Evidence of Upcoding in Pay-for-Performance Programs

Hamsa Bastani
Stanford University Department of Electrical Engineering, hsridhar@stanford.edu

Joel Goh
Harvard Business School, jgoh@hbs.edu

Mohsen Bayati
Stanford University Graduate School of Business, bayati@stanford.edu

Recent Medicare legislation has been directed at improving patient care quality by financially penalizing providers for hospital-acquired infections (HAIs). However, Medicare cannot directly monitor HAI rates, and instead relies on providers accurately self-reporting HAIs in claims data to correctly assess penalties. Consequently, the incentives for providers to improve service quality may disappear if providers *upcode*, i.e. mis-report HAIs (possibly unintentionally) in a manner that increases reimbursement or avoids financial penalties. Identifying upcoding in claims data is challenging due to unobservable confounders such as patient risk. Our approach leverages state-level variations in adverse event reporting regulations and instrumental variable techniques to discover contradictions in HAI and present-on-admission (POA) infection reporting rates that are strongly suggestive of upcoding. We conservatively estimate that over 10,000 out of nearly 60,000 annual reimbursed claims for POA infections (18.5%) were upcoded HAIs, resulting in an added cost burden of $200 million to Medicare. Our findings suggest that self-reported quality metrics are unreliable and thus, recent legislation may result in unintended consequences. In particular, contrary to widely-held beliefs, increasing financial penalties may not reduce HAI incidence and may even exacerbate the problem. We make several policy recommendations based on our results, including a new measure for targeted HAI auditing and suggestions for effective adverse event reporting systems.

Key words: Medicare, pay-for-performance, upcoding, asymmetric information, quality control and detection, hospital-acquired infections, strategic behavior

1. Introduction

The United States is one of the highest per-capita healthcare spenders in the world, surpassing annual expenditures of $2.5 trillion (Schoen 2013, Martin 2012). Yet, there are serious concerns about the quality of care, particularly due to the prevalence of medical errors (Green 2012). Recent Medicare legislation has aimed to improve patient outcomes and reduce costs through the gradual introduction of pay-for-performance policies, which create financial penalties for providers based on quality of care. In principle, these penalties would incentivize providers to modify their operations and improve their quality of service. Since Medicare cannot directly monitor patient outcomes, pay-for-performance policies rely on providers self-reporting accurate information in order to correctly

assess penalties. However, this information asymmetry creates financial incentives for providers to *upcode*, i.e., bias their claims (possibly unintentionally) towards collecting greater reimbursement (Silverman and Skinner 2004). Prior literature on contract design (Fuloria and Zenios 2001) suggests that such distortion of reported information may cause pay-for-performance contracts to fail since they reduce providers' incentives to improve quality of care. Anecdotal evidence suggests that such upcoding may occur frequently in practice (Himmelstein and Woolhandler 2015).

In this paper, we empirically study upcoding in response to Medicare's efforts at reducing hospital-acquired infection (HAI) rates, as well as its economic and policy implications.

Background. HAIs are infections developed by patients as a consequence of medical treatment in a hospital. On any given day, about 1 in 25 hospital patients in the US has at least one HAI, leading to the deaths of an estimated 75,000 hospital patients a year (Magill 2014). In addition, official estimates by the Centers for Disease Prevention and Control (CDC) estimate the direct economic cost of HAIs to be between $28 to $34 billion annually (Scott 2009).

Evidence has shown that most HAIs are preventable through the use of better clinical practices (see, e.g., Berenholtz 2004, Berriel-Cass 2006). However, until recently, Medicare's fee-for-service model reimbursed healthcare providers for these infections regardless of whether or not they were due to an avoidable lapse in the provider's quality of care. Furthermore, Hsu (2014) found that providers could increase their margins over eight-fold for a given ICU patient if he or she incurred a HAI, since the patient would require an extended stay and more services. This created perverse incentives for providers to *increase* HAI rates.

This issue was addressed by the Centers of Medicare & Medicaid Services (CMS) through the hospital-acquired condition (HAC) nonpayment policy (starting on October 1, 2008), which aimed in part to incentivize providers to invest in reducing HAI incidence (a subset of HACs) by placing the financial burden of treating HAIs on the provider rather than on Medicare. The policy targeted selected high cost and high volume HAIs that were considered to be reasonably preventable through better healthcare practices. When providers submitted reimbursement claims diagnosing patients with one or more of these infections, they could indicate whether the infection was present-on-admission (POA) or not. An infection qualifies as POA if the provider detected it within a certain time window of the patient's hospital admission; we will refer to this as a *POA time window*, which is typically 48 hours (Meddings et al. 2010). If the infection was not POA, it was deemed a preventable HAI and would not be reimbursed, causing a large financial loss to the provider for the resulting treatment (Center for Medicare & Medicaid Services 2014). Furthermore, CMS began publicly reporting provider-specific risk-adjusted HAI rates on Hospital Compare[1] in order to create added reputational incentives and to help route patients to higher-quality providers.

[1] http://www.medicare.gov/hospitalcompare

Unfortunately, multiple sources of evidence suggest that the HAC nonpayment policy has had little impact on the true rate of HAIs (Lee 2012, Schuller et al. 2014). It has been hypothesized that this may be because the financial impact of the policy was too small to influence significant change in practice (McNair et al. 2009). Consequently, public organizations that promote patient safety have called for stronger financial penalties (see, e.g., Health Watch USA 2011). In response, further Medicare legislation was issued in the form of the HAC Reduction Program, which created additional financial penalties (starting on October 1, 2014) for providers with high HAI incidence.

Upcoding. We investigate another explanation for the lack of improvement in HAI incidence: upcoding (inaccurate claims reporting that results in higher reimbursement or reduced penalties) may have diminished providers' financial incentives for reducing true HAI rates. In this paper, we focus on two types of upcoding that can occur when encountering a patient with a true HAI:

1. *POA over-reporting*, i.e., reporting a HAI as present-on-admission (POA) infection, and
2. *HAI under-reporting*, i.e., failing to report a HAI.

Even though both types of erroneous reporting can be considered upcoding, we distinguish them in our exposition and analysis as they are likely to occur through different mechanisms and have different policy implications. Our principal objective of this paper is to assess whether reliable evidence can be found for the existence of either type of upcoding.

Presently, there is mixed evidence for whether HAI upcoding exists. One study manually reviewed eighty medical records from the University of Michigan Health System and found that both types of upcoding were rampant:

> *"It is concerning that cases of CA-UTI (the most common type of hospital-acquired infection) are rarely identified in [claims data]... In addition, coders often listed [infections] as present on admission, although the medical record indicated that it was hospital acquired... Because coding ... seems to be fraught with error, nonpayment according to CMS policy may not reliably occur."* - Meddings et al. (2010)

On the other hand, the Office of the Inspector General (OIG) conducted a manual review of a few hundred medical records across the nation and found that HAIs were indeed reliably reported and that there was very little evidence of upcoding (Snow 2012). These conflicting results illustrate the drawbacks of identifying upcoding through manual auditing of claims data: manual review is a time-consuming and expensive process that produces high-variance results due to small sample sizes and the rarity of HAIs. Yet, detecting upcoding from available observational data (i.e., hospital claims records) is challenging because a patient's true diagnosis is unobservable. Moreover, standard econometric techniques such as diff-in-diff estimates of reporting rates before and after the nonpayment policy do not apply because the distinction between HAIs and POAs in claims reporting did not exist prior to the nonpayment policy.

It is important to gauge the extent of upcoding at a national scale in order to better understand its implications for Medicare policy and for providers. In particular, significant upcoding (if present) can erode the effectiveness of the current regulation at reducing HAI incidence. Additionally, it raises questions about the veracity of self-reported HAI rates. This is especially concerning since financial penalties from the HAC Reduction Program are determined on the basis of self-reported data; thus, current policy may unfairly penalize providers who report HAIs accurately. Furthermore, Medicare publishes these self-reported HAI rates to inform patients so as to guide their choice of providers. Thus, upcoding could ultimately lead to the undesirable outcome of patients choosing low-quality providers who upcode over high-quality providers.

We note that there is a division of opinion on whether upcoding occurs intentionally or as a consequence of ineffective quality management. For example, Silverman and Skinner (2004) suggest that upcoding may be intentional profit-maximizing behavior, while Meddings et al. (2010) suggest that it may be a result of miscommunication between nurses and medical coders (specialized hospital staff who translate medical records to claims reports). We restrict the focus of our paper to finding evidence of upcoding rather than the question of intent. However, in §7, we discuss some insights derived from conversations with hospital quality control staff about our results.

Main Contributions. We use national claims reporting data to estimate the extent of upcoding after the nonpayment policy went into effect. Our identification strategy is driven by variations in existing state-level regulation on adverse event reporting, and we address endogeneity concerns through the use of instrumental variables. We find that the differential impact of state-level regulation on HAI and POA reporting rates is strongly suggestive of upcoding. In particular, under some mild assumptions, we find that providers in weakly-regulated states are either (1) over-reporting POAs, or (2) under-reporting HAIs relative to providers in strongly-regulated states. If we further assume that the omitted variable bias from unobservable patient risk in our empirical analysis is negligible, we can make the stronger claim that weakly-regulated providers over-report POAs. We strengthen the validity of this assumption by using an extensive set of patient risk controls (derived from patient claims histories and demographics) that have been validated in the medical literature.

In order to estimate the financial impact of upcoding, we make conservative estimates of the rate of upcoding in Medicare inpatient claims for the two most common and important infections – central line-associated bloodstream infections (CLABSIs) and catheter-associated urinary tract infections (CAUTIs) – that have been targeted by the HAC legislation. We estimate that there are over 10,000 over-reported POAs a year, resulting in an added annual cost burden of $200 million to Medicare for reimbursing these HAIs. While this cost inefficiency is small compared to other Medicare expenditures, it is important to note that this money was intended as a penalty to providers to incentivize them to reduce HAI incidence. The practice of upcoding has therefore eroded this

financial incentive, thereby reducing the effectiveness of the policy. Medicare's current plan to increase penalties through the HAC Reduction Program does not address these concerns, and may in fact exacerbate the problem since providers with high HAI rates will face even greater financial pressure to engage in upcoding. Moreover, providers who are trying to report more accurately than others will be unfairly penalized, both financially and reputationally.

Our results suggest that in order for HAI reduction policies to be effective and fair, federal regulation must be introduced to induce accurate reporting. To this end, we provide two policy recommendations: (1) targeted audits based on a new measure we introduce for identifying potentially upcoding providers, and (2) federal implementation of certain features of current state-level regulations that we find to be effective at eliciting truthful reporting. More broadly, we emphasize the importance of ensuring the veracity of self-reported data as Medicare moves towards adopting a growing number of data-driven pay-for-performance policies (HHS 2015).

1.1. Related Literature

Our work relates to the literature on incentivizing provision of high-quality service under asymmetric information. This typically falls under the umbrella of principal-agent problems (Bolton and Dewatripont 2004). As motivated by Arrow (1963), one of the distinguishing features in healthcare settings is the nature of the information asymmetry between Medicare (principal) and the provider (agent). Particularly in our setting, Medicare does not directly observe the provider's chosen action (i.e. level of service quality), making it difficult to design contracts that can successfully improve quality of care. For example, Fuloria and Zenios (2001) develop an optimal outcomes-based reimbursement contract for healthcare providers; however, they acknowledge that the new payment model can only achieve significant gains in quality of care if the payer has access to accurate information about patient characteristics. In fact, they show that if providers distort reported information, outcomes-based contracts can perform worse than standard payment models.

Providers have several levers through which they can strategically take advantage of asymmetric information. One well-studied example is that hospitals may strategically choose which patients they admit. For instance, KC and Terwiesch (2011) find empirical evidence that specialized hospitals cherry-pick easy-to-treat patients. Similarly, Ata et al. (2013) show how the current hospice reimbursement policy may cause providers to engage in adverse selection by preferentially admitting short-lived patients. Brown et al. (2014) empirically find that Medicare overpays capitation payments to private Medicare Advantage plans due to risk-selection by private insurers despite recent efforts to employ patient-level risk-adjustment in deciding payment levels. In these papers, providers take advantage of Medicare's (relative) lack of knowledge of patient risk; in our work, we illustrate that providers (possibly unintentionally) take advantage of Medicare's lack of knowledge of infection occurrence by upcoding, either through poor infection detection or inaccurate claims.

The issue of strategically choosing poor detection levels of low quality (or detection of infections in our setting) has been studied in the supply chain management literature, particularly with respect to social and environmental responsibility (Baiman et al. 2000). The closest work is by Plambeck and Taylor (2015) who study how increased auditing pressure may motivate suppliers to exert greater effort to pass the buyer's audit by hiding information and less care to improving quality and safety. Similarly, in our setting, increased financial penalties may incentivize providers to strategically maintain low detection levels of HAIs to avoid liability and penalties.

