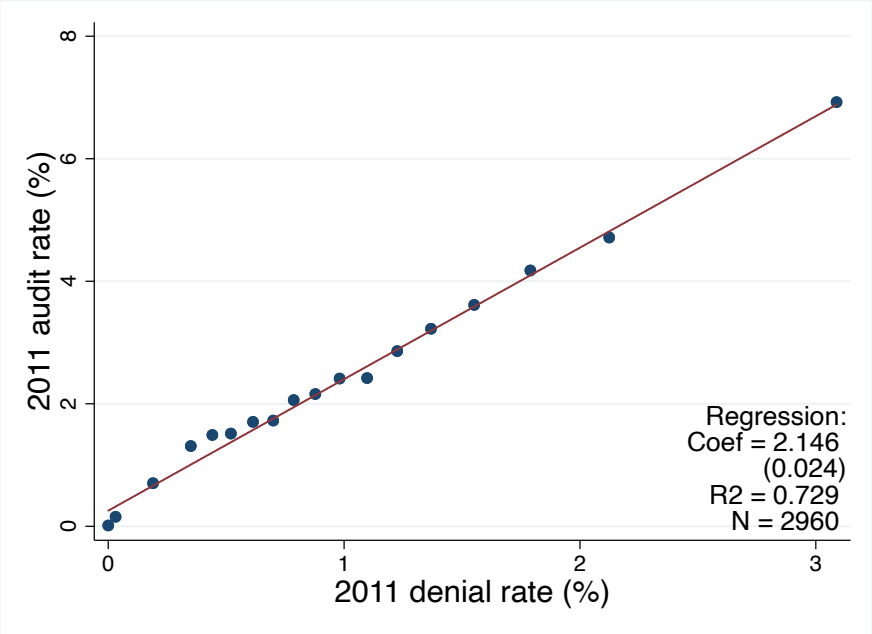


(f) Indicator for installing medical necessity software



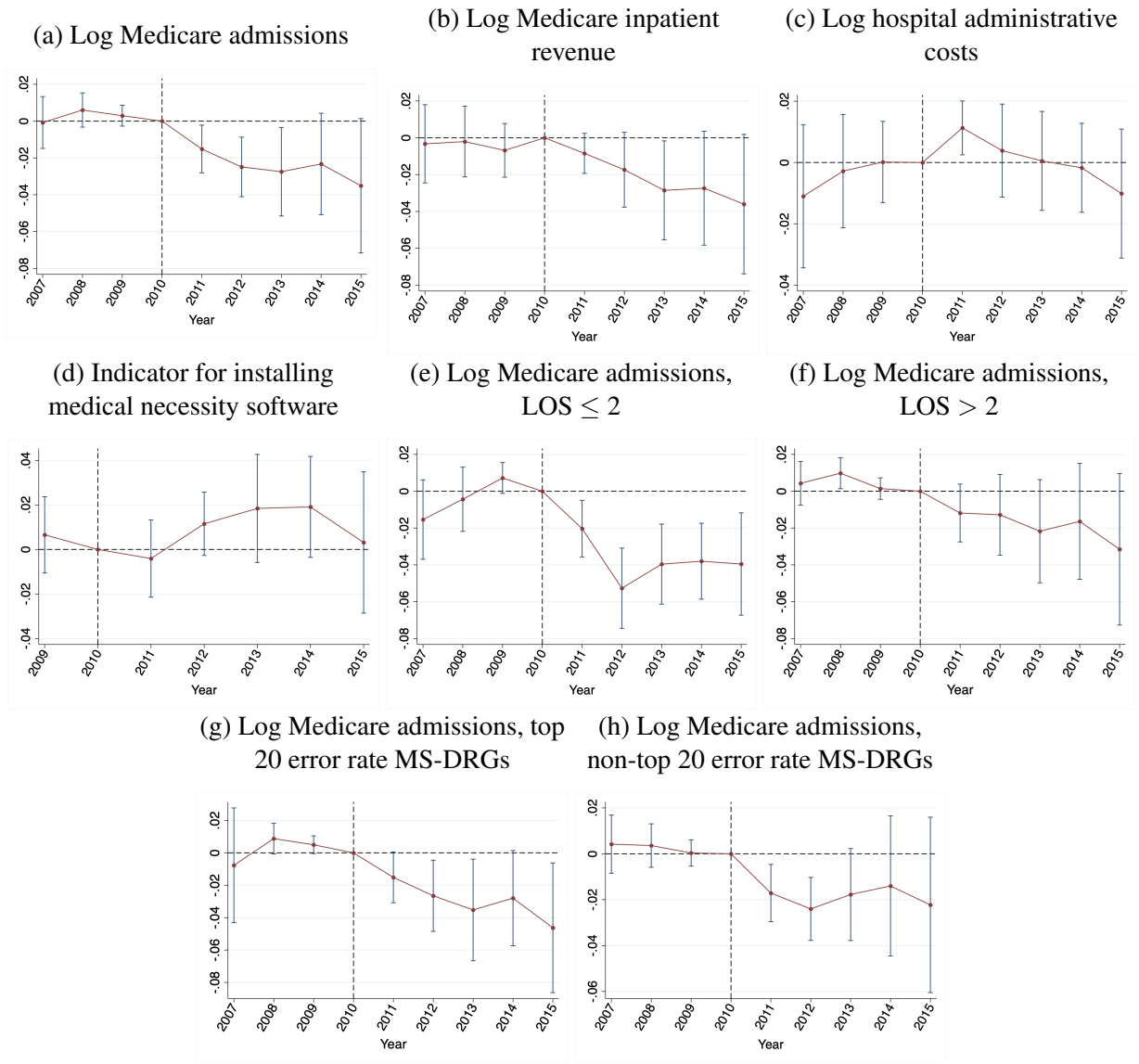
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5 using the denial rate rather than the audit rate, scaled by the correlation between the leave-one-out 2011 denial rate and the actual 2011 denial rate in the weighted border hospital sample. Denial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

