HOW DO HOSPITALS RESPOND TO PAYMENT INCENTIVES?

Gautam Gowrisankaran
Keith A. Joiner
Jianjing Lin

WORKING PAPER 22873

NBER WORKING PAPER SERIES

HOW DO HOSPITALS RESPOND TO PAYMENT INCENTIVES?

Gautam Gowrisankaran
Keith A. Joiner
Jianjing Lin

Working Paper 22873
http://www.nber.org/papers/w22873

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2016, Revised June 2022

We have received helpful comments from David Chan, Sebastian Fleitas, Mike Geruso, Josh Gottlieb, Ben Handel, Ashley Langer, Kelli Marquardt, Jeff McCullough, Jessamyn Schaller, Bob Town, and seminar participants at several institutions. This paper includes material from and supersedes "Does Hospital Electronic Medical Record Adoption Lead to Better Information or Worse Incentives?" and "Does Health IT Adoption Lead to Better Information or Worse Incentives?" The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How do Hospitals Respond to Payment Incentives?
Gautam Gowrisankaran, Keith A. Joiner, and Jianjing Lin
NBER Working Paper No. 22873
December 2016, Revised June 2022
JEL No. H51,I11,I13,O33

ABSTRACT

A recent literature finds that hospitals "upcode" when doing so increases revenues, suggesting that incomplete information creates substantial distortions. However, reporting complete information is itself costly. We examine the impact of both revenues and coding costs on hospital billing practices for Medicare inpatients. Following the literature, we investigate the fraction of patients top coded as the revenues from top coding vary. We then examine how this fraction changes following Medicare reforms—which increased the requirements and complexity to justify top codes—interacted with hospital electronic medical record adoption—which may decrease coding costs. We find evidence that coding costs drive top coding behavior.

Gautam Gowrisankaran
Department of Economics
Columbia University
420 West 118th St.
Rm. 1026 IAB, MC 3308
New York, NY 10027
and HEC Montreal and CEPR
and also NBER
gautamg2@gmail.com

Keith A. Joiner
College of Medicine and
Department of Economics
University of Arizona
kjoiner@email.arizona.edu

Jianjing Lin
Rensselaer Polytechnic Institute
110 8th St
Troy, NY 12180
linj17@rpi.edu

1 Introduction

Over the past several decades, the complexity of reimbursement mechanisms for health care has increased dramatically in the U.S. and many other countries. In 1983, Medicare introduced the inpatient Prospective Payment System, which reimbursed hospitals based on the primary diagnosis or procedure performed, instead of on a fee-for-service basis. The goal of this system was to move towards reimbursement for optimal care. More recently, Medicare and other payors have implemented a series of payment reforms including bundled payments,[1] hospital readmission reduction programs, and hospital value-based purchasing programs, all with the intention of reimbursing quality over quantity.[2]

These payment innovations have enormously increased the costs to providers of documenting and coding patient visits accurately (Dunn et al., 2021; Shi, 2021). Membership in the American Academy of Professional Coders has more than doubled in recent years,[3] and the U.S. Bureau of Labor Statistics lists medical billers and coders as one of the fastest growing occupations between 2016 and 2026.[4] The increases in the complexity of medical coding suggests that coding costs may be an important factor that is driving coding behavior by healthcare providers. In particular, in order to maximize revenues net of costs, hospitals need to strategically consider how to deploy resources to help them document patient charges

[1]Bundled payment systems may be complex due to a variety of factors including the contract design and negotiations between providers and payors, the separation of performance and insurance risk, and the distribution of payments across providers (Steenhuis et al., 2020).

[2]While the initial focus of complex, prospective payments had been for inpatient stays, Medicare and other payors have shifted towards such reimbursement incentives for outpatients, including under the 2015 Medicare Access and CHIP Reauthorization Act (MACRA).

[3]By 2017, there were more than 170,000 members (https://www.nytimes.com/2017/03/29/magazine/those-indecipherable-medical-bills-theyre-one-reason-health-care-costs-so-much.html).

[4]See https://www.medicalbillingandcoding.org/qnas/the-coder-is-in-demand-medical-billers-coders/.

completely and maximally capture revenues.

A recent literature shows that medical providers document diagnoses more thoroughly when revenues from coding are higher (Brown et al., 2014; Geruso and Layton, 2018), a practice that is often called *upcoding*. Other papers show that upcoding may specifically be due to bill inflation (Silverman and Skinner, 2004; Dafny, 2005; Jürges and Köberlein, 2015). The literature argues that upcoding may create substantial distortions to firm and consumer behavior. It also implicitly recognizes that regulators have limited information about patient health risk, suggesting that generating and reporting this information accurately is costly. Despite this, the literature largely does not consider the fact that medical providers may bear costs from coding. This may lead it to present an incomplete understanding of the distortions generated by current and alternative reimbursement practices. In particular, in the presence of costly coding, hospitals need to consider both the revenues that they would gain from coding more completely and the costs of obtaining those extra revenues. Costly coding may, in some cases, provide an alternative explanation to upcoding for hospitals' practices.

The purpose of this paper is to provide evidence on the extent to which coding costs and upcoding may be important drivers of hospital incentives, in the context of Medicare inpatient admissions. It is important to separate these two explanations both because very different policy implications underlie them and also because of the large dollar values at stake: Medicare admissions account for $146 billion in revenues in 2018 and, more generally, many private payors follow Medicare in their billing practices (Clemens and Gottlieb, 2017). If upcoding is important, hospitals as well as the Center for Medicare and Medicaid Services (CMS) may want to increase billing enforcement and/or penalties for inaccurate reporting. If costly coding is important, then the continued increase in reimbursement complexity that stems from insurers' goal of rewarding quality over quantity is likely to generate more distortions and costs over time.[5]

The idea of our analysis is to use changes to Medicare payments, interacted with hospital electronic medical record (EMR) adoption to separate the costly coding and upcoding explanations. During our 2005-9 sample, Medicare revenues from more complete coding changed across patient illnesses while Medicare billing increased in complexity in two discrete ways. EMRs can facilitate the task of capturing the hospital course of illness and treatment that in turn can be translated into bills or claims, through a process called *coding*. The value of EMRs in facilitating billing increases with the complexity of the reimbursement system. Thus, EMR adoption together with the increases in the complexity of Medicare billing provide variation that allows us to separately identify the importance of the costly coding and upcoding explanations.

We develop a simple model of *upcoding* and *costly coding* to motivate our empirical analysis. Upcoding focuses on the extra revenues that hospitals would gain from coding at the high level. With this explanation, providers need to decide whether to code a patient at a low or high level, with the high reported bill code requiring more specificity in coding but also generating more revenues. Changes in the extra revenue over time provide identifying variation for the upcoding explanation. Costly coding focuses on the costs of completely documenting

[5] The concern about costly coding is relevant well outside the U.S. context that we consider. Many countries have increased the complexity of their hospital reimbursement systems by moving to payment systems based on diagnosis related group (DRG) (Mathauer and Wittenbecher, 2013; Hopfe et al., 2018). This has led to concerns that changes in billing incentives were leading to changes in coding practices in Switzerland (Fässler et al., 2015) and France (Or, 2014).

the conditions necessary to generate the high level, recognizing that this documentation may require substantial effort. Unlike with revenues, we do not directly observe the extra costs of coding completely, but EMR adoption provides identifying variation because EMRs are a tool that may help reduce the costs of complete coding under complex payment mechanisms.

We estimate four main specifications that together help us distinguish between upcoding and costly coding. Our specifications consider the percent of discharges that are reported to be a top code, focusing on the variation over time in top codes at a hospital within a narrow set of diagnoses—called a base DRG.

First, we examine upcoding using variation in the extra revenues between the top and bottom codes across time within the narrow set of diagnoses, similarly to Silverman and Skinner (2004) and Dafny (2005). Because a Medicare payment reform in Q4:2007 increased billing complexity by tightening the requirements for top codes, we estimate this specification separately before and after the reform. We find evidence that prior to Q4:2007, increases in the extra revenue from the top code led to increases in hospitals reporting the top code. Following Q4:2007, increases in revenues had a negative—and marginally statistically significant—effect on reported top codes. This suggests that upcoding was an important driver of coding prior to the payment reform but not afterwards. The difference in results may be due to the increased complexity of billing post-reform leading coding costs to be more important.

Second, we use the variation in costs generated by the Q4:2007 Medicare payment reform. Specifically, before the reform, many chronic diseases qualified for a complicating/comorbid condition (CC)—the more severe and thus higher-paid DRG within a base DRG. The reform increased the stringency and complexity of coding for CCs, requiring the acute manifestation of the chronic disease or a new acute disease.[6] Since EMRs facilitate the specificity of coding medical information and since the reform resulted in this specific information being necessary to obtain a top code, we hypothesize that hospitals with EMRs had lower costs of complete coding in the post-reform period relative to the pre-reform period. We estimate whether hospitals with EMRs had an increase in reported top codes after the reform relative to before the reform.[7] We find that EMR adopters had 2.0 percentage points more top codes post-reform than pre-reform, relative to non-EMR adopters. This is consistent with costly coding and EMRs lowering the cost of coding.

Third, we consider the impact of EMRs in documenting preventable conditions, which are often called *never events*. In Q4:2008, CMS started penalizing certain preventable complications, when acquired during the course of hospital treatment. An example of a never event is a catheter-associated urinary tract infection acquired in-hospital. To understand whether a condition was acquired during the course of a hospital treatment, Medicare mandated the disclosure and underlying documentation of a "present on admission" field. Never events work in the opposite direction from most payment mechanisms in that, with never events, the bill code with more specificity actually leads to lower reimbursements in the short run.[8] We find that EMR adopters responded to the penalization of never events by reporting more never events. This occurred through better coding of the "present on admission" indicators

[6] There are exceptions to this rule including, most prominently, for heart failure (Sacarny, 2018). The reform also further splits CCs into major complication conditions (MCCs) and other CCs.

[7] To avoid having heterogeneity in the comparison groups over time, we use hospitals which had adopted EMRs prior to the start of our estimating sample for these regressions (Borusyak and Jaravel, 2017; De Chaisemartin and d'Haultfoeuille, 2020; Sun and Abraham, 2021).

[8] Hospitals were incentivized to code never events accurately over the medium run, as not doing so could lead to financial and reputation losses due to oversight by Medicare including audits.

for secondary diagnoses by EMR hospitals, as these hospitals reported fewer inconclusive values for these indicators. Our result here is consistent with coding never events being costly and with EMRs increasing the completeness of coding, despite short-run revenue losses.

Finally, we separately consider the impact of the 2007 payment reform on medical and surgical DRGs. It is well appreciated that surgeons interact with EMR systems less than medical physicians (see, e.g., Gawande, 2018). Moreover, the extra reimbursement from the top codes all relates to diagnoses, rather than to procedures performed. Specifically, the factors justifying a CC or MCC are exclusively medical conditions (such as heart failure, pneumonia, or kidney failure). For these reasons, EMRs will have a different impact for surgeons than for medical physicians, with them lowering the cost of coding more for medical DRGs. We therefore evaluate whether bill codes are more impacted by this reform for medical than for surgical DRGs. We find that top codes for EMR adopters increase only for medical DRGs, with EMR adopters reporting 2.67 percentage points more top codes for medical DRGs in the post-reform period than in the pre-reform period, relative to non-EMR adopters. In contrast, the impact on surgical DRGs is negative, with EMR adopters reporting 1.87 percentage points fewer top codes. This result is again consistent with EMRs *increasing* the coding costs for surgical DRGs, likely due to surgeons' relative lack of interaction with the EMR systems and fixed costs of using EMRs to document completely.

Overall, our findings suggest that costly coding is more important than upcoding in explaining hospitals' decisions to report top codes for Medicare hospital discharges, in the period following 2007. The magnitude of our results suggests that the 2007 reform would lead to $1.03 billion annually in extra Medicare hospital claims costs if all hospitals acted like early EMR adopters. This figure did not enter into the payment reform calculations, which were intended to be budget neutral. Thus, the extent of variation in billing created by costly coding is large as an absolute number, though moderate as a function of overall Medicare hospital billing.

The remainder of the paper is structured as follows. Section 2 provides background on the market. Section 3 discusses our analytic framework and testable hypotheses. Section 4 discusses our data. Section 5 provides our results and implications. Section 6 concludes.

2 Background

2.1 Medicare billing for hospitalized patients

In 1983, the Health Care Financing Administration (now CMS, the Centers for Medicare and Medicaid Services), developed a flat-rate payment system for Medicare Part A inpatient admissions, known as the inpatient Prospective Payment System (PPS). Under PPS, a hospital assigns a single diagnostic related group (DRG) for each patient stay using the primary diagnosis, additional diagnoses, primary procedure, additional procedures, and discharge status. Each DRG has a weight, which is set by CMS to reflect the average resources used to treat Medicare patients in that DRG. Medicare then reimburses the hospital a flat rate for the admission, calculated as the hospital's base rate multiplied by the DRG weight. Each hospital's base rate varies based on the costs in the area. For instance, in 2008, a hospital may receive anywhere from $2,991.79 to $7541.39 for treating a patient with DRG weight

1, depending on its area cost factor.[9] PPS was the first of many Medicare payment reforms that rewarded efficiency but also increased coding complexity.

DRGs can be either medical or surgical. Essentially, surgical DRGs are for patients who underwent surgery and medical DRGs are for patients who did not undergo surgery.[10] The coding of an inpatient admission into a DRG uses the following logic: diagnoses are identified by ICD-9 diagnosis codes. Surgical procedures are identified by ICD-9 procedure codes.[11] Using the ICD-9 codes, an admission is first coded into a base DRG using the primary diagnosis code (for medical DRGs) or the primary procedure code (for surgical DRGs). An example of a medical base DRG is "Heart Failure and Shock" while "Spinal Fusion Except Cervical" is an example of a surgical base DRG. Subsequently, the admission is coded to an exact DRG, based exclusively on the presence or absence of complicating/comorbid conditions (CCs) and major CCs (MCCs). Each base DRG has one to three associated DRGs (also called *severity subclasses*), which differ only in the presence of CCs and MCCs. In some cases, a CC will be lumped with an MCC into one severity subclass. CCs and MCCs all indicate the presence of secondary *diagnoses*, all of which are medical conditions, and not secondary procedures. This is an important distinction, since medical physicians will have a lower cost than surgeons of diagnosing and documenting the CCs and MCCs, given their role in the healthcare production process.