Previous studies in the medical and economics literature have studied a different form of upcoding where providers report higher-paying diagnoses under Medicare's traditional fee-for-service system. Silverman and Skinner (2004) found that for-profit hospitals in particular bias their claims reports towards higher-paying diagnoses (DRGs) in order to maximize reimbursement; however, this form of upcoding has greatly declined after increased auditing pressure from Medicare. In fact, concurrent work by Heese et al. (2015) finds that upcoding occurs more frequently among non-profit providers since Medicare preferentially avoids auditing them; based on these results, the authors hypothesize that Medicare allows "beneficient" non-profit hospitals to make some profit from upcoding in order to recover losses from other factors (e.g., treating poorer patients). From a methodological standpoint, both these papers use a provider's fraction of claims that correspond to the highest-paying DRG as a proxy for that provider's level of upcoding. However, a key limitation of this proxy (acknowledged by the authors) is that it can be biased by patient selection based on unobservable risk factors (e.g., as suggested by empirical work in KC and Terwiesch (2011)). This issue was resolved by a large-scale manual review of medical records conducted by the OIG, which definitively established evidence for upcoding (Silverman and Skinner 2004). Similarly, two studies took this manual approach for evaluating the extent of HAI upcoding, but as mentioned earlier, they yielded conflicting estimates (Meddings et al. 2010, Snow 2012). This may be because such studies involve hiring costly medical experts and are thus limited to small sample sizes, leading to high-variance estimates. Our approach uses observational claims data and helps resolve this conflict by finding evidence of HAI upcoding as well as conservative estimates of its magnitude. In contrast to prior methodologies, we account for patients' unobservable risk factors by employing a double regression. We note that concurrent work by Geruso and Layton (2015) also uses observational data to study yet a different form of upcoding by private insurers among Medicare Advantage patients. However their approach relies on comparing risk scores of patient populations under different insurers, which cannot be used to identify upcoding within the Medicare population (since all patients have the same insurer). Thus, to the best of our knowledge, our paper is the first to show and quantify upcoding behavior among Medicare-only patients using observational data.

The remainder of the paper is organized as follows. In §2, we outline our identification strategy and present intuition behind our argument that providers are engaging in HAI upcoding. In §3, we provide a formal model that maps our upcoding hypothesis to a hypothesis that is empirically verifiable from claims data under certain assumptions. We describe our various sources of data on patients, providers, and state regulations in §4, and establish our key empirical results for upcoding in §5 along with estimates of the monetary losses incurred by Medicare. We perform robustness checks to justify our model assumptions in §6. We conclude with insights from conversations with hospital staff about our results, as well as a discussion of the policy implications of this work in §7.

2. Empirical Strategy for Identifying Upcoding

A naive approach would use national claims data to identify upcoding providers as those who have high POA reporting rates (suggestive of over-reporting POAs) and/or low HAI reporting rates (suggestive of under-reporting HAIs). One may further wish to risk-adjust these reporting rates to account for the variation in infection susceptibility among individual patients. However, these effects may also be caused by variations in:

1. *Provider quality*: higher-quality providers are likely to cause fewer HAIs and would thus report lower (risk-adjusted) HAI rates.

2. *Provider's POA infection detection*: providers who successfully detect more POAs (by finding the infection within the POA time window) would report higher (risk-adjusted) POA rates. Analogously, they may report relatively lower (risk-adjusted) HAI rates since providers with poor POA detection may incorrectly identify POAs as HAIs.

3. *Unobservable patient risk confounders*: if we fail to properly risk-adjust for infection susceptibility, providers with relatively riskier patients will appear to report higher (risk-adjusted) POA rates and providers with less risky patients will appear to report lower (risk-adjusted) HAI rates.

We address these concerns by exploiting variations in existing state-level adverse event regulation. Many states passed laws that mandated the reporting of various HAIs prior to the federal nonpayment policy in 2008. As documented by the Office of Inspector General (OIG) of the Department of Health and Human Services, regulations on the contents of these reports varied significantly from state to state (Levinson 2008), thereby creating a natural quasi-experiment. The primary aim of these reporting systems was to track HAI incidence across providers, and a subset of states included measures to ensure accurate reporting (e.g., detailed patient and event information monitoring and root cause analysis). We will refer to this subset of states as *strongly-regulated* states and all other states as *weakly-regulated* states. It is important to note that this state-level regulation was independent of Medicare legislation and had no outcomes-based financial incentives (unlike Medicare policies). However, states could conduct on-site audits and exact financial penalties if providers

were caught purposefully mis-reporting. Strongly-regulated states in particular required provider accountability for accurate HAI reporting as well as implementation of follow-up corrective strategies; thus, the regulations indirectly mandated that providers in such states had to improve their operational capabilities to correctly detect and prevent these targeted infections.

Thus, providers in strongly-regulated states should have both (1) higher quality and (2) better POA infection detection. We provide further empirical evidence in §6 to support these claims in the context of general medical care:

1. *(Provider quality)* We compare a variety of risk-adjusted provider quality metrics reported on Hospital Compare, including mortality rates and process of care measures (§6.1). We find overwhelming evidence that strongly-regulated providers offer higher quality care, as shown through better scores on nearly all process of care measures (particularly those targeted towards preventing infections), as well as reduced downstream mortality rates.

2. *(POA infection detection)* Much of the medical literature (see, e.g., Meddings et al. 2010, Mark and Harless 2010, Duffin 2014) as well as our conversations with hospital staff suggests that timely infection detection and attribution is a nurse-centric task. Unfortunately, evidence shows that US hospitals have severe nurse understaffing; for example, 33% of surveyed nurses report inadequate staffing levels and half report insufficient time with patients (ANA 2016). Consequently, several states have introduced regulation to ensure adequate hospital nurse staffing levels. Using data from the American Nurses Association (ANA), we find that the majority (75%) of strongly-regulated states (versus only 8% of weakly-regulated states) have adopted such regulation, suggesting that strongly-regulated states have improved nurse staffing levels and would thus have higher POA detection. However, since nurse staffing levels are not publicly reported in most states, we examine some alternate nurse-centric metrics (§6.2). First, we compare *time-sensitive* process of care measures. We find that strongly-regulated providers are significantly more successful at performing necessary tasks within a pre-specified time window, suggesting that they would be more successful at identifying true POAs by testing for the infection within the allowed POA time window (typically 48 hours). Second, we examine payment rates (adjusted for variations in demographic differences in pay, patient risk, and provider mortality rates) reported on Hospital Compare, as well as our own measure of "billing aggressiveness" computed from claims data. We find evidence that strongly-regulated providers are significantly more aggressive in claims billing, and furthermore, succeed at receiving higher payments for equal quality of services provided. This suggests that they are more likely to discover and report all reimbursable complications (including POA infections). Furthermore, our conversations with hospital staff support the claim that strongly-regulated providers have established additional infrastructure for ensuring compliance with state-level guidelines on accurate HAI (and therefore, POA) detection and prevention (see discussion in §7.1).

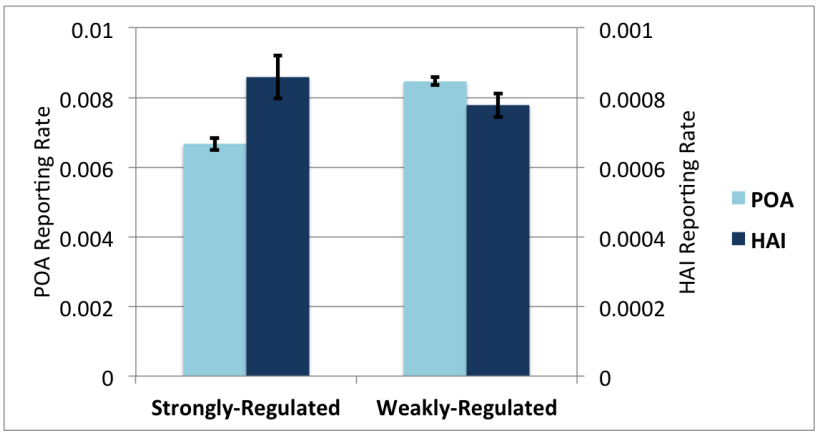

Figure 1 Average (unadjusted) POA and HAI reporting rates for strongly- and weakly-regulated states in a random sample of almost a million Medicare inpatient stays from 2009-10. Providers in strongly-regulated states have relatively lower POA and higher HAI reporting rates.

Yet, we empirically find that providers in strongly-regulated states have *lower* (risk-adjusted) POA reporting rates and *higher* (risk-adjusted) HAI reporting rates (see Fig. 1 for the unadjusted reporting rates in a random sample of almost a million Medicare inpatient stays from 2009-10). Since strongly-regulated providers have both (1) higher provider quality and (2) better POA infection detection, these two confounders cannot explain the observed effects. Furthermore, (3) unobservable patient risk confounders cannot simultaneously explain *both* effects. In particular, if weakly-regulated providers report higher POA rates because they treat relatively riskier patients (who are more susceptible to infections), then their patients would also contract more HAIs since the infection is the same and weakly-regulated providers have equal or lower quality of care; they should thus report higher HAI rates as well. Similarly, if weakly-regulated providers report lower HAI rates because they treat relatively less risky patients, they should also report lower POA rates.

Another potential explanation is that state adverse event regulation may be endogenous to HAI reporting rates. Specifically, strongly-regulated states may have introduced adverse event reporting regulation as a response to low provider quality with respect to HAIs (although they have overall higher quality of care). In this case, it would be reasonable that providers in strongly-regulated states report relatively higher (risk-adjusted) HAI rates. We address this issue by using an instrumental variable approach: our instruments are various measures of state taxation levels (known as the Economic Freedom Index (Ashby et al. 2010)) which are correlated with the strength of a state's regulatory environment but bear no direct relationship with HAI-specific provider quality. We find our results remain consistent despite accounting for this endogeneity.

Thus, the differential impact of state-level regulation on HAI and POA reporting rates strongly suggests that weakly-regulated providers engage in upcoding. If we further assume that the bias from unobservable patient risk confounders in our analysis is negligible, we can make the stronger

claim that weakly-regulated providers are over-reporting POAs by at least their excess (risk-adjusted) POA reporting rates. To support this hypothesis, we use an extensive set of patient risk controls (§4.4) that have been validated in the medical literature. These arguments are formalized in Proposition 1 using a model of provider reporting in §3.

3. Model

Accurate reporting signifies reporting a POA in the case of a true POA infection, and similarly reporting a HAI in the case of a true HAI occurrence. According to Medicare policy, all detected infections must be reported. There are several mechanisms through which providers can deviate from accurate reporting by *upcoding* (claiming more reimbursement than allowed) and *downcoding* (claiming less reimbursement than allowed), which we detail in a simple model below.

3.1. Definitions

We describe all possible provider reporting behaviors conditional on a patient's infection type. Model parameters are defined for strongly-regulated states (denoted by superscript S); analogous definitions hold for weakly-regulated states (denoted by superscript W).

3.1.1. Reporting Mechanisms.

Given a patient with a true POA infection, strongly-regulated providers report:

- a POA with probability $1-\epsilon_1^S-\epsilon_2^S$ (accurate reporting)
- a HAI with probability ϵ_1^S (downcoding)
- no infection with probability ϵ_2^S (downcoding)

Note that failing to report a POA accurately is considered downcoding since reporting the infection would likely increase the provider's reimbursement. Downcoding occurs as a result of the hospital's failure to detect the infection within the POA time window (thus, forcing it to report the infection as a HAI) or failure to detect the infection entirely. It may also occur if the hospital fails to communicate the early detection or existence of the infection to the medical coder who files claims.

Similarly, given a patient with a true HAI infection, strongly-regulated providers report:

- a HAI with probability $1-\delta_1^S-\delta_2^S$ (accurate reporting)
- a POA with probability δ_1^S (upcoding via POA over-reporting)
- no infection with probability δ_2^S (upcoding via HAI under-reporting)

Both upcoding mechanisms may be a consequence of poor provider quality: POA over-reporting may be due to poor coder training, and HAI under-reporting may be due to poor infection detection ability. Of course, both mechanisms may also occur through intentional claims manipulation.

See Table 1 for a summary of parameters. We note that we only model provider reporting behavior in cases where the patient truly has either a POA or a HAI infection. In principle, providers could report a POA or a HAI even if the patient did not have any infection. However, there is no evidence that this occurs in practice (Snow 2012, Meddings et al. 2010).

	Reported POA	Reported HAI	No Report	(Risk-Adjusted) Infection Prob.
True POA	$1-\epsilon_1^R-\epsilon_2^R$	ϵ_1^R	ϵ_2^R	p^R
True HAI	δ_1^R	$1-\delta_1^R-\delta_2^R$	δ_2^R	$\alpha^R \cdot p^R$

Table 1 Summary of model parameters. The superscripts $R \in \{S, W\}$ denote strong vs weak regulation.