Figure H6. 2011 Audit Rate vs. 2011 Denial Rate



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

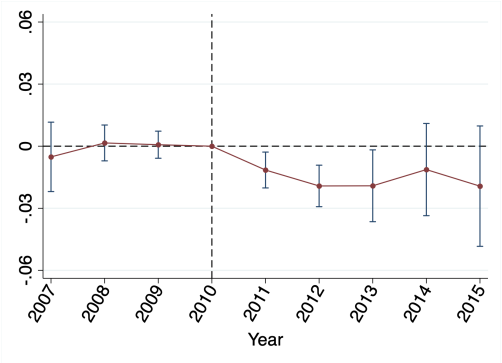
Figure H7. Event Studies on Effect of 2011 Audit Rate on Hospital Outcomes, including Controls



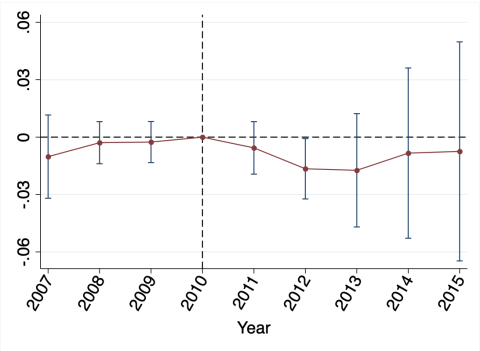
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5 including control variables, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010.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

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

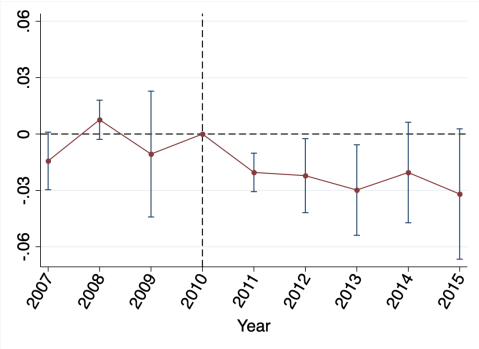
(a) Sample: within 100 miles (main)



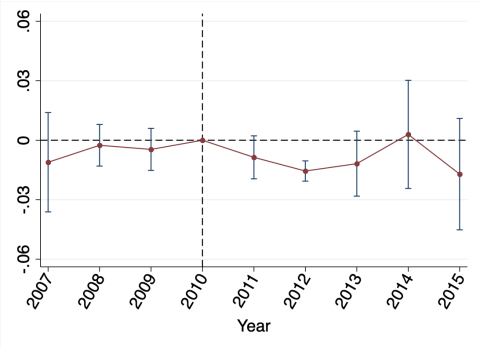
(b) Sample: within 50 miles



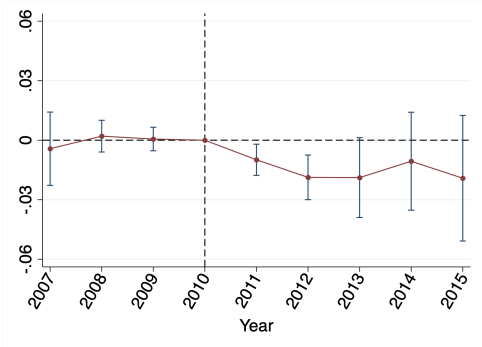
(c) Sample: within 150 miles



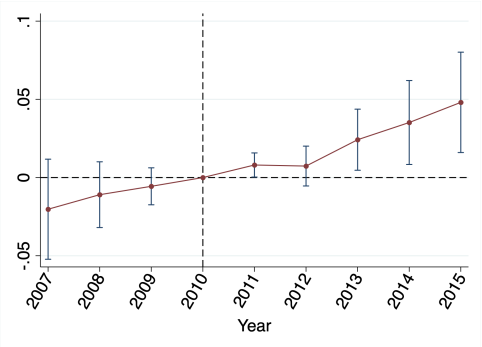
(d) Sample: w/in 100 miles, outside 10



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



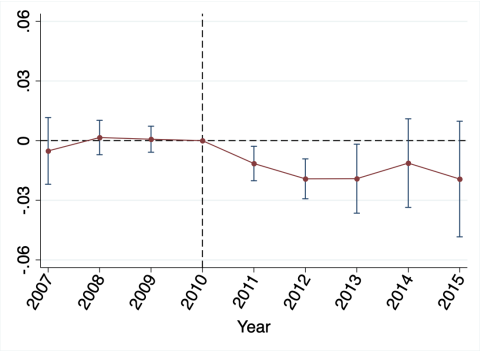
(f) Sample: non-border, all hospitals



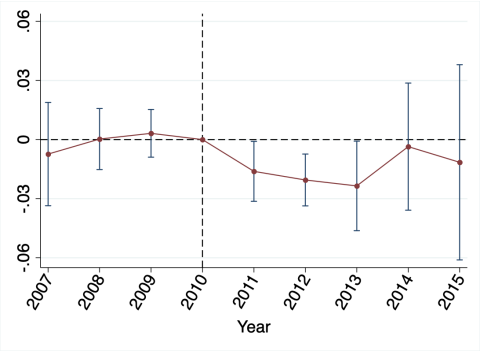
This figure plots robustness analysis event studies of the scaled reduced form coefficients and 95% confidence intervals of the specification in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted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

Figure H9. Robustness to Instrument Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

