Hospital Upcoding Decisions under Medicare Audits

Jianjing Lin
Rensselaer Polytechnic Institute
Department of Economics

Juan Pantano*
University of Arizona
Department of Economics

March 9, 2023

Abstract

We develop a general model to explain hospitals' coding practices, that encompasses some of the major mechanisms advanced in the literature and propose a new channel that moderates the role of financial incentives in driving these upcoding decisions. We then derive testable implications and leverage aggregate Medicare Part A claims data from the past decade (2012-2019) to explore which of these mechanisms best explains hospitals' admissions and billing behavior in this recent period. We also develop a novel empirical strategy based on a random sample of audits that further helps us distinguish among the alternative mechanisms. This idea, of using audits data to sort out competing hypotheses in settings with potentially inappropriate but hidden actions has broader appeal, and could be applied in other contexts. Using the evidence from both aggregate claims and audit microdata we conclude that hospital decision-makers do engage in upcoding practices, but also seem to derive some utility from engaging in partially-compliant behavior. In particular, their demand for compliance is subject to substitution and income effects arising from changes in reimbursement, with the income effect likely being dominant during our study period.

*Corresponding Author: juanpantano@gmail.com. We thank Amanda Cook, Keith Ericson, John Friedman, Keith Joiner, Kelli Marquardt and Anthony Lo Sasso for helpful comments.

1 Introduction

Medicare, on average, processes 4.5 million claims and makes $1 billion in payments every day.[1] Given the increasing healthcare expenditure and a growing beneficiary population, it is important to protect the integrity of the Medicare program. Hospitals inpatient care is the largest contributor to Medicare expenditure and therefore, it is important to examine the prevalence of improper payments. The Centers for Medicare and Medicaid Services define improper payments as "any payment that should not have been made or that was made in an incorrect amount (including over-payments and under-payments)" (CERT Program, 2016). A strand of literature has been devoted to exploring to what extent hospitals engage in inappropriate billing practices and what are the mechanisms underlying their behavior. Prior studies found mixed evidence of improper billing and provided different explanations. We revisit this topic by proposing a general model that characterizes hospital coding decisions and seek to identify the most prevalent mechanism by combining the analysis from aggregate claim counts and more novel data from a random sample of audits.

The primary source for hospital revenue comes from the provision of inpatient care, where hospitals consider what services are provided to treat the patient. The practice by which a hospital seeks to enhance its revenue by coding services at a level higher than it is medically justified, is known as *upcoding*. Hospitals that are found to engage in these activities may undergo financial and other consequences. In this paper, we explore the prevalence of these activities and the various incentives behind them.

Following the theoretical framework by McGuire and Pauly (1991) and Gruber and Owings (1996), we construct a model where hospital decision makers must decide on admission and coding assignment for a group of patients sharing a common primary diagnosis or procedure—a *base* diagnostic related group (DRG). These models encompass profit maximization as a special case, but in general allow the utility function of decision-makers to play a role. While these models are typically used to model the behavior of individual physicians, we argue that it is worthwhile to consider similar ideas in the context of understanding how hospitals make decisions, assuming hospital administrators and clinical staff share a common goal. In particular, we propose a model in which hospitals' key decision-makers aim to maximize a utility function that depends positively on consumption and negatively on the level of non-compliance associated with their improper admissions and billing practices. Consumption is funded by (legitimate and improper) profit generated from treating patients, and the utility from compliance arises from the satisfaction of an "internal conscience" as in

[1]See https://www.aarp.org/money/scams-fraud/info-2016/federal-strategy-to-fight-medicare-fraud.html.

the work of McGuire and Pauly (1991) and Gruber and Owings (1996).

In the context of the "internal conscience" model of induced demand in Gruber and Owings (1996), physicians are reluctant to provide patients unnecessary care, which is presumably due to concerns about patients' health and the desire to adhere to medical guidance. We believe that the hospital's key decision-makers value coding compliance for similar reasons. In particular, coding a service to a higher level than it is supposed to be may result in dis-utility for hospital decision makers due to the violation of coding protocols and also hurt patients' health. The diagnosis and procedure codes of a patient's visit will be documented into the patient's medical records and affect the clinical decisions for the patient's future care. If a patient's condition is improperly coded, the providers who access the medical record down the line might have inaccurate account of the patient's condition, which may cause severe patient harm.[2]

In this model, compliance can arise through two channels. First, compliance could be an optimal profit-maximizing decision that generates higher profits, after accounting for the financial expected costs associated with the inappropriate behavior being detected. But it could also arise simply from decision maker's utility from doing the "right thing" even when at the expense of extra ill-begotten profits. To our knowledge this feature is novel in the hospital upcoding literature. It resembles the concept of "tax morale" in the tax compliance literature (e.g. Luttmer and Singhal (2014)).[3]

Starting from this general model, we then develop four special cases that feature each of the mechanisms that could underlie hospital coding behavior. For each specialized model, we derive testable implications regarding the level of inappropriate admissions and coding as financial incentives vary.[4] While our focus is on upcoding, we derive the implications for inappropriate admissions because it helps identify the main underlying mechanism.

The first special case is inspired by the seminal work by Dafny (2005). We introduce a model where the financial incentive is the predominant factor in the utility function (*without* considering compliance in the utility function). We allow hospitals to a) admit patients unnecessarily and b) upcode rightly admitted patients into a higher billing tier than warranted by their medical condition. This model predicts no effect on total admissions and more top

[2]See https://bok.ahima.org/doc?oid=107687#.Y_6ZOezMJBV. For instance, if a patient was upcoded to have diabetes with major complications or comorbidities but actually has diabetes without complications or comorbidities, the provider who treats the patient later on may order medications that are different from what should have been given to the patient.

[3]We thank John Friedman for bringing to our attention the connection across the two literatures.

[4]Note that we will separately allow for unnecessary admissions and upcoding as two distinct choices that hospitals can engage in. These decisions are made at different margins of the distribution of potential patients. In other words, only properly admitted patients will possibly be upcoded whereas unnecessarily admitted ones will not.

bill coding as the reimbursement rate for the top bill code increases. Then, in line with the empirical evidence from Sacarny (2018), we incorporate stricter coding requirements into this model and discuss how admission and coding intensity respond to the increasing reimbursements for sicker patients that are tied-in with more stringent documentation requirements for top tier billing justification.

Next, we present a model specialization where both financial incentives and compliance play a role in the payoff. Similar to the model in Gruber and Owings (1996), there is a substitution effect and an income effect. The substitution effect captures the finding in prior studies that hospitals tend to upcode more when the revenue from it gets higher, because compliance becomes more "expensive." In contrast, if the income effect is the primary force, hospital decision-makers tend to engage in fewer inappropriate activities, because they can "afford" to be more compliant. The ultimate effect of a change in reimbursement rate depends on the interplay between the substitution and income effects.

Finally, we consider a class of models where hospitals do not consider financial incentives in their decisions, but changes in treatment technology or care-management approaches could drive changes in admission and coding patterns and reimbursement rates. For example, if treatment for a complex medical condition becomes less invasive leading to less costly, shorter hospital stays, it would be possible to observe more admissions (as patients who would otherwise stay out of hospitals with their condition managed with medications now receive the treatment at hospital) and lower reimbursements (as the new treatment takes up less hospital input). In other words, hospitals can treat more patients and do so in a more efficient manner by applying the advanced treatment technology. As fewer resources are required to care for patients with the same condition, the reimbursement rates set by the Centers for Medicare and Medicaid Services (CMS) could also get lower.[5] In this environment, technological changes will simultaneously increase the usage of hospital services in some DRGs and decreases in the reimbursement rates for those DRGs. The observable implication is that the number of admitted and top-coded patients with a certain condition is negatively correlated with the change in the weight for the top bill code of those conditions.

Understanding which is the prevailing mechanism provides potentially important implications for hospital payment policy and other initiatives designed to curb down improper billing. For instance, if revenue enhancing is the predominant incentive that leads hospitals to engaging in upcoding, greater fraud enforcement might be needed to deter such activities. However, if coding behavior responds to the increasing coding requirements to a greater

[5] An important reason for the annual review of the weights of DRG is to ensure the reimbursements for treatment of patients admitted under various DRGs reflects the right amount of resource usage to treat a patient with a particular diagnosis in the hospital.

extent, policymakers might consider offering healthcare providers more instructive guidance, or even rewarding providers' efforts in producing accurate documentation. Alternatively, if a dominant income effect from the compliance utility model is the primary mechanism, policymakers might consider reinforcing the legal consequences from proper/improper billing (that could make compliance play a more important role in hospitals' decision-making) and ensure adequate compensation for treating patients (so hospitals are less likely to substitute improper billing practices for legitimate behavior). Finally, if technological changes constitute the main driver explaining the volume of hospitals' admissions in certain DRGs and in particular in those DRGs with complicating conditions, minimum government intervention is likely to be sufficient.

We explore these mechanisms by examining the testable implications in both the aggregate Medicare Part A hospital inpatient claims data and a random sample of audit data during 2012 – 2019. We first exploit the longitudinal nature of the data to estimate fixed effect models using aggregate claims from Medicare Part A. By using fixed effects we rely on more convincing variation in reimbursements within a given DRG over time, relative to what could be done in a purely cross-sectional approach. We find that *fewer* admissions are reported with the top bill code as the reimbursement rate for this code *goes up*. While this could be interpreted as inconsistent with a model where financial incentives lead to improper coding, our general model shows that the financial incentives model can still be valid, when augmented with either compliance utility or an assumption that increases in reimbursement are tied-in with stricter coding and documentation requirements.

Next, we examine the effect on another outcome—total admissions within a base DRG. We also find a negative effect on this outcome from an increase in the corresponding top DRG weight. This finding is less consistent with the predictions from the financial incentive model augmented by a data generating process assumption that increases in reimbursement tend to be tied in with stricter coding requirements. On the other hand, we show that a financial incentives model supplemented with a compliant utility feature remains consistent with the two lines of evidence. The evidence up to this point is also consistent with the models of fully compliant behavior where the negative effects from reimbursement on admissions are driven by technological changes that simultaneously affect both the treatment cost (and thus reimbursement) for a particular condition and the patient volume with that condition.

The results based on the aggregate claims data then leave us with two specialized models: the one accounting for compliance utility and the one emphasizing changes in treatment technology. A common route at this point would be to look for instruments or other exogenous sources of variation in reimbursement to subject the models that ignore financial incentives and assume full compliance to a more stringent test. This approach would remove the technologically-driven endogeneity from our testable implications. An alternative that we pursue is to rely on the different implications these two alternative models have when examining a random sample of audits that regulators use to investigate upcoding. The key idea is that if our tests from aggregate Medicare claims are fully driven by hospital-technology endogeneity of reimbursements, we should not see changes in the *probability* of audit detections of upcoding cases as reimbursement within a DRG changes over time.

We distinguish these two models using a random sample of audits coming from the Comprehensive Error Rate Testing (CERT) program, one of the auditing strategies adopted by CMS. The CERT program is unique and distinct from other national audit programs, as its review process is based on a random sample of Medicare Parts A and B claims. To the extent that even during an audit detection is not perfect, its results can provide a lower bound for the prevalence of Medicare improper billing. Moreover, we base our analysis on audit outcomes that clearly distinguish improper from proper billing by audit reviewers, enabling us to identify the key mechanism driving our results.

We define upcoding-related errors from the CERT data and estimate the effect of financial incentives on the probability of observing these errors using a similar fixed effects specification. We find that the probability of claims being flagged with errors associated with upcoding is lower when DRGs are subject to higher reimbursement rates, suggesting that the specialized model that focuses on technology changes is not the primary mechanism. Instead, these findings support the mechanism where financial incentives matter but they are modulated by a demand for compliant behavior which is itself driven by a strong income effect in the compliance utility model. Taken together, the findings in the aggregate claims data and the random sample of audit data suggest that the financial incentives model augmented with compliance utility and a strong income effect in hospital decision makers' demand for compliant practices seems to explain coding patterns during the period covered in this study.

Related literature. Upcoding and related practices that seek to improperly increase revenue for healthcare providers and/or private health insurers have been analyzed for different segments of the U.S health care system. For example Geruso and Layton (2020) focus on upcoding in the Medicare Advantage program, Fang and Gong (2017) explore physician over-billing in Medicare Part B, and Finkelstein et al. (2017) explore potential issues with the risk adjustment system in the Affordable Care Act exchanges where there might be differential coding intensity across providers in different regions. We focus on the subset of this literature that examines the appropriateness of hospitals' coding practices in the Medicare Part A program following the seminal work of Dafny (2005).[6]

[6]Note that upcoding by hospitals in Part A, which we focus on, operates primarily through selection of

Following Dafny (2005), a growing literature in economics explores how hospitals respond to price changes. Dafny (2005) found substantial upcoding but no increase in medical admissions for conditions that underwent large increases in reimbursement as a result of a 1988 payment reform. Silverman and Skinner (2004) further examined hospital upcoding behavior by ownership and found that for-profit hospitals experienced the largest increase in patients assigned to the most lucrative DRG for pneumonia and respiratory infections. A more recent paper by Cook and Averett (2020) used HCUP data from 2005 to 2010 and found that at least 3% of reimbursement could be attributable to upcoding, exploiting the 2007 re-structuring of the DRG payment system. These findings have been influential and along with other evidence, have triggered policy responses aimed at curbing the extent of upcoding. We contribute to this literature by exploring data from the 2010s after these reforms have been implemented. In addition, a novel feature of our work is the use of data from random audits, which are used by the Centers for Medicare and Medicaid Services (CMS) to monitor the prevalence of upcoding and other improper payments in the Medicare program. Because we use a more recent time period and our methodology differs, our work builds on this previous literature but does aim to re-litigate its findings, that showcase the importance of upcoding under the more lax regulatory environment prevalent in previous decades, since the inception of the DRG payment system until 2010.

Our paper also complements the literature that investigates how healthcare providers respond to regulators' non-financial policy changes, such as stricter requirements for documenting care provision and increasing monitoring and detection of fraud and abuse. Sacarny (2018) has found that, after a payment reform in 2007, hospitals could have increased revenues by specifying a patient's heart failure condition but only capture about half of them due to substantial institution-level frictions. Shi (2021) finds that monitoring healthcare providers has reduced the rendering of unnecessary care but also imposed a large amount of compliance costs; in response, providers adopt new information technology to identify potentially medical unnecessary care. Leder-Luis (2020) and Howard and McCarthy (2021) have found significant deterrence effects of anti-fraud and abuse enforcement under the False Claims Act. Based on these findings, we study how hospitals perform coding decisions given policy-makers' oversight and fraud-detection efforts.

Finally, our paper contributes to the literature on healthcare providers' decision-making about the provision of care. A strand of this literature specifically studied the physicians' decision to induce demand for their services (Dranove, 1988; McGuire and Pauly, 1991; Gruber

more profitable DRG codes during the inpatient admission. As discussed in detail in Geruso and Layton (2020), upcoding in the Medicare Advantage program operates in a completely different way, with insurers looking for ways to increase patients' risk scores in legal (and perhaps illegal) ways to obtain higher capitations.

and Owings, 1996). We build on these "inducement" models by introducing a preference for compliance into a hospital decision-maker's payoff. Key players in hospitals' decision-making, whether clinicians or administrators, are likely to care about this "internal conscience" to some extent. Starting from this general model, we are able to develop specialized models that feature specific mechanisms that have been advanced in the literature and motivate the empirical analysis.