3.1.2. Infection Risk. Let X denote the patient's observed risk covariates. Denote $p^S(X)$ as the patient's risk-adjusted probability of a true infection outside the hospital (i.e. the probability of a POA infection conditioned on X) in strongly-regulated states. (We omit the X-dependence when it is clear from context.) Note that this quantity does not depend on hospital-specific factors (such as quality or detection ability). Thus, if our patient-level risk-adjustment is unbiased (i.e., if there is no omitted variable bias), then $p^S(X) = p^W(X)$; however, we will not assume that this is the case as there are many unobservable patient risk factors in healthcare that can create bias.

Similarly, let $\alpha^S \cdot p^S(X)$ denote the patient's risk-adjusted probability of a true HAI (i.e., the probability of a HAI conditioned on X) in strongly-regulated states. Intuitively, $p^S(X)$ captures the patient's propensity for infection and α^S captures the impact of the hospital's quality of care on the patient's risk for infection. In other words, we assume that the overall risk for a true HAI is multiplicative in patient and provider risk factors.

3.1.3. Reporting Rates. Given the model above, the observed per-visit probabilities of POA and HAI reports in strongly-regulated providers are, respectively,

$$r^S_{POA} := p^S(1-\epsilon_1^S-\epsilon_2^S) + \alpha^S p^S \delta_1^S \,,$$
$$r^S_{HAI} := \alpha^S p^S(1-\delta_1^S-\delta_2^S) + p^S \epsilon_1^S \,.$$

Analogous expressions hold for weakly-regulated providers.

3.2. Assumptions

We now formalize the assumptions we stated earlier (in §2) in the language of our model:

1. *(Provider Quality)* We assume that strongly-regulated providers cause equal or fewer true HAIs compared to weakly-regulated providers given patients that are equally susceptible to infection (i.e., when fixing p). In other words, we assume

$$\alpha^S \leq \alpha^W \,.$$

We justify this assumption by comparing a variety of risk-adjusted provider quality metrics reported on Hospital Compare, including mortality rates and process of care measures (§6.1).

2. *(POA Infection Detection)* We assume that strongly-regulated providers report true POAs more accurately (i.e., they downcode less). In particular,

$$\epsilon_1^S \leq \epsilon_1^W \quad \text{and} \quad \epsilon_2^S \leq \epsilon_2^W \,.$$

We justify this assumption by comparing *time-sensitive* process of care measures (§6.1) and adjusted payment rates reported on Hospital Compare, as well as our own measure of "billing aggressiveness" computed from claims data (§6.2).

3.3. Hypotheses

Let the rate of POA over-reporting (i.e., the risk-adjusted probability of a reported POA in the event of a true HAI) in strongly-regulated states be denoted by $O^S = \alpha^S p^S \delta_1^S$ and similarly, the rate of HAI under-reporting (i.e., the risk-adjusted probability of no reported infection in the event of a true HAI) be denoted $U^S = \alpha^S p^S \delta_2^S$. The total rate of mis-reported HAIs is then simply $O^S + U^S$. We consider the following hypothesis:

> Compared to strongly-regulated providers, weakly-regulated providers have higher rates of upcoding, either through (i) POA over-reporting or (ii) HAI mis-reporting.

We can write this as hypothesis H_1 (with corresponding null hypothesis H_0):

$$H_0 : O^S \geq O^W \quad \text{and} \quad O^S + U^S \geq O^W + U^W$$
$$H_1 : O^S < O^W \quad \text{or} \quad O^S + U^S < O^W + U^W$$

This hypothesis cannot be directly evaluated since we do not observe true POAs/HAIs. However, using our assumptions (§3.2), we can map this hypothesis to the following empirically verifiable hypothesis H_1' (with corresponding null hypothesis H_0'):

$$H_0' : r_{POA}^S \geq r_{POA}^W \quad \text{or} \quad r_{HAI}^S \leq r_{HAI}^W$$
$$H_1' : r_{POA}^S < r_{POA}^W \quad \text{and} \quad r_{HAI}^S > r_{HAI}^W$$

PROPOSITION 1. *Hypothesis H_1' implies H_1. If we further assume $p^S = p^W$, then weakly-regulated providers over-report POAs by at least the excess risk-adjusted POA reporting rate ($r_{POA}^W - r_{POA}^S$):*

$$O^W \geq O^W - O^S \geq r_{POA}^W - r_{POA}^S \,.$$

Proof of Proposition 1 : See Appendix B.

Thus, if we empirically verify H_1', Proposition 1 implies that weakly-regulated providers have higher rates of upcoding compared to strongly-regulated providers. Moreover, if our patient risk-adjustment is unbiased (i.e., $p^S = p^W$), then weakly-regulated providers over-report POAs by at least as much as their excess risk-adjusted POA reporting rate.

4. Datasets

In this section, we describe our various sources of data and define our key variables for treatment effect estimation. We also discuss potential confounders and our approach to control for these effects. We reproduce state reporting system features in Table 2, and we report summary statistics of selected variables in Table 10 (in Appendix A).

4.1. Data Sources

Medicare Patient Data. Our main source of data was Medicare inpatient claims data from the MedPAR Research Identifiable Files (RIF) made available by CMS. This dataset contains information on every inpatient stay between 2007 and 2010 of a randomly selected 5% sample of all Medicare beneficiaries in the United States. Our dataset spans 3,865,733 inpatient stays by 492,218 unique beneficiaries. Each record includes provider IDs, diagnoses (ICD-9 codes) and procedures, patient demographics, and billing information. Beneficiaries can be tracked across multiple inpatient stays, allowing us to compute health risk measures for individual patients based on their claims histories during this period. We use a rolling six-month window of claims histories to compute various measures of patient risk for each inpatient stay.

Our unit of observation is an individual Medicare inpatient stay, and we perform a cross-sectional analysis on inpatient stays in 2009-10. We limit our sample to short stays under the prospective payment system served by providers in the United States (which is the healthcare setting that was targeted by the nonpayment policy). Finally, if a patient is newly enrolled in Medicare (e.g., from Medicare Advantage), their past visits may be censored (thus, creating bias in controls computed on past visits); we address this by limiting our sample to patients with at least one prior Medicare inpatient stay in the past 24 months. We note that these filters affect all states uniformly, and therefore do not create bias in our analysis; see Appendix C.

State Reporting System Classification. As of January 2008, 26 states had implemented adverse event reporting systems in the absence of federal guidelines. The OIG performed a detailed comparison of these systems based on telephone interviews with the staff responsible for each state's reporting system (Levinson 2008). The OIG report describes key features of the state reporting systems, including the type of information that must be reported by each state regarding (1) the affected patient, (2) the adverse event, and (3) the root cause of the adverse event. All 26 states with reporting systems enforced at least reporting the identity of the hospital and the adverse event that had occurred. We reproduce the information reported in each category and the number of states that had implemented each requirement in Table 2.

Other Sources of Data. We used data from the American Community Survey (2008-12) by the US Census Bureau to control for patient demographics. Similarly, we obtained data on county-level

Category	Information	# States
Any Reporting	Event and Hospital	26
Patient-Specific	Impact of Event on Patient	12
	Patient Age or Date of Birth	19
	Patient Diagnosis	16
	Patient Medical Record Number	5
	Patient Billing Number	2
Event-Specific	Type of Event	26
	Location within Hospital	20
	Date of Event	24
	Date of Discovery	10
	Summary Description	18
	Detailed Description	11
Root Cause Analysis	Root Cause Analysis Team Name	7
	Identified Cause	12
	Contributing Factors	16

Table 2 Different types of information reporting requirements used in state adverse event reporting systems and the number of states that had implemented each requirement. Reproduced from Levinson (2008).

life expectancies from the Institute of Health Metrics and Evaluation. For our instrumental variable analysis, we used the state-level 2010 North American Economic Freedom indices (Ashby et al. 2010). Finally, we used Hospital Compare's provider-level quality metrics for robustness checks.

4.2. Treatment Variable

One possible definition of the treatment variable is simply having an adverse event reporting system. Interestingly, our results show that merely having an adverse event reporting system did not have a significant effect on POA and HAI claims reporting rates for CLABSIs and CAUTIs (see Table 5). This is because the quality of the reporting systems varied widely. Instead, our approach is to look for states that impose meaningful requirements on the quality of reporting. We construct a treatment variable that is an indicator for whether the provider is located in a state that had strong regulations on adverse event reporting prior to the federal nonpayment policy in 2008.

As previously noted, we use data from an OIG report which lists each state's information reporting requirements (§4.1). We are particularly interested in regulation that enforced accurate reporting. The OIG report claims that states identified cases of mis-reporting by

> "analyzing reported data, comparing hospital reports against complaints, referrals, and administrative databases, and conducting onsite audits" (Levinson 2008, pg. 4).

These methods are greatly aided by the availability of more detailed data. In particular, we argue that the more data a state has regarding the circumstances of an adverse event, the harder it is for a provider to mis-report the event without being detected. Thus, we used the amount of required information reported to states in each category as a proxy for increased regulatory pressure for

accurate reporting. For simplicity, we chose the most informative reporting requirement from each of the three information categories (see Table 2), namely,

1. Patient-specific: patient medical record number or billing number
2. Event-specific: detailed description of the adverse event
3. Root cause analysis: identified cause of adverse event

We define our treatment variable based only on these three reporting requirements, which helps us better interpret our results and make concrete policy suggestions. Since there are many ways to define the treatment variable, we perform a robustness check (§5.4) by considering several alternate definitions of the treatment variable to alleviate the concern that a particular definition of the treatment variable gave rise to our results by chance.

In order to construct the treatment variable, we compute a binary "strength" for each state's regulation of its adverse event reporting system based on the number of these three features adopted. The median number of features adopted among states with reporting systems was one; thus, we considered the strongly-regulated states to be those with two or more features. According to this definition, the strongly-regulated states are CT, FL, MA, MN, NJ, NY, RI, and SD. (We define a binary treatment variable to improve the interpretability of our results; in §6.3, we perform a robustness check to ensure our results are consistent if the treatment variable is continuous.)

Thus, we define the binary treatment variable S for providers as:

- $S = 0$: Provider is located in a weakly-regulated state, i.e., either had no adverse event reporting system, or had an adverse event reporting system that had zero or one of the reporting requirements described above.
- $S = 1$: Provider is located in a strongly-regulated state, i.e., had an adverse event reporting system with two or more of the reporting requirements described above.

4.3. Outcome Variables

We focus on CLABSIs and CAUTIs, the only two HAIs directly targeted by both the HAC non-payment policy and the recent HAC Reduction Program. We define two outcome variables:

- POA_i is an indicator variable for whether either a CLABSI or a CAUTI was diagnosed along with the present-on-admission indicator in the claims record for inpatient stay i.
- HAI_i is an indicator variable for whether either a CLABSI or a CAUTI was diagnosed without the present-on-admission indicator in the claims record for inpatient stay i.

4.4. Controls

We define a variety of controls to account for potential confounders.

Patient Risk. States that implement strong regulation for HAIs are likely to have also implemented other measures towards improving population health; this may, in turn, affect downstream

patient infection rates. To account for this effect, we control for an extensive list of patient-specific factors that are computed from their claims histories. Age, sex, and race are obtained from MedPAR's summarized beneficiary demographic information. We use a rolling window of 6 months of each patient's claims history to identify risk-associated quantities such as the number of days since the patient's last admission, the number of prior admissions, the number of prior procedures performed on the patient during those admissions, the number of previous CLABSI and/or CAUTI infections sustained during that time, and the total length of hospital stay days.

We also use 6-month patient history to compute the Charlson comorbidity index (measure of patient risk that predicts patient mortality within 6 months) and 29 Elixhauser comorbidities (scores that capture patient comorbidities outside of the primary reason for hospitalization). These measures have been frequently validated and are widely accepted in the medical community (Deyo et al. 1992, Elixhauser et al. 1998). We compute the Elixhauser comorbidities as recommended by the Agency of Healthcare Research & Quality (AHRQ)[2]. Finally, we control for the patient's current type of diagnosis using DRG groupings used in computing the Elixhauser scores.

Demographic Factors. States that did not implement strong HAI regulation may generally be poorer or more resource-constrained. This may, in turn, affect the completeness of patient claims data; in particular, poor patients may not have access to frequent healthcare due to lack of health insurance or other resource constraints, and thus their health risks may not be completely captured from claims histories. We address this by using health-related controls from census data based on the patient's listed zipcode. These controls (aggregated by zipcode) included the average household income as well as fractions of individuals in the population who were above 65, uninsured, unemployed, near the poverty line, foreign-born and/or had not completed high school.

Provider's Billing Aggressiveness. Providers that code their claims more aggressively to achieve the highest possible reimbursement rates may generally have more complete patient claims data. Note that this is different from upcoding, since the codes may be accurate. We control for variations in coding practices by defining a provider-specific measure of billing aggressiveness:

$$\text{billing aggressiveness}_j = \frac{\sum_{i \in T_j} \text{charges to Medicare for inpatient stay } i}{\sum_{i \in T_j} \text{Medicare payment for inpatient stay } i}$$

where j is the index of the provider and T_j is the set of all inpatient stays under the care of provider j. This heuristic – the ratio between average payment requested and payment received – would likely be higher for aggressive providers who attempt to charge Medicare much more for each patient stay than the typical reimbursement level.