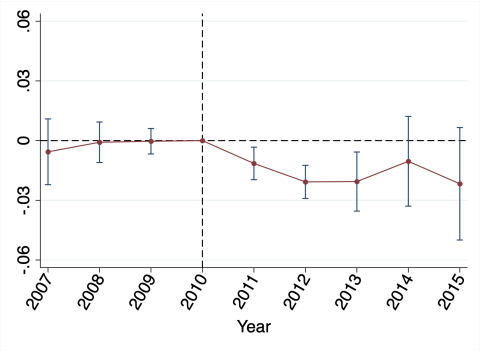
(a) Instrument: state audit rate



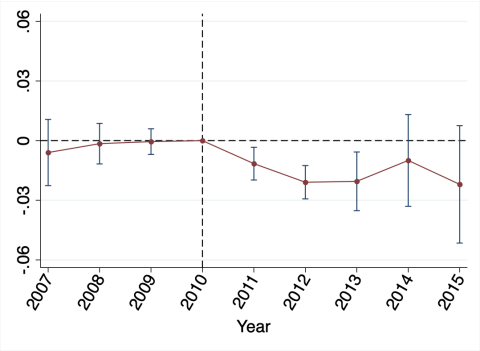
(b) Instrument: state rate, outside HRR



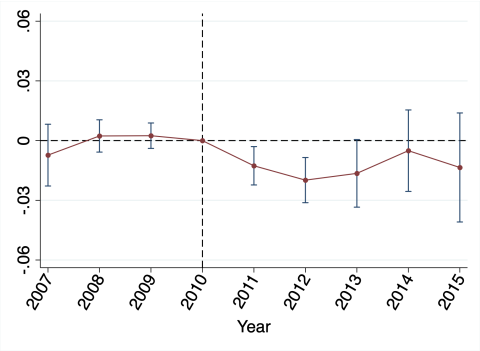
(c) Instrument: leave-one-out RAC region rate



(d) Instrument: RAC rate, outside HRR

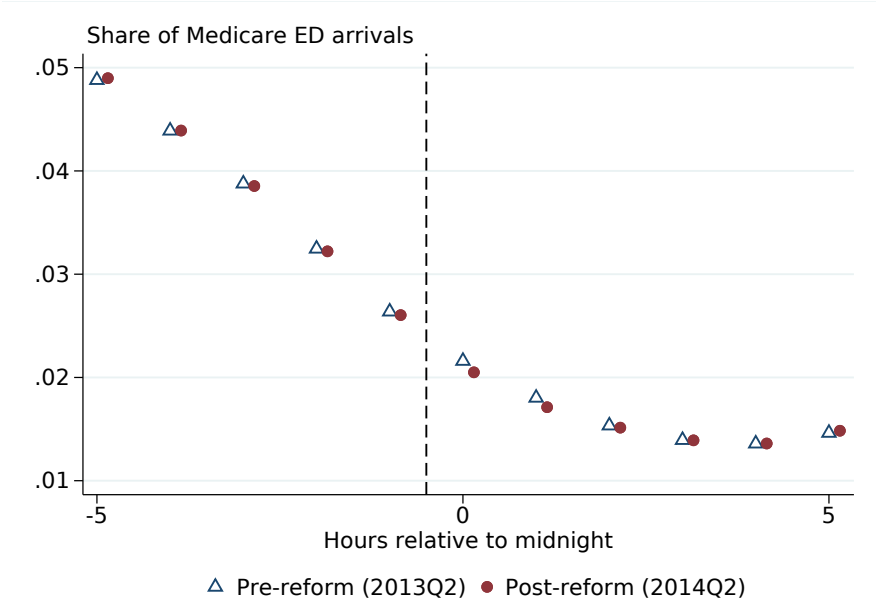


(e) Instrument: state rate, outside system



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

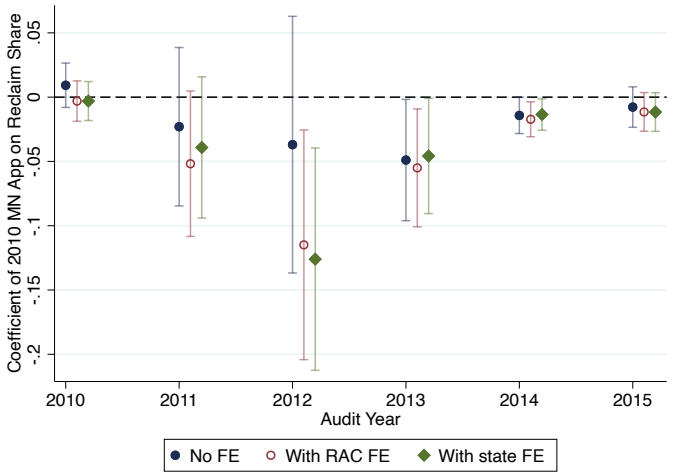
Figure H10. Share of Medicare ED Patients By Hour of ED Arrival



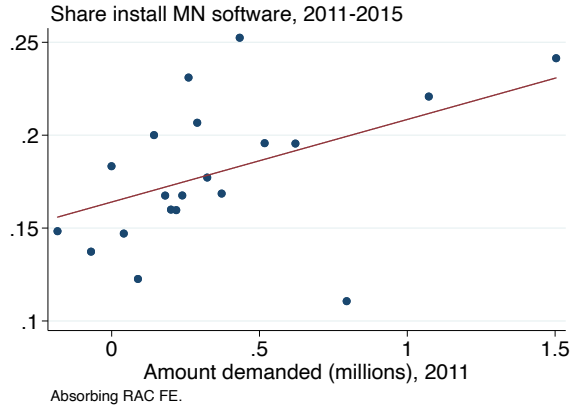
This figure plots the share of Medicare patients that arrive at the ED at each hour (relative to midnight) pre- and post-reform, among traditional Medicare patients who arrived in the ED within 5 hours of midnight in Florida. Data: HCUP SID/SEDD.