The rest of the paper proceeds as follows. Section 2 introduces the institutional background. Section 3 discusses the general model and the various specializations that feature specific mechanisms. In this section we also derive testable implications that can be taken to the aggregate claims data. Section 4 describes the datasets and reports summary statistics. Section 5 presents the empirical strategy that is used to test the implications from the various models along with results based on the aggregate data for Medicare Part A IPPS claims. Section 6 derives the testable implications for different mechanisms based on the CERT data. Section 7 discusses the empirical analysis based on the CERT data. The last section concludes.

2 Background

2.1 Medicapre Part A IPPS

Medicare pays each inpatient admission based on a flat rate payment system, called the inpatient prospective payment system (IPPS). The payment for each inpatient admission depends on the assigned DRG, which is determined according to the patient's primary diagnosis/procedure, additional diagnoses/procedures, and discharge status. Each DRG is associated with a weight that is set by CMS, representing the average resources required to care for Medicare patients in that particular DRG. The flat rate paid for each admission is equal to the DRG weight multiplied by a base rate that depends on an area cost factor to account for geographic heterogeneity in the cost of hospital inputs.

Typically, an admission is first assigned to a *base DRG*, based on the patient's primary diagnosis or primary procedure. Then, the admission is assigned to an exact DRG, depending on the presence/absence of complication or comorbidity (CC) or major CC (MCC). A base DRG could include one to three associated DRGs. For instance, DRGs 88 – 90 are "concussion w/ MCC," "concussion w/ CC," and "concussion w/o CC/MCC," respectively, all belonging to the same base DRG. We call each DRG *within* a base DRG a tier, a level, or a severity subclass interchangeably, and specifically refer to the most severe DRG as the top bill code. CMS publish a list of CCs and MCCs such that, whenever a patient is coded with a secondary diagnosis from that list, s/he will qualify for a more severe DRG. The one with more complications implies a higher severity level and results in greater reimbursements, but it also imposes stricter requirements on the justifying documentation that needs to be submitted with the claim.

The DRG-based IPPS was first established in 1983 and undergoes an annual review of its DRG classification and adjustment to the DRG weights every year. DRG weights are set by the Medicare Payment Advisory Commission. The formulae to calculate the DRG weight has been revised over time, but the core process remains the same. Simply speaking, for a given DRG, an average standardized charge is obtained by first calculating the total charges of all cases in that DRG and then dividing the sum by the total number of cases in that DRG. The DRG weight for that DRG is the ratio of the average standardized charge for that DRG to the overall average charge across all DRGs, reflecting the relative cost to treat the patients (Office of Evaluation and Inspections, 2001).

The periodic recalculation of DRG weights aims to capture the changes (increases or decreases) in the necessary use of resources and inputs, which might arise from the presence of new treatments/technology, improved efficiency in care management, and any other factors that will affect the resources needed to care for the patients.[7] A significant payment reform to IPPS occurred in October 2007, in which CMS reclassified some of the DRGs, revised the CC/MCC list, and adjusted DRG weights, with the goal to better align payments with the resources used by hospitals. One of the most dramatic outcomes from this reform was that the percentage of patients who qualified for a DRG with (at least) CCs dropped from 77.7% (based on the pre-reform criteria) to 40.3% (based on the post-reform criteria).[8] Recently, CMS again suggested severity-level changes in the proposed rule for IPPS in Fiscal Year (FY) 2020, with the majority of adjustments downgrading MCCs to CCs or CCs to non-CCs.[9] All of these efforts reflect a consistent CMS goal to offer adequate compensation for treating ill patients while maintaining the accuracy of Medicare billing.

2.2 Coding hospitalized patients and CERT Audits

Upon admission, all the clinical information on treating a patient will be recorded in the patient chart. A patient chart includes an admission note, a patient progress note, and a discharge summary that provides a brief synopsis of the patient's stay and disposition. Cod-

[7] See https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/MS-DRG-Classifications-and-Software.

[8] See Office of the Federal Register and National Archives and Records Service (2007), p. 47,153 – 47,154.

[9] See https://www.federalregister.gov/documents/2019/05/03/2019-08330/medicare-program-hospital-inpatient-prospective-payment-systems-for-acute-care-hospitals-and-the.

ing refers to the process of translating the information on patients' charts into standardized, billable codes that healthcare providers submit to insurance companies for payments. It is mainly done by coders. While it is the coders who are mainly responsible for coding, the success of coding also requires efforts from clinicians (who document the care provided) and administrators (who oversee and coordinate the entire process). After verifying the substantiation of each billed code, a coder enters the codes and other information into grouper software, which generates the appropriate DRG for billing purposes. Upcoding usually reflects preferential selection of codes with higher reimbursement rates. In some cases upcoded claims might be submitted without sufficient documentation. Either because the documentation does not exist or when it does, it would facilitate the upcoding determination by auditors. Alternatively, claims might be submitted with complete documentation that actually shows, that the higher code was inappropriate.

CMS have adopted multiple audit strategies to ensure the accuracy of Medicare payments, including the Comprehensive Error Rate Testing (CERT) program, the Recovery Audit Program, Medicare Administrative Contractors, Supplemental Medical Review Contractors, and the Zone Program Integrity Contractor audits. The following are the processes that could result in upcoding and thus be scrutinized by auditors: (1) selecting a base dignosis with a higher weight than justified, (2) coding a CC or MCC modifying the base diagnosis that is not present/not sufficiently documented, (3) coding selected diagnoses as present on admission when they were not, and (4) unbundling services/procedures that are bundled under a single DRG. There is no immediate penalty involved in CERT in case of an determination of improper payments, but hospitals are subject to other consequences such as reputation loss, and the potentially higher probability of getting audits in the future.[10]

The CERT program is unique and distinct from other programs because it is the only one that reviews claims based on a *random* sample. The results provide a reasonable estimate of the overall prevalence of improper bills in Medicare, and thus, CMS rely on the results from CERT to identify program vulnerabilities and also use them as guidance to improve reimbursement mechanisms and related policies. To the extent that auditors fail to detect some of the upcoding that hospitals engage in (even when those cases are randomly selected for audit), the estimates from CERT provide a lower bound for the true prevalence of upcoding.

The CERT program randomly select approximately 50,000 Medicare Parts A and B claims for review, following a stratification strategy.[11] For each claim selected for review, the

[10]Other audit programs, such as the Recovery Audit Program, may prioritize reviewing claims flagged by the CERT data (Recovery Audit Program, 2013).

[11]The selection is based on claims types, including Part A IPPS, Part A excluding IPPS, Part B, and Part B durable medical equipment, prosthetics, orthotics and supplies (DMEPOS). For each claim type, CMS further divide the claims into approximately 100 service levels except for Part A excluding IPPS, which

CERT contractors send a request to providers for all related documents. If the documentation is available, providers have incentives to respond as failure to submit these documents could lead to CMS recouping the Medicare payments for that claim.

CERT auditors evaluate all supporting documentation before making a determination. Claims that are found with improper billing will be categorized into one of the following payment error types: medical necessity, no documentation, insufficient documentation, incorrect coding and others.[12] We group the last four error types together and refer to them as upcoding-related errors in our analysis based on the CERT data, as each of them most likely reflects upcoding in one way or another. Medical necessity emphasizes the level of services billed meeting the medical needs according to Medicare coverage guidelines, but it is mainly related to where the services should be performed—in the ambulatory setting or hospitals—in the context of hospitalization care (CERT Program, 2013). The other four errors involve inaccurate coding for inpatient services or inadequate justification due to the lack of documentation, but the service considered was provided in the right setting.

3 Theoretical Framework

In this section, we first describe a general model that characterizes hospital decisions on admission and tier coding assignment. We then discuss a comparative set of specialized mechanisms that generate different predictions on the admission and coding behavior. Note that our main focus is on the implications on coding behavior, but we also analyze the impact on admissions, which help distinguish between different mechanisms.

3.1 A General Model

We develop the model by adapting the theoretical insights about physician behavior from McGuire and Pauly (1991) and Gruber and Owings (1996) to a hospital decision-making context. In our model, hospital decision makers do not simply seek to maximize profits, but rather a utility function that depends positively on consumption and compliance with the billing and coding protocols. The level of non-compliance in admission and coding practices is defined as the extent of deviation from the appropriate levels.

Given a patient with a particular disease, hospital decision makers decide whether to admit the patient, and if they do, which severity subclass the patient is assigned to. Note that the process involves a series of decisions made by clinicians and administrators. For instance,

contains fewer than 20 strata due to the difficulty in service level stratification.

[12]See CERT Program (2020) for more detail on the definition of each error type.

physicians determine diagnosis or treatment choice and document care into the patient's chart; coders generate bill codes from the patient's medical records; and administrators generally coordinate and oversee the entire process. We assume all the stakeholders share a common goal to maximize the overall expected utility.[13]

For a given base DRG, we let s index the severity of a patient's condition. s follows a uniform distribution between 0 and 1. Also, we let $\underline{s}^a$ denote the level of severity above which a patient should be admitted following the protocol, and $\underline{s}^u$ denote the condition severity above which a patient should be coded into the top billing tier following the proper coding guidelines. For each visit i, hospital decision-makers choose whether to admit the patient and whether to top-code a rightly admitted patient. Hospitals always admit every patient with $s \geq \underline{s}^a$ and, among admitted patients, always code into the upper tier those who, in addition, have $s \geq \underline{s}^u$. For the remaining cases, decisions must be made on admission for those medically unnecessary cases ($s < \underline{s}^a$) and tier coding assignment for those who are rightly admitted but should be coded into the lower tier ($\underline{s}^a \leq s < \underline{s}^u$). The model is decentralized at the DRG level, effectively assuming that a different decision maker with similar utility function is in charge of admissions and tier coding for each DRG.[14]

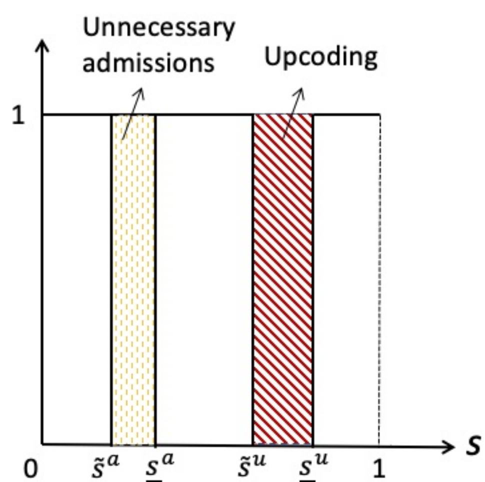


Figure 1: Distribution of Potential Patients' Condition for a Base DRG
Medically-Appropriate Thresholds ($\underline{s}^a, \underline{s}^u$) vs. Threshold Choices ($\tilde{s}^a, \tilde{s}^u$)

As mentioned above, assuming patient severity, s, at a given base DRG follows a uniform

[13] This is equivalent to assuming that they have different goals but the executive could structure clinician and coder incentives to align with his. Prior studies indicate that there could be incentive misalignments between hospital administrators and clinical staff (Kuhn et al., 2015; Sacarny, 2018). We assume that they share a common goal here for modelling purposes, and also given the increasing trend of hospitals integrating with physician groups.

[14] We abstract from a more complex modeling where a single decision maker would decide across all DRGs within a hospital but the main insights we derive do carry over to that model.

distribution between [0, 1], Figure 1 displays graphically the key elements of the model. In particular, it is easy to see, for each base DRG, we have $0 < \tilde{s}^a < \underline{s}^a < \tilde{s}^u < \underline{s}^u < 1$. The dark-shaded area captures the extent of upcoding, whereas the light-shaded area captures the extent of medically unnecessary admissions.

Hospitals get reimbursed for each admission belonging to a given base DRG by CMS with a baseline amount B multiplied by the DRG weight. ω^h and ω^ℓ, respectively, denote the DRG weight for the bottom and top tier within a base DRG. For simplicity, we model base DRGs with only two tiers, though our empirical work extends to base DRGs with three tiers. Note that we examine coding behavior within a base DRG, although improper billing might occur via substitution between base DRGs. Following the majority studies in the upcoding literature, our theoretical model and empirical analysis focus on coding changes *within* a base DRG (Dafny, 2005; Li, 2014; Cook and Averett, 2020; Gowrisankaran et al., 2022).

Decision-makers at the hospitals may decide to admit patients unnecessarily (inpatient admissions not justified by medical necessity) or upcode properly admitted patients, that is, improperly classify patients into the upper tier when they should not, as they stand to profit by doing so. They do however expose themselves to being audited and having to return the extra reimbursement and suffer additional future costs. Thus, hospital decision-makers must choose the extent to which they allow unnecessary admissions and inappropriately top-code patients. In doing do, they trade-off profit motives with the expected costs from audits and the dis-utility from non-compliance.

Let $\tilde{s}^a$ denote the *least* severe patient that does *not* qualify for admission but gets admitted. Similarly, let $\tilde{s}^u$ denote the *least* severe patient that is top-coded but does *not* meet the threshold for top-tier coding. Note that implicit in the definition of $\tilde{s}^u$ is an assumption we view as a reasonable approximation to the set of opportunities the hospital faces to engage in improper practices. While mild cases can be unnecessarily admitted ($\tilde{s}^a \leq s < \underline{s}^a$), they cannot also be upcoded. In other words, some cases are at the margin of being admitted unnecessarily and some cases are at the margin of being coded inappropriately into the top tier, but no case can be both, unnecessarily admitted *and* upcoded.[15] Similarly, we assume that hospital decision-makers only consider marginal cases—those close to $\underline{s}^a$—for unnecessary admission given that the cost of doing so is too high among the cases farther away from the cutoff.

We assume that the decision-maker in charge of hospital admission and coding practices

[15]We test this assumption by comparing the incidence of medical necessity error among the top bill codes with that among the bottom codes, using the CERT audit data. We find only about 3.6% claims reported with top tier DRGs were subject to medical necessity error, whereas the percentage is 14% among the bottom DRGs. This supports our assumption here.

is risk-neutral[16] and solves the following problem:

$$
\begin{aligned}
&\max_{\{\tilde{s}^a,\, \tilde{s}^u\}} E\left[U(c,m)\right] \\
\text{s.t. } c &= \int_{\tilde{s}^a}^{\tilde{s}^u} \Big[B\omega^\ell - c^\ell - E[\zeta^\ell(s)]\Big] ds + \int_{\tilde{s}^u}^{1} \Big[B\omega^h - c^h - E[\zeta^h(s)]\Big] ds; \\
\tilde{s}^a &\in (0, \underline{s}^a); \quad \tilde{s}^u \in (\underline{s}^a, \underline{s}^u); \\
m &= \tilde{s}^a + \tilde{s}^u - \underline{s}^a; \\
E[\zeta^\ell(s)] &= \mathbb{1}\{s < \underline{s}^a\} \times \phi^{\text{AUDIT}} \times \phi^a(\underline{s}^a - s) \times \xi^a; \\
E[\zeta^h(s)] &= \mathbb{1}\{s < \underline{s}^u\} \times \phi^{\text{AUDIT}} \times \phi^u(\underline{s}^u - s) \times \xi^u;
\end{aligned}
\tag{1}
$$

where m denotes the level of compliance; c^h and c^l denote the cost of treating the patient at the corresponding tier the patient is assigned to.