The severity subclass system with separate CCs and MCCs described above was implemented by CMS starting in Q4:2007, and is known as Medicare Severity DRGs (MS-DRGs).

[9] Authors' calculations based on FY2008 data.

[10] While a patient who underwent a surgical procedure can qualify for a medical DRG based on her illness, the surgical DRG will almost always have a higher payment.

[11] Starting in October, 2015, both diagnoses and procedures are identified with ICD-10 codes.

CMS reformed the reimbursement system in 2007 in order to better align payments with the resources used by a hospital.[12] A realignment was deemed necessary because many conditions that previously needed costly and lengthy hospitalizations could then be managed in an out-patient setting using drug or other therapies. Prior to the reform, the presence of a chronic disease was sufficient to justify a CC. Following the reform, a new acute manifestation of a chronic disease or a new acute disease—both of which reflect a more severe illness—generally became necessary to justify a CC or MCC.[13] Overall, the intent of the 2007 reform was to lower the fraction of admissions that would qualify for a CC or MCC. Using the universe of 2006 patients, 77.7% of admissions had at least one CC under the pre-reform criteria, while only 40.3% had a CC or MCC under the post-reform criteria.[14]

The process of coding CCs and MCCs for Medicare Part A relates to the coding process for Medicare Part C, also called Medicare Advantage, studied by Geruso and Layton (2018) among others. In Medicare Part C, insurance companies assume the financial risk for providing health care for patients who enroll with them for Medicare coverage. The reimbursement amount for a patient covered by Medicare Part C is based on the patient's Hierarchical Condition Category (HCC) score, deriving from the patient's chronic conditions. HCCs are intended to capture the costs of all patient care over a year-long period while CCs/MCCs seek to capture the cost of a single episode of inpatient care. Accordingly, HCCs focus on chronic conditions (Yeatts and Sangvai, 2016) while CCs/MCCs focus on acute conditions.

[12] Gross et al. (2021) find that hospitals for which Medicare revenues increased from this reform increased their Medicare patient volume following the reform.

[13] See Office of the Federal Register and National Archives and Records Service (2007) p. 47,153 and Sacarny (2018).

[14] See Office of the Federal Register and National Archives and Records Service (2007), p. 47,153-4.

In general, it will be easier to document individual chronic conditions that enter into the HCC score than to document conditions that result in a CC/MCC. However, the HCC score depends on multiple conditions and these conditions need to be documented annually.

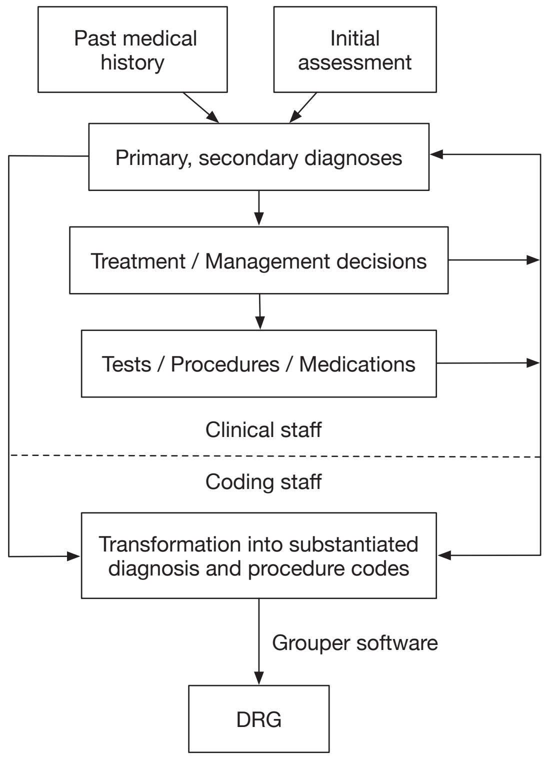
2.2 Coding and EMRs

We now discuss the process of coding patient stays into bills and how this is affected by EMRs, with a diagrammatic representation in Figure 1. Based on an initial assessment and past medical history, the clinical staff at a hospital will note the primary and secondary diagnoses upon admission. They will also make treatment and management decisions and order tests, procedures, and medications. Results from these tests and interventions will in turn lead them to potentially update the primary and secondary diagnoses and treatment and management decisions.

All of the clinical information from these processes will be recorded on the patient chart. The patient chart starts with the admission note, which describes the status of the patient and the diagnoses that are known upon admission. The chart also includes patient progress notes, which are made on a daily basis. These list the patient's course, test results, changes in medication, and other relevant information. For surgical admissions, the patient chart also includes an operative note. This note documents in significant detail the procedure(s) that were performed.

Finally, the chart includes the discharge summary, which provides a brief synopsis of the patient's stay and disposition. Usually dictated after discharge by the attending physician or by a resident who was involved in the care of the patient, the discharge summary is mostly

Figure 1: An Overview of the Coding Process



based on the patient progress notes. The discharge summary lists the primary and secondary diagnoses, summarizes the status of the patent on admission, the hospital course, and the disposition, including medications, pertinent laboratory data, and plans for follow-up care.

The coding of an inpatient admission is then done by the hospital's coding staff (or outsourced to coders at a health analytics firm). The primary role of the coding staff is to create *substantiated* diagnosis and procedure codes that can be used in billing, and to remove unsubstantiated codes from consideration. The coding staff will rely principally on the discharge summary and operative note. CMS (and other payors) require substantiation from the patient chart for each billed secondary diagnosis, that typically includes a combination of results from the patient history, physical examination, laboratory tests, medical imaging, specialty consultations, hospital course, and more. Finally, since roughly the 1990s, coders feed the substantiated codes and other information into *grouper software*. For each admission, this software outputs the highest-weight DRG that can be billed to Medicare.

As an example of the necessary substantiation, to document a secondary condition of diabetes, the patient record would generally need to show results in a clinically defined range for fasting blood glucose, a glucose tolerance test (GTT), or a marker of elevated glucose (HbA1c) (American Diabetes Association, 2015). The test would need to be done during the current admission unless the results were in the medical record of the patient from a prior clinical visit. If none of these diagnostic tests nor prior documentation of diabetes were present in the medical record, coders would be instructed to disallow the diagnosis of diabetes. One role of EMRs is to provide order sets that are specific to particular patients based on their known conditions. These order sets both help ensure medical compliance and optimize reimbursements by uncovering and documenting related conditions.

Previous research has noted that coding staff may try to code more aggressively when faced with financial incentives. For instance, Dafny (2005) interviewed a medical resident who was asked by coding personnel "to reconsider her diagnosis of 'urinary tract infection' and replace it with 'septicemia' ... as the hospital is 'underpaid' and 'needs' the funds to provide care for the uninsured." Medicare has required much more stringent documentation in the interim. In the current environment, the coder would have to document septicemia with specific criteria or face penalties and/or reversal upon audit. Under the False Claims Act, both civil and criminal penalties can be assessed for billing fraud and penalties can be, and often are, severe.[15] Providers and coders are now acutely aware of the these consequences, which serve as strong disincentives for bill inflation. Moreover, as Figure 1 illustrates with a dotted line, the coding staff is separate from the clinical staff and generally will not communicate with the former except for clarification requests.

Having discussed the coding process in general, we now consider how EMRs affect coding. According to the Healthcare Information and Management Systems Society (HIMSS), the following components are key for the meaningful use of EMRs: Clinical Data Repository (CDR), Clinical Decision Support Capabilities (CDS), and Computerized Physician/Provider Order Entry (CPOE) (McCullough et al., 2010). CDR is a centralized database that collects, stores, accesses, and reports health information, including demographics, lab results, radiology images, admissions, transfers, and diagnoses. Its goal is to provide a full picture of the care that is received by a patient.

CDS uses individual data including biometric information to guide and simplify patient management. It assists clinicians with diagnostic support and setting treatment plans. For instance, it provides prompts for specific interventions and assessments and for the documentation that is necessary to justify particular diagnoses. CPOE is a more advanced type of electronic prescribing. It is generally connected with CDS to offer more sophisticated drug safety features such as checks for drug allergies, cross-drug interactions, or dosage adjustments.[16] Both CDS and CPOE require physician training and involvement to provide

[15]In 2018, the Department of Justice collected $2.5 billion in fines in the healthcare sector. See https://www.justice.gov/opa/pr/justice-department-recovers-over-28-billion-false-claims-act-cases-fiscal-year-2018.

[16]See https://psnet.ahrq.gov/primers/primer/6/computerized-provider-order-entry.

real-time support.

Typically, physicians interact with CDS and CPOE—rather than CDR, for instance—when making treatment decisions or ordering tests. Nonetheless, CDS and CPOE rely on CDR for the underlying databases of patient information. For this reason, we rarely see CDS or CPOE adoption without CDR adoption.

The EMR system will generally record the hospital course, providing templates to aid the physician in documentation. At the time of admission, assuming the patient being admitted has previously been seen in the system, a list of pre-existing diagnoses populates a window in the EMR. The admitting physician, or person entering information on her behalf, can choose any or all of those diagnoses, along with any new diagnoses prompting the admission. The latter are chosen from a pop-up list organized by organ system or functional abnormality, which appears after text is entered by the physician. The EMRs can also be used to "clone" information, including diagnoses and patient status, across different notes for a given patient, so that the physician does not need to reenter the information.[17] Returning to the diabetes example, test results from previous encounters with the medical system may be more likely to be accessible by the provider with EMRs.

In the absence of EMRs, the patient chart will be on paper. In this case, the attending physician or resident preparing progress notes must refer back to previous notes to ensure that diagnoses are carried through the record so that they end up in the discharge summary. The EMR largely obviates the need to refer back to progress notes to ensure completeness of the list of diagnoses in the discharge summary.

[17]See http://www.hcca-info.org/Portals/0/PDFs/Resources/Rpt_Medicare/2016/rmc022216.pdf.

With or without EMRs, it is costly to obtain and accurately document information on secondary diagnoses. Sometimes, the admitting physician can learn about comorbidities (such as conditions that justify a CC or MCC) from previous medical encounters, but this information is not always available, and particularly undependable, laborious, and prone to omissions with paper charts. Consultations from specialty services, often related to comorbidities, are included in the body of the patient chart, but may or may not be entered into the patient's list of diagnoses. Even if the physician knows of a secondary diagnosis, the substantiation of this diagnosis in a way that conforms to CMS guidelines can require substantial effort.

2.3 Never events

Medicare began penalizing hospitals for preventable adverse events in 2008. This followed a series of reports, including the landmark Institute of Medicine (IOM) study "To Err is Human: Building a Safer Health System" (Donaldson et al., 2000), that documented the large number of medical errors during hospitalizations. Initial estimates in the IOM study suggested that up to 98,000 deaths from medical errors occurred annually.[18] A separate analysis suggested that one quarter of Medicare beneficiaries were harmed during their hospitalizations (Levinson, 2010). In 2002, the National Quality Forum generated a list of 28 serious events that were unambiguously defined and usually preventable when they occurred during a hospitalization; these were coined *never events*. These never events generated additional expenses during hospitalization and were estimated to cost at least $4.5 billion annually.

[18]A more recent study puts the number at 250,000, accounting for 10% of all deaths and the third leading cause of death in the U.S. (Makary and Daniel, 2016).

In 2007, CMS initially selected eight of these events for non-payment during hospitalization, arguably the first time that Medicare withheld payments in an effort to improve the quality of care. Subsequent and continued expansion of this initiative now includes the hospital value-based purchasing program (VBP), the hospital acquired conditions (HAC) program, and the hospital readmission reduction program (HRRP).[19] These programs are all intended to improve the quality of care delivered in hospitals. They contain a mixture of carrots and sticks, with hospitals at risk of losing up to 6% of total Medicare reimbursements in any given year for non-compliance (Kahn et al., 2015).

Starting in Q4:2007, Medicare mandated that claims data designate whether diagnoses on a defined HAC list were present on admission (POA) or not. Non-compliant claims were supposed to be returned to the provider for clarification.[20] The main decision-maker for determining whether a secondary diagnosis can be coded as POA is the coder. Coders designate a condition as POA if it is either a) explicitly documented as such by the provider, b) a chronic condition documented prior to admission, or c) determined at some point following admission to have been POA, based on additional information collected during the hospitalization. Improper coding of a never event as POA when in fact it was hospital acquired can lead to *higher* reimbursement that is not justified (Saint et al., 2009).[21]

In situations where the above information is missing and uncertainty exists, coders can query the provider, who are then supposed to provide appropriate documentation of a POA

[19]Ibrahim et al. (2018) and Ody et al. (2019) find that incomplete coding of secondary diagnoses has led to a substantial overstating of the benefits of the HRRP.

[20]See https://www.optum360coding.com/CodingCentralArticles/?id=699.

[21]While Saint et al. (2009) exclusively analyze one diagnosis on the HAC list, catheter-associated urinary tract infections, the points that they discuss hold for the HAC list more broadly.

condition in the record. This process imposes additional costs and is "burdensome for both the coder and the clinician" (Saint et al., 2009, p. 5), which may ultimately lead to coders providing inaccurate or incomplete information. As is the case for coding more generally, CMS instructs coders to be vigilant in verifying that clear and complete documentation is included in the record and to not submit claims without such documentation.[22]

3 Analytic and empirical framework

3.1 Model

We develop a model to characterize a hospital's decision regarding the assignment of the top code for each of its Medicare patients. For ease of notation, our model conditions on base DRGs with exactly two bill codes, though some of our empirical work extends to base DRGs with three bill codes. Top codes correspond to more specific diagnoses, which are the high severity subclass DRG or never events, depending on the context. The former represents a more lucrative code within a base DRG, and the latter results in a penalty, but both require more specificity in documentation and thus more costly efforts in coding.

We model two mechanisms: upcoding and costly coding. Upcoding focuses on the role of revenues in the assignment of the patient to the top code while costly coding focuses on the role of the costs of documentation. To distinguish these two mechanisms, we rely on the variations in the costs relative to the revenues from coding the top code.

[22] See https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/fraud_and_abuse.pdf.