Figure H11. Cross-sectional Correlations with Medical Necessity Checking Software

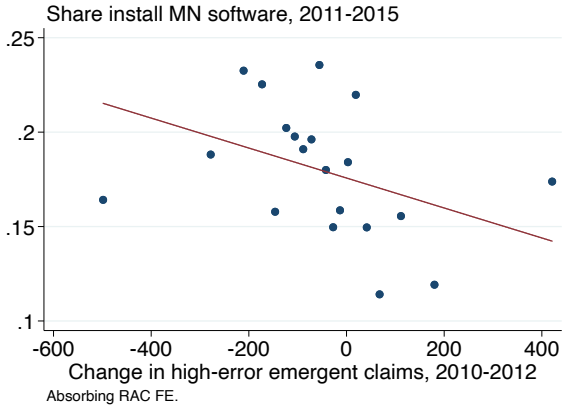
(a) Denial rate and MN software already installed in 2010



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

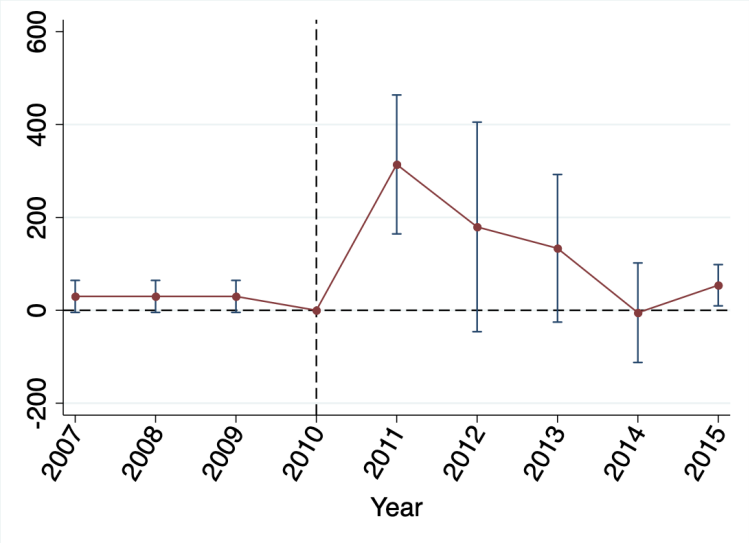


(c) Binscatter of 2011-12 emergent stay change & share install MN software, 2011-15



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

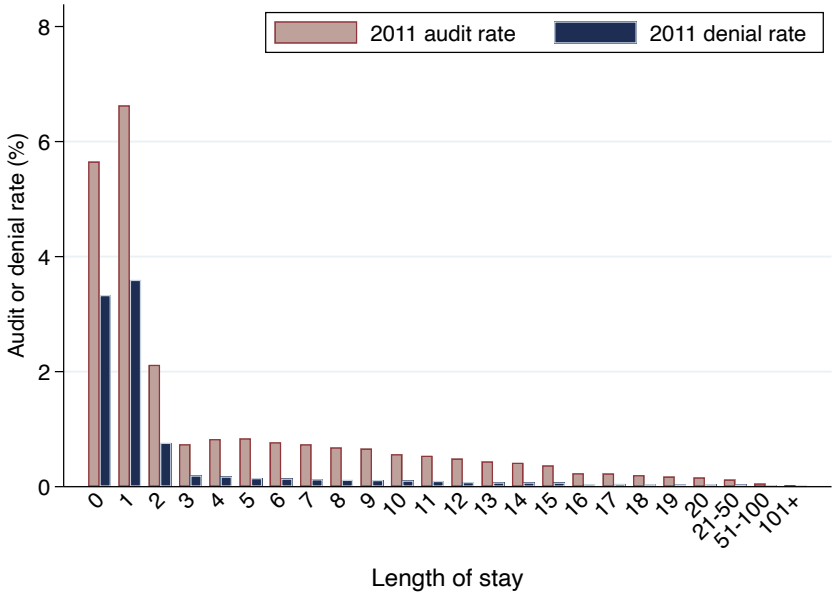
Figure H12. Event Study on Effect of 2011 Audit Rate on Payment Demanded ($1000s) from RAC Audits



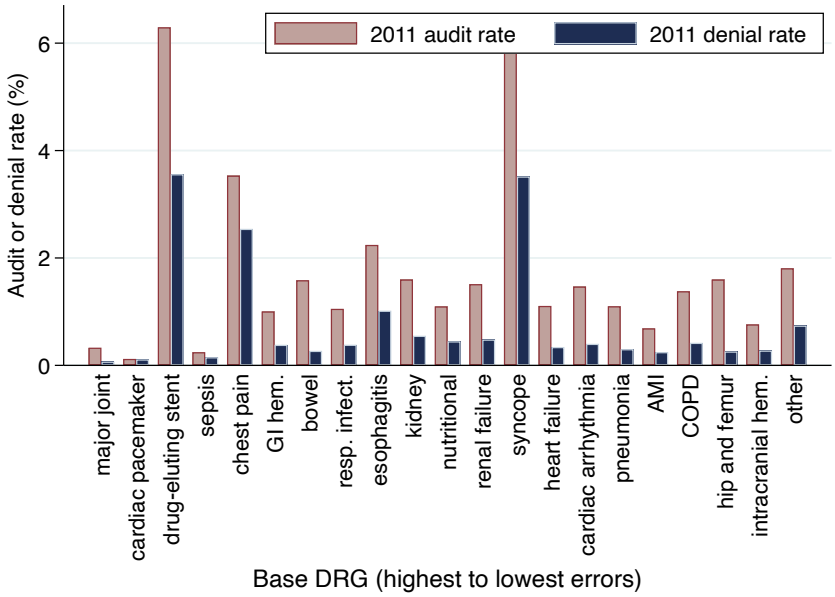
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents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

