How do Hospitals Respond to Payment Incentives?

Gautam Gowrisankaran, PhD* Keith A. Joiner, MD, MPH† Jianjing Lin, PhD‡

November 2, 2019

Abstract

A literature has found that medical providers inflate bills and report more conditions given financial incentives. We evaluate whether Medicare reimbursement incentives are driven more by bill inflation or coding costs. Increases in reimbursements from reporting more diagnoses led to more high bill codes before a 2007 payment reform only. We test for costly coding by comparing electronic medical records (EMR) hospitals to others. EMR adopters reported relatively more diagnoses when coding is costly—even when that lowered revenues—contrary to bill inflation but consistent with costly coding. Reducing coding costs may increase inpatient Medicare costs by $1.04 billion annually.

JEL Codes: I11, I13, H51, O33
Keywords: bill inflation, payment reforms, Medicare, electronic medical records

We have received helpful comments from Sebastian Fleitas, Mike Geruso, Josh Gottlieb, Ben Handel, 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?"

*University of Arizona, HEC Montreal, and NBER
†University of Arizona
‡Rensselaer Polytechnic Institute

1 Introduction

Over the past several decades, the complexity of reimbursement mechanisms for healthcare 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, hospital readmission reduction programs, and hospital value-based purchasing programs, all with the intention of reimbursing quality over quantity.[1] These payment innovations have enormously increased the costs to providers of documenting and coding patient visits accurately. Over the past ten years, in the U.S., membership in the American Academy of Professional Coders has more than doubled,[2] while in Australia, the demand for clinical coders has risen 62% (Healthcare Australia, 2017).

Starting more recently, most hospitals in the U.S. have adopted electronic medical records (EMRs), spurred in part by the HITECH Act of 2009. EMRs facilitate patient quality and safety (Parente and McCullough, 2009; Miller and Tucker, 2011; McCullough et al., 2016; Freedman et al., 2018), improve productivity (Lee et al., 2013), and can reduce costs (Dranove et al., 2014). Yet, a fundamental reason why hospitals adopt EMRs is to optimize reimbursements.[3] EMRs can facilitate the task of capturing the hospital course 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.

A recent literature shows that medical providers document diagnoses more thoroughly when they have a financial incentive to do so (Brown et al., 2014; Geruso and Layton, 2018). However, an important issue is whether this difference in reporting is more due to

[1] 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).

[2] 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).

[3] `https://www.revenuexl.com/does-emr-ehr-increase-revenues`

bill inflation—where providers report inaccurate or inadequately documented diagnoses in order to add revenues—or to *costly coding*—where providers report diagnoses incompletely when coding costs are high. In the presence of bill inflation, providers inappropriately code conditions that are not present when there are incentives to do so. In the case of costly coding, providers appropriately code conditions that are present but were previously too costly to code completely in the absence of tools or incentives. A number of papers provide evidence of bill inflation (Silverman and Skinner, 2004; Dafny, 2005; Jürges and Köberlein, 2015) by hospitals and evidence that EMRs lead to bill inflation (Li, 2014; Ganju et al., 2018),[4] while others find evidence of costly coding (Ibrahim et al., 2018; Sacarny, 2018; Ody et al., 2019) facilitated by EMRs (Edwardson et al., 2017).

The purpose of this paper is to test for and separate the bill inflation and costly coding models for Medicare inpatient admissions. We use changes to Medicare payments and hospital EMR adoption to separate these two explanations. While it is accepted that hospitals aim to maximize their revenues from Medicare, net of the costs of complete coding and the negative financial and reputational consequences from bill inflation, the balance between these two effects is less understood. Is bill inflation widespread or have updated payment mechanisms reduced bill inflation and instead resulted in hospitals leaving revenue on the table due to the high costs of coding completely?

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 bill inflation is important, hospitals as well as the Center for Medicare and Medicaid Services (CMS) should increase billing enforcement and/or penalties for inaccurate reporting. If costly coding is important, then this is likely to generate more distortions and costs over time as payors continue to increase reimbursement complexity in order to reward quality over quantity. Many countries have increased the complexity of their hospital reimbursement systems by

[4] A number of hospitals have been found guilty of fraudulent Medicare billing under the False Claims Act (`https://oig.hhs.gov/fraud/enforcement/criminal/`), implying that some bill inflation does exist.

moving to DRG-based payment systems (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).