The term $E[\zeta^\ell(s)]$ ($E[\zeta^h(s)]$) denotes the expected cost from inappropriate admissions (upcoding) if the hospital is found engaging in these practices during an audit. ϕ^{AUDIT} refers to the probability of an admission being audited, whether by the CERT program, recovery auditors, or any other audit programs. For simplicity, we treat ϕ^{AUDIT} as a parameter for the ease of notation. In principle, we could let ϕ^{AUDIT} vary with (ω^h, ω^l). For instance, auditors may have greater incentives to review the top bill codes which experienced the largest increases in reimbursement rates. We discuss this case in more detail in Section 3.2.1 and Appendix II. Also, it would be possible to let ϕ^{AUDIT} vary according to the distance between s and the thresholds ($\underline{s}^a$ or $\underline{s}^u$) to capture that a claim could be more likely to be selected for review if auditors suspect that the patient's severity level is faraway from either one of the thresholds. The implications we derive later hold in these more general cases.

ϕ^a and ϕ^u denote, respectively, the probabilities that, in the event of an audit, auditors detect medically unnecessary admissions and upcoding. Moreover, we assume that the probability of getting detected increases with the deviation from the cutoff, that is, $\phi^a(\cdot)$ ($\phi^u(\cdot)$) increases with $|\underline{s}^a - s|$ ($|\underline{s}^u - s|$). ξ^a and ξ^u denote, respectively, the costs associated with an audit determination of improper payment due to a medically unnecessary admission and upcoding, including penalty, reputation loss, or greater scrutiny by Medicare in the future.

The decision makers aim to maximize the expected utility taking into account the probability of getting audited (ϕ^{AUDIT}) and the probability of getting detected for improper billing if an audit is initiated (ϕ^a or ϕ^u). The utility increases with consumption (c) that is financed by the revenue hospitals earn from treating patients and with the level of compliance (m) from properly admitting and coding patients. On the one hand, excessively admitting pa-

[16] It is possible that the decision maker is risk-averse. We make this assumption for modelling purposes.

tients or coding patients to the top tier could increase revenue but result in dis-utility from violating Medicare payment guidelines. Moreover, it also subjects the hospital to higher risk of getting caught in audits and other adverse outcomes such as penalty, reputation loss, and greater scrutiny in the future (captured by ξ^a or ξ^u). The hospital decision makers have to balance these tradeoffs in making the admission and coding decisions.

Several points worth noting. First, our model incorporates both the cost and benefit to improper billing, with the former captured by $E[\zeta^\ell(s)]$ and $E[\zeta^h(s)]$ and the latter captured by the financial payoff, ω^ℓ and ω^h. Second, the expected cost of either unnecessarily admitting patients or upcoding them is an increasing function of the deviation from the appropriate cutoff, i.e., $|s - \underline{s}^a|$ or $|s - \underline{s}^u|$. Therefore, hospitals will always find it optimal to first improperly admit or upcode the marginal patients, that is, those whose severity level is below the threshold but relatively close to it. These cases are less egregious, and therefore stand a higher chance of passing an auditor's review, in the off chance that the case is selected for an audit. Moreover, since our model applies to a given base DRG, we allow the opportunity of successfully upcoding $(1 - \phi^u)$ to vary by diagnosis. For instance, there are certain DRGs where the opportunity for upcoding is present and others where no such behavior is available (e.g. a solid organ transplant). Our model captures this possibility.

3.2 Specialized Models

In the following, we discuss four specialized models each of which features a specific mechanism from the general model we developed above. For each of them, we describe the specific setup and predict the admission and coding decisions as the reimbursement rate varies. Note that the general model works as a foundation for these mechanisms. Our goal is to compare and test the implications from the mechanisms. We start with the two mechanisms that have been advanced in the literature. We then introduce the mechanism we propose as well as an alternative model that could generate similar implications.

3.2.1 Financial Incentive Mechanism

This specialized model is inspired by the seminal work by Dafny (2005), who studied hospital coding behavior in response to a fee change in the IPPS in 1988. In this specialized model, hospitals are profit-maximizing entities, without considering the dis-utility of non-compliance. In other words, the utility for a hospital decision-maker, which does not involve the level of compliance, depends on $\tilde{s}^a$ and $\tilde{s}^u$ only through their effects on profits and ultimately consumption.[17] We revise the model as follows:

$$
\begin{aligned}
&\max_{\{\tilde{s}^a,\, \tilde{s}^u\}} E[U(c)] \\
&\text{s.t. } c = \int_{\tilde{s}^a}^{\underline{s}^a} \left(\pi^\ell - \phi^{\text{AUDIT}} \times \phi^a(\underline{s}^a - s) \times \xi^a\right) ds + \int_{\tilde{s}^u}^{\underline{s}^u} \left(\pi^h - \pi^\ell - \phi^{\text{AUDIT}} \times \phi^u(\underline{s}^u - s) \times \xi^u\right) ds \\
&\quad + (\underline{s}^u - \underline{s}^a)\pi^\ell + (1 - \underline{s}^u)\pi^h
\end{aligned}
\tag{2}
$$

where $\pi^\ell = B\omega^\ell - c^\ell$ and $\pi^h = B\omega^h - c^h$ denote the operating profits from coding a hospital admission into a low or high billing tier not taking yet into account the possible financial costs from a potential audit.

To condense notation, let

$$
\begin{aligned}
f(x_1, x_2) &= \int_{x_1}^{x_2} \left[\phi^{\text{AUDIT}} \times \phi^a(\underline{s}^a - s) \times \xi^a \times \frac{1}{x_2 - x_1} ds\right] \\
g(y_1, y_2) &= \int_{y_1}^{y_2} \left[\phi^{\text{AUDIT}} \times \phi^u(\underline{s}^u - s) \times \xi^u \times \frac{1}{y_2 - y_1} ds\right]
\end{aligned}
\tag{3}
$$

where $x_1,\ x_2 \in [0, \underline{s}^a]$ and $y_1,\ y_2 \in [\underline{s}^a, \underline{s}^u]$. $f(x_1, x_2)$ captures the average expected costs of detection per inappropriately admitted patient whose severity level falls into the range of $[x_1, x_2]$. Likewise, $g(y_1, y_2)$ captures the average expected costs of detection per upcoded patient whose severity level falls into the range of $[y_1, y_2]$. For instance, $f(\tilde{s}^a, \underline{s}^a)$ and $g(\tilde{s}^u, \underline{s}^u)$ measure the average expected costs from engaging in improper admissions and coding practices, respectively. Thus, the maximization problem becomes

$$
\begin{aligned}
&\max_{\tilde{s}^a,\, \tilde{s}^u} E[U(c)] \\
&\text{s.t. } c = \underbrace{(\underline{s}^a - \tilde{s}^a)\left(\pi^\ell - f(\tilde{s}^a, \underline{s}^a)\right)}_{\text{Total profit from unnecessary admissions}} + \underbrace{(\underline{s}^u - \tilde{s}^u)\left(\pi^h - \pi^\ell - g(\tilde{s}^u, \underline{s}^u)\right)}_{\text{Total profit from upcoding patients}} \\
&\qquad + \underbrace{(\underline{s}^u - \underline{s}^a)\pi^\ell + (1 - \underline{s}^u)\pi^h}_{\text{Total profit from proper admissions and proper coding}}
\end{aligned}
$$

Prior studies have found that hospitals tend to upcode more when the revenue from coding becomes higher (Silverman and Skinner, 2004; Dafny, 2005; Jürges and Köberlein, 2015). Thus, we consider an increase in ω^h, which leads to an increase in π^h. Note that in

[17]For simplicity, we are assuming this decision-maker is the owner in the sense that he or she consumes all of the profit. But similar implications arise if he or she is only entitled to a fraction of the hospital profits.

practice c^h may also change in the same direction as ω^h. For instance, an important reason for the adjustment of DRG weights arises from the change in the treatment cost.

Throughout the paper, we assume the absolute change in a DRG weight times the baseline reimbursement B is greater than the variations in the treatment cost. This assumption is reasonable. Hospitals may change the way they treat patients, say, at time t, primarily for two reasons. First, they may follow new treatment guidelines that result in a more costly course of treatment. These hospitals incur a loss at t, but get some relief when reimbursement ω is increased at $t+1$ to reflect that higher treatment cost. In this case the increase in reimbursement from t to $t+1$ eliminates the temporary loss and thus leads to higher profits. Second, without changing guidelines, hospitals may find more cost-efficient ways of treating patients in a certain DRG at time t. This would generate a temporary profit. At time $t+1$ however, reimbursement for that DRG will be reduced to match that lower treatment cost. In this case too, the sign of the change from t to $t+1$ in ω will be the same as the change in profit π.

We present the first testable proposition as follows:[18]

Proposition 1 As ω^h and thus π^h gets higher, we expect no changes in total admissions within the base DRG, i.e., $\partial \tilde{s}^a / \partial \pi^h = 0$, and more admissions to the high-severity level, i.e., $\partial \tilde{s}^u / \partial \pi^h < 0$.

Proof. See Appendix I.

An alternative mechanism that might have the same implication as Proposition 1 is that the increase in ω^h (and hence π^h) reduces the extent of "down-coding." Downcoding refers to the situation where hospitals' documentation costs for the top tier is so high that it is optimal for them to bill the lower-tier code. As the returns to reporting the top code increases, hospitals might be more incentivized to completely document patients' conditions and diagnoses for the top code. In recent years, CMS started to refine the classification of patients based on severity of illness and increased the reimbursements made for patients with the most severe conditions, such as the DRG reclassification in 2007. Thus, hospitals might have greater incentives for complete documentation, so as to recover more billable services than before (Gowrisankaran et al., 2022).

With the evolving changes in healthcare delivery, such as the shift to outpatient care and increasing availability of post-acute care facilities, inpatient admissions are on average more

[18] Additional technical assumptions that we maintain include: $\frac{\partial f(\tilde{s}^a, \underline{s}^a)}{\partial \tilde{s}^a} = f_1 < 0$, $\frac{\partial f(\tilde{s}^a, \underline{s}^a)}{\partial \underline{s}^a}|_{d\tilde{s}^a=0} = f_2 > 0$, $\frac{\partial^2 f(\tilde{s}^a, \underline{s}^a)}{(\partial \tilde{s}^a)^2} = f_{11} \geq 0$, $\frac{\partial g(\tilde{s}^u, \underline{s}^u)}{\partial \tilde{s}^u} = g_1 < 0$, $\frac{\partial g(\tilde{s}^u, \underline{s}^u)}{\partial \underline{s}^u}|_{d\tilde{s}^u=0} = g_2 > 0$, and $\frac{\partial^2 g(\tilde{s}^u, \underline{s}^u)}{(\partial \tilde{s}^u)^2} = g_{11} \geq 0$

severe, and subject to more acute conditions. In response to that, CMS have been constantly revising DRG classification and payments, with the goal of better aligning reimbursements with treatment costs and ensuring the accuracy of Medicare billing.[19] Thus, an increased severity of illness adjustment payment could usually come with stricter requirements in coding and documentation, as can be seen in the 2007 DRG payment reform (Sacarny, 2018; Gowrisankaran et al., 2022) and the recent proposal of revising CC/MCC list by CMS.[20]

In the following, we consider the effect of a change in ω^h with a simultaneous change in $\underline{s}^u$ to capture the presence of increasing reimbursement for severe illness tied with stricter coding requirements. In other words, we now consider an increase in both the financial pay-off from top tier billing *and* change in the relevant threshold for what it takes in terms of severity to be classified into the top tier. Specifically, we assume $cov(\omega^h, \underline{s}^u) > 0$.[21] Since two parameters, ω^h and $\underline{s}^u$, vary at the same time, we first derive how $\tilde{s}^u$ and $\tilde{s}^a$ vary in response to the change in $\underline{s}^u$ and then discuss the combined effect from a simultaneous change in both ω^h and $\underline{s}^u$.[22] As is shown in Appendix I, a higher ω^h leads to more top-tier coding but a simultaneous increase in $\underline{s}^u$ produces the opposite effect. The ultimate effect depends on the force that dominates. Thus, we propose the second testable implication:

Proposition 2 If the effect from greater $\underline{s}^u$, i.e., stricter coding requirements, dominates the effect from ω^h, we expect no changes in total admissions within the base DRG and fewer admissions to the high-severity level.

Proof. See Appendix I.

Preposition 2 suggests that the predicted change in top bill coding might be in the opposite direction to that in Proposition 1. Indeed, if the response to the increase in ω^h is smaller than the deterring effect from greater $\underline{s}^u$, then the model implies that an increase in ω^h will lead to a *decline* in admissions into the top tier within a DRG. Prior studies have

[19]See https://www.cms.gov/icd10m/version37-fullcode-cms/fullcode_cms/Design_and_development_of_the_Diagnosis_Related_Group_(DRGs).pdf.

[20]For the proposed change in the CC/MCC list, see https://www.federalregister.gov/documents/2019/08/16/2019-16762/medicare-program-hospital-inpatient-prospective-payment-systems-for-acute-care-hospitals-and-the.

[21]We use the increasing $\underline{s}^u$ as a proxy for the stricter requirements for documentation. The idea is that a stricter documentation requirement in our model can be seen as a more stringent threshold for what cases can be legitimately classified into the top tier. For instance, according to Sacarny (2018), the disease code "congestive heart failure, unspecified" that was treated as high severity before the policy change was recategorized into low severity after the reform, because patients with this (somewhat vague) code were not found to result in greater treatment costs than average as patients with codes that gave more information about the nature of the condition. The recent revision of the CC/MCC list shares a similar idea.

[22]Appendix I provides more detail on the effect on $\tilde{s}^u$ and $\tilde{s}^a$ as $\underline{s}^u$ varies.

shown empirical evidence consistent with this prediction for specific DRGs and at particular points in time. For example, Sacarny (2018) found that hospitals only captured half of the revenue in a code reclassification on heart failure, suggesting that hospitals may face large barriers in meeting the documentation requirements. Gowrisankaran et al. (2022) found relatively fewer top bill codes among DRGs that experienced greater increments in the DRG weight of the high severity subclass, after the 2007 payment reform.

Note that our discussion so far assumes that audit probability, ϕ^{AUDIT}, is a parameter, which is fixed. If we allow the scrutiny of top-tier claims to vary with DRG weights, the predicted effects on total admissions within a base DRG and the admissions to the high-severity subclass are the same as those in Proposition 2, under certain conditions. We refer to this specialized model as financially-driven and augmented by variable audit scrutiny. We provide more details about it in Appendix II.

3.2.2 Compliance Utility Model

The model in Section 3.2.1 shuts down the preference for compliance in the general model (presented in Equation (1)). We now consider the general model that accounts for compliance in the utility function and derive the implications on admission and coding practices in a similar way to Section 3.2.1.