Figure H13. 2011 Audit and Denial Rates by Stay Characteristics

(a) By Length of Stay



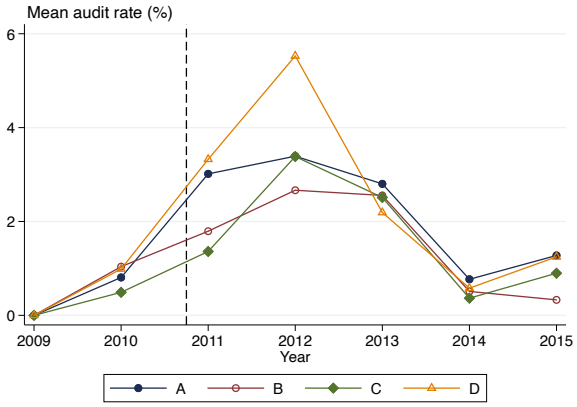
(b) By Base Diagnosis Related Group



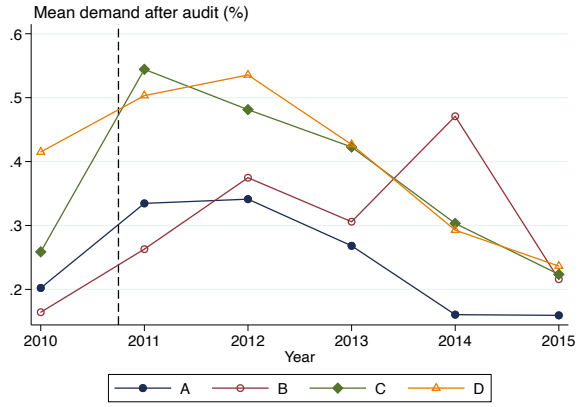
This figure plots the count of 2011 audit and denial rates by (a) an admission's length of stay and (b) its base DRG. Panel (b) shows the top 20 base DRGs with highest improper payments identified in the 2010 CERT report, in descending order of estimated improper payments (Centers for Medicare and Medicaid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

Figure H14. Audit Outcomes and Admissions by RAC Region and Year

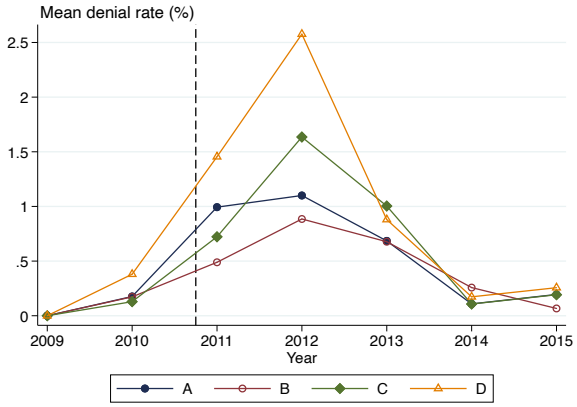
(a) Mean Audit Rate



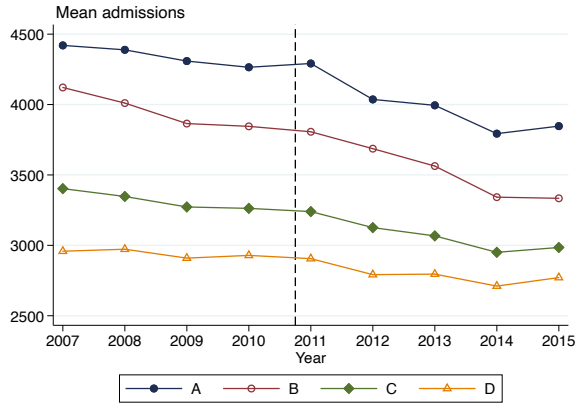
(b) Mean Demand After Audit Rate



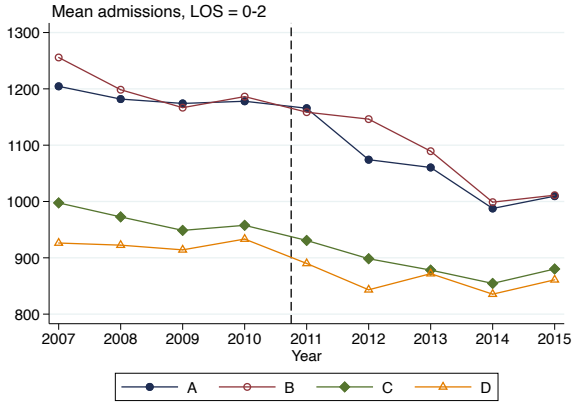
(c) Mean Denial Rate



(d) Mean Admissions

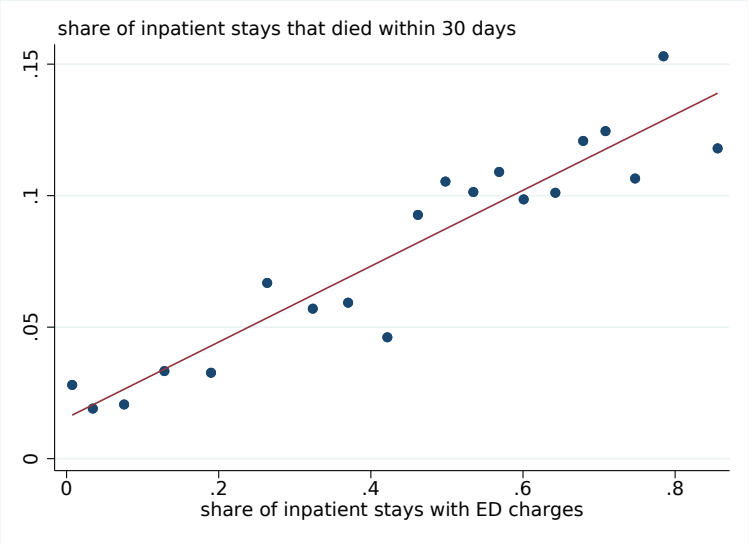


(e) Mean Admissions (LOS = 0-2)