[2] https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp

4.5. Instrumental Variables

Our treatment variable is potentially endogenous if states with poor provider quality (with respect to HAIs) chose to pass adverse event reporting regulation. While strongly-regulated states have higher provider quality on many standard metrics (see §6.1), HAI-specific provider quality (the quantity of interest in this context) is unobservable. We address this issue through the use of instrumental variables. We focus on two factors that drive increased state-level regulation that are not caused by high HAI rates: (i) a state's capacity to issue and enforce costly regulation, and (ii) voters' preferences for increased regulation. Following the example of Mukamel (2012), we use the Economic Freedom Indices as our instruments (see Table 10 in Appendix A for summary statistics):

1. *Area 1 (Size of Government)*: This measure captures the government's (both federal and state) spending in the state as a percentage of the state's GDP. While weakly-regulated states tend to have lower tax earnings and thus lower state-level spending, they are poorer and have higher total government expenditures due to intergovernment transfers (i.e., incoming funds from federal assistance programs). In contrast, strongly-regulated states are more affluent and have less federal government intervention (indicating financial ability to support sovereign regulation). For example, the federal government spent more in West Virginia (weakly-regulated) than they raised through taxation, while the opposite held for Connecticut (strongly-regulated) (Ashby et al. 2010).

2. *Area 2 (Takings & Discriminatory Tax)*: This measure captures the government's (both federal and state) tax revenue from this state as a percentage of the state's GDP. Strongly-regulated states tend to have higher tax rates to financially support more stringent government regulation.

3. *Area 3 (Labor Market Freedom)*: This measure scores the government's stringency of regulation with respect to labor issues (e.g. the minimum wage, union density) as well as bureaucracy (e.g. the proportion of employed individuals who work for the government). Voters in strongly-regulated states seem to prefer increased labor market regulation.

Summarizing, strongly-regulated states have higher Area 1 index and lower Area 2 and 3 indices: in particular, they tend to have stronger state-level government presence (Areas 2 and 3) and stronger financial resources to support regulation (Areas 1 and 2). We verify that these instruments are sufficiently correlated with the treatment variable through a weak identification test (see §5.2).

Second, we believe the instruments are uncorrelated with the error term in the HAI regression, i.e., the instruments are not predictive of reported HAI rates after conditioning on the treatment and controls. Unlike the original treatment variable, which may have been directly caused by poor provider quality with respect to HAIs, there is no apparent causal relationship between poor HAI-specific provider quality and our instruments. We perform an overidentification test (§5.2), which suggests that the instruments are, in fact, not correlated with the error term in our regression.

Furthermore, our instruments capture state affluence; since affluent states have more financial resources to improve provider quality, we expect that the instrumented treatment variable would be positively correlated with HAI-specific provider quality. To support this hypothesis, we show that both the instruments and the instrumented treatment variable are positively correlated with higher provider quality on a number of standard quality metrics (see Appendix D.2, specifically Tables 13 and 14). Recall that our instrumental variable approach is motivated by the concern that the treatment variable may be correlated with (unobserved) poor HAI-specific provider quality, i.e., the error term of the HAI regression. Therefore, this hypothesis suggests that even if the instruments are correlated with the error term, we expect the correlation to be negative. As described in Appendix D.3, this ensures that our treatment effect estimates are conservative; i.e., the causal effect of the original treatment variable on HAI reporting rates can only be higher than our estimate.

Finally, we note that it is highly implausible that a state's decision to regulate was influenced by its POA rates, because adverse event reporting systems targeted HAIs and, to the best of our knowledge, there were no state agencies that even collected information on present-on-admission infection rates. Thus, we only use an instrumental variable approach for our analysis of HAI rates.

5. Estimation & Results

We perform treatment effect estimation to determine the causal effects of strong state-level adverse event reporting regulation on POA and HAI claims reporting rates for CLABSIs and CAUTIs.

Our primary analysis uses standard regression techniques under a linear probability model. Although our outcomes are binary, we use a linear probability model (see Section 15.2 of Wooldridge (2010) for a justification) rather than a logit or probit model so that we can perform instrumental variable validity tests in the presence of clustered errors (e.g., see Cameron and Miller (2015)). We perform robustness checks in §6.3 to show that our results remain consistent under alternate specifications. We find that the presence of strong state regulation of adverse event reporting was associated with decreased POA rates and increased HAI rates. Finally, since we seek to jointly establish statistical significance for both POA and HAI regressions, we can conservatively apply the Bonferroni correction by summing the p-values for both regressions as our joint p-value. We note that in all our regressions (including robustness checks), the sum of the two corresponding p-values is less than 0.05, showing that we have joint significance. As argued in §3, this suggests that providers in weakly-regulated states have statistically significant rates of upcoding.

5.1. POA Regression

Let S_i be the treatment variable (as defined earlier), C_i denote the vector of controls (including an intercept term) for inpatient stay i. We use a linear model with the econometric specification:

$$POA_i = \beta_S^{POA} S_i + \beta^T C_i + \epsilon_i$$

where ϵ_i is the error term. The coefficient of interest is β_S^{POA}, which represents the effect of strong state regulation on POA reporting rates. Specifically, if β_S^{POA} is negative, this would indicate that after controlling for potential confounders, providers in states with strong regulations have a lower probability of reporting POAs than providers in states with little or no regulation.

The standard OLS estimator makes the assumption that all errors in the POA model are homoskedastic and independent. However, this is unlikely to be the case as hospital stays served by the same provider may have correlated heteroskedastic errors due to unobserved provider-specific variables. To account for this, we cluster our data at the provider (i.e., hospital) level, and use cluster-robust standard errors that relax our assumptions to allow both arbitrary heteroskedasticity and arbitrary within-provider correlation. (In §6.3, we perform a robustness check with coarser state-level clustering and confirm that our results remain significant.)

The regression results with cluster-robust standard errors are shown in Table 3. We find that, after controlling for patient risk through claims histories and demographic factors, strong state regulation is associated with significantly lower POA reporting rates ($p = 1.0 \times 10^{-4}$).

5.2. HAI Regression

Let I_i denote the vector of instrumental variables for inpatient stay i. We use standard two-stage least squares (2-SLS) for estimation (implemented in the ivreg2 Stata package). In the first stage, we fit our endogenous variable

$$\hat{S}_i = \beta_1^T C_i + \beta_I^T I_i + \epsilon_{i,1}$$

In the second stage, we fit our outcome variable using the predicted $\hat{S}_i$ from the first stage

$$HAI_i = \beta_S^{HAI} \hat{S}_i + \beta_2^T C_i + \epsilon_{i,2}$$

where $\epsilon_{i,1}, \epsilon_{i,2}$ denote error terms. In this case, if β_S^{HAI} is positive, then after controlling for potential confounding variables and the endogeneity of regulation, providers in strongly-regulated states have a higher probability of reporting HAIs than providers in weakly-regulated states.

Once again, we use cluster-robust standard errors clustered at the provider level. We also perform weak- and over-identification tests to support the validity of our chosen instruments.

The regression coefficients, and robust standard errors clustered by provider are shown in Table 3. We find that, after controlling for patient risk through claims histories and demographic factors, strong state regulation is associated with significantly higher HAI reporting rates ($p = 1.6 \times 10^{-2}$).

Tests of Instrument Validity. Our first-stage regression produced $R^2 = 0.54$, and the instruments (economic freedom indices) alone had a partial $R^2 = 0.17$ (see results in Appendix D.1). We performed the standard IV validity tests under robust provider-level clustering:

Variable	(1) POA Reports		(2) HAI Reports	
	Estimate	**SE**	**Estimate**	**SE**
(Intercept)	3.76×10^{-3}	8.17×10^{-3}	4.74×10^{-3}*	2.85×10^{-3}
sex: female	-2.86×10^{-3}***	2.34×10^{-4}	1.30×10^{-4}**	5.96×10^{-5}
age	-1.31×10^{-4}***	9.91×10^{-6}	-1.06×10^{-5}***	2.79×10^{-6}
race: white	1.84×10^{-3}*	1.03×10^{-3}	3.27×10^{-4}	3.36×10^{-4}
race: black	3.35×10^{-3}***	1.09×10^{-3}	4.87×10^{-4}	3.46×10^{-4}
race: asian	5.99×10^{-4}	1.38×10^{-3}	5.72×10^{-4}	4.70×10^{-4}
race: hispanic	2.61×10^{-3}**	1.30×10^{-3}	2.56×10^{-4}	3.87×10^{-4}
race: other	1.90×10^{-3}	1.39×10^{-3}	-8.23×10^{-6}	3.98×10^{-4}
race: unknown	4.49×10^{-4}	2.17×10^{-3}	-1.17×10^{-4}	5.62×10^{-4}
elixhauser score: x	(omitted)	(omitted)	(omitted)	(omitted)
charlson score	-5.20×10^{-5}	1.11×10^{-4}	-8.48×10^{-5}**	3.40×10^{-5}
diagnosis: cardiac	-4.66×10^{-3}***	1.79×10^{-4}	-3.29×10^{-4}***	6.88×10^{-5}
diagnosis: renal	5.38×10^{-2}***	1.73×10^{-3}	-3.98×10^{-4}***	1.28×10^{-4}
diagnosis: nervous	-6.82×10^{-3}***	5.80×10^{-4}	-4.58×10^{-4}	2.90×10^{-4}
diagnosis: pulmonary	-4.71×10^{-3}***	2.13×10^{-4}	-7.40×10^{-4}***	7.58×10^{-5}
diagnosis: diabetes	-5.93×10^{-3}***	6.31×10^{-4}	-1.46×10^{-4}	3.23×10^{-4}
diagnosis: hypothyroidism	-2.46×10^{-3}**	1.14×10^{-3}	6.26×10^{-4}	7.15×10^{-4}
diagnosis: renal_failure	-5.87×10^{-2}***	1.76×10^{-3}	5.11×10^{-4}**	2.42×10^{-4}
diagnosis: liver	-6.73×10^{-3}***	4.77×10^{-4}	-5.53×10^{-4}**	2.40×10^{-4}
diagnosis: ulcer	-4.86×10^{-3}***	3.72×10^{-4}	-3.35×10^{-4}**	1.62×10^{-4}
diagnosis: cancer	-1.02×10^{-2}***	4.25×10^{-4}	7.11×10^{-4}**	2.85×10^{-4}
diagnosis: nutrition	-6.78×10^{-3}***	3.14×10^{-4}	-2.81×10^{-4}	1.82×10^{-4}
diagnosis: alcohol	-5.78×10^{-3}***	6.67×10^{-4}	-9.31×10^{-4}***	2.30×10^{-4}
diagnosis: hypertension	-4.15×10^{-3}***	5.63×10^{-4}	-8.87×10^{-4}***	5.55×10^{-5}
diagnosis: blood disorders	-5.37×10^{-3}***	5.59×10^{-4}	1.34×10^{-4}	2.86×10^{-4}
diagnosis: mental disorders	-6.89×10^{-3}***	4.08×10^{-4}	-9.50×10^{-4}***	1.70×10^{-4}
days since last admit	-1.71×10^{-5}***	1.83×10^{-6}	-3.33×10^{-6}***	6.10×10^{-7}
# past admits	-7.22×10^{-4}***	1.28×10^{-4}	-9.12×10^{-5}**	3.63×10^{-5}
# past procedures	4.90×10^{-4}***	5.28×10^{-5}	3.08×10^{-5}**	1.36×10^{-5}
# past cauti	1.25×10^{-1}***	5.22×10^{-3}	4.62×10^{-4}	4.72×10^{-4}
# past clabsi	6.87×10^{-2}***	4.58×10^{-3}	1.26×10^{-3}**	5.51×10^{-4}
total past length of stay	6.00×10^{-5}***	6.08×10^{-6}	3.25×10^{-6}**	1.36×10^{-6}
billing aggressiveness	1.27×10^{-4}**	5.14×10^{-5}	5.01×10^{-6}	1.28×10^{-5}
demographics: % uninsured	-6.30×10^{-4}	2.29×10^{-3}	1.10×10^{-3}	6.74×10^{-4}
demographics: % above 65	-1.96×10^{-3}	1.51×10^{-3}	-1.47×10^{-4}	5.43×10^{-4}
demographics: % foreign-born	-3.28×10^{-3}**	1.43×10^{-3}	-3.73×10^{-4}	4.19×10^{-4}
demographics: % unemployed	2.20×10^{-3}	2.86×10^{-3}	-3.67×10^{-4}	8.44×10^{-4}
demographics: % near poverty	2.09×10^{-3}	1.73×10^{-3}	1.39×10^{-4}	5.11×10^{-4}
demographics: % no high school	-3.60×10^{-3}**	1.80×10^{-3}	-2.77×10^{-4}	5.07×10^{-4}
local household income	1.68×10^{-8}**	6.83×10^{-9}	-7.01×10^{-10}	2.04×10^{-9}
local female life expectancy	3.35×10^{-4}	2.29×10^{-4}	-1.51×10^{-4}*	7.77×10^{-5}
local male life expectancy	-1.62×10^{-4}	1.68×10^{-4}	1.14×10^{-4}**	5.43×10^{-5}
S	$\mathbf{-1.21 \times 10^{-3}}$***	$\mathbf{3.26 \times 10^{-4}}$	$\mathbf{5.23 \times 10^{-4}}$**	$\mathbf{2.16 \times 10^{-4}}$

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 3 Results of regressions. Point estimates and cluster-robust standard errors (SE) of coefficients for (1) OLS regression of POA reports and (2) 2-SLS regression of HAI reports against strength of state reporting system and controls (Elixhauser coefficients omitted due to space constraints).