Let $m^a = \tilde{s}^a$ and $m^u = \tilde{s}^u - \underline{s}^a$, where m^a and m^u denote the level of compliance in admission and tier coding assignment, respectively. In Figure 1, the white area to left of the light-shaded area captures compliance in admissions, m^a. Note that m^a includes a) units of compliance driven strictly by profit maximization, where, given the expected cost of an audit, complying is the choice that generates higher expected profit and b) additional units of compliance that stem from a desire to further increase utility from compliance even at the cost of forgone profits. The white area in between the two shaded regions captures compliance in bill coding, m^u. Again, compliance demand for m^u includes both profit-driven and compliance utility-driven units. Therefore, equilibrium compliance in admissions and coding is higher in this model than it is in the model without compliance utility.

After some simple re-arrangements of the terms in the budget constraint of Equation (1), we can rewrite the model as follows:

$$
\begin{aligned}
&\max_{\{m^a, m^u\}} E[U(c, m^a, m^u)] \\
&\text{s.t. } c + m^a \underbrace{\left(\pi^\ell - f(0, m^a)\right)}_{\text{Average “price” for } m^a} + m^u \underbrace{\left(\pi^h - \pi^\ell - g(\underline{s}^a, m^u + \underline{s}^a)\right)}_{\text{Average “price” for } m^u} \\
&= \underbrace{\underline{s}^a \times \left(\pi^\ell - f(0, \underline{s}^a)\right)}_{\text{Profit from admitting everyone below } \underline{s}^a} + \underbrace{(\underline{s}^u - \underline{s}^a) \times \left(\pi^h - \pi^\ell - g(\underline{s}^a, \underline{s}^u)\right)}_{\text{Profit from upcoding everyone within } [\underline{s}^a, \underline{s}^u]} \\
&+ \underbrace{(\underline{s}^u - \underline{s}^a)\pi^\ell + (1 - \underline{s}^u)\pi^h}_{\text{Total profit from proper admissions and proper coding}} .
\end{aligned}
\tag{4}
$$

The formulation in Equation (4) provides additional insight as it reflects “expenditures” in the two types of compliance on the left hand side of the budget constraint. Further, each of these compliance expenditures is factored into the corresponding quantities times the average price per unit. Note that the average “price,” $\pi^\ell - f(0, m^a)$, for a unit of compliance in admission (i.e., a unit of m^a) equals the profit π^ℓ that the hospital could have received by admitting a patient that is not supposed to be admitted, minus $f(0, m^a)$, the average unit expected cost from potential detection in audits associated with the (hypothetical) inappropriate admission of patients in the range $[0, m^a]$. Similarly, the “price,” $\pi^h - \pi^\ell - g(\underline{s}^a, m^u + \underline{s}^a)$, per unit of compliance with the coding assignment protocols (i.e., a unit of m^u) is the forgone incremental profits from upcoding a patient $(\pi^h - \pi^\ell)$ net of the average expected cost associated with audit detection from hypothetical non-compliance in the compliant coding range $[\underline{s}^a, m^u + \underline{s}^a]$, given by $g(\underline{s}^a, m^u + \underline{s}^a)$.

The right-hand side of the budget constraint represents the sum of the proper profits and the improper profits the hospital would make if it admitted everyone and coded everyone into the top tier net of the costs from the audits. Moreover, the average compliance “prices” associated with m^a and m^u—the ones on the left-hand side of the budget constraint—are lower than the ones associated with a maximum level of unnecessary admissions and upcoding—the ones on the right-hand side. This is because compliance is always on the units that are more likely to be detected and expected detection costs subtract from the compliance prices, making compliance on units with high expected detection costs effectively cheaper. Therefore, the units of compliance that the hospital decision maker chooses to “buy” are always cheaper than the ones it chooses not to “buy.”

Having formulated the optimization problem in the model augmented with compliance in the utility function, we now consider the effects of increases in the reimbursement for top tier coding ω^h which implies an increase in π^h. On the one hand, it becomes more “expensive” to follow coding guidelines and this leads to more inappropriate coding behavior (*substitution* effect). But on the other hand, reimbursement rate becomes higher, and thus, decision makers in hospitals can "afford" to engage in more legitimate coding and indulge their preference for compliance (*income* effect). Ultimately, the effect of a higher ω^h on the consumption of m^u or the level of upcoding (i.e., $\underline{s}^u - \tilde{s}^u$) depends on the interplay between the income and substitution effects: If the substitution effect is smaller than the income effect, hospitals will tend to be more compliant, (i.e., upcode less).

Note that variation in ω^h might also affect the consumption of m^a or the number of inappropriate admissions (through a *cross* price effect), which also contains the income and substitution effects. Similarly, if the income effects dominates, a higher ω^h will result in greater consumption of m^a. However, if the income effect does not dominate, the prediction of the cross price effect would be ambiguous. We summarize how ω^h affects $\tilde{s}^a$ and $\tilde{s}^u$ in the following proposition:

Proposition 3 With greater ω^h, if the income effect dominates the substitution effect, we expect fewer total admissions within the base DRG and fewer admissions to the high-severity level.

Proof. See Appendix I.[23]

3.2.3 Models with full compliance and technological changes

In the following, we consider a set of models with the following two features: (1) Hospitals comply with all the admission and coding protocols, i.e., they engage in no unnecessary admissions or inappropriate coding whatsoever; and (2) the change in treatment technologies leads to simultaneous changes in both the usage of hospital services and the reimbursement for these services. For instance, the emergence of minimally invasive surgery is usually associated with less pain, fewer complications, and shorter hospital stay.[24] Patients who suffer from certain severe conditions and used to be treated outside hospitals to avoid open surgeries may now receive these new treatments, resulting in an increasing number of hospital stays. In the mean time, the reimbursement rates for the same conditions may drop as fewer resources are required to treat these patients (Xu et al., 2015).

Since hospitals are fully compliant with the admission and coding guidelines, patients

[23]Note that we illustrate Proposition 3 using a constant elasticity of substitution (CES) utility function with two goods. It is a simplified version of our model, but we expect the implications hold in a more general setup.

[24]See `https://www.mayoclinic.org/tests-procedures/minimally-invasive-surgery/about/pac-20384771`.

Table 1: Testable Implications of $\uparrow \omega^h$ on Aggregate DRG-level Claims Count

Model	Admissions	Top-tier coding
Financial-driven mechanism	No effect	More
Financial-driven mechanism enhanced by coding requirements	No effect	Fewer if the effect from the change in $\underline{s}^u$ dominates or more otherwise
Compliance utility model	Fewer if income effect dominates	Fewer if income effect dominates
Models with full compliance & technological changes	Fewer	Fewer

who are qualified for admission—that is, whose severity levels go beyond $\underline{s}^a$—will be admitted and those meeting the top-tier threshold—that is, those with severity levels above $\underline{s}^u$—will be top coded. However, the technological changes might generate a negative correlation between the number of patient admitted to these severe conditions (i.e., the top-tier DRGs) and the reimbursement rates for treating patients with these diagnoses.[25] The effect on total admissions in a given DRG will follow the same direction as that on the top tier. This leads to the last testable implication:

Proposition 4 In the presence of technological advancement and hospitals' fully compliant behavior, we expect more admissions within the base DRG and/or more patients coded into the top tier as ω^h drops.

To summarize, the various special cases have different testable implications in the number of total admissions and top-coded patients. Table 1 shows the implications of these 4 specialized mechanisms:

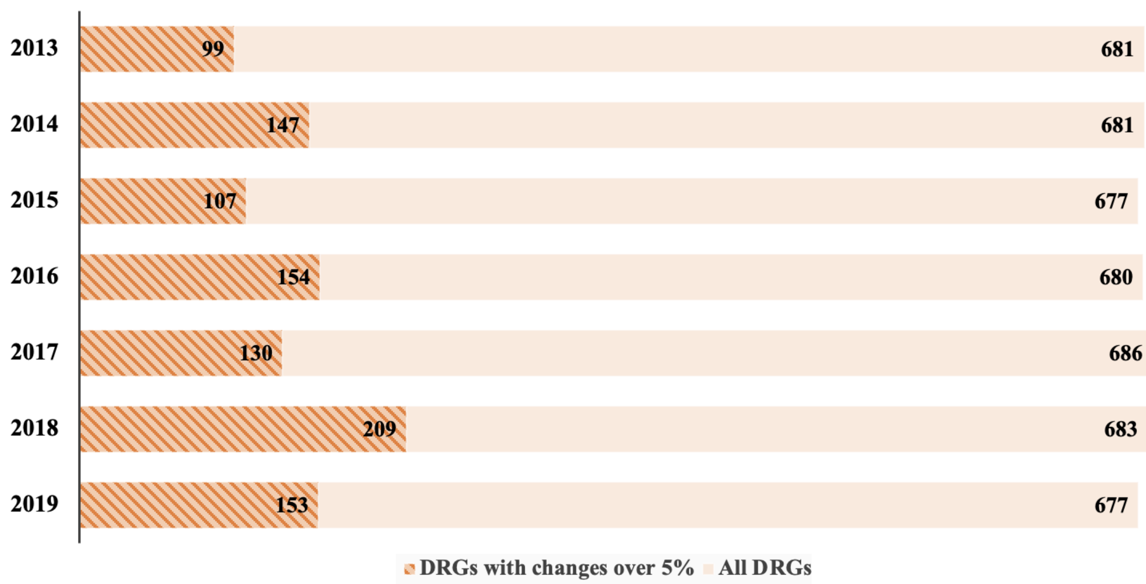
[25]Proposition 4 only shows the directional changes due to technology advancement, but these models can also capture the mechanism driven by the emergence of more extensive hospital treatments that result in longer and more complicated hospital stays. We expect to observe the same negative correlation between the resulting hospital admissions and reimbursement rates: Fewer patients may desire these treatments (and thus, fewer admissions to the top-tier DRGs and the given base DRGs) and higher reimbursement weights in the corresponding DRGs because of more resources involved to treat the patients.

4 Data

4.1 Aggregate Medicare Part A Claims

Our first dataset comes from the files for the IPPS final rules and correction notice published by CMS in 2012-2019, which include wage indices, the list of DRGs, DRG weights, mean length of stay, the number of discharges for each DRG, and IPPS operating and capital statewide average cost-to-charge-ratios. We construct a base DRG by grouping together DRGs sharing a common primary diagnosis/procedure. We use the aggregate counts of discharges at the DRG level to construct the two dependent variables: total admissions within a base DRG and the number of admissions to the top severity subclass. The key variable of interest is the DRG weight, which approximately measures the revenue hospitals receive from treating a patient that is assigned to the particular DRG.

Figure 2: DRGs with weight changes over 5% from last year



Our analysis focuses on base DRGs with at least two severity subclasses. On average, there are 748 DRGs, about 682 of which belong to base DRGs with multiple tiers. Table 2 shows the summary statistics for these base DRGs, separately by severity subclass. On average, we observe more discharges among the non-top severity subclasses. Almost all DRGs experience annual weight adjustments. Figure 2 shows the number of DRGs whose weight changes over 5% relative to the previous year. These DRGs account for 15% – 30% of all the DRGs with multiple tiers. Table 2 also shows that the mean DRG weight is relatively stable over time, but the changes in weights between a pair of successive years—denoted by Δwt—vary a lot across DRGs and years, creating important variation for our identification.

Table 2: Summary statistics for aggregate claims data

	2012	2013	2014	2015	2016	2017	2018	2019
Top severity subclasses								
DRG wt	2.88	2.89	2.89	2.91	2.92	2.94	2.91	2.95
	(2.29)	(2.34)	(2.33)	(2.36)	(2.38)	(2.41)	(2.30)	(2.33)
Δwt	–	0.00286	-0.00455	-0.00104	-0.00206	0.01383	-0.02253	0.04661
	–	(0.167)	(0.160)	(0.131)	(0.176)	(0.153)	(0.224)	(0.178)
# discharges	10,708	11,382	11,323	11,652	11,613	12,408	12,719	13,260
	(26,159)	(28,946)	(29,690)	(32,714)	(34,270)	(39,302)	(41,144)	(46,280)
# base DRGs	263	263	264	263	265	265	265	268
Bottom severity subclasses								
DRG wt	1.32	1.34	1.37	1.40	1.43	1.45	1.48	1.46
	(1.29)	(1.29)	(1.36)	(1.42)	(1.38)	(1.44)	(1.39)	(1.32)
Δwt	–	0.01573	0.02737	0.01813	0.01871	0.02385	0.02239	-0.00950
	–	(0.062)	(0.112)	(0.064)	(0.108)	(0.115)	(0.149)	(0.143)
# discharges	16,612	15,969	14,754	13,741	12,667	12,373	11,897	11,431
	(41,130)	(39,864)	(37,961)	(37,060)	(36,255)	(36,142)	(36,924)	(37,045)
# base DRGs	263	263	263	263	265	265	264	267
Middle severity subclasses								
DRG wt	1.84	1.84	1.83	1.85	1.87	1.85	1.85	1.82
	(1.09)	(1.11)	(1.13)	(1.18)	(1.16)	(1.15)	(1.10)	(1.10)
Δwt	–	-0.00216	-0.00608	0.01537	0.00945	-0.00026	-0.00226	0.00845
	–	(0.079)	(0.078)	(0.093)	(0.077)	(0.053)	(0.174)	(0.082)
# discharges	17,030	17,145	16,644	16,158	15,326	15,272	15,090	14,139
	(33,804)	(34,352)	(32,623)	(31,940)	(29,654)	(29,778)	(27,991)	(24,945)
# base DRGs	153	153	153	154	155	154	153	159

Note: Table reports the mean value, with the standard deviation in parenthesis. The number of base DRGs is smaller for the group of middle severity subclasses because only base DRGs with three tiers contain the middle level.

4.2 CERT Microdata

The second dataset we rely on is the Medicare CERT improper payment data in 2012-2019, a *random* sample of Medicare Parts A and B claims that are reviewed by CERT auditors. In particular, we focus on Part A IPPS claims. For each claim, the data contains information on provider type, type of bill, assigned DRG and procedure codes, and whether a determination of improper payment was made due to one of the five error types specified by CMS. We focus on upcoding-related errors, which occur if the claim is determined to have incurred in one of the following four errors: (1) incorrect coding, (2) insufficient documentation, (3) no documentation, and (4) others. The definition is made based on these error types, as the occurrence of any of them—such as inadequate justification for the reported code due to no or insufficient documentation, assigning the incorrect codes, or other potential miscellaneous causes—could all be associated with upcoding practices.[26]

Table 3 reports the summary statistics based on the CERT data for the base DRGs with multiple severity subclasses. Despite a relatively small number of claims in the data compared with the entire Part A IPPS population, the sampling strategy in CERT follows a stratification plan that enables a good representation of the overall population. Claims with upcoding-related errors account for 7% – 9% of the entire sample.

Table 3: Summary statistics for CERT microdata

	2012	2013	2014	2015	2016	2017	2018	2019
Upcoding (%)	8.71	9.27	8.83	5.45	7.54	7.50	7.11	8.71
# claims	10,744	12,519	10,936	12,577	12,362	11,166	11,926	11,768

Note: Table reports the proportion of audited claims that were flagged with upcoding-related, improper payment errors.