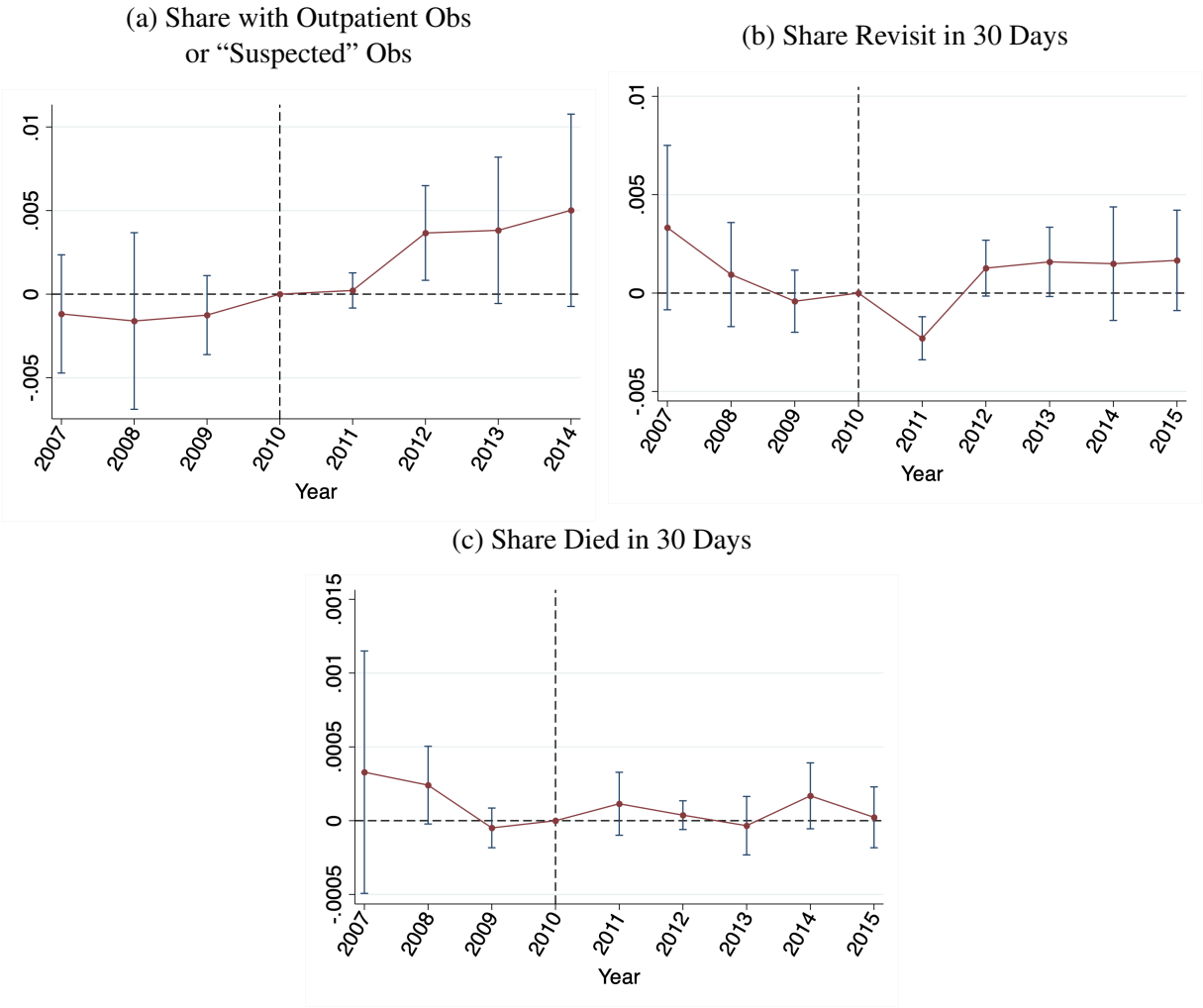
These figures plot time series by RAC region of the (a) mean hospital audit rate, (b) demand after audit rate, (c) denial rate, (d) mean number of admissions, and (e) mean number of admissions with LOS ≤ 2. Denial and demand after audit rate are defined as follows: $Denial\ Rate_{ht} = P(Audit)_{ht}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

Figure H15. DRG ED Visit Share vs. Died Share



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

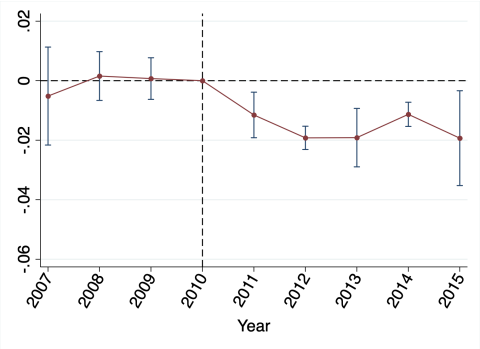
Figure H16. Event Studies on Effect of 2011 Audit Rate on Medicare ED Visits



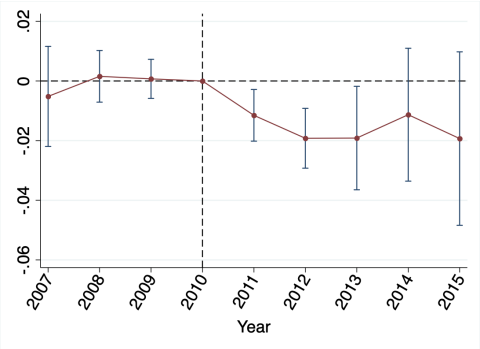
This figure plots event studies of the reduced form coefficients and 95% confidence interval in Equation 5, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient represents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