3.1.1 Upcoding mechanism

Under upcoding, a hospital chooses the assignment of patients to the top code based on the characteristics of the patient and the extra revenues that it would gain from this assignment. Denote patient by i, hospital by j, base DRG by d, time (quarter) by t, and the number of patients in each cell—hospital/base DRG/time—by N_{jdt}. Specifically, let u_{ijdt}^{up} denote the net utility from coding patient i to the top code; u_{ijdt}^{up} will depend on the patient characteristics. Most importantly, hospitals will have a higher utility from reporting the top code for patients who appear to have conditions that are consistent with this code. This will help the hospital avoid penalties and stigma from unjustified reporting of top codes. However, u_{ijdt}^{up} will also be increasing in the revenue gained from obtaining the top code.

The hospital will report a top code if the net utility of this action is positive. Thus, we write

$$D_{ijdt}^{up} = \mathbb{1}\left\{u_{ijdt}^{up} \geq 0\right\},$$

where D_{ijdt}^{up} is an indicator for reporting the top code. We let the expectation of D_{ijdt}^{up} take the following form:

$$E\left[D_{ijdt}^{up}\right] = f^{up}\left(Rev_{jdt}\right) + g^{up}\left(Pat_{ijdt}\right). \tag{1}$$

In (1), $f^{up}(\cdot)$ is a function of Rev_{jdt}, which denotes the extra revenue from a top code for patients within the same base DRG, hospital, and time period; $g^{up}(\cdot)$ is a function of Pat_{ijdt}, which denotes the characteristics of patient i who belongs to base DRG d at hospital j at period t. Rev_{jdt} mainly comes from the DRG reimbursement and thus does not vary across patients. We aggregate equation (1) across patients within a hospital, base DRG, and time period:

$$Topcode_{jdt} = \sum_{i=1}^{N_{jdt}} \left[f^{up}\left(Rev_{jdt}\right) + g^{up}\left(Pat_{ijdt}\right)\right], \tag{2}$$

where $Topcode_{jdt}$ is the total expected number of patients who are coded with the top code.

3.1.2 Costly coding mechanism

Costly coding pertains to both high severity subclass DRGs and never events. The reporting of a high severity subclass DRG and a never event are similar as they both require more involved coding—of secondary diagnoses in both cases and of the presence on admission field in the latter case. Unlike reporting high severity subclass DRGs, hospitals are subject to penalties by reporting never events. Under costly coding, a hospital will choose whether to report a CC/MCC or a never event based in part on the hassle cost of reporting the specific information. Let u_{ijdt}^{comp} denote the net utility from assigning patient i to the top code after complete documentation of the patient's diagnoses and conditions. u_{ijdt}^{comp} will depend on the patient characteristics and will be decreasing in the cost of complete coding, which includes the time and effort that hospitals spend to document completely.[23] A hospital will completely document a patient's diagnoses and conditions for the top code if the net utility is positive. We write

$$D_{ijdt}^{comp} = \mathbb{1}\left\{u_{ijdt}^{comp} \geq 0\right\},$$

[23] Dunn et al. (2021) find both substantial monetary and time costs of coding, across a number of payors.

where D^{comp}_{ijdt} is an indicator for complete coding. As with upcoding, we let the expectation of D^{comp}_{ijdt} take the following form:

$$E\left[D^{comp}_{ijdt}\right] = f^{comp}\left(Cost_{ijdt}\right) + g^{comp}\left(Pat_{ijdt}\right), \tag{3}$$

where $Cost_{ijdt}$ denotes the average cost of complete coding for patient i belonging to base DRG d at hospital j at period t. We aggregate equation (3) across patients within a hospital, base DRG, and time period to obtain:

$$Topcode_{jdt} = \sum_{i=1}^{N_{jdt}} \left[f^{comp}\left(Cost_{ijdt}\right) + g^{comp}\left(Pat_{ijdt}\right)\right]. \tag{4}$$

Both upcoding and complete coding focus on the set of patients coded with the top code, but they emphasize different reasons for this reporting.

3.2 Empirical specifications

3.2.1 Specification 1: test of upcoding using variation in revenue

We first test for upcoding by examining variation in revenues from assigning patients to the top code. In general, Rev_{jdt} is proportional to the *spread*—the difference in DRG weights between the top and bottom codes—plus some constant. Let

$$f^{up}\left(Rev_{jdt}\right) + g^{up}\left(Pat_{ijdt}\right) = \gamma Spread_{dt} + g^{up}\left(Pat_{ijdt}\right) + \bar{c}^{up}_{jdt}, \tag{5}$$

where γ measures the effect from spread on extra billing; $\overline{c}_{jdt}^{up}$ includes the fixed revenue and the residual according to the upcoding mechanism. Thus, the fraction of top coded patients within a hospital, base DRG, and quarter cell has the following form:

$$\frac{Topcode_{jdt}}{N_{jdt}} = \gamma Spread_{dt} + \frac{\sum_{i=1}^{N_{jdt}} g^{up}\left(Pat_{ijdt}\right)}{N_{jdt}} + \overline{c}_{jdt}^{up}, \tag{6}$$

We condition on the period before and after the 2007 DRG reform, so that the average patient characteristics are relatively homogeneous across time. Thus, we parameterize the last two terms in (6), which are

$$\frac{\sum_{i=1}^{N_{jdt}} g^{up}\left(Pat_{ijdt}\right)}{N_{jdt}} + \overline{c}_{jdt}^{up},$$

as $\overline{c}_{jd}^{H} + \overline{c}_{t}^{Q} + \delta X_{jt} + \varepsilon_{jdt}$. This results in:

$$\frac{Topcode_{jdt}}{N_{jdt}} = \gamma Spread_{dt} + \overline{c}_{jd}^{H} + \overline{c}_{t}^{Q} + \delta X_{jt} + \varepsilon_{jdt}, \tag{7}$$

where $\overline{c}_{jd}^{H}$ denotes the hospital / base DRG fixed effects; $\overline{c}_{t}^{Q}$ denotes the time (quarter) fixed effects; X_{jt} denotes a set of other hospital controls; δ measures the effect of these controls; and ε_{jdt} is the unobservable term.

Equation (7) forms our first estimating equation. We perform all regressions with OLS. Our unit of observation is a unique hospital, base DRG, and quarter cell. Our main dependent variable is the fraction of patients within this unit who are assigned to the top codes. Variation in revenues in our context will be given by the changes in the spread. With upcoding, hospitals should move more patients to the top code if there is an increase in the spread for a base DRG, all else equal, i.e., $\gamma > 0$. We estimate this regression separately before and after the reform, during each of which the average patient characteristics can be captured by the hospital/base DRG and time fixed effects.

Our fixed effects capture the fact that different hospitals may have different fractions of patients who qualify for top codes and for variations across hospitals in the penalties from inappropriate coding. For instance, Medicare's Recovery Audit Program, which aims to identify and recover improper payments for the Medicare Fee-For-Service program, was active in six states during 2005-2008. Its impact on the cost of upcoding will be captured by the fixed effects for hospitals in those six states.[24]

Our regressions all report standard errors calculated with two-way clustering at the hospital and base DRG levels (Cameron et al., 2012; Thompson, 2011). This allows for dependence in the residuals for different base DRGs across the same hospital and for different hospitals across the same base DRG. We also weight our regressions by the mean number of patients over time within a hospital/base DRG. Our controls X_{jt} include bed size, total outpatient visits, total admissions, total number of births, the number of full-time physicians and dentists, percentage of Medicare and Medicaid patients, profit status, and a teaching hospital indicator.

Following the prior literature (Silverman and Skinner, 2004; Dafny, 2005; Li, 2014), we further examine the effect from $Spread_{dt}$ across different types of hospitals. Dafny (2005) found that for-profit hospitals upcode more than government or not-for-profit hospitals and

[24]The program was expanded from three to six states in July 2007. We view its shifting effect on costs as minimal, because there are only three months overlapped with the pre-reform period. The national program was mandated by January 2010, which is outside the post-reform period we examine.

that hospitals with a high debt-asset ratio exhibited a larger increase in percent top codes. She concluded these subsets of hospitals respond more actively to the upcoding incentive. Based on equation (7), we additionally interact $Spread_{dt}$ with measures of financial health—*financially distressed* (the 25% with the highest debt-asset ratio) and *financially healthy* (the 25% with the lowest debt-asset ratio)—or with whether the hospital is for-profit or not-for-profit, with the omitted category being public hospitals.

3.2.2 Specification 2: test of costly coding from DRG reform and EMR adopters

We next test the costly coding mechanism from variation in coding costs. Similarly to Equation (6), we obtain the fraction of patients with top codes after complete documentation as follows:

$$\frac{Topcode_{jdt}}{N_{jdt}} = \frac{\sum_{i=1}^{N_{jdt}} \left[f^{comp}\left(Cost_{ijdt}\right) + g^{comp}\left(Pat_{ijdt}\right)\right]}{N_{jdt}}. \tag{8}$$

Unlike for revenues, we do not directly observe costs. However, we do observe two events that generate variation in costs across different types of hospitals. First, the 2007 payment reform increased the cost of coding by refining the distinction between DRG severity subclasses and revising the CC/MCC list. Second, we hypothesize that the costs of documentation are lower for EMR hospitals in the post-reform period than for non-EMR hospitals, since EMRs help with the specificity that was necessary to code completely during this period. Thus, we let the first term in equation (8) take the following form:

$$\frac{\sum_{i=1}^{N_{jdt}} f^{comp}\left(Cost_{ijdt}\right)}{N_{jdt}} = \beta_t \, EMR_j + \overline{c}_t^Q, \tag{9}$$

where EMR_j indicates the presence of an EMR system at the hospital before 2007 and β_t is a separate coefficient on this indicator for each quarter t. Here, we allow the effect from EMRs to vary across time in order to capture the shifting effect on the cost of documentation due to the payment reform. Similar to the above, we parameterize the average patient characteristics, which is

$$\frac{\sum_{i=1}^{N_{jdt}} g^{comp}\left(Pat_{ijdt}\right)}{N_{jdt}},$$

as $\delta X_{jt} + \overline{c}_{jd}^H + \varepsilon_{jdt}$.[25] Thus, an implication of the costly coding mechanism is that the percent top codes will be relatively higher in the post-reform period from in the pre-reform period among EMR hospitals. We define a regression specification based on this idea:

$$\frac{Topcode_{jdt}}{N_{jdt}} = \beta_t \, EMR_j + \overline{c}_t^Q + \overline{c}_{jd}^H + \delta X_{jt} + \varepsilon_{jdt}. \tag{10}$$

Since the main policy change occurs in Q4:2007, we expect β_t starting in Q4:2007 to be positive if EMR hospitals report relatively more top codes following the reform.

For Specification 2, we only include hospitals that had adopted EMRs prior to 2007—which we call EMR hospitals or early adopters—and hospitals that had not adopted EMRs

[25] As we noted above, $\overline{c}_t^Q$ will also capture variations in revenues across time.

by the end of 2009—which we call non-EMR hospitals or non adopters. We focus on these two sets of hospitals to avoid imprecise estimates due to the shifting control group over time (Borusyak and Jaravel, 2017; De Chaisemartin and d'Haultfoeuille, 2020; Sun and Abraham, 2021). Our unit of observation, dependent variable, controls X_{jt}, and clustering strategy are the same as in Specification 1.

3.2.3 Specification 3: test of costly coding from penalization of never events

The penalization of never events provides further identification of the costly coding mechanism, as the reporting of never events requires more specificity of documentation—hence increasing the probability that EMR hospitals generate a top code—but lowers hospital revenue. Similar to (9), we allow for the change in coding costs from the 2008 penalization program by interacting EMR adoption with time dummies. As with Specification 2, we assume that the average patient characteristics can be captured by hospital/base DRG fixed effects, time fixed effects, and hospital characteristics. Thus, with the same parameterization as above, we obtain the third main specification. It is almost the same as equation (10) except that the dependent variable becomes the percent of never events within a hospital/base DRG/quarter cell.

Since an increase in reported top codes lowers revenue, a finding of positive β_t during the post-reform period is both consistent with the costly coding mechanism, and also shows that the upcoding mechanism may be relatively unimportant. EMR hospitals are expected to report relatively fewer top codes post reform if the financial incentive dominates, but they are able to report more never events if the lower documentation cost due to EMR adoption is the main driving mechanism.

3.2.4 Specification 4: test of costly coding from medical and surgical DRGs

Finally, to further substantiate our analysis, we also investigate differences in cost variation for medical versus surgical base DRGs. We expect that EMRs may lower documentation costs more for medical physicians than surgeons given that CCs/MCCs are generated from medical conditions and not procedures. Moreover, for medical admissions, physicians must make a continuous series of decisions regarding patient care and management that will allow them to learn and document these conditions. To test this proposition, we allow a more flexible specification for the costs terms than in (9). Specifically, we let the cost terms from equation (9) be:

$$\mathbb{1}\{d = MED\}(\beta_{mt}\, EMR_j + \bar{c}^{Q}_{mt}) + \mathbb{1}\{d = SURG\}(\beta_{st}\, EMR_j + \bar{c}^{Q}_{st}), \tag{11}$$

where MED and $SURG$ are indicators for the base DRG being medical and surgical respectively. In equation (11), both the base costs of coding the top code and the incremental costs of coding the top code for EMR adopters post-reform vary across MED and SURG base DRGs.

We estimate separate specifications of (10) for MED and SURG base DRGs. This allows us to examine whether hospitals' response to the 2007 payment reform differs between these two types of base DRGs. If costly coding is the primary mechanism underlying variation in billing top codes, we expect that the increase in reporting the top code will be relatively greater among medical DRGs.

3.3 Identification

All specifications detailed in Section 3.2 include fixed effects at the hospital/base DRG level. Hence, identification is based on variations within a hospital/base DRG across time. The key identifying assumption for these analyses is that the main variables of interest—$Spread_{dt}$ and EMR adoption prior to the start of our sample interacted with quarter dummies—are exogenous to the unobservables after controlling for a rich set of hospital characteristics and the fixed effects.