5 Empirical Analysis of Aggregate Medicare Part A Claims

In this section, we examine the testable implications summarized in Table 1 using the aggregate Medicare Part A claims data, to identify which mechanism was the most prevalent during our study period. We first describe our fixed effects specification[27] and then discuss the empirical results.

5.1 Empirical strategy

We estimate the following model, by exploiting the variation in reimbursements over time within a base DRG:

$$Y_{dt} = \alpha + \beta^h \omega_{dt}^h + \beta^\ell \omega_{dt}^\ell + \lambda_d + \lambda_t + \mu_d \cdot t + \varepsilon_{dt} \tag{5}$$

[26] As a robustness check, we also define upcoding-related errors only based on the error type "incorrect coding"—the one that is most related to upcoding. We present the results in Appendix Tables A3, and the main findings hold.

[27] Note that it could be more natural to use a multinomial model to examine coding responses—which of the three tiers a patient is assigned to—within a base DRG. However, it is difficult to control for various fixed effects in a non-linear model. Our linear probability model approach allows us to easily control for various fixed effects, enhancing the credibility of our identification strategy.

where Y_{dt} denotes the outcome of interest, which could be total admissions within a base DRG, d, or the number of admissions to the top severity subclass in this base DRG. In the model notation from Section 3.1, total admissions at the base DRG level are captured by $1 - \tilde{s}^a$ and admissions into the top tier DRG correspond to $1 - \tilde{s}^u$.

The key variables of interest are the ω's. We use the DRG weight at the *top* tier for ω^h_{dt}. In the analysis of total admissions, we use the DRG weight at the *bottom* level to measure ω^ℓ_{dt}. In the analysis of admissions into the top tier, we use the weight at the *next lower* level to measure ω^ℓ_{dt}.[28] The rationale for these choices is that the bottom-tier DRG weight is more relevant to the admission decision. The DRG weight of the next lower level serves as the baseline for the incremental gain of top-coding and thus, matters more in deciding whether to code a patient into the true severity level or the next (higher) severity level.

The rest of the variables in Equation (5) represent various fixed effects or interactions with time trends. λ_d and λ_t, respectively, denote base DRG and year fixed effects. Note that including base DRG fixed effects here is equivalent to using DRG fixed effects, because we consider only the top severity subclass for each base DRG. Thus, including the (base) DRG fixed effects controls for the time-invariant unobserved heterogeneity at this level, such as pre-existing differences in coding patterns or in the fraction of severe patients between DRGs. The year fixed effects flexibly capture changes over time for the dependent variable. $\mu_d \cdot t$, denotes (base) DRG-specific linear trends. They capture secular changes over time at the (base) DRG level, such as he long-term secular changes in patient population with a particular disease. For instance, patients who are less severe with coronary heart disease might no longer receive care in hospitals due to medical treatments being available in the outpatient setting. In this case, one might observe smaller patient volume and increasing reimbursement rates for DRGs related to this disease. The DRG-specific trends could also capture the increasing trend of Medicare Part A beneficiaries switching into Medicare Advantage plans.

Our regression analysis first explores how the number of admissions to the top bill code varies in response to the change in revenues, captured by DRG weights. Our specification differs from that in related studies (Dafny, 2005; Li, 2014; Ganju et al., 2021; Gowrisankaran et al., 2022) in the following aspects. First, we include the DRG weights of both the top and lower levels in the regression instead of using the spread—the difference in DRG weights between the top and low levels. While we understand that spread captures the extra gains from upcoding, our specification allows for a more flexible functional form to capture the effect of revenue changes on coding intensity. Using spread imposes the assumption that a

[28]For example, if the base DRG has three tiers, the next lower level would be the middle tier. If the DRG has only two tiers, then the next lower level would be the bottom tier.

unit increase in ω_{dt}^h has the same impact as a unit decrease in ω_{dt}^ℓ on coding, but in fact, the marginal effect of each of them could be different in magnitude. As a robustness check, we replace the DRG weights (ω_{dt}^h and ω_{dt}^l) with spread in the main regression and present the results in Appendix Tables A1 and A2. The finding is consistent.

Second, we use the (log of the) *number* of admissions to top severity subclass as the dependent variable, whereas prior studies use the fraction of patients admitted to the high severity subclass among all patients within the same base DRG. We use the count instead of the share, because ω_{dt}^ℓ could affect both the numerator and denominator in the fraction. Assume that hospitals have no incentives for upcoding and only for admitting unnecessary patients for simplicity. An increase in ω_{dt}^ℓ—a reduction in spread—leads to more admissions into the given base DRG and relatively fewer admissions into the top level within the base DRG, resulting in a decrease in the fraction top-coded. In this case, the estimated ω_{dt}^ℓ would be negative if we use the share as the dependent variable, which would be interpreted as evidence of upcoding. However, in fact, the incentive for upcoding is absent here.

The key coefficients of interest are β^ℓ and β^h, both measuring the marginal effects of the weights. Our identification relies on variation within a base DRG across time.[29] If upcoding is mainly driven by financial incentives as suggested in the literature, we expect β^h to be positive and/or β^ℓ to be negative, which is equivalent to a positive effect of spread. In contrast, a negative β^h and/or a positive β^ℓ might suggest other mechanisms proposed in Section 3.2. While we mainly focus on coding intensity in response to the changes in financial payoff, we also examine another outcome—total admissions within a base DRG, which helps us distinguish across mechanisms. Our identifying assumption is that the *changes* in DRG weights are exogenous to admission/coding decisions after controlling for the various fixed effects and base DRG-specific trends. Below in Section 7 we supplement our current analysis in a way that supports the validity of this assumption. We weight our regressions by the mean number of discharges over time within a base DRG and cluster standard errors at the same level.

5.2 Results

We present the estimated coefficients for the key variables of interest in Table 4. The first coefficient column reports the estimates for β^h and β^l based on the entire sample. It suggests

[29] It could be interesting to test upcoding behavior by examining whether hospitals move patients across base DRGs. For instance, Silverman and Skinner (2004) found hospitals upcode patients by substituting DRGs associated with pneumonia with DRGs associated with respiratory infections and inflammations. Following the majority of the literature though, we use all DRGs in a fixed effect model to estimate the average response across all DRGs but limiting upcoding opportunities to be only within a base DRG.

that the *greater* the DRG weight for the top tier is, the *fewer* admissions there will be assigned to that tier. Specifically, we expect an average reduction of 5.8% in the admissions to the top tier, as the corresponding DRG weight increases by 0.2, approximately 10% of the average DRG weights across all DRGs in multiple-tier base DRGs. The coefficient for the DRG weight at the next low level DRG is positive but not significant. This is different from the prediction in the financial-driven mechanism in Table 1, suggesting that while the financial incentive to both, unnecessarily admit and upcode patients, is still present and prevents a fully compliant behavior, other mechanisms could explain why we see fewer admissions into a DRG when its reimbursement rate increases.

Table 4: Effect of DRG weights on Log(Admissions) to the top tier within a base DRG

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Top-tier DRG weight	-0.2906**	-0.1552**	-2.0508*	-0.1912***	-0.6948*	-0.1117*
	(0.1228)	(0.0602)	(1.1446)	(0.0720)	(0.3806)	(0.0579)
Next lower-tier DRG weight	0.0039	-0.0766*	6.0525**	-0.1493	0.2047	-0.0218
	(0.1061)	(0.0449)	(2.7375)	(0.2505)	(0.4485)	(0.0968)
N	2110	209	190	1711	526	1584
R^2	0.998	0.998	0.993	0.999	0.997	0.979
MeanDepVar	8.155	8.088	8.618	8.112	10.118	7.504

Note: Analysis based on Medicare Part A Population data. Unit of observation is DRG/year. Other regressors include DRG fixed effects, year fixed effects, and DRG-specific time trends. Regression weighted by total number of discharges per base DRG. Standard errors in parentheses, clustered at base DRGs.

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

The rest of the columns in Table 4 report results that explore heterogeneity across two dimensions. We first decompose the sample into DRGs with high weights (the DRGs with the top 10% highest weights), DRGs with low weights (the bottom 10%), and the remaining DRGs (those with weights within the 10% to 90% range). The general finding holds: fewer admissions will be assigned to the top tier when the corresponding weight becomes higher. Interestingly, among the bottom 10% DRGs, more top bill codes are reported when the weight of the lower level DRG rises, suggesting that a greater spread—that could arise from the increase in DRG weight from the top level and/or the reduction from the lower level—results in fewer top bill codes. The last two columns show the results separately for common DRGs (the top 30% in terms of the average number of discharges across years) and the less common ones (the remaining DRGs), and the main results hold.[30]

[30]Note that the number of observations among common DRGs is smaller because the unit of observation is a unique DRG and year cell. The analysis is based on DRGs ranked top 30% in terms of average discharges.

To sum up, the negative effect of top-tier DRG weights on admissions into the top tier DRGs suggests that mechanisms might also underlie the coding behavior during our sample period. As our model proposed, this could be related to a tied increase in coding requirements, the dominating income effect in the demand for compliance among hospital decision-makers that care about some "internal conscience," or the technological advancement absence of hospital financial incentive. In the following, we explore these other mechanisms by examining the effect of financial incentives on a different outcome.

Table 5: Effect of DRG weights on Log(Total Admissions) in a base DRG

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Top-tier DRG weight	-0.117*	-0.141	0.073	-0.080	-0.267**	-0.034
	(0.069)	(0.100)	(0.553)	(0.070)	(0.106)	(0.072)
Bottom-tier DRG weight	-0.170	-0.132	2.656	-0.324	-0.200	-0.174
	(0.127)	(0.133)	(1.633)	(0.292)	(0.138)	(0.163)
N	2110	209	190	1711	526	1584
R^2	0.998	0.996	0.996	0.999	0.998	0.989
MeanDepVar	9.292	9.031	9.607	9.288	11.157	8.672

Note: Analysis based on Medicare Part A Population data. Other regressors include base DRG fixed effects, year fixed effects, and linear time trends by base DRGs. Each observation is a combination of a base DRG and year. Regression weighted by total number of discharges per base DRG. Standard errors in parentheses, clustered at base DRGs.

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

Table 5 reports the estimated effect of DRG weights on total admissions in a similar format to Table 4. The analysis based on the entire sample shows a negative effect of the top-tier DRG weight on the total admissions, yet with less significance and magnitude. The coefficients for the weight of the top bill code turn insignificant when we break down the sample by DRG weight magnitude, but it is still large and significantly negative in the subsample of common DRGs, suggesting that most of the effect occurs to the common DRGs. The finding here suggests that an increase in financial incentives that is tied with increased stringency in coding requirements cannot be the only explanation behind hospitals' admission and coding decision.[31] This leaves us with the model of financial incentives augmented with compliance utility and the model with full compliance and technological changes. We discuss how we separate these two models in the next section.

[31] Note that the model with the simultaneous increase in both reimbursements and coding requirements could still be a dominant mechanism if hospitals' measure of patient severity upon admission is noisy. We implicitly assume that hospitals can precisely figure out a patient's severity of disease at the time of admission, which could be a restricted assumption.

6 Testable Implications in a Random Sample of Audits

To distinguish the financial incentives model augmented with compliance utility from the models of technological change where financial incentives are absent, one possibility is to look for instruments that generate exogenous variation in reimbursements within a DRG over time. An alternative to which we now turn, is to explore the different implications these models have in a random sample of audits. Specifically, we focus on the different predictions these two models have on the probability that upcoding is found among top-coded claims, in the random audit data.

To better explain the idea, suppose the model emphasizing technological change is the primary underlying mechanism. Hence, all the billing errors arise from random mistakes, such as missing signature. An increase in DRG weights should *not* alter the probability of detecting incidental errors from a random sample of claims. But in contrast, if the financial incentive is the main driving force, hospital decision makers who are concerned about the dis-utility of noncompliance should vary the extent of inappropriate coding in response to the change in DRG weights. This should, in turn, be reflected in a random sample of audits as a change in the probability hat a top-billed claim is flagged as upcoded. To separate these two models, we exploit the fact that these models generate different implications in the case of CERT audits.

All individuals whose severity level s is such that $s > \tilde{s}^a$ are admitted to hospitals and constitute the universe of admissions. The CERT program takes a random sample from this universe and examines whether each case is subject to improper billing after auditors evaluate all the relevant information. The extent to which hospitals upcode cases can be represented by $|\tilde{s}^u - \underline{s}^u|$. Since $\underline{s}^u$ is simply another parameter and we only consider the change in financial incentives here, i.e., $cov(\underline{s}^u, \omega^h) = 0$, the compliance utility model with a dominant income effect implies that we should expect to see fewer claims flagged with upcoding-related errors in the CERT data when reimbursements increase within a DRG.

In the other mechanism that is purely driven by technological changes, hospitals top-code all qualified patients. One should not expect any changes in the incidence of upcoding-related errors as the top-tier DRG weight varies. Note that the CERT audit data may not detect all upcoding but simply reflects the level of upcoding that was detected. This is captured in our model by the fact that $\phi^u < 1$ in the range of s where the hospital may upcode, $(\underline{s}^a, \underline{s}^u)$. And while by looking at audit records we are one step removed from the details of the medical record, given that it is a random sample of the population claims, it could still inform the *directional changes* in coding intensity as the financial payoff varies. The following table summarizes the testable implications in CERT for both models:

Table 6: Summary of Testable Implications of $\uparrow \omega^h$ in Audits Data

Model	Upcoding-related errors
Compliance utility model	Fewer if income effect dominates
Models with full compliance & technological changes	No effect

7 Empirical Analysis of Audit Microdata

To test the additional implications available in a sample of audits, we estimate a similar specification as in Equation (5) but now using the CERT microdata. The dependent variable is a *binary indicator* which equals one when an audited claim is flagged with upcoding-related errors. In essence, the model is a linear probability model for upcoding. This model estimates how the probability that a randomly selected claim gets flagged with an "upcoding" determination depends on the DRG weight.

The advantage of using the CERT data—a random sample of audits—is that it allows us to distinguish between the "compliance utility" and "technology endogeneity" hypotheses. As discussed above, these hypotheses have different implications with respect to how the upcoding rates in audits should change with reimbursement weights. A threat to identification would occur, if, for instance, the selection of claims for review is a function of certain technology changes that might also affect the setting of reimbursement rates. However, we believe the likelihood is small, because the CERT program has been using the current random sampling strategy since the early 2000s, which lends credibility to the approach.

We identify the effect of financial incentives based on the changes in the top-tier DRG weights and the changing incidence of audited claims with upcoding flags for this DRG. We use the DRG fixed effects specification. Identification comes from within-DRG variation in reimbursements over time. The unit of observation is a claim-year combination. To account for stratification, we weight each observation with the relative frequency with which claims from a given DRG in the CERT sample compare with the aggregate claims data.

Table 7 summarizes the results in a similar format to Table 4. The first column suggests that relatively fewer upcoding-related errors are detected among DRGs with an increasing weight. It is equivalent to a reduction of 2.47 percentage point in the incidence of these errors as the corresponding weight increases by 0.2. When we break down the DRGs by DRG weights, both of those standing above the 90 percentile and below the bottom 10 percentile experience fewer upcoding-related errors as the top-tier DRG weight gets higher.