Figure H17. Robustness to Border Segment Definition: Event Studies on Effect of 2011 Audit Rate on Log Medicare Admissions

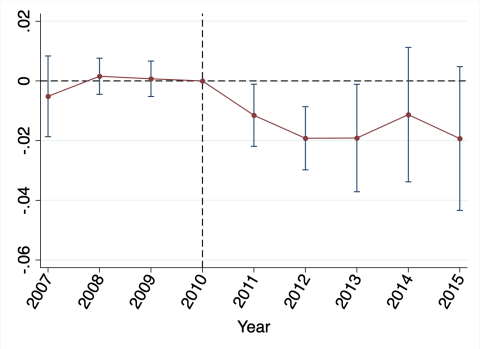
(a) 50-mile Border Segments



(b) 100-mile Border Segments (main)



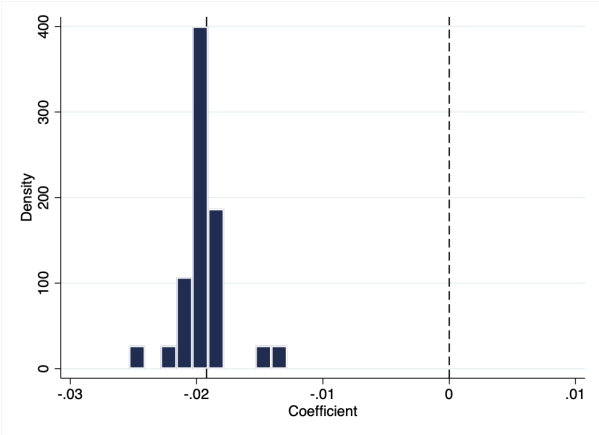
(c) 150-mile Border Segments



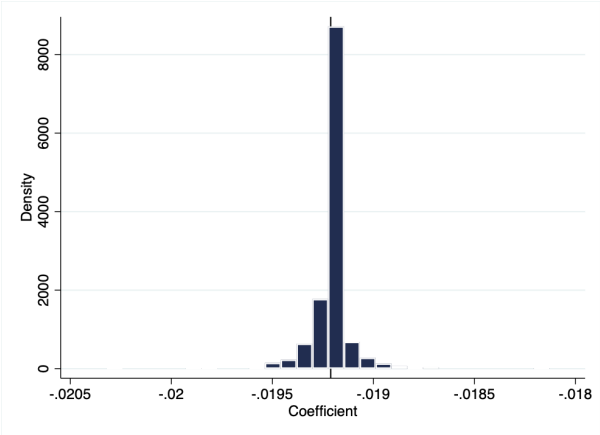
This figure plots event studies of the reduced form coefficients and 95% confidence intervals of the specification in Equation 5, scaled by the correlations between the instruments and the actual 2011 audit rate in the weighted border hospital sample. The omitted year is 2010. Each coefficient estimates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

Figure H18. Robustness Test: Leave-One-Out Coefficients of 2012 Effect of 2011 Audit Rate on Log Medicare Admissions

(a) Leaving 1 state out at a time



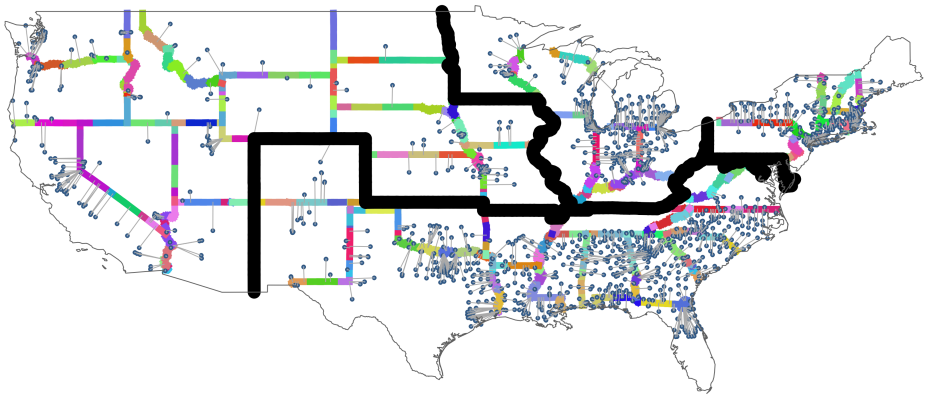
(b) Leaving 1 comparison group out at a time



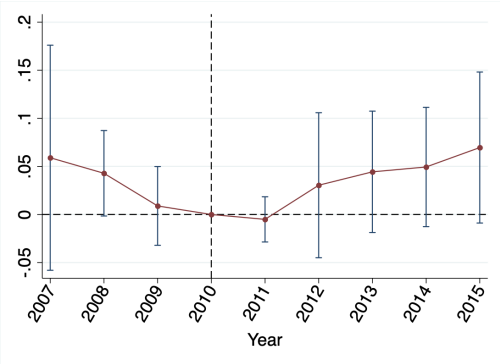
This figure plots distributions of the 2012 coefficient of the reduced form event study specification in Equation 5 on log Medicare admissions, scaled by the correlation between the leave-one-out 2011 audit rate and the actual 2011 audit rate in the weighted border hospital sample the outcome. Panel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

Figure H19. Falsification Test: Interior State Borders

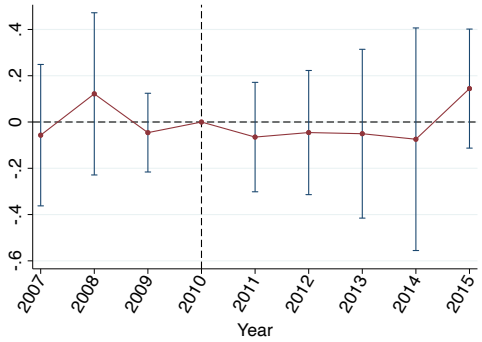
(a) Falsification Test Border Segments and Hospitals Within 100 Miles