Our study uses a 100% sample of Medicare inpatient claims from 2005-9. We base our empirical specifications on the differences in reimbursement within a *base DRG*—which records the patient's primary diagnosis or procedure—and the presence of hospital EMRs. To explain the concept of a base DRG further, Medicare reimburses hospitals a single amount for each admission. The amount is based on the patient's diagnosis related group (DRG), which is a set of diagnoses or procedures with similar treatment costs. Specifically, each DRG has an associated weight, and the payment to the hospital is proportional to this weight. Medicare groups DRGs into base DRGs based on the patient's *primary* diagnosis or procedure. Medicare then issues lists of secondary diagnoses that, when present, allow coding the admission as a "with complications/comorbid conditions (CC)" or a "with major complications/comorbid conditions (MCC)" DRG, both of which have higher weights and hence higher reimbursements. Documenting CCs and MCCs requires more specificity than documenting the base diagnosis or procedure used in the bottom bill code, which may incur costs.

We develop a simple model of bill inflation and costly coding that we then take to data. Our model conditions on a patient at a given hospital and with a given base DRG, who can be coded into one of two bill codes. We distinguish between the bottom and top bill codes, where the bottom bill code requires less specificity.

Our bill inflation model regards the subset of patients who *do not* qualify for the top bill code. Hospitals may choose to report the top bill code for some of these patients, thereby inflating their bills.[5] The decision to inflate bills for these patients is driven by the extra revenue that the hospital would gain from bill inflation net of the extra costs. The extra revenue is a function of the *spread*, or difference in DRG weights between the top and bottom codes. The costs are largely driven by expected penalties and stigma from being discovered to have inflated bills. Changes in the spread over time for a given base DRG provide identifying variation for this model.

[5] We view bill inflation as any coding that would be reversed by a sufficiently thorough audit.

Our costly coding model regards the remainder of the patients, who *do* qualify for the top bill code. The decision to correctly select the top bill code for these patients is again driven by the extra revenue that the hospital would gain from coding completely net of the extra costs. This model emphasizes that coding is complex, requires substantial effort, and that in the absence of financial incentives and tools such as EMRs, hospitals may only capture a part of the allowable charges. Thus, hospitals may code incompletely for some of these patients if the documentation costs are higher than the revenue gain.[6] 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 distinguish between bill inflation and costly coding. All our main specifications have the same unit of observation, the hospital / base DRG / quarter, and all include hospital / base DRG fixed effects, so that our specifications test for differences within a hospital and base DRG. Our dependent variables all generally indicate the percent of discharges with a top bill code within each cell.

First, we use variation in the spread between bottom and top bill codes across time within base DRGs. Specifically, CMS changes the weights of the DRGs within a base DRG on an annual basis. Similarly to Silverman and Skinner (2004) and Dafny (2005), this variation provides differing financial incentives to report top bill codes for a given hospital and base DRG across time. We evaluate whether a higher spread leads to more reported top bill codes, controlling for hospital / base DRG fixed effects and time fixed effects. A Medicare payment reform in Q4:2007 increased billing complexity. We estimate this specification separately before and after the reform. We find evidence that prior to Q4:2007, increases in spread led to increases in the top bill code, with a unit increase in spread corresponding to an increase of 0.6 percentage points in reported top codes. Following Q4:2007, increases in spread had no significant effect (and indeed the point estimate is negative). This suggests that changes in revenues were an important driver of coding prior to the payment reform but not afterwards.

[6] This model includes the possibility that hospitals may perform more tests and procedures on patients in order to better document secondary conditions.

The difference in results may be due to the increased complexity of billing post-reform leading coding costs to be more important. While the pre-reform positive impact of spread provides evidence that hospitals responded to revenues in coding, it is consistent with both our bill inflation and costly coding models, since under both models, hospitals may invest more in coding top bill codes when the revenues for doing so increase.

Second, we directly use the Q4:2007 Medicare payment reform. Before the reform, many chronic diseases qualified for a CC, while after the reform, the acute manifestation of the chronic disease or a new acute disease were often necessary to obtain a CC or MCC. Since EMRs facilitate the specificity of coding medical information and since the reform resulted in this specific information being necessary to obtain a top bill code, we hypothesize that hospitals with EMRs had relatively lower costs of complete coding in the post-reform period to the pre-reform period. We estimate whether hospitals with EMRs had an increase in top bill codes after the reform relative to before the reform.[7] We find that EMR adopters had 1.64 percentage points more top bill codes post-reform than pre-reform, relative to non-EMR adopters. This is consistent with complete coding being costly and with EMRs lowering this cost. While a positive impact of EMR hospitals on top bill codes post-reform provides evidence that hospitals respond to the coding costs in their billing decisions, it is still possible that this response is part of a bill inflation strategy, i.e., that those additional patients did not quality for a top bill code.