Table 7: Effect of DRG weights on upcoding-related determinations

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Top-tier DRG weight	-0.1235**	-0.1736*	-0.3610**	-0.0433	-0.2711**	-0.0273
	(0.0558)	(0.0949)	(0.1322)	(0.0875)	(0.1094)	(0.0822)
Next lower-tier DRG weight	0.0142	0.0953	0.6314	-0.1591	0.2333	-0.0289
	(0.0571)	(0.0841)	(0.4641)	(0.1418)	(0.2052)	(0.0489)
N	25276	2779	3263	19234	18539	6737
R^2	0.052	0.152	0.022	0.057	0.030	0.181
MeanDepVar	0.083	0.091	0.082	0.083	0.081	0.091

Note: Analysis based on CERT microdata. Other regressors include DRG fixed effects, year fixed effects, and DRG-specific time trends. Each observation is a claim-year combination. Regression weighted by the relative frequency per DRG per year. Standard errors in parentheses, clustered at base DRGs.

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

Also, such a negative effect seems to mainly arise from the most common DRGs among all admissions.

To summarize, the significantly negative effect of financial incentives on the occurrence of upcoding-related errors from the audit data lends support to the compliance-utility model. It is consistent with the mechanism where hospital decision-makers pay attention to and care about financial incentives but show a preference for compliance in admission and billing practices. This occurs when the income effect out-weights the substitution effect in their decision-making. This desire for compliance only limits, but does not eliminate the role financial incentives play in inducing upcoding.

Robustness checks. We conduct the following robustness checks. First, we replace the DRG weights with the variable "spread"—the difference in DRG weight between the top and bottom severity subclasses within a base DRG. It captures the incremental gains if a patient is moved from the bottom to the top tier. Appendix Tables A1 and A2 report the estimates, which suggest the same findings as our main results. Second, we adopt a more narrow definition of upcoding-related errors, only based on the error type "incorrect coding." Appendix Table A3 presents the coefficients and shows that our results are robust to this alternative definition. Finally, the opportunity for successfully upcoding could be higher for certain DRGs but not all of them. For instance, it is probably hard to upcode patients with a disease related to a solid organ transplant. To verify whether our findings are driven by DRGs where patients are hard to be upcoded, we rerun our main specifications after dropping DRGs with low upcoding rate—whose incidence of upcoding-related error is below the median in the overall sample. We report the results in Appendix Tables A4 and

A5. The main findings still hold.

8 Conclusion

We study to what extent inappropriate coding practices respond to increases in reimbursement rates in recent years in the context of Medicare Part A inpatient hospital care. Our paper features both theoretical and empirical contributions. On the theoretical side we bring to bear models typically associated with individual physician behavior (e.g. Gruber and Owings (1996)) adding to the complex forces driving billing decisions by hospital decision-makers. To guide our empirical investigation and explore the possible underlying mechanisms, we consider a general model characterizing hospitals' decisions regarding the admission of patients and their billing coding Our model incorporates both the financial incentives and a preference for compliant behavior into hospital decision makers' utility function. This feature, while novel in the hospital upcoding literature, is now standard in the tax evasion/tax compliance literature, where it is referred to as "tax morale" (e.g. Luttmer and Singhal (2014)). Moreover, we develop several specialized models that feature particular mechanisms from the general model which have been emphasized in the literature.

We test these specialized models by analyzing claim counts from Medicare at the base DRG level using changes in reimbursement within DRGs over time as the key source of identifying variation. We find a *negative* effect of the top-tier DRG weight on the number of admissions into the top tier of those DRGs and on the overall number of admissions into the base DRG. While the aggregate claims data allows us to distinguish across some models, it is in principle consistent with a model where hospital decision-makers are influenced by financial incentives but show a preference for compliance and a set of models that emphasize technological changes but lack the financial incentive.

To further test these two types of models, we rely on a novel empirical strategy based on microdata from a random sample of audited claims. We find that fewer claims are flagged with upcoding-related determinations as the top-tier DRG weight increases. It suggests that a model with financial incentives and a "utility from coding compliance" feature where demand for compliance is dominated by a strong income effect might be consistent with our evidence from both aggregate claims and the audit sample.

Our findings support the use of a well-known mechanism from physician behavior models in an important literature that seeks to explain how hospitals engage in coding practices in response to changing financial incentives. This new mechanism seems important to understand hospital behavior in recent years. Our finding also imply that the decisions on admission and coding by hospitals could be quite sophisticated. As a result, policymakers might consider imposing more comprehensive reward/penalty mechanisms to encourage proper billing.

References

CERT Program (2011-2020). Medicare fee-for-service improper payments report. *Centers for Medicare and Medicaid Services*. 1, 2.2, 12

Cook, A. and S. Averett (2020). Do hospitals respond to changing incentive structures? Evidence from medicare's 2007 drg restructuring. *Journal of Health Economics 73*, 102319. 1, 3.1

Dafny, L. (2005). How do hospitals respond to price changes? *American Economic Review 95*(5), 1525–1547. 1, 3.1, 3.2.1, 3.2.1, 5.1

Dranove, D. (1988). Demand inducement and the physician/patient relationship. *Economic inquiry 26*(2), 281–298. 1

Fang, H. and Q. Gong (2017). Detecting potential overbilling in medicare reimbursement via hours worked. *American Economic Review 107*(2), 562–91. 1

Finkelstein, A., M. Gentzkow, P. Hull, and H. Williams (2017). Adjusting risk adjustment–accounting for variation in diagnostic intensity. *The New England journal of medicine 376*(7), 608. 1

Ganju, K. K., H. Atasoy, and P. A. Pavlou (2021). Do electronic health record systems increase medicare reimbursements? The moderating effect of the recovery audit program. *Management Science*. 5.1

Geruso, M. and T. Layton (2020). Upcoding: Evidence from medicare on squishy risk adjustment. *Journal of Political Economy 128*(3), 984–1026. 1, 6

Gowrisankaran, G., K. Joiner, and J. Lin (2022). How do hospitals respond to Medicare payment reforms? *NBER Working Paper No. 22873*. 3.1, 3.2.1, 5.1

Gruber, J. and M. Owings (1996). Physician financial incentives and cesarean section delivery. *RAND Journal of Economics 27*(1), 99–123. 1, 3.1, 8

Howard, D. H. and I. McCarthy (2021). Deterrence effects of antifraud and abuse enforcement in health care. *Journal of Health Economics 75*, 102405. 1

Jürges, H. and J. Köberlein (2015). What explains drg upcoding in neonatology? The roles of financial incentives and infant health. *Journal of health economics 43*, 13–26. 3.2.1

Kuhn, T., P. Basch, M. Barr, T. Yackel, and M. I. C. of the American College of Physicians* (2015). Clinical documentation in the 21st century: executive summary of a policy position paper from the american college of physicians. *Annals of internal medicine 162*(4), 301–303. 13

Leder-Luis, J. (2020). Can whistleblowers root out public expenditure fraud? evidence from medicare. 1

Li, B. (2014). Cracking the codes: Do electronic medical records facilitate hospital revenue enhancement. *Working paper*. 3.1, 5.1

Luttmer, E. and M. Singhal (2014). Tax morale. *Journal of Economic Perspectives 28*(4), 149–168. 1, 8

McGuire, T. G. and M. V. Pauly (1991). Physician response to fee changes with multiple payers. *Journal of health economics 10*(4), 385–410. 1, 3.1

Office of Evaluation and Inspections (2001). Medicare Hospital Prospective Payment System: How DRG Rates Are Calculated and Updated. *Office of Inspector General*. 2.1

Office of the Federal Register and National Archives and Records Service (2007). Medicare program; changes to the hospital inpatient prospective payment systems and fiscal year 2008 rates. *Federal Register (Wednesday, August 22, 2007) 72*(162), 47129–48175. From the Federal Register Online via the Government Publishing Office [www.gpo.gov], FR Doc No: 07-3820. 8

Recovery Audit Program (2011-2016). Recovery auditing in Medicare for Fiscal Year 2011 – 2016. *Centers for Medicare and Medicaid Services*. 10

Sacarny, A. (2018). Adoption and learning across hospitals: The case of a revenue-generating practice. *Journal of Health Economics 60*, 142–164. 1, 13, 3.2.1, 21

Shi, M. (2021). The costs and benefits of monitoring providers: Evidence from medicare audits. *Working paper*. 1

Silverman, E. and J. Skinner (2004). Medicare upcoding and hospital ownership. *Journal of health economics 23*(2), 369–389. 1, 3.2.1, 29

Xu, T., S. M. Hutfless, M. A. Cooper, M. Zhou, A. B. Massie, and M. A. Makary (2015). Hospital cost implications of increased use of minimally invasive surgery. *JAMA surgery 150*(5), 489–490. 3.2.3

Appendix I

Proof of Proposition 1

In this specialized model, the hospital decision-maker aims to

$$\begin{aligned}
&\max_{\tilde{s}^a,\, \tilde{s}^u} E[U(c)] \\
\text{s.t. } c = &(\underline{s}^a - \tilde{s}^a)\left(\pi^\ell - f(\tilde{s}^a, \underline{s}^a)\right) + (\underline{s}^u - \tilde{s}^u)\left(\pi^h - \pi^\ell - g(\tilde{s}^u, \underline{s}^u)\right) \\
&+ (\underline{s}^u - \underline{s}^a)\pi^\ell + (1 - \underline{s}^u)\pi^h
\end{aligned}$$

Consider an increase in ω^h, which leads to an increase in π^h. We are interested in how the admission and coding decision varies as π^h increases. In the following, we describe and prove the first proposition.

Proposition 1 As ω^h and thus π^h gets higher, we expect no changes in total admissions within the base DRG, i.e., $\partial \tilde{s}^a / \partial \pi^h = 0$, and more admissions to the high-severity level, i.e., $\partial \tilde{s}^u / \partial \pi^h < 0$.

Proof. The first order conditions are

$$\begin{aligned}
\frac{\partial U}{\partial \tilde{s}^a} &= U_c\left[(\underline{s}^a - \tilde{s}^a)(-f_1) + (\pi^\ell - f(\tilde{s}^a, \underline{s}^a))(-1)\right] = 0; \\
\frac{\partial U}{\partial \tilde{s}^u} &= U_c\left[(\underline{s}^u - \tilde{s}^u)(-g_1) + (\pi^h - \pi^\ell - g(\tilde{s}^u, \underline{s}^u))(-1)\right] = 0
\end{aligned}$$

where $f_1 = \partial f(\tilde{s}^a, \underline{s}^a)/\partial \tilde{s}^a$ and $g_1 = \partial g(\tilde{s}^u, \underline{s}^u)/\partial \tilde{s}^u$. After simplification,

$$\begin{aligned}
f - \pi^\ell - f_1(\underline{s}^a - \tilde{s}^a) &= 0; \\
g - \pi^h + \pi^\ell - g_1(\underline{s}^u - \tilde{s}^u) &= 0.
\end{aligned} \tag{6}$$

By assuming that $\Delta B \omega^h > \Delta c^h$, we know that the net change leads to higher π^h. Take derivatives w.r.t. π^h on both sides:

$$\begin{aligned}
f_1 \frac{\partial \tilde{s}^a}{\partial \pi^h} - f_{11}(\underline{s}^a - \tilde{s}^a)\frac{\partial \tilde{s}^a}{\partial \pi^h} + f_1 \frac{\partial \tilde{s}^a}{\partial \pi^h} &= 0; \\
g_1 \frac{\partial \tilde{s}^u}{\partial \pi^h} - 1 - g_{11}(\underline{s}^u - \tilde{s}^u)\frac{\partial \tilde{s}^u}{\partial \pi^h} + g_1 \frac{\partial \tilde{s}^u}{\partial \pi^h} &= 0.
\end{aligned}$$

$\Longrightarrow$

$$\begin{aligned}
&\frac{\partial \tilde{s}^a}{\partial \pi^h}\left[2f_1 - f_{11}(\underline{s}^a - \tilde{s}^a)\right] = 0 \;\rightarrow\; \frac{\partial \tilde{s}^a}{\partial \pi^h} = 0 \\
&\frac{\partial \tilde{s}^u}{\partial \pi^h}\left[2g_1 - g_{11}(\underline{s}^u - \tilde{s}^u)\right] = 1 \;\rightarrow\; \frac{\partial \tilde{s}^u}{\partial \pi^h} = \frac{1}{2g_1 - g_{11}(\underline{s}^u - \tilde{s}^u)} < 0
\end{aligned}$$

As a result, the testable implications for a higher ω^h are no change in admissions and more top-tier coding in claims.

Proof of Proposition 2

Based on the same setup as the previous proposition, we consider the change in $\underline{s}^u$. Take derivatives w.r.t. $\underline{s}^u$ on both sides of Equations (6):

$$f_1 \frac{\partial \tilde{s}^a}{\partial \underline{s}^u} - f_{11}(\underline{s}^a - \tilde{s}^a)\frac{\partial \tilde{s}^a}{\partial \underline{s}^u} + f_1 \frac{\partial \tilde{s}^a}{\partial \underline{s}^u} = 0;$$

$$g_1 \frac{\partial \tilde{s}^u}{\partial \underline{s}^u} + g_2 - g_{11}(\underline{s}^u - \tilde{s}^u)\frac{\partial \tilde{s}^u}{\partial \underline{s}^u} - g_{12}(\underline{s}^u - \tilde{s}^u) - g_1 + g_1 \frac{\partial \tilde{s}^u}{\partial \underline{s}^u} = 0.$$

$\implies$

$$\frac{\partial \tilde{s}^a}{\partial \underline{s}^u}\left[2f_1 - f_{11}(\underline{s}^a - \tilde{s}^a)\right] = 0 \ \rightarrow \ \frac{\partial \tilde{s}^a}{\partial \underline{s}^u} = 0;$$

$$\frac{\partial \tilde{s}^u}{\partial \underline{s}^u}\left[2g_1 - g_{11}(\underline{s}^u - \tilde{s}^u)\right] = g_1 - g_2 + g_{12}(\underline{s}^u - \tilde{s}^u) \ \rightarrow \ \frac{\partial \tilde{s}^u}{\partial \underline{s}^u} = \frac{g_1 - g_2 + g_{12}(\underline{s}^u - \tilde{s}^u)}{2g_1 - g_{11}(\underline{s}^u - \tilde{s}^u)} > 0 \text{ if } g_{12} \leq 0.$$

For instance, $g_{12} \leq 0$ if $\phi^u(\cdot)$ is a constant or a polynomial function of $(\underline{s}^u - \tilde{s}^u)$, with at least one degree.
Since $cov(\omega^h, \underline{s}^u) > 0$, a higher ω^h leads to more top-tier coding/upcoding but a simultaneous increase in $\underline{s}^u_j$ produces an opposite effect. The ultimate effect depends on the force that dominates.

Illustration of Proposition 3

Below we use a CES utility function for illustration with the following simplifications: (i) Only two goods are included; and (ii) the price of compliance does *not* vary by the level of compliance. Next, we go through the same exercise after relaxing (ii) by using a linear function for the probability of getting caught, i.e., $\phi^a(\underline{s}^a - s)$ is linear in $(\underline{s}^a - s)$ and $\phi^u(\underline{s}^u - s)$ is linear in $(\underline{s}^u - s)$.