(b) Event Study on All Interior Borders



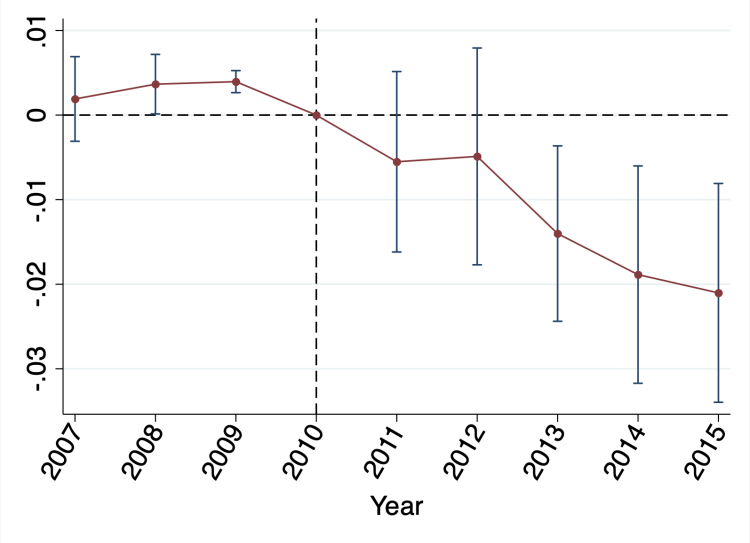
(c) Event Study on MAC Interior Borders



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

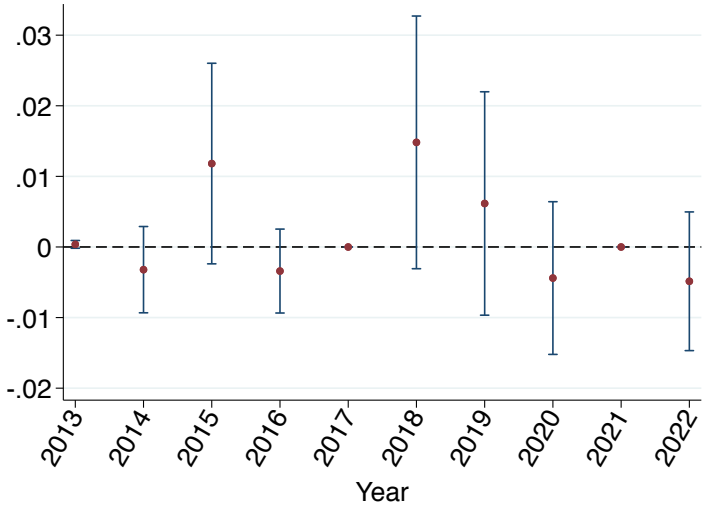
Figure H20. Event Studies of Effect of 2011 Audit Rate on Coding

(a) Log Mean Diagnoses per Admission



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

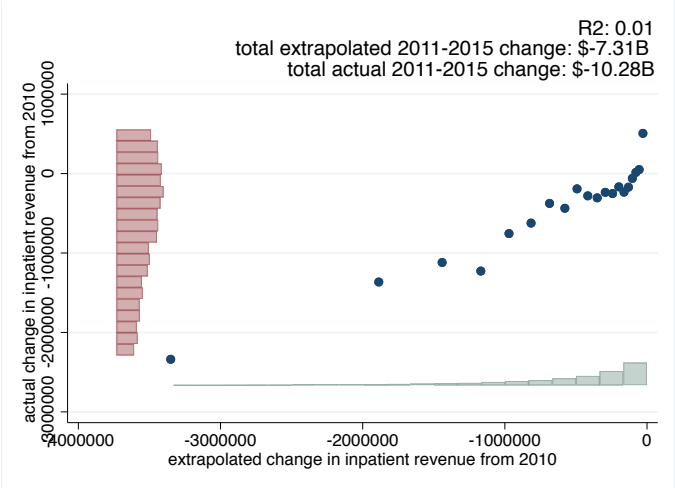
Figure H21. Coefficients of Effect of 2011 Audit Rate on Rural Hospital Closure in a Given Year



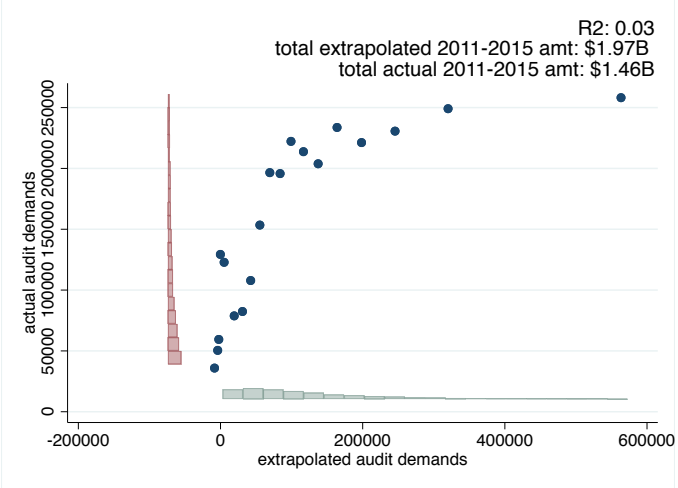
This figure plots the coefficients from individual regressions of the instrumented 2011 audit rate on a dummy for whether a hospital closed in a given year, for rural hospitals in the border sample. There are no closures prior to 2013 and no closures in 2017 and 2021 in the border hospital sample. Data: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

Figure H22. Extrapolation Exercise: Actual vs. Extrapolated Savings

(a) Savings from changes in Medicare inpatient spending



(b) Savings from audit demands



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