For *Specification 1* in equation (7), identification is from changes in the spread and relative changes in percent top codes within a hospital/base DRG cell. The central threats to identification would come from correlations between changes in the spread and changes in the unobservable. For example, a threat to identification would occur if base DRGs where the spread increased were associated with relatively more complicating conditions for those patients due to technological change. We believe that such association is unlikely because changes in CCs/MCCs would occur slowly over time, if at all. Another threat would come if CMS increased fraud enforcement for a base DRG when it increased the spread for this base DRG. However, CMS enforcement initiatives, such as Medicare's Recovery Audit Program, occur more broadly across base DRGs rather than changing targeting to base DRGs in response to changes in spread.[26]

[26] The claims that were flagged to trigger review by auditors were selected based on the auditors' proprietary algorithm (Shi, 2021) or a set of claims and providers which it had determined had a high propensity for error under a program called Comprehensive Error Rate Testing (CERT) that dated back to 2003 (`https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/`

For *Specification 2*, we identify costly coding from changes in the fraction of top codes among EMR hospitals relative to changes in this fraction among the comparison group,[27] from the pre- to post-reform period. A threat to identification here derives from changes in patient mix at EMR hospitals relative to other hospitals over time. For instance, if relatively more patients with CCs seek care at EMR hospitals over time, this would make EMR hospitals report relatively more top codes following the reform, even if EMRs did not help with complete coding. We investigate whether there were differential pre-reform or post-reform trends for EMR hospitals relative to other hospitals or whether changes following the payment reform were sudden. We also examine other measures of patient mix, such as distance traveled to a hospital, to see whether any effects reflect differential changes in patient demand rather than coding, at EMR hospitals post-reform. In the absence of such effects, we believe that the reform is the most likely causal impact of any effect that we find.

Even if the relative difference in top codes was caused by the reform, another threat to identification comes from non-time-varying differences across hospitals leading to coding differences in the post-reform period. For instance, if EMR hospitals have relatively more patients who merit a CC post-reform relative to the fraction of patients who merit a CC pre-reform than does the comparison group, this would yield positive post-reform EMR coefficients, even if EMRs did not help with complete coding. However, Specification 3 is not vulnerable to this threat, since this specification considers the same criteria pre- and post-reform. Also, Specification 4, which breaks down extra top codes by MED and SURG,

`Medicare-FFS-Compliance-Programs/CERT/index.html?redirect=/Cert`). In general, the set of these claims did not vary with changes in the DRG spread for a given base DRG.

[27]Because we compare early EMR adopters to a comparison group of hospitals that had not adopted by the end of our sample, our results are not vulnerable to having different treatment groups over time.

can mitigate this concern by adding supporting evidence to the costly coding story.

For *Specification 3*, we identify the impact of never events based on how EMR hospitals act differently in the penalization period from the pre-penalization period. Thus, the central threat to identification will be similar to the first threat to identification from Specification 2. We will find evidence of upcoding if there is a drop in reported never events for EMR hospitals relative to other hospitals in the penalization period relative to the pre-penalization period. However, as with Specification 2, a decreasing number of reported never events could also result from different time trends at EMR hospitals relative to other hospitals. For instance, if EMR hospitals improved their prevention of never events during the penalty period relative to other hospitals, this would look similar to upcoding for EMR hospitals. To consider the possibility of differential changes in care, we will consider the source of the changes in reported never events.

For *Specification 4*, we identify costly coding for MED relative to SURG base DRGs from changes in the fraction of top codes among EMR hospitals minus changes in this fraction among non-EMR hospitals for MED base DRGs relative to this difference for SURG base DRGs. Thus, the threats to identification are similar to Specification 2, but they would have to occur from EMR hospitals having a different fraction of medical patients with CCs/MCCs post-reform to pre-reform relative to surgical patients with CCs/MCCs. As with Specification 2, we further examine these threats by considering pre-trends, presenting results graphically, and examining whether there were sharp effects.

4 Data and Summary Statistics

4.1 Data sources

Our primary dataset is the Medicare Provider Analysis and Review (MedPAR) data. For our purposes, this dataset contains information on all inpatient hospital stays for Medicare beneficiaries. Each observation in these data represents one patient stay and contains information on the hospital, the beneficiary's home zip code, age, gender, dates of service, reimbursement amount, dates of admission and discharge, DRG, and principal and secondary diagnosis and procedure codes. We drop admissions that are not paid under PPS, such as those from Critical Access Hospitals (CAHs) (Gowrisankaran et al., 2018).[28] We construct our main dependent variable, the percent of patients with documented top codes within a particular base DRG, hospital, and quarter, from the MedPAR data. Our discharge data extend from Q1:2005 through Q4:2009. However, most of our analyses use data over a shorter time period, as we discuss below.[29]

We merge our base data with information on DRGs from the CMS. This information indicates whether the DRG is medical or surgical. It also indicates the weight for each DRG. DRG weights change at the beginning of each fiscal year, which corresponds to the fourth quarter of a calendar year. We use the DRG weight data to calculate our measures of spread.

We also create never event indicators using the MedPAR data. We first find whether any

[28]CAHs receive cost-based reimbursements from Medicare. Other PPS-exempt hospitals include swing-bed short-term/acute care hospital, swing-bed long-term hospital, inpatient psychiatric hospitals, etc.

[29]We also construct other dependent variables using this dataset, including the distance traveled, length of stay, mean DRG weight, and numbers of diagnoses and procedures. We calculate the distance between each patient and the hospital based on the latitude and longitude of the patient and hospital zip codes.

of the secondary diagnoses reported in our main analysis are on the list of hospital acquired conditions (HACs). If so, we then define a never event if the event is definitively not present on admission (POA). According to the final rule published in 2009,[30] the HAC list broadly includes 12 conditions, each of which is associated with a series of ICD-9 codes and some with procedures. For most of these conditions, the presence of a specified ICD-9 code as a secondary diagnosis determines inclusion in the HAC list. For others, multiple conditions are required to identify inclusion. For instance, the condition "Surgical Site Infection Following Bariatric Surgery for Obesity" requires the presence of specified ICD-9 codes, procedures, as well as the primary diagnosis being morbid obesity (ICD-9 code: 278.01).

To categorize a condition in the HAC list to be POA or not, the CMS mandated a new code, "present on admission indicator," which requires hospitals to document whether each diagnosis occurred before or after hospital admission. This indicator can be coded as "Y" (present at the time of admission), "N" (acquired at hospital), "W" (providers unable to determine), "U" (insufficient documentation), "1" (unreported/not used, or undesirable blanks), and other values. Note that CMS will *not* pay the CC/MCC DRG for the HAC coded with non-POA indicators, that is, all values other than "Y" or "W."[31] The dependent variable for never events is the percent of discharges with the POA field coded as "N" within a hospital/base DRG/quarter cell.

To further investigate whether EMR hospitals engage in different coding of the POA field,

[30] See https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2008-Fact-sheets-items/2008-08-042.html. Also see Office of the Federal Register and National Archives and Records Service (2008), p. 48,471-48,482.

[31] See https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Coding.

we define an admission as *inconclusive* POA, if any of the HACs is coded as "U," "1," missing, or other values that are not specified by CMS, due to insufficient or missing documentation regarding the POA status. The dependent variable for these specifications is the percent of discharges with inconclusive POA within a hospital/base DRG/quarter cell.

Our second main dataset provides information on EMR adoption from the Healthcare Information and Management Systems Society (HIMSS) Analytics Database, which is the most comprehensive national source of hospital IT adoption data. We use the Medicare provider number as a crosswalk between this dataset to the MedPAR data. This dataset is the most complete, detailed, and longest-running survey recording the choice and evolution of a hospital's IT capacities.

As noted in Section 2.2, there are several components of EMRs. We use the presence of a live and operational CDS or CPOE within the organization as our measure of EMR adoption, since physicians will interact most with these components. This is roughly consistent with what has been done in the literature. For instance, Jha et al. (2009) divide EMR systems into 32 functionalities, of which they view eight (including some parts of CPOE) as necessary for "basic" EMR operation. Miller and Tucker (2009) measure EMR adoption by whether a hospital has installed an "enterprise EMR" system, which they state is a "basic" system that underlies CDR, CDS, and CPOE. Recent studies defined EMR capabilities by either enterprise EMR, CDS, or CPOE (Lee et al., 2013; Agha, 2014; Dranove et al., 2014; McCullough et al., 2016; Ganju et al., 2021). We also examine the robustness of our results to alternate definitions of EMR adoption based on different components, and the results are basically consistent.

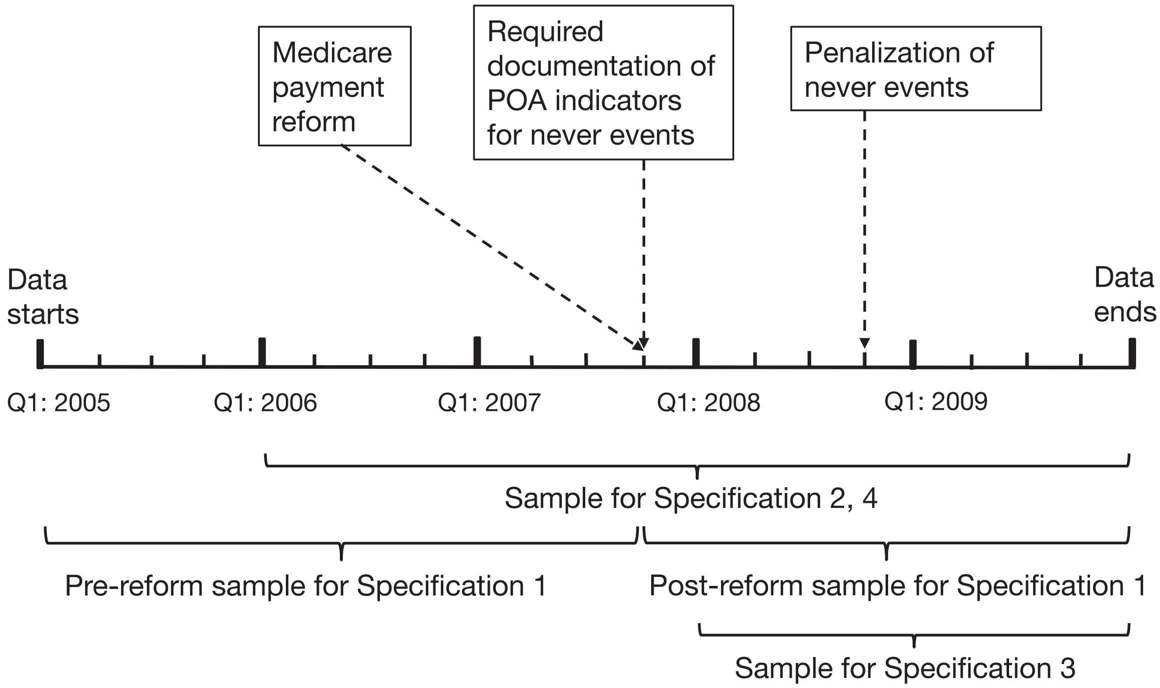
Finally, we use two other datasets. First, we merge in the American Hospital Association (AHA) Annual Survey data, using the Medicare provider number as the primary crosswalk. In cases where the Medicare provider number was missing, we merge the databases using the hospital's name and exact address. We match approximately 3,200 non-CAH hospitals across the three datasets. The AHA data provide us with hospital characteristics such as number of beds, system affiliation, profit status, etc. In addition, to understand the impact of hospital financial status on billing, we merge financial status data from the Medicare Cost Reports, using the Medicare Provider Number field as the crosswalk. Following the literature (Dafny, 2005; Li, 2014), we use the debt-to-asset ratio as a measure of financial health. We construct this measure by dividing current liabilities by total assets, both of which are listed in the cost reports. We define a hospital as financially distressed if its debt-to-asset ratio is above the 75th percentile and as financially healthy if this ratio is below the 25th percentile.

4.2 Timeframe for analysis

Figure 2 provides timelines for the data used in different specifications and the underlying policy changes that motivate the timelines. In Specification 1, we examine how hospitals respond to changes in revenues, ceteris paribus. This limits our data to the pre- or post-reform period separately, so as to minimize the change in average patient characteristics. Here, we only use base DRGs with two severity subclasses for a clear definition of spread. We also perform separate regressions for the pre- and post-reform period. Not shown in Figure 2 for clarity, spread changes in the fourth quarter of every year.

Specification 2, which considers the impact of the payment reform on EMR hospitals

Figure 2: Timeline of data for different hypotheses



relative to non-EMR hospitals, uses four years of data (2006-2009) and all matched base DRGs from before and after the reform with multiple severity subclasses.[32] For base DRGs with three severity subclasses, we use the fraction of patients with the top of the three severity subclasses as the dependent variable.

For Specification 3, which considers never events, although mandatory POA reporting for the HAC list started in Q4:2007, POA reporting was not used in the claim processing system until January 2008.[33] Hence, we use Q1:2008 as the start of the sample for Specification 3. Our data do not include the POA field for 2010 and hence this sample ends in Q4:2009.

Specification 4 replicates the analysis in Specification 2, separately for MED and SURG

[32]In this specification, we skip the year 2005 due to a large number of missing observations on EMR adoption status for that year.

[33]See the change request (CR) 5679: https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/Downloads/R289OTN.pdf.

base DRGs. Hence, the timeline for this specification is the same as for Specification 2.

4.3 Summary statistics on data

Table 1 provides summary statistics on overall patient samples and the samples used in the different specifications. Panel 1 of Table 1 shows our overall data separated by year. On average, there were more than 13 million Medicare discharges each year in our data.[34] The mean age of a Medicare patient discharged from a hospital was 73 years in our sample and the mean DRG weight was rising over time, from 1.44 in 2005 to 1.55 in 2009.

Panel 2 shows the sample by fiscal year (FY) for Specification 1, since this specification separately examines the pre- and post-reform periods and the reform is implemented on a fiscal year basis. For our samples, which contain base DRGs with two severity subclasses, the percent top codes drops substantially following the reform, from over 73% before the reform to 28% immediately after the reform.[35] Note that the number of discharges is relatively small in FY2005, as the data in Q4:2004 are not available.

Panel 3 shows the sample by year for Specification 2. Our sample here contains patients in base DRGs with multiple severity subclasses pre- and post-reform for which the definition of the base DRG before and after the reform was identical. Note that the percent top codes in 2007 is a bit lower than that in 2006, as it is a mix of the percent top codes before and after the reform.[36] Specification 4 also decomposes this sample by whether the base DRG

[34]This includes admissions that are not paid on PPS.

[35]The 40.3% in the Introduction is cited from the Federal Registry. It is calculated using the 2006 MedPAR data based on the revised CC list, whereas the 28% here is based on the sample for Specification 2—the MedPAR data for a subset of base DRGs with two subclasses in FY2008.

[36]The 2007 payment reform was not implemented until Q4:2007.