1. *(Weak Identification Test)* Our analysis yielded a Kleinberg-Paap Wald F-statistic of 138, which is well above the Stock-Yogo weak ID test critical values for the maximal IV relative bias

(13.91 at the 5% level) and for the maximal IV size (22.30 at the 10% level). This indicates that our instruments are not weak (Baum 2007).

2. *(Overidentification Test)* We computed a Hansen J statistic of 0.228 with a χ^2 p-value 0.89. Thus, we do not reject the null hypothesis that our model is correctly specified, suggesting that our instruments are valid, i.e., economic freedom indices are uncorrelated with HAI reporting rates except through the treatment variable and controls.

3. *(Endogeneity Test of Treatment Variable)* We found evidence ($p = 0.03$) rejecting the null hypothesis that the treatment variable is exogenous with respect to HAI reporting outcomes; this result justifies our instrumental variable approach.

5.3. Loss Estimates

In order to estimate the annual number of upcoded infections as well as their associated costs to Medicare, we assume that the patient risk adjustment in our empirical analysis is unbiased. Following the notation introduced in §3, this translates to $p^S = p^W$. Thus, by Proposition 1, weakly-regulated providers over-report POAs by at least as much as their excess risk-adjusted POA reporting rate, $r^W_{POA} - r^S_{POA}$. Using this estimate implicitly makes two conservative assumptions:

1. Providers in strongly-regulated states have an upcoding rate of zero
2. All providers have similar capabilities for infection detection

We believe these estimates are conservative since providers in strongly-regulated states likely have better infection detection due to their increased infrastructure in response to reporting requirements as discussed earlier. In this case, the number of over-reported POAs by weakly-regulated providers is larger than what we estimate. Secondly, it is unlikely that providers in strongly-regulated states have zero upcoding; in this case, the overall amount of upcoding is again larger than our estimate. Furthermore, we do not consider losses through under-reporting HAIs in this analysis, since the cost of under-reporting is indirect and therefore harder to measure.

We perform two linear regressions on CLABSI-POA and CAUTI-POA outcomes respectively. We find the absolute value of the treatment effects, i.e., excess POA reporting rates, of:

- CLABSI-POA: 2.26×10^{-4} with standard error 1.08×10^{-4}
- CAUTI-POA: 1.34×10^{-3} with standard error 2.50×10^{-4}

Our data comprises 690,743 inpatient stays in weakly-regulated states over 2 years. Since we have a random 5% sample of all Medicare inpatient stays, we estimate that there are 6,907,430 Medicare inpatient stays per year in weakly-regulated states that meet our criteria. We compute the number of over-reported POAs for each infection (see Table 4) as:

$$[\#\text{ patients in weakly-regulated states per year}] \times [\text{excess POA rate}]$$

We also obtain estimates of Medicare's added reimbursement cost for these infections from Umscheid et al. (2011). They find that CLABSIs result in a mean estimated incremental cost of $110,800 (95% CI: $22,700 - $327,000), and CAUTIs result in a mean estimated incremental cost of $2950 (95% CI: $1200 – $4700). We use these inputs to estimate the cost burden to Medicare from upcoding (see Table 4).

Infection	Estimated # Upcoded Cases		Estimated Added Cost to Medicare	
	Estimate	95% CI	Estimate	95% CI
CLABSI	1,561	[99, 3023]	$173 million	$2.2 million – $989 million
CAUTI	9,256	[5871, 12641]	$27 million	$7.0 million – $59.4 million

Table 4 Conservative estimates are shown for the number of upcoded cases per year and the associated cost burden to Medicare for both CLABSIs and CAUTIs.

Thus, we estimate a total of 10,817 over-reported POAs (out of 58,520 annually reported POA infections from weakly-regulated states) with an associated cost burden of approximately $200 million in annual Medicare reimbursements. We note that these estimates only account for direct healthcare service costs, and do not include broader societal costs, e.g., long-term impact of HAIs on patient health and the loss of patient productivity as a result of their extended hospital stay.

5.4. Policy Comparison

We defined our original treatment variable using three reporting requirements that we considered informative. We now alter the definition of the treatment variable based on reporting requirements along three dimensions: patient, event, and cause (see Table 2). This serves two purposes:

1. We show that our results are robust to the choice of treatment variable as long as it captures the stringency of regulations on truthful reporting.

2. We draw inferences about which types of reporting requirements may be most effective at reducing upcoding rates in order to make policy recommendations.

Alternative Definitions of Treatment Variable. We construct alternative definitions of the treatment variable through the following procedure. For every combination of patient/event/cause, we consider the relevant set of reporting requirements and compute the median number implemented by states with adverse event reporting systems (see Appendix E and Table 15 for details). We define all states with more than the median number of requirements as "strongly regulated."

We also investigate an alternative definition where a strongly-regulated state is one that simply has an adverse event reporting system. These states include CA, CO, CT, DC, FL, GA, IN, KS, ME, MD, MA, MN, NJ, NV, NY, OH, OR, PA, RI, SC, SD, TN, UT, VT, WA, and WY.

For each of these definitions of the treatment variable, we ran a linear regression and a 2-SLS regression for POA and HAI outcomes respectively, as described in Sections 4.1–4.2. We list the

estimated treatment effect along with cluster-robust standard errors and p-values in Table 5. The "Original" definition refers to the measure that was defined and used earlier in the paper.

Results & Observations. First, we find that our results are largely consistent for different definitions of strong regulation that capture the magnitude of the providers' reporting burden in that state. In particular, stringent regulation on adverse event reporting is associated with reduced upcoding levels. On the other hand, merely having regulations for adverse event reporting is not associated with significant changes in upcoding rates. These findings support the hypothesis that laws cannot create proper incentives without sufficient accountability.

Treatment	**(1) POA Reports**			**(2) HAI Reports**		
Definition	**Estimate**	**SE**	p**-value**	**Estimate**	**SE**	p**-value**
Patient	-1.02×10^{-3}	2.97×10^{-4}	0.001	5.85×10^{-4}	2.38×10^{-4}	0.014
Event	-1.03×10^{-3}	3.43×10^{-4}	0.003	2.76×10^{-4}	2.42×10^{-4}	0.255
Cause	-7.01×10^{-4}	3.18×10^{-4}	0.027	6.91×10^{-4}	2.91×10^{-4}	0.018
Patient & Event	-9.97×10^{-4}	2.95×10^{-4}	0.001	5.91×10^{-4}	2.39×10^{-4}	0.013
Patient & Cause	-9.53×10^{-4}	3.09×10^{-4}	0.002	8.56×10^{-4}	3.51×10^{-4}	0.015
Event & Cause	-1.37×10^{-3}	3.51×10^{-4}	0.000	1.85×10^{-4}	2.62×10^{-4}	0.479
Patient, Event & Cause	-1.43×10^{-3}	3.15×10^{-4}	0.000	7.04×10^{-4}	3.07×10^{-4}	0.022
Has Reporting System?	-4.88×10^{-4}	2.83×10^{-4}	0.084	5.30×10^{-4}	2.86×10^{-4}	0.064
Original	-1.21×10^{-3}	3.26×10^{-4}	0.000	5.23×10^{-4}	2.16×10^{-4}	0.016

Table 5 Point estimates and cluster-robust standard errors for the coefficient of the treatment variable are shown for alternative definitions of strong state regulation.

Second, we infer that reporting patient information is most valuable, while only reporting information on the event has limited value. This may be because reporting patient information (such as the medical record number) may allow state entities to more easily audit hospital records. Our findings also suggest that reporting along all three dimensions is best; in particular, reporting patient, event, and cause information was associated with the highest reduction in upcoding.

6. Robustness Checks

We perform several robustness checks to provide further evidence justifying our assumptions and to show that that our empirical results are consistent under alternate specifications.

6.1. Provider Quality

As noted in §2, one of our key assumptions is that strongly-regulated providers have higher quality of care with respect to HAIs. We perform two robustness checks to demonstrate that strongly-regulated providers have higher quality of care even in a general medical context using the December 2010 release of Hospital Compare's risk-adjusted provider quality metrics.

Mortality. We compare risk-adjusted mortality rates (which are the most direct measure of provider quality) between providers in strongly- and weakly-regulated states (see Table 6). During this time period, Medicare reported these rates only for three conditions: heart attack, heart failure, and pneumonia. Using a t-test, we find that providers in strongly-regulated states have lower risk-adjusted mortality rates across all 3 conditions with high statistical significance. In Appendix D.2, we verify that a similar relationship holds for our instrumental variables as well as the instrumented treatment variable (see Table 13).

Condition	**Mean Mortality (Strong States)**	**Mean Mortality (Weak States)**	**95% CI of Difference**	**p-value**
Heart Attack	15.75%	16.27%	[-0.69%, -0.36%]	6.7×10^{-10}
Heart Failure	10.82%	11.36%	[-0.66%, -0.41%]	2.1×10^{-16}
Pneumonia	11.19%	11.77%	[-0.73%, -0.44%]	6.2×10^{-15}

Table 6 T-test results are shown comparing Medicare providers' risk-adjusted mortality rates in strongly vs. weakly regulated states for heart attack, heart failure, and pneumonia patients.

Process of Care. We compare all reported (risk-adjusted) process of care quality measures, except outpatient and pediatric ones (since our study is focused on the adult inpatient setting) (see Table 7). We find that strongly-regulated providers outperform weakly-regulated providers on all but one measure (where the difference is not statistically significant). The improvement in performance is statistically significant for 75% of the measures. We particularly draw attention to measures related to the appropriate administration of antibiotics (PN_5c, PN_6, SCIP_INF1, SCIP_INF2, and SCIP_INF3), which aids in infection prevention. Even more relevant, the measure SCIP_INF9 captures the appropriate and timely removal of urinary catheters after surgery, which is instrumental to CAUTI prevention (Saint 2009). Strongly-regulated providers perform significantly better on these measures compared to weakly-regulated providers. In Appendix D.2, we verify that a similar relationship holds for the instrumented treatment variable (see Table 14).

6.2. Provider POA Infection Detection

Another key assumption (discussed in §2) is that strongly-regulated providers are more likely to detect and accurately report POA infections. We perform two robustness checks to demonstrate that strongly-regulated providers exhibit (i) better performance in *time-sensitive* process of care measures (suggesting that they are more likely to correctly identify POAs by detecting the infection promptly within the allowed POA time window), and (ii) more aggressive billing behavior (which suggests that they are better at detecting and reporting reimbursable complications such as POAs).