Consider the following problem:

$$\begin{aligned}
&\max_{\{\mathbb{m}_a,\, \mathbb{m}_u\}} U(\mathbb{m}_a, \mathbb{m}_u) = (\mathbb{m}_a^\rho + \mathbb{m}_u^\rho)^{1/\rho} \\
&\text{s.t. } p_a \mathbb{m}_a + p_u \mathbb{m}_u = p_a + p_u - \eta, \\
&\text{where } \mathbb{m}_a \in (0,1),\ \mathbb{m}_u \in (0,1),\ \text{and } \eta > 0.
\end{aligned}$$

In this model, hospitals decide on $\mathbb{m}_a$ and $\mathbb{m}_u$, with p_a and p_u denoting the price for each of them, respectively. Although the model does not include consumption in the utility function, it is implicitly captured in η. Unlike in the original model where hospitals balance the trade-off between consumption, compliance in admission decision, and compliance in top-coding, we show the decision making on the two compliance levels by holding consumption constant in this model.

Write down the Lagrangian: $\mathcal{L} = U(\mathbb{m}_a, \mathbb{m}_u) - \lambda(p_a \mathbb{m}_a + p_u \mathbb{m}_u - p_a - p_u + \eta)$. The first order conditions are

$$\begin{aligned}
&U_a - \lambda^* p_a = 0 \\
&U_u - \lambda^* p_u = 0 \\
&- p_a \mathbb{m}_a^* - p_u \mathbb{m}_u^* + p_a + p_u - \eta = 0.
\end{aligned}$$

Take the derivative w.r.t p_u on the first-order conditions:

$$\begin{bmatrix} U_{aa} & U_{au} & -p_a \\ U_{ua} & U_{uu} & -p_u \\ -p_a & -p_u & 0 \end{bmatrix} \times \begin{bmatrix} \frac{\partial \mathbb{m}_a^*}{\partial p_u} \\ \frac{\partial \mathbb{m}_u^*}{\partial p_u} \\ \frac{\partial \lambda^*}{\partial p_u} \end{bmatrix} = \begin{bmatrix} 0 \\ \lambda^* \\ \mathbb{m}_u^* - 1 \end{bmatrix}$$

Solving the system of equations above, the sign of $\partial \mathbb{m}_u^* / \partial p_u$ depends on the sign of

$$-\lambda^* p_a^2 - (1 - \mathbb{m}_u^*) \frac{p_u}{U_u} (U_u U_{aa} - U_a U_{au}),$$

and the sign of $\partial \mathbb{m}_a^* / \partial p_u$ depends on

$$\lambda^* p_a p_u + (1 - \mathbb{m}_u^*) \frac{p_u}{U_u} (U_u U_{au} - U_a U_{uu}).$$

If the income effect of the change in p_u dominates the substitution effect, then $\partial \mathbb{m}_u^* / \partial p_u > 0$, and thus, $U_u U_{aa} - U_a U_{au} < 0$, i.e., U_a / U_u decreases in $\mathbb{m}_a$. Since $U_a / U_u = (\mathbb{m}_u / \mathbb{m}_a)^{1-\rho}$, U_a / U_u decreases in $\mathbb{m}_a$ as long as $\rho < 1$. When $\rho < 1$, U_a / U_u increases in $\mathbb{m}_u$, i.e., $U_u U_{au} - U_a U_{uu} > 0 \implies \partial \mathbb{m}_a^* / \partial p_u > 0$.

Now we relax the assumption in (ii), and assume that $\phi^a(\underline{s}^a - s) = \gamma^a(\underline{s}^a - s) + \gamma^{a0}$ and $\phi^u(\underline{s}^u - s) = \gamma^u(\underline{s}^u - s) + \gamma^{u0}$, where $\gamma^a > 0$ and $\gamma^u > 0$. Rewrite the average expected cost of receiving a negative audit determination (i.e., Equation (3)):

$$
\begin{aligned}
f(0, m^a) &= \frac{\phi^{\text{AUDIT}}\xi^a}{m^a}\int_0^{m^a}\left[\gamma^a(\underline{s}^a - s) + \gamma^{a0}\right]ds = \frac{\phi^{\text{AUDIT}}\xi^a}{m^a}\times\left[\gamma^a\underline{s}^a m^a - \frac{\gamma^a(m^a)^2}{2} + \gamma^{a0}m^a\right] \\
&= \phi^{\text{AUDIT}}\xi^a\gamma^{a0} + \phi^{\text{AUDIT}}\xi^a\gamma^a\underline{s}^a - \frac{\phi^{\text{AUDIT}}\xi^a\gamma^a m^a}{2}; \\
f(0, \underline{s}^a) &= \frac{\phi^{\text{AUDIT}}\xi^a}{\underline{s}^a}\int_0^{\underline{s}^a}\left[\gamma^a(\underline{s}^a - s) + \gamma^{a0}\right]ds = \frac{\phi^{\text{AUDIT}}\xi^a}{\underline{s}^a}\times\left[\gamma^a(\underline{s}^a)^2 - \frac{\gamma^a(\underline{s}^a)^2}{2} + \gamma^{a0}\underline{s}^a\right] \\
&= \phi^{\text{AUDIT}}\xi^a\gamma^{a0} + \phi^{\text{AUDIT}}\xi^a\gamma^a\underline{s}^a - \frac{\phi^{\text{AUDIT}}\xi^a\gamma^a\underline{s}^a}{2}; \\
g(\underline{s}^a, m^u + \underline{s}^a) &= \frac{\phi^{\text{AUDIT}}\xi^u}{m^u}\int_{\underline{s}^a}^{m^u+\underline{s}^a}\left[\gamma^u(\underline{s}^u - s) + \gamma^{u0}\right]ds \\
&= \frac{\phi^{\text{AUDIT}}\xi^u}{m^u}\times\left[\gamma^u\underline{s}^u m^u - \frac{\gamma^u((m^u)^2 + 2m^u\underline{s}^a)}{2} + \gamma^{u0}m^u\right] \\
&= \phi^{\text{AUDIT}}\xi^u\gamma^{u0} + \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^u - \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^a - \frac{\phi^{\text{AUDIT}}\xi^u\gamma^u m^u}{2}; \\
g(\underline{s}^a, \underline{s}^u) &= \frac{\phi^{\text{AUDIT}}\xi^u}{\underline{s}^u - \underline{s}^a}\int_{\underline{s}^a}^{\underline{s}^u}\left[\gamma^u(\underline{s}^u - s) + \gamma^{u0}\right]ds \\
&= \frac{\phi^{\text{AUDIT}}\xi^u}{\underline{s}^u - \underline{s}^a}\times\left[\gamma^u\underline{s}^u(\underline{s}^u - \underline{s}^a) - \frac{\gamma^u((\underline{s}^u)^2 - (\underline{s}^a)^2)}{2} + \gamma^{u0}(\underline{s}^u - \underline{s}^a)\right] \\
&= \phi^{\text{AUDIT}}\xi^u\gamma^{u0} + \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^u - \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^a - \frac{\phi^{\text{AUDIT}}\xi^u\gamma^u(\underline{s}^u - \underline{s}^a)}{2};
\end{aligned}
$$

Notice that $f(0, m^a) > f(0, \underline{s}^a)$ and $g(\underline{s}^a, m^u + \underline{s}^a) > g(\underline{s}^a, \underline{s}^u)$. Using the equations above to replace all the $f(\cdot)$ and $g(\cdot)$, the budget constraint in Equation (**??**) becomes

$$
\begin{aligned}
& c + m^a\left(\pi^\ell - \phi^{\text{AUDIT}}\xi^a\gamma^{a0} - \phi^{\text{AUDIT}}\xi^a\gamma^a\underline{s}^a + \frac{\phi^{\text{AUDIT}}\xi^a\gamma^a m^a}{2}\right) \\
& + m^u\left(\pi^h - \pi^\ell - \phi^{\text{AUDIT}}\xi^u\gamma^{u0} - \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^u + \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^a + \frac{\phi^{\text{AUDIT}}\xi^u\gamma^u m^u}{2}\right) \\
& = \underline{s}^a\left(\pi^\ell - \phi^{\text{AUDIT}}\xi^a\gamma^{a0} - \phi^{\text{AUDIT}}\xi^a\gamma^a\underline{s}^a + \frac{\phi^{\text{AUDIT}}\xi^a\gamma^a\underline{s}^a}{2}\right) \\
& + (\underline{s}^u - \underline{s}^a)\left(\pi^h - \pi^\ell - \phi^{\text{AUDIT}}\xi^u\gamma^{u0} - \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^u + \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^a + \frac{\phi^{\text{AUDIT}}\xi^u\gamma^u(\underline{s}^u - \underline{s}^a)}{2}\right) + \Pi^p.
\end{aligned}
$$

The price of compliance level in admission decision (tier coding assignment), i.e., m^a (m^u), is a linear function of the consumption of m^a (m^u). Note that the marginal effect of consuming m^a (m^u) on the price of m^a (m^u) is strictly positive, as $\phi^{\text{AUDIT}}\xi^a\gamma^a > 0$ ($\phi^{\text{AUDIT}}\xi^u\gamma^u > 0$).

Let the average compliance prices have the following forms:

$$\pi^\ell - \phi^{\text{AUDIT}}\xi^a\gamma^{a0} - \phi^{\text{AUDIT}}\xi^a\gamma^a\underline{s}^a + \frac{\phi^{\text{AUDIT}}\xi^a\gamma^a m^a}{2} = \alpha_0 + \alpha_1 m^a;$$
$$\pi^h - \pi^\ell - \phi^{\text{AUDIT}}\xi^u\gamma^{u0} - \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^u + \phi^{\text{AUDIT}}\xi^u\gamma^u\underline{s}^a + \frac{\phi^{\text{AUDIT}}\xi^u\gamma^u m^u}{2} = \beta_0 + \beta_1 m^u.$$

An increase in ω^h leads to higher β_0.

Now let us illustrate how ω^h affects compliance levels, using the CES utility function:

$$\begin{aligned}
&\max_{\{\mathbb{m}_a,\, \mathbb{m}_u\}} U(\mathbb{m}_a, \mathbb{m}_u) = (\mathbb{m}_a^\rho + \mathbb{m}_u^\rho)^{1/\rho} \\
&\text{s.t. } (\alpha_0 + \alpha_1\mathbb{m}_a)\mathbb{m}_a + (\beta_0 + \beta_1\mathbb{m}_u)\mathbb{m}_u = (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1) - \eta, \\
&\text{where } \mathbb{m}_a \in (0,1),\ \mathbb{m}_u \in (0,1),\ \alpha_1 > 0,\ \beta_1 > 0,\ \text{and } \eta > 0.
\end{aligned}$$

The Lagrangian becomes

$$\mathcal{L} = U(\mathbb{m}_a, \mathbb{m}_u) - \lambda\left[(\alpha_0 + \alpha_1\mathbb{m}_a)\mathbb{m}_a + (\beta_0 + \beta_1\mathbb{m}_u)\mathbb{m}_u - (\alpha_0 + \alpha_1) - (\beta_0 + \beta_1) + \eta\right].$$

The first order conditions are

$$\begin{aligned}
&U_a - \lambda^*(\alpha_0 + 2\alpha_1\mathbb{m}_a^*) = 0 \\
&U_u - \lambda^*(\beta_0 + 2\beta_1\mathbb{m}_u^*) = 0 \\
&-(\alpha_0 + \alpha_1\mathbb{m}_a^*)\mathbb{m}_a^* - (\beta_0 + \beta_1\mathbb{m}_u^*)\mathbb{m}_u^* + (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1) - \eta = 0.
\end{aligned}$$

Take the derivative w.r.t β_0 on the first-order conditions:

$$\begin{bmatrix} U_{aa} - 2\lambda^*\alpha_1 & U_{au} & -(\alpha_0 + 2\alpha_1\mathbb{m}_a^*) \\ U_{ua} & U_{uu} - 2\lambda^*\beta_1 & -(\beta_0 + 2\beta_1\mathbb{m}_u^*) \\ -(\alpha_0 + 2\alpha_1\mathbb{m}_a^*) & -(\beta_0 + 2\beta_1\mathbb{m}_a^*) & 0 \end{bmatrix} \times \begin{bmatrix} \frac{\partial\mathbb{m}_a^*}{\partial\beta_0} \\ \frac{\partial\mathbb{m}_u^*}{\partial\beta_0} \\ \frac{\partial\lambda^*}{\partial\beta_0} \end{bmatrix} = \begin{bmatrix} 0 \\ \lambda^* \\ \mathbb{m}_u^* - 1 \end{bmatrix}$$

Let

$$\begin{aligned}
\tilde{U}_{aa} &= U_{aa} - 2\lambda^*\alpha_1; \\
\tilde{U}_{uu} &= U_{uu} - 2\lambda^*\beta_1; \\
\tilde{p}_a &= \alpha_0 + 2\alpha_1\mathbb{m}_a^*; \\
\tilde{p}_u &= \beta_0 + 2\beta_1\mathbb{m}_u^*.
\end{aligned}$$

Notice that the effective price of $\mathbb{m}_a$ ($\mathbb{m}_u$) becomes $\tilde{p}_a$ ($\tilde{p}_u$), which is strictly greater than the original p_a (p_u). Thus, the system of equations has the following form:

$$\begin{bmatrix} \tilde{U}_{aa} & U_{au} & -\tilde{p}_a \\ U_{ua} & \tilde{U}_{uu} & -\tilde{p}_u \\ -\tilde{p}_a & -\tilde{p}_u & 0 \end{bmatrix} \times \begin{bmatrix} \frac{\partial \mathbb{m}_a^*}{\partial \beta_0} \\ \frac{\partial \mathbb{m}_u^*}{\partial \beta_0} \\ \frac{\partial \lambda^*}{\partial \beta_0} \end{bmatrix} = \begin{bmatrix} 0 \\ \lambda^* \\ \mathbb{m}_u^* - 1 \end{bmatrix}$$

The sign of $\partial \mathbb{m}_u^* / \partial \beta_0$ depends on the sign of

$$-\lambda^* \tilde{p}_a^2 - (1 - \mathbb{m}_u^*) \frac{\tilde{p}_u}{U_u} (U_u U_{aa} - U_a U_{au} - 2\lambda^* \alpha_1 U_u),$$

and the sign of $\partial \mathbb{m}_a^* / \partial \beta_0$ depends on

$$\lambda^* \tilde{p}_a \tilde{p}_u + (1 - \mathbb{m}_u^*) \frac{\tilde{p}_u}{U_u} (U_u U_{au} - U_a U_{uu} + 2\lambda^* \beta_1 U_a).$$

As a result, we have the same implications as those in the previous exercise.