Table 1: Summary statistics on patient sample

Panel 1: universe of Medicare patients in sample					
	2005	2006	2007	2008	2009
Mean age	73.1	73.1	73.0	72.9	72.9
Mean DRG weight	1.44	1.45	1.46	1.51	1.55
# discharges	13,512,853	13,223,495	12,988,668	14,238,372	14,593,058
Panel 2: samples for Specification 1 (FY 2005-7: base DRGs with 2 subclasses pre-reform; FY2008-9: 2 subclasses post-reform)					
	FY2005	FY2006	FY2007	FY2008	FY2009
% top codes	77.4	73.1	73.8	28.1	30.8
# discharges	2,621,588	3,953,525	3,773,419	4,441,733	4,764,406
Panel 3: samples for Specifications 2 and 4 (matched base DRGs with multiple severity subclasses)					
	2005	2006	2007	2008	2009
% top codes	–	72.2	61.9	27.8	29.9
# discharges	–	3,047,422	2,900,321	3,267,855	3,284,647
MED					
% top codes	–	81.3	68.4	29.7	32.3
# discharges	–	1,729,764	1,682,775	1,934,347	1,958,632
SURG					
% top codes	–	60.3	52.9	25.1	26.4
# discharges	–	1,317,658	1,217,546	1,333,508	1,326,015
Panel 4: sample for Specification 3 (all base DRGs)					
	2005	2006	2007	2008	2009
% discharges w/ HACs	4.72	4.73	5.20	5.05	4.35
% discharges w/ never events	–	–	–	0.176	0.311
% discharges w/ inconclusive POA	–	–	–	2.88	0.059

Note: Panel 2 reports statistics in the fiscal year, which is the accounting period for the federal government, from Q4 of the previous year to Q3 of the current year. Data in 2005 not included in the analysis for Specifications 2 and 4. "POA" indicates present on admission and "HAC" indicates hospital-acquired conditions.

is medical or surgical. The percent top codes is higher among medical than surgical DRGs, both before and after the reform.

Panel 4 shows the sample by year for Specification 3. The percent of never events was much higher in 2009, whereas the proportion of inconclusive POAs dropped substantially from 2008 to 2009.

Table 2: Summary statistics on hospital characteristics by EMR use

	EMR adopters	EMR non-adopters
Bed size	264	111
Total outpatient visits	208,992	79,201
Total admissions	12,479	4646
FTE physicians and dentists	28	7
Total number of births	1,449	486
% teaching hospital	12.6	2.34
% Medicare discharge	44.5	48.9
% Medicaid discharge	18.7	18.1
% for-profit	19.3	30.0
% not-for-profit	69.8	42.6
% public hospitals	10.9	27.4
Debt-asset ratio	0.631	0.696
Number of hospitals	1,522	256

Note: For each set of hospitals in our final data, table reports the mean value of statistics over years in our data.

Our analysis uses hospitals that adopted EMRs during or before 2006 and that had not adopted EMRs by 2009. Table 2 provides summary statistics for the main hospital characteristics separated by adoption status. Hospitals that adopted EMRs in 2006 or earlier are on average larger and more likely to be teaching and not-for-profit hospitals. For instance, the bed size for early adopters is more than twice that of hospitals without adoption through 2009. Early adopters also had a slightly lower debt-to-asset ratio than hospitals without adoption through 2009.

Figure 3 displays the raw trends of percent top codes separately for EMR and non-EMR hospitals based on the sample for Specification 2. We include a vertical line right before Q4:2007 to indicate the implementation of the 2007 payment reform. Visually, there is almost no difference in top codes between early adopters and non adopters prior to the reform. The payment reform reduced the incidence of top codes substantially, but the reduction was relatively smaller among early adopters during the post-reform period.

Figure 3: Percent top codes by EMR adoption status

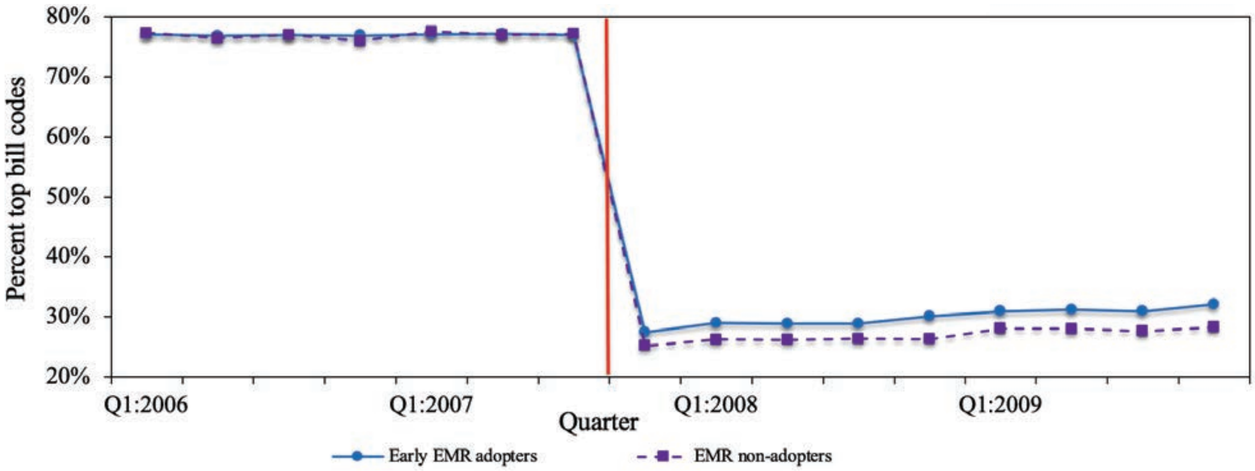


Table 3: Summary statistics on DRG weights

	Panel 1: All DRGs		
Variable	# DRGs	Mean	Std. Dev.
DRG weight, FY 2005	518	1.53	1.90
DRG weight, FY 2006	559	1.47	1.86
DRG weight, FY 2007	579	1.48	1.84
DRG weight, FY 2008	743	1.99	1.93
DRG weight, FY 2009	744	2.02	2.00
	Panel 2: $\Delta Spread$ in sample for Specification 1		
Variable	# base DRG/hospital /quarter cells	Mean	Std. Dev.
$\Delta Spread$, FY2005 to FY2006	630,378	−0.029	0.240
$\Delta Spread$, FY2006 to FY2007	635,570	−0.0027	0.046
$\Delta Spread$, FY2008 to FY2009	570,516	0.328	0.243
$\Delta Spread$, FY2009 to FY2010	141,952	−0.0149	0.093

Note: *Spread* measures the difference between the weight in the top and bottom codes within a base DRG. Only the first quarter of FY2010 (i.e., Q4:2009) is included in the analysis, so the number of observations in this fiscal year is smaller than any others.

Table 3 provides summary statistics on DRG weights with the mean DRG weights in the top panel. The 2007 reform resulted in many more DRGs and in a higher and increasing mean DRG weight, when taken as a simple average across DRGs. The bottom panel of Table 3 shows the change in spread for the base DRGs in the analysis for Specification 1. There were large changes in spread across fiscal years, with standard deviations in the change in spread of 0.046 to 0.243 depending on the year. Thus, there is substantial variation to identify this specification.

5 Results

5.1 Specification 1: test of upcoding using variation in revenue

We present the results for the specifications we developed in Section 3.2. We start with Specification 1, which evaluates the impact of the spread on the probability of reported top codes within a base DRG. We estimate this specification separately before and after the reform, with the percent top codes within a base DRG/hospital/quarter cell as the dependent variable. We focus on all base DRGs with two severity subclasses so that the spread unambiguously defines the incentives to document a top code.

Table 4 presents the coefficient for the key variable of interest: *Spread*, for both the pre- and post-reform samples.[37] The eight coefficients in this table derive from eight separate specifications. For each time period, we estimate four specifications, progressively adding more controls. Our preferred specifications are in column (4). These specifications include

[37] We also tried including both lagged Spread and Spread into the regression. The coefficients for Spread is no longer significant, but the overall effect of Spread is qualitatively similar.

Table 4: Extra top codes with spread (Specification 1)

	Dependent variable: Percent top bill codes within a base DRG			
	(1)	(2)	(3)	(4)
Pre-reform	−2.46	.652***	.63***	.631***
	(4.82)	(.101)	(.101)	(.101)
Post-reform	−8.67	-2.87*	-2.89*	-2.90*
	(6.00)	(1.53)	(1.55)	(1.55)
Quarter dummies	yes	yes	yes	yes
Base DRG FEs	no	yes	no	no
Hospital/base DRG FEs	no	no	yes	yes
Other controls	yes	no	no	yes

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percentage point form. Pre-reform sample is base DRGs with two severity subclasses from Q1:2005-Q3:2007, and post-reform sample is base DRGs with two severity subclasses from Q4:2007-Q4:2009. Columns (1)-(4) in the upper panel report the coefficient for *Spread* from four specifications, separately for the pre- and post-reform sample. The four specifications progressively add more controls, as specified in the lower panel. Other controls, in the last row, include bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, and total number of births. Standard errors are clustered at both hospital and base DRG levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

spread, quarter dummies, hospital/base DRG fixed effects, and other controls noted in the table note.

Focusing first on the pre-reform period, column (4) shows that before Q4:2007, top coding occurs with high frequency when the financial incentive to report the top code increases. A unit increase in spread corresponds to an increase of 0.631 percentage points in the reporting of a top code. This is consistent, though not dispositive, with the finding in prior studies that provided evidence of upcoding (Silverman and Skinner, 2004; Dafny, 2005). Documentation of top codes increasing with spread implies that revenues were an important driver of coding in the pre-reform period.

Turning now to the post-reform period, from Q4:2007 onwards, a larger spread no longer predicts a higher fraction of top codes. For this time period, the column (4) point estimate on spread is negative at the 10% level of significance.[38] This difference in results may be due to a more complex reimbursement system making the cost of coding more important in the billing decision than the revenue gained. The post-reform evidence is not consistent with the upcoding explanation where higher revenues lead to more top codes. While reducing upcoding is not one of the stated policy goals of the 2007 payment reform, our findings show that it is an important side effect.

Table 5: Extra top codes with spread, by hospital types

	Dependent variable: Percent top bill codes within a base DRG			
	Pre-reform		Post-reform	
	Coefficient	S.E.	Coefficient	S.E.
By profit status				
Spread	.712***	(.158)	-1.78	(1.57)
ForProfit×Spread	−.125	(.154)	.342	(1.54)
NotForProfit×Spread	−.085	(.135)	−1.7**	(.728)
By financial health status				
Spread	.677***	(.106)	−3.21**	
FinanciallyDistressed×Spread	−.0473	(.104)	.206	(.50)
FinanciallyHealthy×Spread	−.0709	(.0977)	.998	(.863)

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percentage point form. Pre-reform sample is base DRGs with two severity subclasses from Q1:2005-Q3:2007, and post-reform sample is from Q4:2007-Q4:2009. For the analysis by profit status, the omitted category is public hospitals. For the analysis by financial health status, the omitted category is hospitals whose debt-asset ratio is above 25 and below 75 percentiles. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, profit/financial health status interacting with quarter dummies, quarter dummies, and hospital/base DRG fixed effects. Standard errors are clustered at both hospital and base DRG levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Following prior studies (Silverman and Skinner, 2004; Dafny, 2005; Li, 2014), we also examine the variation in responsiveness to revenues by hospital ownership and financial

[38] The Z statistic for testing the join equality between the pre- and post-reform coefficients is 2.27, suggesting that we can reject the null hypothesis that both coefficients are the same at the 5% level of significance.

health status. The upper panel in Table 5 reports the results where *Spread* is interacted with hospital ownership status, separately for the pre- and post-reform sample using the preferred specification. The coefficient for *Spread* remains significantly positive for the pre-reform sample, but there is no significant difference between for-profit and not-for-profit hospitals. After the reform, the interaction term for not-for-profit and spread is significantly negative, while the other coefficients are not statistically significant. The lower panel presents the results when we add the hospital's financial health status as a regressor. The interaction terms between spread and financial health measurers are not statistically significant.

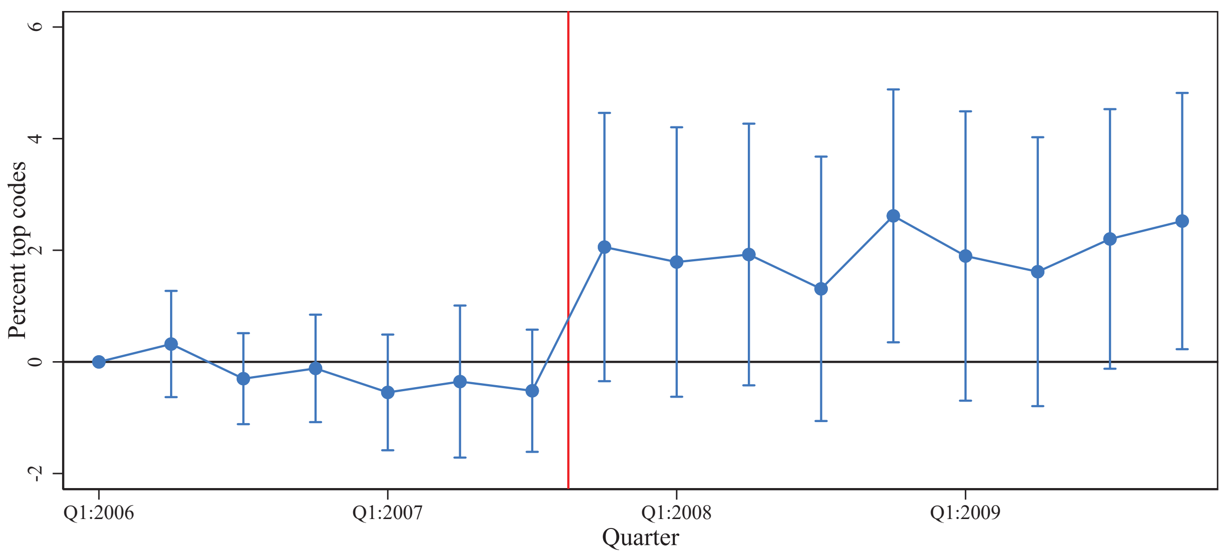
We also examine whether EMR and non-EMR hospitals respond differentially to changes in spread. Table A1 in Appendix A reports the overall effect of changes in spread on percent top codes, as well as the effect by different types of hospitals, separately for early EMR adopters and non-EMR hospitals. EMR hospitals share almost the same patterns as the entire sample, except that the interaction between spread and for-profit becomes significantly negative. The estimates among non adopters are qualitatively similar, yet noisier especially during the post-reform period.