Time-sensitivity. We compare all Hospital Compare (risk-adjusted) process of care measures that are time-sensitive, i.e. require a task to be completed within a given window of time. These

Measure	Definition	Mean (Strong)	Mean (Weak)	Better Quality?	p-value
AMI_1	Patiens given aspirin at arrival	97.3	95.3	Yes	0.00
AMI_2	Patiens given aspirin at discharge	96.3	93.8	Yes	0.00
AMI_3	Patients given ACE inhibitor for Left Ventricular Systolic Dysfunction (LVSD)	94.6	93.2	Yes	0.04
AMI_4	Patients given smoking cessation counseling	97.8	97.1	Yes	0.24
AMI_5	Patients given beta blocker at discharge	96.8	93.4	Yes	0.00
AMI_7a	Patients given fibrinolytic medication within 30 minutes of arrival	49.6	45.6	Yes	0.43
AMI_8a	Patients given PCI within 90 minutes of arrival	87.1	83.9	Yes	0.00
HF_1	Patients given discharge instructions	83.7	81.0	Yes	0.00
HF_2	Patients given an evaluation of Left Ventricular Systolic Dysfunction (LVSD)	93.5	91.3	Yes	0.00
HF_3	Patients given ACE inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)	92.2	90.1	Yes	0.00
HF_4	Patients given smoking cessation counseling	94.2	93.5	Yes	0.35
PN_2	Patients assessed and given pneumococcal vaccination	90.7	88.4	Yes	0.00
PN_3b	Patients whose initial ER blood culture was performed prior to the administration of the first hospital dose of antibiotics	93.5	93.4	Yes	0.87
PN_4	Patients given smoking cessation counseling	93.8	91.7	Yes	0.00
PN_5c	Patients given initial antibiotic(s) within 6 hours after arrival	94.2	94.0	Yes	0.41
PN_6	Patients given the most appropriate initial antibiotic(s)	90.6	89.4	Yes	0.01
PN_7	Pneumonia patients assessed and given influenza vaccination	89.8	87.3	Yes	0.00
SCIP_CARD_2	Percentage of patients who were taking beta blockers before coming to the hospital that were kept on the beta blockers before and after their surgery	92.3	87.8	Yes	0.00
SCIP_INF_1	Surgery patients who received preventative antibiotic(s) one hour before incision	94.9	93.0	Yes	0.00
SCIP_INF_2	Percentage of surgery patients who received the appropriate antibiotic(s) for their surgery	96.2	95.1	Yes	0.00
SCIP_INF_3	Surgery patients whose preventative antibiotic(s) are stopped within 24 hours after surgery	93.7	91.7	Yes	0.00
SCIP_INF_4	Cardiac surgery patients with controlled 6am post-operative blood glucose	90.3	91.8	No	0.29
SCIP_INF_6	Surgery patients with appropriate hair removal	98.5	98.4	Yes	0.65
SCIP_INF_9	Percentage of surgery patients whose urinary catheters were removed on the first or second day of surgery	87.8	85.6	Yes	0.00
SCIP_VTE_1	Surgery patients whose doctors ordered treatments to prevent blood clots for certain types of surgeries	92.8	89.3	Yes	0.00
SCIP_VTE_2	Surgery patients who received treatment to prevent blood clots within 24 hours before or after selected surgeries	91.4	87.9	Yes	0.00

Table 7 T-test results are shown comparing Medicare's risk-adjusted process of care quality measures in strongly vs. weakly regulated states for heart attack, heart failure, pneumonia, and surgical care improvement.

include AMI_7a, AMI_8a, PN_5c, SCIP_INF_1, SCIP_INF_3, SCIP_INF_9, and SCIP_VTE_2 (see Table 7). We find that strongly-regulated providers outperform weakly-regulated providers in all of these metrics, and the difference is statistically significant in over 70% of the measures.

Billing Aggressiveness. First, we compare Medicare's "payment and value of care" metrics (adjusted for variations in demographic differences in pay, patient risk, and provider mortality rates) using the December 2015 release of Hospital Compare's risk-adjusted provider quality metrics (see Table 8); we note that payment information was not made available in Hospital Compare data during the study period (2009-10). We find that strongly-regulated states receive much larger adjusted payments with high statistical significance; this difference persists after we stratify providers by quality (i.e. we restrict the comparison to providers with risk-adjusted mortality rates that are comparable to the national average). Second, we compare our own measure of "billing aggressiveness" (defined in §4.4) computed from the claims data. Using a t-test, we find that providers in strongly-regulated states have higher billing aggressiveness (mean: 4.42) compared to weakly-regulated states (mean: 4.10) with high statistical significance (p-value: 0.003). Thus, we find significant evidence that strongly-regulated providers are more aggressive in claims reporting, and furthermore, succeed at receiving higher payments for equal quality of services provided.

Condition	**Mortality**	**Mean Payment**		p**-value**
		Strong	**Weak**	
Heart Attack	All	\$22,492	\$22,005	2.5×10^{-8}
	Average	\$22,494	\$21,997	2.2×10^{-8}
Heart Failure	All	\$15,815	\$15,231	4.1×10^{-16}
	Average	\$15,738	\$15,183	7.6×10^{-14}
Pneumonia	All	\$14,411	\$14,248	2.8×10^{-3}
	Average	\$14,381	\$14,220	5.7×10^{-3}
All infections	All	1.00	0.98	1.8×10^{-9}

Table 8 The t-test results compare Medicare's adjusted payments to strongly vs. weakly-regulated providers for selected infections over both (i) all providers and (ii) providers with mortality rates comparable to the national average. The last row shows the ratio of adjusted payments to the national average for all Medicare patients.

6.3. Regression Specification

In addition to checking the robustness of our results to different definitions of the treatment variable (§5.4), we also ensure that our results are consistent under alternative regression specifications:

1. Since the outcomes are binary, we use a probit model specification rather than a linear model.

2. We use a continuous (rather than binary) definition of the treatment variable, i.e. the number of total reporting requirements (out of 14) adopted by each state (see Fig 2 in Appendix A).

3. We use a two-level definition of the treatment variable, i.e. $S = \{0, 1, 2\}$ corresponds to no adverse event reporting regulation, weak regulation, and strong regulation respectively.

4. We employ coarser state-level (rather than provider-level) clustering of standard errors. Coarser clustering is believed to yield more conservative estimates (Cameron and Miller 2015). However, since there are only 50 states and we have over 70 controls, we exclude the Elixhauser and diagnosis controls to ensure that the estimation is not rank deficient.

We redo our POA and HAI analyses under each of these alternative specifications. Again, we find that our results are consistent (see Table 9).

Change in Specification	**(1) POA Reports**			**(2) HAI Reports**		
	Estimate	**SE**	p**-value**	**Estimate**	**SE**	p**-value**
Probit Model	-0.07	0.019	0.000	0.19	0.076	0.015
Continuous Treatment	-8.35×10^{-5}	3.22×10^{-5}	0.010	5.03×10^{-5}	2.22×10^{-5}	0.023
Two-level Treatment	-5.71×10^{-4}	1.81×10^{-4}	0.002	2.82×10^{-4}	1.27×10^{-4}	0.026
State-Level Clustering	-1.66×10^{-3}	5.65×10^{-4}	0.005	5.29×10^{-4}	2.35×10^{-4}	0.024

Table 9 Point estimates and cluster-robust standard errors for the coefficient of the treatment variable are shown for alternative specifications of the POA and HAI regressions.

7. Discussion & Concluding Remarks

In summary, our results show that providers in states with stronger regulations on adverse event reporting have (1) lower (risk-adjusted) POA reporting rates and (2) higher (risk-adjusted) HAI reporting rates for CLABSIs and CAUTIs. This effect is statistically significant even after controlling for a wide range of patient risk factors and demographic characteristics, as well as arbitrary intra-provider correlations and endogeneity of regulation for HAI outcomes. The differential impact of state-level regulation on HAI and POA reporting rates strongly suggests that weakly-regulated providers upcode. If we further assume that the bias from unobservable patient risk confounders in our analysis is negligible, we can make the stronger claim that weakly-regulated providers are over-reporting POAs; in particular, we conservatively estimate that over 10,000 POA infections are over-reported annually, i.e. 18.5% of POAs reported by weakly-regulated states are actually HAIs. This result is similar to that reported by Meddings et al. (2010); they find that out of 80 manually reviewed CAUTI medical records, 18 cases (22.5%) were over-reported as POAs. We estimate that the resulting reimbursement burden for Medicare is approximately $200 million a year.

7.1. Why does upcoding occur?

So far, we have focused on detecting whether upcoding happens, and have ignored the question of why and how it may occur. While a rigorous understanding is beyond the scope of this paper, we gained some insight based on discussions with 17 hospital staff that are part of a quality reporting

team at a nearby major hospital[3]. This team, which consists primarily of nurses, is tasked with catching potential coder errors and relaying the information to compliance teams. They described some difficulties that make coding a particularly error-prone process:

1. Coders do not have the medical training to interpret medical records. Rather, they rely on doctor's notes, which are often sparse and incomplete. This claim is supported by the discussion in Meddings et al. (2010) based on their conversations with hospital coders:

"In discussion with hospital coders, we learned that hospital coders are instructed to obtain diagnosis information for payment purposes only from provider notes and not from nursing notes. If a hospital coder does review nursing notes and suspects a diagnosis that is not apparent from provider notes, the hospital coder must then verify the diagnosis with a provider, and the provider would need to change the provider documentation to reflect this additional diagnosis. However, reviewing nursing notes for potential diagnoses that then necessitate communication and additional documentation from a provider is a resource-intensive step."

2. Exacerbating the previous issue, the definitions used for differentiating and reporting HAIs and POAs vary by organization. For example, the same hospital must report the occurrence of each HAI to Medicare, the National Healthcare Safety Network (NHSN), and other patient safety organizations (e.g., state adverse event reporting system); each of these organizations may have different definitions for HAIs vs. POAs, making accurate reporting both onerous and confusing.

3. In some cases, doctors skip or delay the step of performing (blood or urine) cultures (which are required to definitively claim that a patient has an infection), and simply place patients on antibiotics. This may result in failure to attribute these infections correctly (with respect to claims reporting) even if the infection has been treated appropriately. In such cases, doctors may feel that they are unfairly blamed under the HAI legislation. Prior medical literature has argued that such policies may lead providers to "game" the system (e.g., upcode) if they feel that the penalties are unfair (Morreim 1991, Werner et al. 2002).

The hospital we visited has invested significantly in improving coding accuracy by hiring a large quality reporting team. They address these issues by double-checking coder reports and encouraging doctors to always perform cultures when a new patient is admitted. However, they pointed out that in general, there is little financial incentive for hospitals to make such investments. For example, hiring and training an employee to oversee quality can cost significantly more per year than paying a HAI penalty. Consequently, investment in coding quality may vary greatly across hospitals.

In hospitals without substantial investment in quality, coders have to make decisions based on very sparse evidence. In principle, according to the regulation, coders should conservatively report

[3] We redact the name of the hospital to preserve the anonymity of our institutional affiliation.

unclear cases as HAIs. However, due to the combination of financial and reputation penalties, coders may face pressure from hospital administration to upcode claims when the data is uncertain. This problem is exacerbated by the prevalent use of software that "optimizes claims" for the highest possible reimbursement; such tools often auto-fill claims reports with higher-paying diagnoses (which the coder may then revise). This may enable upcoding in the presence of uncertainty since the coder may simply choose not to revise the default (higher-paying) option.

7.2. Harmful Effects of Upcoding

Our work suggests that financial incentives alone are not sufficient to reduce HAI incidence; these policies must be accompanied by regulation to enforce truthful reporting. This hypothesis is supported by recent evidence that the nonpayment policy has not reduced HAI rates (Lee 2012). In fact, increasing financial incentives (e.g. HAC Reduction Program) or reputation incentives (e.g. published infection rates on Hospital Compare) may worsen the problem as providers may simply increase their rate of upcoding. Increased upcoding would have a number of negative consequences:

1. Truthful providers are unfairly penalized and face greater financial pressure to upcode as well
2. Upcoding biases medical records resulting in a loss of accurate information. This interferes with tracking harmful infections and evaluating the effectiveness of policies aimed at improving quality (Saint 2009)
3. Publishing biased quality metrics may harm patients by routing them to providers who are engaging in upcoding rather than providing better quality of care

Thus, we recommend that CMS implement measures to enforce truthful reporting by providers.

7.3. Policy Recommendations

To this end, our results suggest two policy recommendations to help mitigate upcoding. First, we suggest that CMS perform targeted audits of providers with high (risk-adjusted) POA-to-HAI reporting ratios. As discussed in §2, providers with higher risk-adjusted POA reporting rates and lower risk-adjusted HAI reporting rates are more likely to be engaging in HAI upcoding. This approach can complement existing audits conducted by Medicare. In general, targeted auditing has been a profitable strategy for the government: the OIG finds that recently, for every $1 spent on health care fraud control, the federal government has returned $6.80 (Taebel 2012). Second, we recommend that the federal government implement certain features of current state-level regulations that seem to be effective at eliciting truthful reporting. Our analysis establishes the effect of stronger regulation on decreased upcoding, and helps isolate some of the state adverse event reporting system features that were successful in reducing upcoding. These include reporting patient-identifying information (medical record number or billing number), a detailed description of the adverse event, as well as the identified root cause of the adverse event. On the other hand, we note that simply

having a reporting system without stringent requirements produced no significant effect on reporting rates; we find that it is crucial that the regulation creates sufficient provider accountability. We hypothesize that simply requiring providers to report detailed information on how and why an adverse event occurred forces providers to implement the necessary infrastructure for detecting and preventing HAIs. Moreover, reporting more detailed information increases the threat of setting off red flags when upcoding, and thus possibly diminishes the rate of upcoding. CMS may benefit by implementing such detailed information reporting requirements in addition to existing financial incentives to help improve hospital infrastructure and truthful reporting nationally.

More broadly, we emphasize the importance of taking measures to mitigate upcoding as Medicare moves towards adopting a growing number of data-driven pay-for-performance policies (HHS 2015). To support these efforts, we recommend that Medicare choose performance criteria that are not only representative of patient outcomes but are also easily and cheaply verifiable.

Acknowledgments

The authors gratefully acknowledge Centers for Medicare & Medicaid Services (CMS) for providing us with our dataset and Stefanos Zenios for assistance with obtaining this data. This material is based upon work supported by a Stanford Graduate School of Business grant and the following National Science Foundation (NSF) grants: Graduate Research Fellowship Grant No. DGE-114747, NSF EAGER Award CMMI:1451037, and NSF grant CCF:1216011.