Appendix II Increasing audit scrutiny as a result of increasing reimbursements

Our model assumes that the probability of auditing an admission, ϕ^{AUDIT}, is a parameter, which is fixed. However, audit scrutiny could in principle be affected by the DRG weights themselves. In particular, DRG weights can influence audit scrutiny in two ways. First, auditors may have more incentives to scrutinize the top bill codes of DRGs experiencing increases in reimbursement rates, due to the larger dollar value at stake. Auditors are typically paid on a contingent fee basis. The amount of the contingency fee is a percentage of the payments recovered, which is proportional to the DRG weight. Second, auditors are likely to be most interested in claims with obvious upcoding errors, and this will lead them to disproportionately looking into the top-tier DRGs for which the reimbursement rate goes up. Note that here we focus on the first mechanism because our model assumes away the interaction between hospitals and auditors. But the prediction we derive below also applies to the second mechanism.

To further examine this mechanism, we allows ϕ^{AUDIT} to vary by severity subclass within a base DRG. In other words, the probability of auditing a claim documented with the low-severity subclass within a base DRG (denoted as $\phi^{\text{AUDIT\_}\ell}$) is different from the probability of auditing a claim with the high-severity subclass (denoted as $\phi^{\text{AUDIT\_}h}$). We re-write the expected cost from inappropriate admissions/coding if the hospital is found engaging in these practices during an audit (defined in Equation (1)) as follows:

$$E[\zeta^{\ell}(s)] = \mathbb{1}\{s < \underline{s}^a\} \times \phi^{\text{AUDIT\_}\ell} \times \phi^a(\underline{s}^a - s) \times \xi^a;$$
$$E[\zeta^{h}(s)] = \mathbb{1}\{s < \underline{s}^u\} \times \phi^{\text{AUDIT\_}h} \times \phi^u(\underline{s}^u - s) \times \xi^u.$$

Thus, the average expected costs of detection per inappropriately admitted patient and the average expected costs of detection per upcoded patient (defined in Equation (3)) have the following form:

$$f(\tilde{s}^a, \underline{s}^a) = \int_{\tilde{s}^a}^{\underline{s}^a} \left[\phi^{\text{AUDIT\_}\ell} \times \phi^a(\underline{s}^a - s) \times \xi^a \times \frac{1}{\underline{s}^a - \tilde{s}^a} ds\right];$$
$$g(\tilde{s}^u, \underline{s}^u) = \int_{\tilde{s}^u}^{\underline{s}^u} \left[\phi^{\text{AUDIT\_}h} \times \phi^u(\underline{s}^u - s) \times \xi^u \times \frac{1}{\underline{s}^u - \tilde{s}^u} ds\right].$$

Consider an increase in audit scrutiny of the high severity subclass, i.e., $\phi^{\text{AUDIT\_}h} \uparrow$. Take derivatives w.r.t. $\phi^{\text{AUDIT\_}h}$ on both sides of Equation (6):

$$f_1 \frac{\partial \tilde{s}^a}{\partial \phi^{\text{AUDIT\_}h}} - f_{11}(\underline{s}^a - \tilde{s}^a)\frac{\partial \tilde{s}^a}{\partial \phi^{\text{AUDIT\_}h}} + f_1 \frac{\partial \tilde{s}^a}{\partial \phi^{\text{AUDIT\_}h}} = 0;$$

$$g_1 \frac{\partial \tilde{s}^u}{\partial \phi^{\text{AUDIT\_}h}} + \frac{\partial g}{\partial \phi^{\text{AUDIT\_}h}} - g_{11}(\underline{s}^u - \tilde{s}^u)\frac{\partial \tilde{s}^u}{\partial \phi^{\text{AUDIT\_}h}} - \frac{\partial g_1}{\partial \phi^{\text{AUDIT\_}h}}(\underline{s}^u - \tilde{s}^u) + g_1 \frac{\partial \tilde{s}^u}{\partial \phi^{\text{AUDIT\_}h}} = 0.$$

$\Longrightarrow$

$$\begin{aligned} \frac{\partial \tilde{s}^a}{\partial \phi^{\text{AUDIT\_}h}} &= 0; \\ \frac{\partial \tilde{s}^u}{\partial \phi^{\text{AUDIT\_}h}} > 0 \text{ if } \frac{\partial g_1}{\partial \phi^{\text{AUDIT\_}h}} &\leq 0. \end{aligned} \tag{7}$$

Because the decision to admit unnecessary patients is unaffected by the reimbursement rate for the high-level DRG or the audit probability at this level, our model predicts no effect on unnecessary admissions when the revenue of reporting the high-level DRG increases. However, the effect from the change in the reimbursement rate of the top code is uncertain. On the one hand, more patients would be admitted to the high severity subclass in response to an increase in ω^h. But on the other hand, with the expectation of a rise in scrutiny among the top-coded admissions, hospitals would be more cautious in upcoding patients to the high level. The ultimate effect on coding patients to the high-level DRG depends on which force dominates.

Appendix Figure A1 shows the two possible cases of the correlation between $\phi^{\text{AUDIT\_}h}$ and $\tilde{s}^u$. Note that we assume here that the change in the average expected costs of detection per upcoded patient due to the change in upcoding decreases with audit probability, i.e., $\frac{\partial g_1}{\partial \phi^{\text{AUDIT\_}h}} \leq 0$, and thus, $\tilde{s}^u$ increases with $\phi^{\text{AUDIT\_}h}$, as shown in Equation (7). Also, in line with our main model, we assume that $\phi^{\text{AUDIT\_}h}$ is a parameter hospitals take as given, and thus, audit probability is represented as a vertical line in Appendix Figure A1. As a result, our model does not capture the case where auditors tend to increase scrutiny of claims which they believe are most likely to have upcoding errors. While it is a limitation, the prediction based on our current setup can be applied to the more general case, which can also be captured by Appendix Figure A1 by making the vertical lines for $\phi^{\text{AUDIT\_}h}$ downward sloping.

Consider Case (i) in the upper figure. As the weight for the high-level DRG increases from ω_0^h to ω_1^h, the cutoff of improper coding drops from $\tilde{s}_A^u$ to $\tilde{s}_B^u$, if audit probability remains the same. However, when the scrutiny of admissions with the high-tier codes rises as well, from $\phi_0^{\text{AUDIT\_}h}$ to $\phi_1^{\text{AUDIT\_}h}$ because of the change in ω^h, hospitals would increase the cutoff of improper coding from $\tilde{s}_B^u$ to $\tilde{s}_C^u$. In this case, the effect of increased audit scrutiny partially offsets the effect of the increased DRG weights, and the ultimate effect on upcoding is positive. The lower figure shows the opposite case where the effect on audit probability dominates, and thus, we observe less upcoding.

Taken together, if auditors tend to scrutinize the high-level claims within a base DRG for which the reimbursement rate increases and if the deterring effect from the increased scrutiny is sufficiently large, we expect no change in total admissions to this base DRG but fewer patients coded to the higher level. This is consistent with the prediction from the financial-driven mechanism enhanced by coding requirements.

Appendix III Extra Tables and Figures

Table A1: Effect of Spread on Log(Admissions) to the top tier within a base DRG

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Spread	-0.2607**	-0.0681	-1.7278	-0.1990**	-0.7922**	-0.0842
	(0.1264)	(0.0669)	(1.2168)	(0.0794)	(0.3722)	(0.0550)
N	2110	209	190	1711	526	1584
R^2	0.998	0.997	0.993	0.999	0.997	0.979
MeanDepVar	8.155	8.088	8.618	8.112	10.118	7.504

Note: Analysis based on Medicare Part A Population data. Unit of observation is DRG/year. Other regressors include DRG fixed effects, year fixed effects, and DRG-specific time trends. Each observation is a DRG-year combination. Regression weighted by total number of discharges per base DRG. Standard errors in parentheses, clustered at base DRGs.

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

Table A2: Effect of Spread on upcoding-related determinations

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Spread	-0.1038*	-0.1375	-0.3380**	-0.0483	-0.2791***	-0.0048
	(0.0596)	(0.0963)	(0.1338)	(0.0846)	(0.0930)	(0.0619)
N	25276	2779	3263	19234	18539	6737
R^2	0.052	0.151	0.022	0.056	0.030	0.181
MeanDepVar	0.083	0.091	0.082	0.083	0.081	0.091

Note: Analysis based on Medicare Part A Population data. Unit of observation is DRG/year. Other regressors include DRG fixed effects, year fixed effects, and DRG-specific time trends. Regression weighted by total number of discharges per base DRG. Standard errors in parentheses, clustered at base DRGs.

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

Table A3: Effect of DRG weights on incorrect-coding determinations

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Top-tier DRG weight	-0.1090**	-0.1229*	-0.4406***	-0.0587	-0.2039***	-0.0609
	(0.0469)	(0.0676)	(0.1013)	(0.0721)	(0.0655)	(0.0755)
Next lower-tier DRG weight	0.0055	0.0304	0.7951	-0.0541	0.1825*	-0.0343
	(0.0372)	(0.0397)	(0.5477)	(0.1112)	(0.0936)	(0.0423)
N	25276	2779	3263	19234	18539	6737
R^2	0.046	0.072	0.021	0.053	0.024	0.178
MeanDepVar	0.083	0.091	0.082	0.083	0.081	0.091

Note: Analysis based on CERT microdata. Other regressors include DRG fixed effects, year fixed effects, and DRG-specific time trends. Each observation is a claim-year combination. Regression weighted by the relative frequency per DRG per year. Standard errors in parentheses, clustered at base DRGs.

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

Table A4: Effect of DRG weights on Log(Admissions) to the top tier within a base DRG, after dropping DRGs with low upcoding rate

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Top-tier DRG	-0.1486*	-0.1730	0.7485	-0.1264	-0.3197**	-0.0360
weight	(0.0857)	(0.0993)	(1.0288)	(0.1114)	(0.1466)	(0.0889)
Next lower-tier	-0.0898	-0.0990**	4.2842	-0.0105	-0.2453	-0.0569
DRG weight	(0.0877)	(0.0350)	(3.9756)	(0.3723)	(0.2720)	(0.1189)
N	1047	116	78	853	211	836
R^2	0.998	0.998	0.992	0.999	0.999	0.981
MeanDepVar	7.976	8.107	7.972	7.958	10.083	7.443

Note: Analysis based on Medicare Part A Population data. Unit of observation is DRG/year. Other regressors include DRG fixed effects, year fixed effects, and DRG-specific time trends. Each observation is a DRG-year combination. Regression weighted by total number of discharges per base DRG. Standard errors in parentheses, clustered at base DRGs.

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

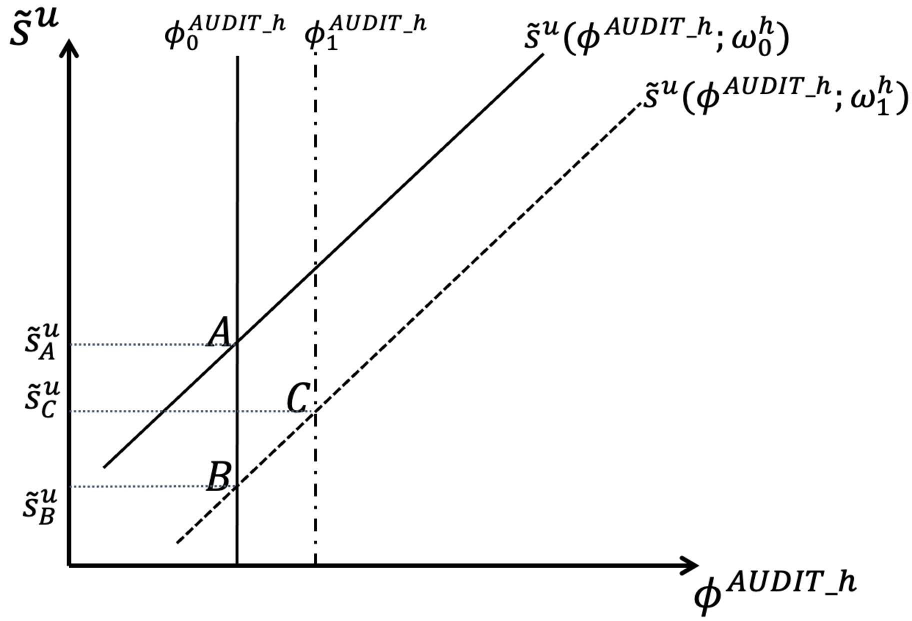
Table A5: Effect of DRG weights on upcoding-related determinations, after dropping DRGs with low upcoding rate

	All	High weights	Low weights	Medium weights	Common DRGs	Less common DRGs
Top-tier DRG weight	-0.1242***	-0.0921**	-0.6057***	-0.0855	-0.2702***	0.0096
	(0.0465)	(0.0415)	(0.1017)	(0.0693)	(0.0720)	(0.0259)
Next lower-tier DRG weight	0.0416	0.0288	0.9148	0.0651	0.2108*	0.0027
	(0.0397)	(0.0218)	(0.6481)	(0.1235)	(0.1173)	(0.0185)
N	12740	1577	1917	9246	9880	2860
R^2	0.021	0.054	0.015	0.023	0.017	0.077
MeanDepVar	0.061	0.051	0.069	0.062	0.066	0.045

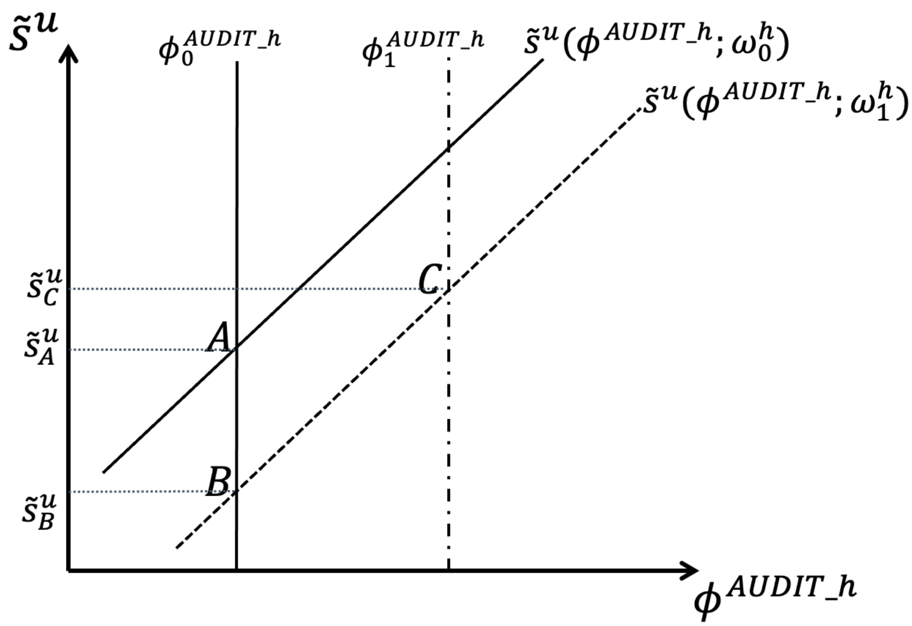
Note: Analysis based on CERT microdata. Other regressors include DRG fixed effects, year fixed effects, and DRG-specific time trends. Each observation is a claim-year combination. Regression weighted by the relative frequency per DRG per year. Standard errors in parentheses, clustered at base DRGs.

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

Figure A1: Correlation between the cutoff of upcoding ($\tilde{s}^u$) and the probability of auditing admissions with the top code ($\phi^{\text{AUDIT\_}h}$)



Case (i)



Case (ii)