To summarize, we find that hospitals responded to financial incentives in coding before the reform, which could be due to upcoding. Prior to the reform, there is little evidence that the impact varied by hospital type. The significant impact on top-coding based on financial incentives does not extend to the post-reform period, and indeed the point estimate is negative. This suggests that the cost of complete coding might play a more important role in the coding decision post-reform, due to the increased complexity in the new payment system. We investigate this hypothesis in the remaining specifications.

5.2 Specification 2: test of costly coding from DRG reform and EMR adopters

Specification 2 evaluates the impact of EMRs on reported top codes. This specification uses variation in costs generated by the 2007 payment reform and in EMR adoption status prior to the start of our sample. Our analysis uses all matched base DRGs from before and after the reform with multiple severity subclasses. Our regressors include quarter dummies, indicators for quarter interacted with early EMR adoption, hospital/base DRG fixed effects, and the same other controls as in Specification 1.

Figure 4: Extra top codes with early EMRs



Note: The line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percentage point. The red line represents the 2007 payment reform. The vertical lines show 95% confidence intervals, based on standard errors clustered at both hospital and base DRG levels. Unit of observation is hospital/base DRG/quarter. Sample is base DRGs with multiple severity subclasses, from Q1:2006-Q4:2009. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects.

Figure 4 presents the coefficients on interactions between quarter and EMR adoption, as well as their 95% confidence intervals. Table A2 in Appendix A provides more details on the same regression. All reported coefficients are relative to Q1:2006 (which has a zero coefficient in the graph); we need to omit one quarter because (by construction) early EMR adoption does not vary during our sample for a given hospital/base DRG. We include a vertical line right before Q4:2007 to indicate that the reform was enacted starting in Q4:2007.

Following the reform, we observe more top codes for EMR adopters relative to non adopters. Visually, there is no trend in either the pre-reform or post-reform periods, but rather a sharp change in the post-reform period relative to the pre-reform period. The distinct change at the point where the reform was enacted suggests that the result is due to the reform rather than other secular trends.[39]

The coefficients for the post-reform interaction terms show a mean increase in top coding of 2.0 percentage points. A joint significance test reveals that these coefficients are together significantly non-zero (P=0.0039). Thus, the results show that early EMR adopters had a significant comparative advantage in reporting top codes post-reform to pre-reform.

Our model specifies that the relative increase in reported top codes for EMR hospitals post-reform is due to coding differences. However, we further test whether this result could be instead due to differences in patient severity of illness or services performed post-reform.

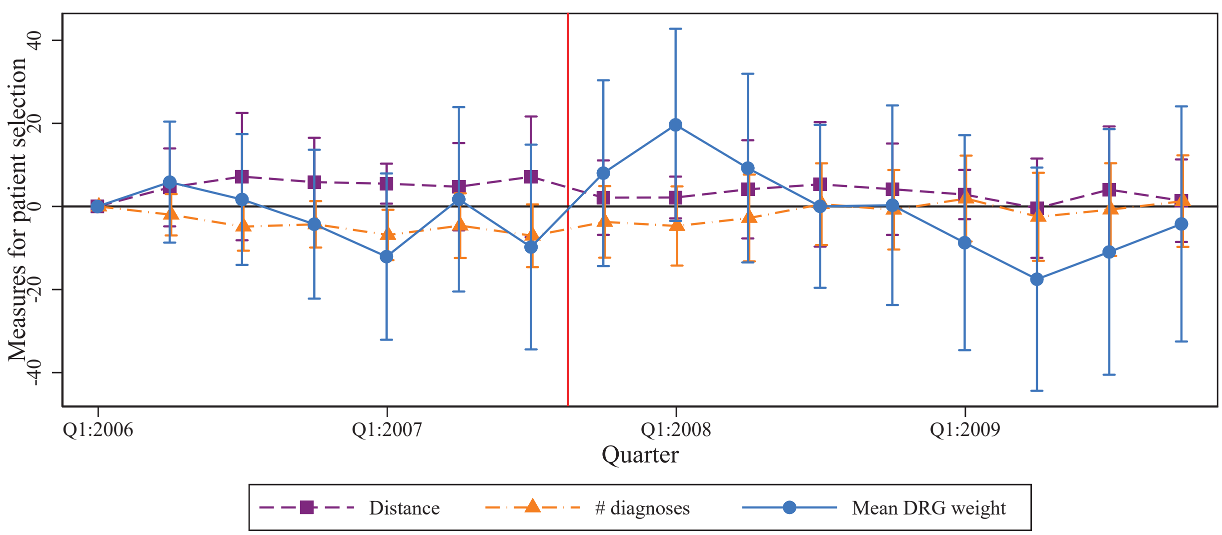
Figure 5 considers three patient characteristics: distance traveled to the hospital, the number of reported diagnoses, and the mean DRG weight based on the base DRG.[40] In

[39]This result contrasts with Geruso and Layton (2018), who find no significant impact of EMR adoption on increased coding for Medicare Advantage. The two settings incorporate a number potential explanatory differences: Geruso and Layton use 2011 EMR adoption (which was at the end of their 5-year sample period, while many entities adopted EMRs during their sample period); they consider Medicare Advantage insurance coding (while we consider Medicare inpatient coding for a single hospital admission); their "meaningful use" definition of EMR adoption is narrower than our definition; and they consider physician office EMR adoption while we consider hospital EMR option.

[40]We use the mean DRG weight for the lowest severity subclass rather than the actual reported DRG weight because we would like this effect to be robust to misreporting of severity subclasses.

none of the three cases is there a sharp change post-reform relative to pre-reform. Moreover, we find that the post-reform indicators are not jointly statistically significant for distance (P=0.657) and for number of diagnoses (P=0.379). For mean DRG weight of the base DRG, the post-reform coefficients are statistically significant (P=0.0605) but this reflects both positive and negative values in different quarters.[41] Hence, it does not appear that the increase in reported top codes post-reform is due to changes in the patient mix at these hospitals relative to others.

Figure 5: Patient characteristics with early EMRs



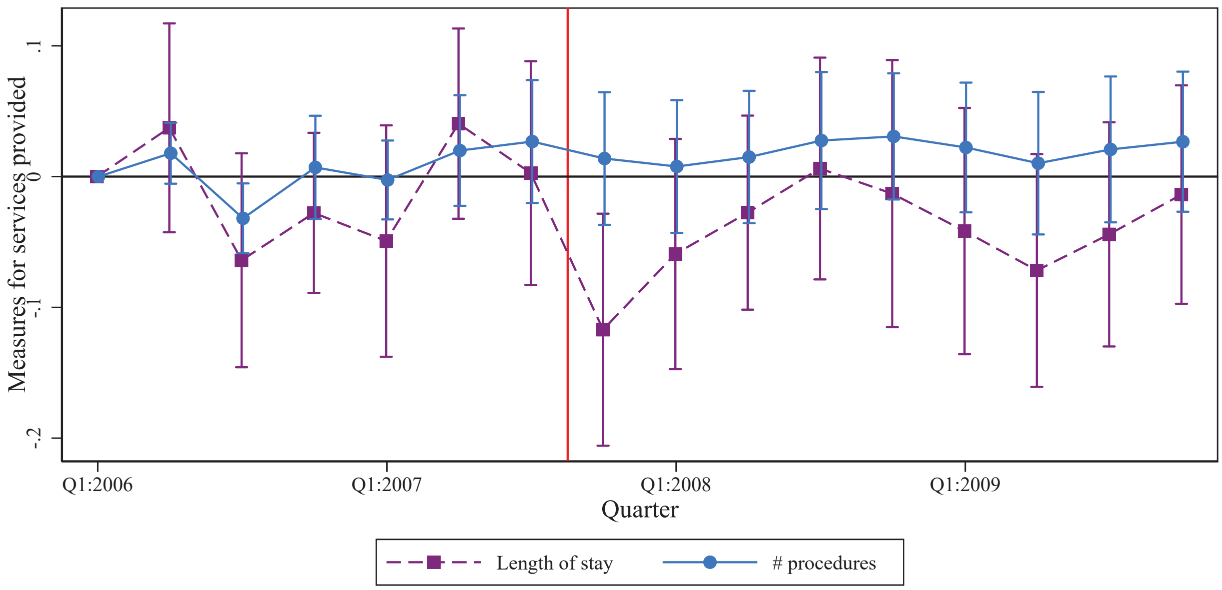
Note: Unit of observation is hospital/base DRG/quarter for travel distance and the number of diagnoses but hospital/quarter for mean DRG weight. The mean DRG weight is calculated using the lowest weight for any base DRG. Coefficients are reported in percent point form, rescaled by 100 times for number of diagnoses and by 1000 times for mean DRG weight. Sample is all matched base DRGs from before and after the reform with multiple severity subclasses, from Q1:2006-Q4:2009. Standard errors are clustered at both hospital and base DRG levels for travel distance and the number of diagnoses but at the hospital level for mean DRG weight. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects.

Figure 6 considers two measures of services performed: length-of-stay and number of procedures performed. Again, neither measure shows a sharp change post-reform relative to pre-reform. The post-reform indicators for number of procedures are not jointly statistically

[41] To test for the presence of monotonic effects, we estimate each of these models with a single post-reform / early EMR interaction coefficient. We find test statistics of P=0.095, P= 0.437, and P=0.765, respectively.

significant (P=0.583). The post-reform indicators for length-of-stay are significantly negative (P=0.00133), which could not be caused by sicker patients at EMR hospitals post-reform but may reflect better treatments and hence shorter length of stay at these hospitals over time.[42]

Figure 6: Services provided with early EMRs



Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percent point form. Sample is all matched base DRGs from before and after the reform with multiple severity subclasses, from Q1:2006-Q4:2009. Standard errors are clustered at both hospital and base DRG levels. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects.

5.3 Specification 3: test of costly coding from penalization of never events

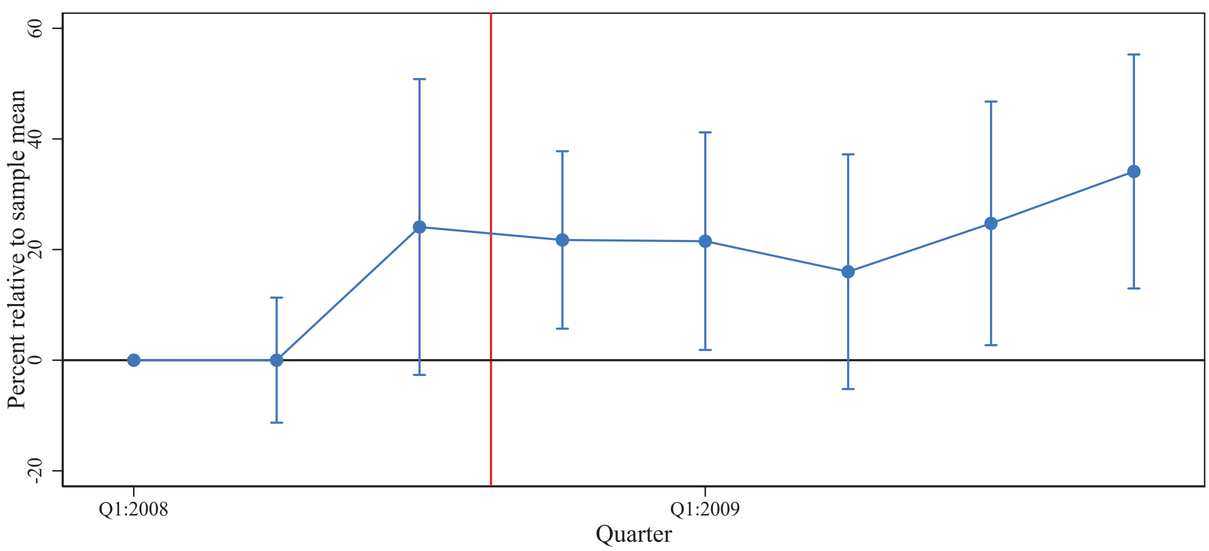
Our findings in Specification 2 show that coding costs play a role in reported top codes post-reform relative to pre-reform. It suggests that costly coding has more support in the post-2007 data than upcoding. Specification 3 seeks to further test the two mechanisms against each other by examining never events. We base our analysis on the 2008 penalization

[42]Here again, we estimate each of these models with a single post-reform / early EMR interaction coefficient, finding test statistics of P=0.380 and P=0.214, respectively.

of never events, which creates variation in costs by increasing the specificity of coding the POA field, but affects revenue in the opposite direction to the 2007 payment reform. For this specification, the dependent variable is the percent of discharges with at least a listed HAC coded as a never event within a hospital/base DRG/quarter cell. Thus, a higher value of the dependent variable indicates the incidence of never events and lowers reimbursements in the post-penalization period. We use the same regressors as in Specification 2.

Similar to Figure 4, Figure 7 shows the interaction coefficients and their 95% confidence intervals, with more details in the left panel of Table A3 in Appendix A. In Figure 7, we normalize the coefficients by the 2008-09 sample mean of the never event probability, which is 0.256%, and include a vertical line right before Q4:2008 to indicate the start of penalization for never events. All reported coefficients are relative to Q1:2008.