This paper has also benefitted from valuable feedback from Jay Bhattacharya, Doug Owens, David Studdert, Stefanos Zenios, various seminar participants, as well as an anonymous review team. They have been instrumental in guiding us to improve the paper, often in ways that would not have happened without them.

References

ANA, American Nursing Association. 2016. Nursing world. URL `http://nursingworld.org/MainMenuCategories/ThePracticeofProfessionalNursing/NurseStaffing`.

Arrow, K. 1963. Uncertainty and the welfare economics of medical care. *The American Economic Review* **53**(5).

Ashby, Nathan J, Fred McMahon, Avilia Bueno. 2010. *Economic Freedom of North America 2010*. Fraser Institute Vancouver, BC.

Ata, Baris, Bradley L Killaly, Tava Lennon Olsen, Rodney P Parker. 2013. On hospice operations under medicare reimbursement policies. *Management Science* **59**(5) 1027–1044.

Baiman, Stanley, Paul E Fischer, Madhav V Rajan. 2000. Information, contracting, and quality costs. *Management Science* **46**(6) 776–789.

Baum, C. F. 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal* **7**(4) 465–506(42).

Berenholtz, Sean M et al. 2004. Eliminating catheter-related bloodstream infections in the intensive care unit. *Critical care medicine* **32**(10) 2014–2020.

Berriel-Cass, Dorine et al. 2006. Eliminating nosocomial infections at ascension health. *Joint Commission Journal on Quality and Patient Safety* **32**(11) 612–620.

Bolton, Patrick, Mathias Dewatripont. 2004. *Contract Theory*. The MIT Press.

Brown, Jason, Mark Duggan, Ilyana Kuziemko, William Woolston. 2014. How does risk selection respond to risk adjustment? new evidence from the medicare advantage program. *The American Economic Review* **104**(10) 3335–3364.

Cameron, A Colin, Douglas L Miller. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* **50**(2) 317–372.

Center for Medicare & Medicaid Services. 2014. Statute regulations program instructions. URL `http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/`. Last Accessed March 14, 2014.

Deyo, Richard, Daniel Cherkin, Marcia Ciol. 1992. Adapting a clinical comorbidity index for use with icd-9-cm administrative databases. *Journal of Clinical Epidemiology* **45** 613–619.

Duffin, Christian. 2014. Increase in nurse numbers linked to better patient survival rates in icu. *Nursing Standard* **28**(33).

Elixhauser, Anne, Claudia Steiner, D Robert Harris, Rosanna M Coffey. 1998. Comorbidity measures for use with administrative data. *Medical care* **36**(1) 8–27.

Fuloria, Prashant C, Stefanos A Zenios. 2001. Outcomes-adjusted reimbursement in a health-care delivery system. *Management Science* **47**(6) 735–751.

Geruso, Michael, Timothy Layton. 2015. Upcoding or selection? evidence from medicare on squishy risk adjustment. *NBER Working Paper* .

Green, Linda V. 2012. Operations management forum-the vital role of operations analysis in improving healthcare delivery. *Manufacturing & Service Operations Management* **14**(4) 488–494.

Health Watch USA. 2011. Medicare's payment policy for hospital acquired conditions. URL `http://www.healthwatchusa.org/HWUSA-Initiatives/PDF-Downloads/20110705-HAC-Medicare-Non-Payment-comments.pdf`. Last Accessed April 6, 2015.

Heese, Jonas, Ranjani Krishnan, Frank Moers. 2015. Regulator leniency and mispricing in beneficent nonprofits. *HBS Working paper.* .

HHS. 2015. Better, smarter, healthier: In historic announcement, hhs sets clear goals and timeline for shifting medicare reimbursements from volume to value. URL `http://www.hhs.gov/about/news/2015/01/26/better-smarter-healthier-in-historic-announcement-hhs-sets-clear-goals-and-timeline-for-shifting-r`
`html`.

Himmelstein, David, Steffie Woolhandler. 2015. Quality improvement: 'become good at cheating and you never need to become good at anything else'. URL http://healthaffairs.org/blog/2015/08/27/quality-improvement-become-good-at-cheating-and-you-never-need-to-become-good-at-anything-else/.

Hsu, Eugene et al. 2014. Doing well by doing good assessing the cost savings of an intervention to reduce central line associated bloodstream infections in a hawaii hospital. *American Journal of Medical Quality* **29** 13–19.

KC, Diwas Singh, Christian Terwiesch. 2011. The effects of focus on performance: Evidence from california hospitals. *Management Science* **57**(11) 1897–1912.

Lee, Grace et al. 2012. Effect of nonpayment for preventable infections in us hospitals. *New England Journal of Medicine* **367**(15) 1428–1437.

Levinson, Daniel R. 2008. Adverse events in hospitals: State reporting systems. URL https://oig.hhs.gov/oei/reports/oei-06-07-00471.pdf.

Magill, Shelley et al. 2014. Multistate point-prevalence survey of health care–associated infections. *New England Journal of Medicine* **370**(13) 1198–1208.

Mark, Barbara A, David W Harless. 2010. Nurse staffing and post-surgical complications using the present on admission indicator. *Research in nursing & health* **33**(1) 35–47.

Martin, Anne et al. 2012. Growth in us health spending remained slow in 2010; health share of gross domestic product was unchanged from 2009. *Health Affairs* **31**(1) 208–219.

McNair, Peter, Harold Luft, Andrew Bindman. 2009. Medicare's policy not to pay for treating hospital-acquired conditions: The impact. *Health Affairs* **28** 1458–1493.

Meddings, Jennifer, Sanjay Saint, Laurence McMahon. 2010. Hospital-acquired catheter-associated urinary tract infection: documentation and coding issues may reduce financial impact of medicare's new payment policy. *Infection Control and Hospital Epidemiology* **31** 627–633.

Morreim, E Haavi. 1991. Gaming the system: dodging the rules, ruling the dodgers. *Archives of Internal Medicine* **151**(3) 443–447.

Mukamel, Dana et al. 2012. The effect of state regulatory stringency on nursing home quality. *Health services research* **47**(5) 1791–1813.

Plambeck, Erica L, Terry A Taylor. 2015. Supplier evasion of a buyer's audit: Implications for motivating supplier social and environmental responsibility. *Manufacturing & Service Operations Management* .

Saint, Sanjay et al. 2009. Catheter-associated urinary tract infection and the medicare rule changes. *Annals of Internal Medicine* **150**(12) 877–884.

Schoen, Cathy et al. 2013. Access, affordability, and insurance complexity are often worse in the united states compared to ten other countries. *Health Affairs* **32**(12) 2205–2215.

Schuller, Kristin, Janice Probst, James Hardin, Kevin Bennett, Amy Martin. 2014. Initial impact of medicare's nonpayment policy on catheter-associated urinary tract infections by hospital characteristics. *Health Policy* **115** 165–171.

Scott, R Douglas II. 2009. *The direct medical costs of healthcare-associated infections in US hospitals and the benefits of prevention*. Division of Healthcare Quality Promotion National Center for Preparedness, Detection, and Control of Infectious Diseases, Centers for Disease Control and Prevention.

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

Snow, Catherine et al. 2012. Accuracy of coding in the hospital-acquired conditions - present on admission program. *Centers of Medicare and Medicaid Services* .

Taebel, Scott. 2012. Oig hospital compliance audits. URL `http://www.ehcca.com/presentations/racsummit6/taebel_2.pdf`.

Umscheid, Craig, Matthew Mitchell, Jalpa Doshi, Rajender Agarwal, Kendal Williams, Patrick Brennan. 2011. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infection Control and Hospital Epidemiology* **32** 101–114.

Werner, Rachel, Caleb Alexander, Angela Fagerlin, Peter Ubel. 2002. The hassle factor: What motivates physicians to manipulate reimbursement rules? *Archives of Internal Medicine* **162**(10) 1134–1139.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT Press.

Appendix

A. Summary Statistics

Variable	Strong			Weak		
	All	**POA**	**HAI**	**All**	**POA**	**HAI**
# Observations	230,794	1,533	198	690,743	5,852	538
sex	59%	47%	65%	60%	50%	62%
race: white	81%	76%	75%	81%	74%	79%
race: black	13%	17%	19%	14%	20%	17%
race: asian	.85%	1.0%	.51%	1.1%	.94%	1.5%
race: hispanic	3.5%	3.7%	4.0%	1.9%	2.5%	1.5%
race: native american	.28%	.20%	0%	.68%	.58%	.56%
race: other	1.4%	1.6%	1.0%	1.0%	1.2%	.56%
race: unknown	.32%	.20%	.51%	.21%	.26%	0.0%
age	75	71	73	74	70	70
	(14)	(16)	(16)	(14)	(16)	(15)
charlson score	1.8	2.7	2.1	1.8	2.6	2.2
	(2.3)	(2.6)	(2.4)	(2.3)	(2.6)	(2.5)
days since last admit	89	56	65	91	61	64
	(75)	(65)	(71)	(75)	(67)	(69)
# past admits	1.7	2.8	2.1	1.6	2.6	2.1
	(2.0)	(2.5)	(2.3)	(1.9)	(2.4)	(2.0)
# past procedures	2.2	4.4	2.9	1.9	3.9	3.0
	(3.6)	(5.1)	(3.8)	(3.3)	(4.8)	(3.9)
# past cauti	.0049	.12	.025	.0068	.17	.011
	(.078)	(.40)	(.19)	(.099)	(.52)	(.12)
# past clabsi	.0065	.12	.045	.0071	.092	.022
	(.095)	(.44)	(.23)	(.098)	(.39)	(.16)
total past length of stay	16	40	28	15	33	22
	(34)	(53)	(48)	(31)	(47)	(33)
billing aggressiveness	4.9	4.9	4.9	4.7	4.8	4.7
	(2.4)	(2.3)	(2.4)	(2.5)	(2.7)	(2.3)
demographics: % uninsured	13%	13%	13%	15%	15%	15%
demographics: % above 65	16%	16%	18%	14%	14%	13%
demographics: % foreign-born	17%	16%	16%	8.9%	9.6%	9.6%
demographics: % unemployed	9.5%	9.9%	9.6%	9.7%	9.9%	9.9%
demographics: % near poverty	68%	68%	69%	68%	69%	68%
demographics: % no high school	13%	13%	14%	15%	15%	15%
local household income	$29,400	$28,600	$27,800	$25,200	$27,000	$25,700
	($19,600)	($18,200)	($18,400)	($17,600)	($17,900)	($16,700)
local female life expectancy	81	81	81	80	80	80
	(1.3)	(1.4)	(1.5)	(1.8)	(1.7)	(1.8)
local male life expectancy	77	76	77	75	75	75
	(1.7)	(1.8)	(1.9)	(2.4)	(2.3)	(2.5)
economic freedom index: area 1	7.4	7.5	7.5	7.1	7.1	7.1
	(.32)	(.34)	(.33)	(.84)	(.86)	(.76)
economic freedom index: area 2	5.7	5.7	5.6	6.2	6.2	6.2
	(.37)	(.38)	(.35)	(.55)	(.54)	(.56)
economic freedom index: area 3	7.0	7.0	7.0	7.2	7.2	7.2
	(.58)	(.55)	(.55)	(.62)	(.63)	(.64)

Table 10 Summary statistics for selected variables. Standard deviations are shown in parentheses.

The heat map below (Fig. 2) illustrates the strength of adverse event reporting regulation in 2008 across the continental United States using data from Levinson (2008).

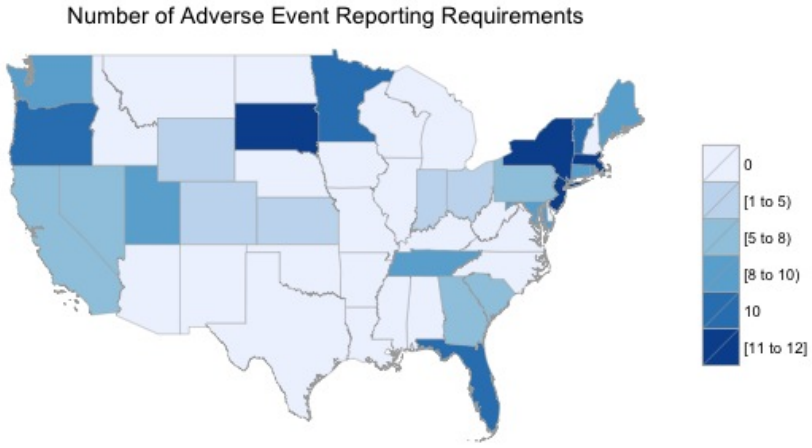

Figure 2 Heat map depicting the number of reporting requirements implemented by each state's adverse event reporting system (before 2008) across the continental US using data from Levinson (2008).

B. Proof of Proposition 1

Assume H_1' is true. Then, $r^S_{POA} < r^W_{POA}$, so we have

$$\begin{aligned} p^S(1-\epsilon_1^S-\epsilon_2^S)+O^S &< p^W(1-\epsilon_1^W-\epsilon_2^W)+O^W \\ &\leq p^W(1-\epsilon_1^S-\epsilon_2^S)+O^W \\ \iff (p^S-p^W)(1-\epsilon_1^S-\epsilon_2^S) &< O^W-O^S \end{aligned}$$

using assumption (2) on POA infection detection that $\epsilon_1^S \leq \epsilon_1^W$ and $\epsilon_2^S \leq \epsilon_2^W$. Thus, since $1-\epsilon_1^S-\epsilon_2^S \geq 0$, we either have that (1) $O^W > O^S$ (there is increased POA over-reporting in weakly-regulated states), or that (2) $p^W > p^S$ (our risk-adjustment is biased and patients in weakly-regulated states are more susceptible to infection).

Consider the case where $O^W \leq O^S$ (which implies $p^W > p^S$). We also have $r^S_{HAI} > r^W_{HAI}$ from H_1', so we can write

$$\begin{aligned} \alpha^W p^W - O^W - U^W + p^W\epsilon_1^W &< \alpha^S p^S - O^S - U^S + p^S\epsilon_1^S \\ &< \alpha^W p^W - O^S - U^S + p^W\epsilon_1^W \end{aligned}$$

Using assumption (1) on provider quality that $\alpha^S \leq \alpha^W$, we can cancel terms to get $O^W + U^W > O^S + U^S$, i.e., the rate of improper HAI reporting is higher in weakly-regulated states. This proves our first statement that H_1' implies H_1.

Now, consider the case where we further assume $p^S = p^W$. Then, applying assumption (1) gives

$$\begin{aligned} r^W_{POA} - r^S_{POA} &= p^W(1-\epsilon_1^W-\epsilon_2^W) - p^S(1-\epsilon_1^S-\epsilon_2^S) + O^W - O^S \\ &\leq O^W - O^S \leq O^W \quad \square \end{aligned}$$

C. Data Sample Selection

The data sample construction for our regression analyses is detailed in Table 11. We are interested in data from 2009-10, since the claims data only begins to distinguish between POAs and HAIs after the policy was implemented in late 2008. We also restrict our sample to US providers under the prospective payment system since billing procedures may vary otherwise. Similarly, we restict our sample to "short stays" (as defined in the claims data) since providers who treat "long stay" patients are typically not hospitals and are subject to different billing procedures as well. Another concern is that if a patient is newly enrolled in Medicare (e.g., switched from Medicare Advantage), their past visits may be censored (thus, creating bias in controls computed on past visits); we address this by restricting our sample to stays where the patient had at least one prior visit in the last 2 years (which ensures that they were recently enrolled in Medicare). We choose a two-year window because it is the longest possible window in our dataset for patients in our sample (2009-10). Finally, we restrict our sample to those patient and providers for whom we have demographic controls (from zipcode-level census data) and provider-level controls (from Hospital Compare data).

Criteria	**# Observations**	**# POA**	**# HAI**
5% Random Sample (2007-2010)	3,865,733	-	-
5% Random Sample (2009-2010)	1,939,552	8,639	1,993
US providers	1,914,704	8,615	1,984
Short stays (as defined by Medicare)	1,570,400	8,458	1,265
Stays under prospective payment	1,473,135	8,372	1,021
Stays with prior visits in last 2 years	948,495	7,578	759
Merge demographic controls	925,397	7,412	738
Merge provider controls	924,380	7,405	738
Merge instrumental variables	921,537	7,385	736

Table 11 Data sample construction for regression analyses.

D. Instrumental Variable Analysis

D.1. First-Stage Regression

The results of the first-stage regression are show in Table 12. The total $R^2 = 0.54$, and the instruments (economic freedom indices) alone had a partial $R^2 = 0.17$. This produced a first-stage F-statistic of 128.12 with a corresponding p-value of 0, indicating that our chosen instruments are not weak.

We note that although the Area 3 economic freedom index has a negative pairwise correlation with the treatment variable, the regression coefficient of this instrument in the first-stage regression is positive. This is because our three instruments are correlated, i.e., the area 3 EFI is positively correlated with the treatment variable after conditioning on the other two economic freedom indices. However, the IV estimator does not require the instruments to be uncorrelated with each other, so this does not affect the correctness of our econometric analysis.

Variable	Estimate	SE
(Intercept)	-4.06^{***}	4.62×10^{-1}
sex	1.26×10^{-3}	1.53×10^{-3}
age	1.49×10^{-5}	1.08×10^{-4}
charlson score	-7.00×10^{-4}	5.77×10^{-4}
(other patient controls)	(omitted)	(omitted)
demographics: % uninsured	-1.36^{***}	1.37×10^{-1}
demographics: % above 65	$2.99 \times 10^{-1}{}^{***}$	8.35×10^{-2}
demographics: % foreign-born	$6.85 \times 10^{-1}{}^{***}$	8.09×10^{-2}
demographics: % unemployed	$6.03 \times 10^{-1}{}^{***}$	1.14×10^{-1}
demographics: % near poverty	1.07×10^{-2}	2.82×10^{-2}
demographics: no high school	3.97×10^{-2}	8.93×10^{-2}
local household income	$-1.19 \times 10^{-6}{}^{***}$	2.83×10^{-7}
local female life expectancy	$1.47 \times 10^{-1}{}^{***}$	1.17×10^{-2}
local male life expectancy	$-9.29 \times 10^{-2}{}^{***}$	8.39×10^{-3}
economic freedom index: area 1	$1.45 \times 10^{-1}{}^{***}$	1.07×10^{-2}
economic freedom index: area 2	$-3.71 \times 10^{-1}{}^{***}$	2.17×10^{-2}
economic freedom index: area 3	$1.10 \times 10^{-1}{}^{***}$	1.98×10^{-2}

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 12 Results of first-stage regression for instrumental variable analysis. Point estimates and cluster-robust standard errors (SE) of coefficients for regression of treatment variable S against economic freedom indices (coefficients of some patient risk variables omitted due to space constraints).

D.2. Quality Comparison

Since, we are using the instrumented regulation variable rather than the true regulation variable in our HAI regression, we repeat our robustness checks to ensure that provider-level quality metrics are positively correlated with the instrumented regulation variable. First, we compare risk-adjusted mortality rates against each of the three instruments and the instrumented strong regulation variable (see Table 13). Using a Pearson correlation test, we find similar results:

1. mortality rates are negatively correlated with the Area 1 economic freedom index (which is positively correlated with strong regulation),

2. mortality rates are positively correlated with the Area 2 and 3 economic freedom indices (which are negatively correlated with strong regulation), and

3. mortality rates are negatively correlated with the instrumented strong regulation variable.

Condition	EFI: Area 1		EFI: Area 2		EFI: Area 3		Instrumented S	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Heart Attack	-0.085***	0.00	0.018	0.36	0.040**	0.03	-0.094***	7×10^{-7}
Heart Failure	-0.066***	0.00	0.009	0.57	-0.001	0.93	-0.072***	8×10^{-6}
Pneumonia	-0.041***	0.01	0.008	0.61	0.028*	0.08	-0.047***	3×10^{-3}

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 13 Pearson correlation test results are shown for Medicare providers' risk-adjusted mortality rates and our three instruments (economic freedom indices) as well as the instrumented treatment variable for heart attack, heart failure, and pneumonia patients.

Measure	Definition	Correlation with Instrumented Reg.	Better Quality?	p-value
AMI_1	Patiens given aspirin at arrival	0.089	Yes	0.00
AMI_2	Patiens given aspirin at discharge	0.085	Yes	0.00
AMI_3	Patients given ACE inhibitor for Left Ventricular Systolic Dysfunction (LVSD)	0.056	Yes	0.00
AMI_4	Patients given smoking cessation counseling	0.053	Yes	0.01
AMI_5	Patients given beta blocker at discharge	0.098	Yes	0.00
AMI_7a	Patients given fibrinolytic medication within 30 minutes of arrival	0.031	Yes	0.49
AMI_8a	Patients given PCI within 90 minutes of arrival	0.019	Yes	0.46
HF_1	Patients given discharge instructions	0.062	Yes	0.00
HF_2	Patients given an evaluation of Left Ventricular Systolic Dysfunction (LVSD)	0.082	Yes	0.00
HF_3	Patients given ACE inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)	0.094	Yes	0.00
HF_4	Patients given smoking cessation counseling	0.035	Yes	0.03
PN_2	Patients assessed and given pneumococcal vaccination	0.038	Yes	0.01
PN_3b	Patients whose initial ER blood culture was performed prior to the administration of the first hospital dose of antibiotics	0.018	Yes	0.25
PN_4	Patients given smoking cessation counseling	0.030	Yes	0.05
PN_5c	Patients given initial antibiotic(s) within 6 hours after arrival	-0.004	No	0.81
PN_6	Patients given the most appropriate initial antibiotic(s)	0.067	Yes	0.00
PN_7	Pneumonia patients assessed and given influenza vaccination	0.048	Yes	0.00
SCIP_CARD_2	Percentage of patients who were taking beta blockers before coming to the hospital that were kept on the beta blockers before and after their surgery	0.064	Yes	0.00
SCIP_INF_1	Surgery patients who received preventative antibiotic(s) one hour before incision	0.001	Yes	0.96
SCIP_INF_2	Percentage of surgery patients who received the appropriate antibiotic(s) for their surgery	0.038	Yes	0.02
SCIP_INF_3	Surgery patients whose preventative antibiotic(s) are stopped within 24 hours after surgery	0.034	Yes	0.04
SCIP_INF_4	Cardiac surgery patients with controlled 6am post-operative blood glucose	-0.053	No	0.06
SCIP_INF_6	Surgery patients with appropriate hair removal	0.007	Yes	0.68
SCIP_INF_9	Percentage of surgery patients whose urinary catheters were removed on the first or second day of surgery	0.012	Yes	0.48
SCIP_VTE_1	Surgery patients whose doctors ordered treatments to prevent blood clots for certain types of surgeries	0.038	Yes	0.02
SCIP_VTE_2	Surgery patients who received treatment to prevent blood clots within 24 hours before or after selected surgeries	0.058	Yes	0.00

Table 14 Pearson correlation test results are shown comparing Medicare's risk-adjusted process of care quality measures and the instrumented propensity to be a strongly-regulated state for heart attack, heart failure, pneumonia, and surgical care improvement.

Second, we compare all reported (risk-adjusted) process of care quality measures against the instrumented strong regulation variable (see Table 14). Similar to our earlier findings for the un-instrumented regulation variable, we find that the instrumented regulation variable is correlated with higher quality in all but two measures (the negative correlation is not statistically significant for both measures), and the improvement in performance is statistically significant for 75% of the measures.

D.3. Conservative Estimates

In this section, we argue that if our instrument is negatively correlated with the error term and positively correlated with our treatment variable, then our treatment effect estimate will be conservative. Since the Area 2 and 3 indices are negatively correlated with the treatment variable, we will consider the negative of these two indices to be the instruments under consideration. Recall that the error term in HAI reporting rates is poor (unobserved) HAI-specific provider quality (since this is positively correlated with the outcome variable of HAI reporting rates). As shown in the previous section (Appendix D.2), we expect our instruments (with sign defined as above) to be negatively correlated with this error term (poor provider quality).

For simplicity, consider a simple one-dimensional model with a single instrument (the argument easily generalizes to higher dimensions). We have:

$$y_i = \beta x_i + \epsilon_i \,,$$

where y_i is the dependent variable, x_i is an independent variable, ϵ_i is an unobserved error term, and β is a scalar coefficient that we seek to estimate. If x and ϵ are correlated, we use an instrument z. This gives us the IV estimator

$$\hat{\beta}_{IV} = \beta + \frac{z^T \epsilon}{z^T x} \,.$$

Clearly, if z is uncorrelated with ϵ, then the IV estimate is a consistent estimator of β (Wooldridge 2010). We consider the case where z may be negatively correlated with ϵ, i.e., the second term

$$\frac{z^T \epsilon}{z^T x} < 0 \,.$$

(Recall that we defined the instruments so that they are positively correlated with the treatment variable, i.e., $z^T x > 0$.) Then, it follows that $\hat{\beta}_{IV} < \beta$. Thus, if $\hat{\beta}_{IV} > 0$ (as we find in the HAI regression), then the true estimate also satisfies $\beta > 0$.

E. Robustness Checks

We now describe the construction of alternative treatment definitions (used for robustness checks in §5.4). For every combination of patient/event/cause reporting categories, we first consider the relevant set of reporting requirements (see Table 15 below). We then compute the median number implemented by the states with adverse event reporting systems, and define all states with more than the median number of requirements as "strongly regulated."

Treatment Definition	Median Requirements	# States
Patient	2 out of 5	11
Event	4 out of 6	11
Cause	1 out of 3	11
Patient & Event	6 out of 11	12
Patient & Cause	3 out of 8	12
Event & Cause	6 out of 9	9
Patient, Event, & Cause	8 out of 14	10

Table 15 Different definitions of the treatment variable based on the number of reporting requirements along three dimensions (patient, event, and cause), as well as the number of states that satisfied this infection.