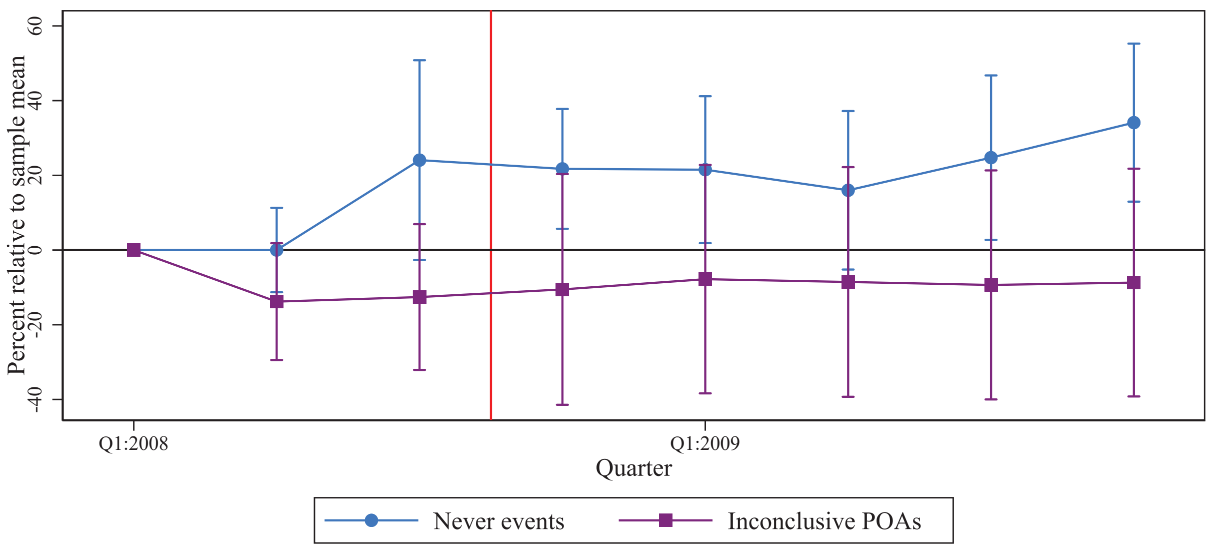
Figure 7: Extra reported never events with early EMRs



Note: Each line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percent relative to the sample mean. The red line represents the 2008 penalization of never events. The vertical lines show 95% confidence intervals, based on standard errors clustered at both hospital and base DRG levels. Unit of observation is hospital/base DRG/quarter. Sample is all base DRGs from Q1:2008-Q4:2009. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects.

All coefficients for the post-penalization periods are positive, implying that early EMR adopters had more never events—0.0617 percentage points on average—in every quarter of the post-penalization period, in the difference-in-difference from their Q1:2008 values and non-EMR adopters. The relative magnitudes, compared to the sample mean, are large: the post-penalization coefficients show that early EMR adoption predicts between 16 and 34 percent more reported never events than the comparison group. The coefficients during the penalty phase are jointly significantly positive (P=0.0385).

Figure 8: Extra never events vs. inconclusive POA with early EMRs



Note: Each line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percent relative to the sample mean. The red line represents the 2008 penalization of never events. The vertical lines show 95% confidence intervals, based on standard errors clustered at both hospital and base DRG levels. Unit of observation is hospital/base DRG/quarter. Sample is all base DRGs from Q1:2008-Q4:2009. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects.

If costly coding were the main mechanism underlying the coding behavior, EMR hospitals would be expected to report more never events even *before* the penalty provision. As is shown in Figure 7, EMR hospitals started improved documentation of never events one quarter before the penalization started, in Q3:2008. Moreover, the costly coding mechanism also predicts that EMR hospitals report relatively more never events than inconclusive POAs both before and during the penalization period. Figure 8, in a similar format to Figure 7, shows the effect of early EMRs on the reporting of never events and inconclusive POA, respectively. For the latter, the dependent variable is the frequency of insufficient or missing documentation in the POA field. The right panel of Table A3 includes more detail on the estimation results. EMR hospitals report relative more never events but fewer inconclusive POA values from Q2:2008 till the end of the sample, which is consistent with the prediction from the costly coding mechanism.

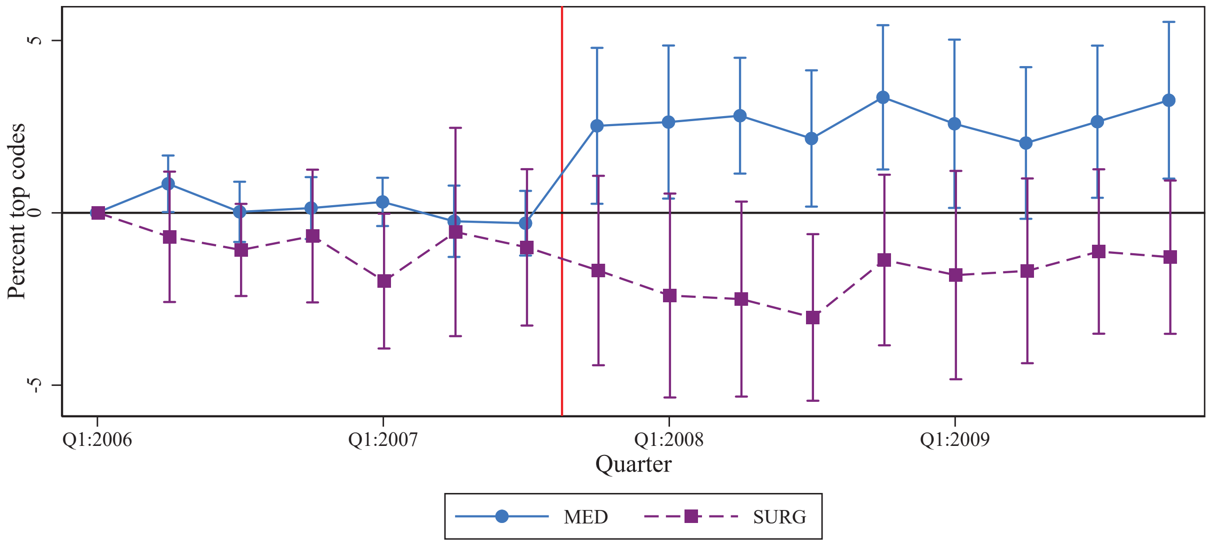
To summarize, we find that EMR hospitals increased their reporting of never events relative to the comparison group when these were penalized (and one quarter below). This occurred because they were relatively better at reporting the POA field. These results suggest that upcoding is not an important factor for these patients, as EMR hospitals would not increase reporting never events in the post-penalization period relative to the comparison group if upcoding were the predominant incentive for reporting these events and EMRs helped with the specificity of coding. However, under costly coding, EMR hospitals may report relatively more never events since EMRs lower the costs of complete coding. In combination with Specification 2, these results lend support to EMRs lowering the cost of complete coding but not leading to upcoding.

5.4 Specification 4: test of costly coding from medical and surgical DRGs

Specification 4 evaluates how EMRs may have differentially affected hospital coding across different types of base DRGs, to further investigate our finding that costly coding is the dominant incentive during our study period. Similarly to Specification 2, we use the variation induced by the 2007 payment reform. If hospitals respond to the increased complexity of billing as predicted by our costly coding mechanism, we expect an increase in reporting top codes where EMRs are more helpful in lowering the cost of coding. As noted above, there could be complementarities between medical admissions and EMR adoption in lowering coding costs. In this case, the effect will be larger among medical than surgical DRGs. We test this with a specification identical to Specification 2, except that we perform separate regressions for medical (MED) and surgical (SURG) base DRGs, thus effectively interacting every coefficient with MED and SURG.

Figure 9 presents coefficients and 95% confidence intervals, with details in Table A4 in the Appendix A. Starting immediately after the payment reform, EMR hospitals report relatively more top codes for MED. The effect is very consistent across time. The magnitudes are large and jointly statistically significant (P=0.00365), with roughly 2.67 percentage points more top codes for early EMR adopters relative to non-EMR hospitals throughout the post-reform period, compared to the mean top coding probability of 31.6% for the post-reform MED sample. In contrast, the post-reform interaction coefficients for surgical DRGs are negative and jointly statistically significant (P=0.00233), with roughly 1.88 percentage points *fewer* top codes for early adopting hospitals in the SURG sample post reform.

Figure 9: Extra top codes with early EMRs, by MED and SURG



Note: The line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percentage point. The red line represents the 2007 payment reform. The vertical lines show 95% confidence intervals, based on standard errors clustered at both hospital and base DRG levels. Unit of observation is hospital/base DRG/quarter. Sample is base DRGs with multiple severity subclasses, from Q1:2006-Q4:2009. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects.

The difference between the MED and SURG coefficients in the post-reform period suggests relatively lower coding costs post-reform at EMR hospitals for MED base DRGs compared to SURG base DRGs. A plausible cause is that the complete documentation of secondary conditions, which generates the top codes, is more integral to care for medical patients than for surgical patients. Typically, medical admissions consist of a whole series of sequential interventions including medications, imaging procedures, and laboratory testing. Secondary conditions—especially those justifying a CC or MCC—play a central role in determining which of these interventions to perform. Documenting the secondary conditions is thus an essential byproduct of appropriate medical management. Moreover, in some cases, documentation of a specific condition is required to prescribe particular medications or order particular imaging procedures with an EMR.

In contrast, the primary intervention with a surgical admission is the surgery itself. Surgeons therefore place relatively more emphasis on the procedure and post-operative management, including side effects from the surgery (such as pain and/or inflammation at the incision site, lack of mobility, constipation, etc.), than on CCs or MCCs. While secondary conditions can complicate the surgery and post-operative management, they are not as central to the typical surgical admission.

The negative SURG coefficients in the post-reform period further imply that surgeons face a relatively *higher* cost of documenting diagnoses completely post-reform at EMR hospitals than at the comparison hospitals. This is likely caused by EMRs having a higher fixed cost but lower marginal cost for optimal use. Surgeons, who have less intensive interactions with EMRs and in particular, with documenting secondary diagnoses in EMRs, may not bear the fixed cost of optimal use. Unlike medical physicians, surgeons can also increase revenues by performing more surgeries, rather than billing more top codes for existing patients. They would then face a higher average cost from reporting diagnoses completely at EMR hospitals post-reform while medical physicians face a lower average cost. This would then result in EMR hospitals reporting relatively fewer top codes post-reform for SURG compared to MED.

5.5 Economic magnitudes of complete coding

Having found that current hospital reimbursement incentives are driven more by variation in the costs of complete coding than by upcoding, we seek to quantify the economic magnitude of this effect. From column 1 of Table A4, early adopters experienced a 2.67 percentage point increase in top coded medical patients in the post-reform period. On average, there are 3.13 million patients per year with the DRGs considered in our paper, about 1.83 million of whom are medical patients, and about 1.13 million of whom are medical patients admitted to early EMR hospitals. The average spread of the DRGs on which we focus is 0.82 and the average DRG price is $6,349 for an admission with weight 1. Therefore, the in-sample extra revenue paid by CMS due to complete coding from the 2007 reform by early EMR adopters is $157 million (=1.13 million × $6,349 × 0.82 × 2.67%) per year for the U.S.

Our sample accounts for about 22% of the Medicare inpatient population. Considering the fact that almost 89% of patients are in DRGs with multiple subclasses, we expect that the costs for early EMR adopters would amount to $634 million[43] when extrapolating to all DRGs. Moreover, given that almost all hospitals have adopted EMRs by 2020, when considering the full Medicare sample, the costs for all hospitals would amount to $1.03 billion[44] per year, which is 0.84 percent of total Medicare hospital claims costs.[45]

Finally, Medicare accounts for about 30% of total spending on hospital care. Many private insurers have DRG-based contracts with hospitals (Gowrisankaran et al., 2014) and generally follow Medicare billing practices. If all hospitals were reimbursed on a DRG basis, the impact of a change to MS-DRGs on extra charges due to EMRs facilitating complete coding would translate into approximately $3.4 billion in annual billed costs.

[43] [$157 million / 22% (percent of Medicare inpatient population in sample)] × 89%.

[44] Assuming that all hospitals behaved like early adopters in 2020, we extrapolate from the medical patients admitted to early adopters in our sample to medical patients admitted to all hospitals: ($634 million / 1.13 million) × $1.83 million.

[45] We obtain the total expenses on paying Medicare hospital claims, $121.73 billion, by calculating the total Medicare payments in our data across years 2008–2010.

The $3.4 billion number is likely conservative for three reasons. First, the number assumes that there were no costs of complete coding in the pre-reform period. Second, because we consider only the difference in coding between non-EMR and EMR hospitals for MED patients, it does not incorporate the revenues from complete coding for non-EMR hospitals. Finally, given the move to complex DRG-based payment systems across many countries, the worldwide impact of complete coding is likely much larger.

Overall, our takeaway is that the extra revenues from complete coding are a small but significant fraction of one of the largest sectors of the economy and hence, substantial in magnitude. This further suggests that the distortions in incentives caused by the costs of complete coding may also be substantial.

6 Conclusion

Over the past decades, U.S. Medicare and healthcare systems in many countries have substantially increased coding complexity. While prior studies have found that providers document more conditions when revenues are higher, the role of coding costs is less understood.

We develop a simple model of upcoding and costly coding, with the former emphasizing the role of revenue in coding and the latter focusing on the role of coding costs. We first estimate how financial incentives to report secondary diagnoses affect the fraction of patients reported as having these diagnoses. Prior to the 2007 Medicare payment reform, hospitals report more top codes when the additional revenues from reporting them increase. This finding is consistent with upcoding. Post-reform, this effect is no longer present. The change may be due to the payment reform making the coding costs more important relative to the revenues from reporting secondary diagnoses.

We test for costly coding using variation across EMR and non-EMR hospitals generated by the 2007 payment reform. We find that EMR hospitals report more top codes following the reform, relative to non-EMR adopters. This finding is consistent with coding being costly post-reform and EMRs lowering coding costs.

Using the 2008 penalization of never events, we are able to separate the costly coding mechanism from the upcoding mechanism. With never events, greater specificity of coding the present-on-admission field leads to *lower* reimbursements. We find that EMR hospitals report relatively *more* never events following the penalization. This is inconsistent with EMRs leading to upcoding but is consistent with EMRs reducing the hassle costs of coding.

Last, we examine whether EMRs affect hospital coding differently for medical and surgical admissions. Medical physicians have more extensive interactions with EMRs and the coding of diagnoses—that justify a top code—is more central to medical than surgical admissions. We find that EMR hospitals reported more top codes following the 2007 payment reform for medical DRGs, while the effect on surgical DRGs is negative. This is consistent with EMRs lowering the costs of coding for medical discharges but raising the costs for surgical discharges, due to the fixed costs of optimal use.

Our calculations suggest that EMR hospitals billed \$1.03 billion more to CMS and \$3.4 billion to all payors annually, from lower coding costs. The worldwide distortions from the costs of complete coding may be much higher as complex DRG-based payment systems for hospitals have become increasingly popular in many countries.

More generally, our paper provides general evidence on incentives in the health sector. On the one hand, our lack of evidence in favor of upcoding after 2007 suggests that there is a limited ability to reduce Medicare hospital expenditures through increased enforcement. On the other hand, our finding that coding is costly suggests that compliance with Medicare billing requirements may be creating distorting incentives in this market. As reimbursements systems move to increase payments for high-quality care, a hidden cost of this move may be increased costs of coding. Policy makers may consider incorporating the costs of coding in their reimbursement formulae to encourage proper documentation among providers.

References

Agha, L. (2014). The effects of health information technology on the costs and quality of medical care. *Journal of Health Economics*, 34:19–30.

American Diabetes Association (2015). Classification and diagnosis of diabetes, *Diabetes care*, 38(Supplement 1):S8–S16.

Borusyak, K. and Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.

Brown, J., Duggan, M., Kuziemko, I., and Woolston, W. (2014). How does risk selection respond to risk adjustment? New evidence from the Medicare Advantage program. *American Economic Review*, 104(10):3335–64.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2012). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*.

Clemens, J. and Gottlieb, J. D. (2017). In the shadow of a giant: Medicare's influence on private physician payments. *Journal of Political Economy*, 125(1):1–39.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review*, pages 1525–1547.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Donaldson, M. S., Corrigan, J. M., Kohn, L. T., et al. (2000). *To err is human: Building a safer health system*, volume 6. National Academies Press.

Dranove, D., Forman, C., Goldfarb, A., and Greenstein, S. (2014). The trillion dollar conundrum: Complementarities and health information technology. *American Economic Journal: Economic Policy*, 6(4):239–270.

Dunn, A., Gottlieb, J. D., Shapiro, A., Sonnenstuhl, D. J., and Tebaldi, P. (2021). A denial a day keeps the doctor away. *Working paper*.

Fässler, M., Wild, V., Clarinval, C., Tschopp, A., Faehnrich, J. A., and Biller-Andorno, N. (2015). Impact of the DRG-based reimbursement system on patient care and professional practise: Perspectives of Swiss hospital physicians. *Swiss medical weekly*, 145(0708).

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

Gawande, A. (2018). Why doctors hate their computers. *New Yorker*.

Geruso, M. and Layton, T. J. (2018). Upcoding or selection: Evidence from Medicare on squishy risk adjustment. *Journal of Political Economy, forthcoming*.

Gowrisankaran, G., Lucarelli, C., Schmidt-Dengler, P., and Town, R. (2018). Can amputation save the hospital? The impact of the Medicare Rural Flexibility Program on demand and welfare. *Journal of health economics*, 58:110–122.

Gowrisankaran, G., Nevo, A., and Town, R. (2014). Mergers when prices are negotiated: Evidence from the hospital industry. *The American Economic Review*, 105(1):172–203.

Gross, T., Sacarny, A., Shi, M., and Silver, D. (2021). Regulated revenues and hospital behavior: Evidence from a medicare overhaul. Technical report, National Bureau of Economic Research.

Hopfe, M., Stucki, G., Bickenbach, J. E., and Prodinger, B. (2018). Accounting for what matters to patients in the G-DRG System: A stakeholder's perspective on integrating functioning information. *Health services insights*, 11:1–10.

Ibrahim, A. M., Dimick, J. B., Sinha, S. S., Hollingsworth, J. M., Nuliyalu, U., and Ryan, A. M. (2018). Association of coded severity with readmission reduction after the hospital readmissions reduction program. *JAMA Internal Medicine*, 178(2):290–292.

Jha, A., DesRoches, C., Campbell, E., Donelan, K., Rao, S., Ferris, T., Shields, A., Rosenbaum, S., and Blumenthal, D. (2009). Use of electronic health records in us hospitals. *New England Journal of Medicine*, 360(16):1628–1638.

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

Kahn, C. N., Ault, T., Potetz, L., Walke, T., Chambers, J. H., and Burch, S. (2015). Assessing Medicare's hospital pay-for-performance programs and whether they are achieving their goals. *Health Affairs*, 34(8):1281–1288.

Lee, J., McCullough, J. S., and Town, R. J. (2013). The impact of health information technology on hospital productivity. *The RAND Journal of Economics*, 44(3):545–568.

Levinson, D. R. (2010). Adverse events in hospitals: National incidence among Medicare beneficiaries. *Department of Health and Human Services, Office of the Inspector General.*

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

Makary, M. A. and Daniel, M. (2016). Medical error–the third leading cause of death in the US. *BMJ*, 353:i2139.

Mathauer, I. and Wittenbecher, F. (2013). Hospital payment systems based on diagnosis-related groups: Experiences in low-and middle-income countries. *Bulletin of the World Health Organization*, 91:746–756A.

McCullough, J., Casey, M., and Moscovice, I. (2010). Critical Access Hospitals and meaningful use of health information technology. *University of Minnesota, Team FM.*

McCullough, J. S., Parente, S. T., and Town, R. (2016). Health information technology and patient outcomes: The role of information and labor coordination. *The RAND Journal of Economics*, 47(1):207–236.

Miller, A. and Tucker, C. (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science*, 55(7):1077–1093.

Ody, C., Msall, L., Dafny, L. S., Grabowski, D. C., and Cutler, D. M. (2019). Decreases in readmissions credited to Medicare's program to reduce hospital readmissions have been overstated. *Health Affairs*, 38(1):36–43.

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.

Office of the Federal Register and National Archives and Records Service (2008). Medicare program; changes to the hospital inpatient prospective payment systems and fiscal year 2009 rates. *Federal Register, (Tuesday, August 19, 2008)*, 73(161):48433–49083.

Or, Z. (2014). Implementation of DRG payment in France: Issues and recent developments. *Health policy*, 117(2):146–150.

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

Saint, S., Meddings, J. A., Calfee, D., Kowalski, C. P., and Krein, S. L. (2009). Catheter–associated urinary tract infection and the Medicare rule changes. *Annals of Internal Medicine*, 150(12):877–884.

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

Silverman, E. and Skinner, J. (2004). Medicare upcoding and hospital ownership. *Journal of Health Economics*, 23(2):369–389.

Steenhuis, S., Struijs, J., Koolman, X., Ket, J., and Van der Hijden, E. (2020). Unraveling the complexity in the design and implementation of bundled payments: A scoping review of key elements from a payer's perspective. *The Milbank Quarterly*, 98(1):197–222.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, 99(1):1–10.

Yeatts, J. P. and Sangvai, D. (2016). HCC coding, risk adjustment, and physician income: What you need to know. *Family Practice Management*, 23(5):24–27.

Appendix A

Table A1: Extra top codes with spread, by hospital types and EMR adoption status

	Dependent variable: Percent top codes within a base DRG			
	EMR adopters		EMR non-adopters	
	Pre-reform	Post-reform	Pre-reform	Post-reform
Overall				
Spread	.687***	-3.22**	.835***	-1.72
	(.112)	(1.46)	(.233)	(2.44)
By profit status				
Spread	.977***	-1.17	.869**	-3.03***
	(.218)	(1.53)	(.362)	(1.15)
ForProfit×Spread	-.332*	-2.82***	.222	3.28
	(.195)	(.901)	(.718)	(2.35)
NotForProfit×Spread	-.328	-2.22***	-.149	-1.45
	(.205)	(.828)	(.425)	(1.92)
By financial health status				
Spread	.749***	-3.49**	.71**	-1.01
	(.12)	(1.44)	(.29)	(2.87)
FinanciallyDistressed×Spread	-.0721	.00986	.264	-1.45
	(.152)	(.626)	(.48)	(3.66)
FinanciallyHealthy×Spread	-.0921	.684	.412	1.67
	(.134)	(.724)	(.53)	(2.64)

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percentage point form. Pre-reform sample is base DRGs with two severity subclasses from Q1:2005-Q3:2007, and post-reform sample is from Q4:2007-Q4:2009. For the analysis by profit status, the omitted category is public hospitals. For the analysis by financial health status, the omitted category is hospitals whose debt-asset ratio is above 25 and below 75 percentiles. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, profit/financial health status interacting with quarter dummies, quarter dummies, and hospital/base DRG fixed effects. Standard errors are clustered at both hospital and base DRG levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Extra top codes with early EMRs (Specification 2)

	Dependent variable: Percent top codes within a base DRG	
	Coefficient	S.E.
EMR×Quarter 1	.32	(.479)
EMR×Quarter 2	-.301	(.41)
EMR×Quarter 3	-.117	(.484)
EMR×Quarter 4 (Q1:2007)	-.547	(.522)
EMR×Quarter 5	-.352	(.685)
EMR×Quarter 6	-.517	(.551)
EMR×Quarter 7	2.06*	(1.21)
EMR×Quarter 8 (Q1:2008)	1.79	(1.21)
EMR×Quarter 9	1.92	(1.18)
EMR×Quarter 10	1.31	(1.19)
EMR×Quarter 11	2.62**	(1.14)
EMR×Quarter 12 (Q1:2009)	1.9	(1.3)
EMR×Quarter 13	1.62	(1.21)
EMR×Quarter 14	2.2*	(1.17)
EMR×Quarter 15	2.52**	(1.15)
Quarter 1	-.656	(.429)
Quarter 2	.0562	(.423)
Quarter 3	-.133	(.467)
Quarter 4 (Q1:2007)	.574	(.412)
Quarter 5	.373	(.604)
Quarter 6	.206	(.52)
Quarter 7	-49.1***	(3.48)
Quarter 8 (Q1:2008)	-47.3***	(3.49)
Quarter 9	-47.7***	(3.49)
Quarter 10	-47.1***	(3.54)
Quarter 11	-47***	(3.5)
Quarter 12 (Q1:2009)	-45.3***	(3.59)
Quarter 13	-45***	(3.62)
Quarter 14	-45.6***	(3.48)
Quarter 15	-44.9***	(3.49)
N	1,271,517	
p-value for joint significance of post-reform EMR coefficients	.00389	

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percentage point form. Sample is base DRGs with multiple severity subclasses, from Q1:2006-Q4:2009. Standard errors are clustered at both hospital and base DRG levels. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, and hospital/base DRG fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Extra never events vs. inconclusive POA with early EMRs (Specification 3)

	Dependent variable: Percent never events within a base DRG		Dependent variable: Percent inconclusive POAs within a base DRG	
	Coefficient	S.E.	Coefficient	S.E.
EMR×Quarter 1	.0000106	(.0147)	-.202*	(.116)
EMR×Quarter 2	.0616*	(.0348)	-.184	(.145)
EMR×Quarter 3	.0556***	(.0209)	-.154	(.23)
EMR×Quarter 4 (Q1:2009)	.0551**	(.0256)	-.114	(.227)
EMR×Quarter 5	.041	(.0276)	-.125	(.229)
EMR×Quarter 6	.0633**	(.0286)	-.137	(.228)
EMR×Quarter 7	.0873***	(.0275)	-.127	(.227)
Quarter 1	.0296**	(.0148)	-.0168	(.115)
Quarter 2	.226***	(.0406)	-2.35***	(.294)
Quarter 3	.146***	(.0299)	-4.52***	(.525)
Quarter 4 (Q1:2009)	.194***	(.0351)	-4.54***	(.522)
Quarter 5	.184***	(.0326)	-4.54***	(.525)
Quarter 6	.214***	(.0378)	-4.48***	(.521)
Quarter 7	.228***	(.0391)	-4.47***	(.52)
N	1,928,450		1,928,450	
p-value for joint significance of all EMR coefficients	.00611		.304	

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percentage point form. Sample is all base DRGs, from Q1:2008-Q4:2009. Standard errors are clustered at both hospital and base DRG levels. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, and hospital/base DRG fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Extra top codes with early EMRs, by MED and SURG (Specification 4)

	Dependent variable: Percent top codes within a base DRG			
	MED		SURG	
	Coefficient	S.E.	Coefficient	S.E.
EMR×Quarter 1	.843**	(.405)	-.696	(.941)
EMR×Quarter 2	.0282	(.431)	-1.08	(.665)
EMR×Quarter 3	.139	(.443)	-.673	(.958)
EMR×Quarter 4 (Q1:2007)	.317	(.345)	-1.98**	(.973)
EMR×Quarter 5	-.245	(.511)	-.556	(1.5)
EMR×Quarter 6	-.301	(.463)	-1	(1.13)
EMR×Quarter 7	2.52**	(1.12)	-1.67	(1.37)
EMR×Quarter 8 (Q1:2008)	2.63**	(1.1)	-2.4	(1.47)
EMR×Quarter 9	2.82***	(.829)	-2.5*	(1.41)
EMR×Quarter 10	2.16**	(.976)	-3.04**	(1.2)
EMR×Quarter 11	3.35***	(1.03)	-1.37	(1.23)
EMR×Quarter 12 (Q1:2009)	2.58**	(1.2)	-1.81	(1.5)
EMR×Quarter 13	2.03*	(1.09)	-1.68	(1.33)
EMR×Quarter 14	2.64**	(1.09)	-1.12	(1.19)
EMR×Quarter 15	3.27***	(1.12)	-1.29	(1.11)
Quarter 1	-1.09**	(.421)	.24	(.791)
Quarter 2	-.306	(.501)	.868	(.614)
Quarter 3	-.369	(.428)	.405	(.964)
Quarter 4 (Q1:2007)	.00965	(.337)	1.66**	(.809)
Quarter 5	.39	(.511)	.479	(1.31)
Quarter 6	.112	(.468)	.58	(1.06)
Quarter 7	-55.8***	(3.93)	-36.7***	(3.46)
Quarter 8 (Q1:2008)	-54.1***	(3.96)	-34.9***	(3.67)
Quarter 9	-54.8***	(3.76)	-34.7***	(3.75)
Quarter 10	-54.3***	(3.81)	-34.1***	(3.7)
Quarter 11	-53.6***	(3.94)	-34.9***	(3.83)
Quarter 12 (Q1:2009)	-51.6***	(4.26)	-33.9***	(3.64)
Quarter 13	-51.2***	(4.24)	-33.8***	(3.88)
Quarter 14	-51.7***	(4.08)	-34.5***	(3.68)
Quarter 15	-51.1***	(4.03)	-33.5***	(3.74)
N	648,795		622,722	
p-value for joint significance of post-reform EMR coefficients	.00365		.00233	

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percentage point form. Sample is base DRGs with multiple severity subclasses from Q1:2006-Q4:2009. Standard errors are clustered at both hospital and base DRG levels. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, and hospital/base DRG fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.