Third, to address this, we further consider the above result could be part of a bill inflation strategy, using *never events*. In Q4:2008, CMS started penalizing certain preventable complications that are specifically acquired during the course of hospital treatment; these are colloquially known as never events, since they are preventable with appropriate care. An example of a never event is a catheter-associated urinary tract infection acquired in-hospital.[8] For never events, the top bill code—which is the code with more specificity—actually leads

[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; Abraham and Sun, 2018; de Chaisemartin and D'Haultfœuille, 2019).

[8] In addition to the lower reimbursement from never events, there was a general fear that the presence of a never event would lead to a greater risk of patient litigation and legal judgment. See, for instance `http://www.neildymott.com/never-events-hospital-aquired-conditions-hacs`.

to lower reimbursements. 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. We find that EMR adopters responded to the penalization of never events by reporting more never events in the post-reform period than the pre-reform period, relative to non-EMR adopters. This was primarily because they were better at documenting whether a condition was present on admission. In other words, EMR hospitals reported that more of these conditions were acquired in-hospital, justifying the CMS penalty. Since EMR hospitals responded to never events by reporting more top bill codes, this implies that their costs from reporting these codes are lower; it cannot be part of a bill inflation strategy because it decreases revenues.

Finally, we separately consider the impact of the 2007 payment reform on medical and surgical DRGs, in order to further substantiate our findings that costly coding is the dominant incentive post-reform. It is well appreciated that surgeons interact with EMR systems less than medical physicians (see, e.g., Gawande, 2018). Moreover, the 2007 payment reform created financial incentives relating 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 top bill codes are more impacted by this reform for medical than for surgical DRGs. We find that top bill codes for EMR adopters increase only for medical DRGs, with EMR adopters reporting 2.21 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 2.03 percentage points fewer top codes. This result is 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.

The magnitude of our results suggests that the 2007 reform would lead to $1.04 billion annually in extra Medicare hospital claims costs if all hospitals acted liked 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 inpatient admissions, known as the 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] By moving hospitals away from cost-based reimbursements, PPS aimed to reward efficiency and lower expenditure growth.

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

[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.

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.[12]

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). CMS reformed the reimbursement system in 2007 in order to better align payments with the resources used by a hospital. A realignment was deemed necessary because many conditions that previously needed costly and lengthy hospitalizations could then be managed in an outpatient 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]

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

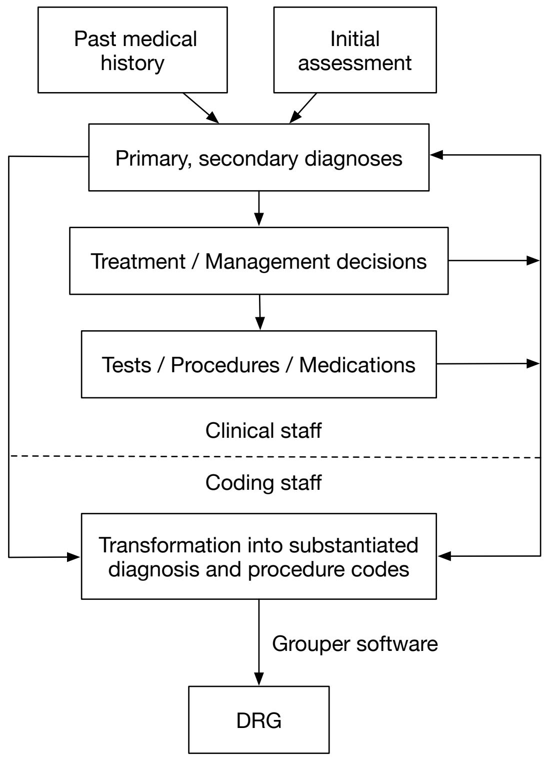
[12]This is an important distinction, since medical physicians are more likely than surgeons to recognize and document the CCs and MCCs, given their role in the healthcare production process.

[13]There are exceptions to this rule including, most prominently, for heart failure. 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.

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.

Figure 1: An Overview of the Coding Process



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 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 to perform 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 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

[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.

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.

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.

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

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.

In 2007, CMS initially selected eight of these events for non-payment during hospitalization, arguably the first time that Medicare withheld payment 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 fol-

[18]1A 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).

[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.

lowing admission to have been POA, based on additional information collected during the hospitalization. Improper coding of a never event 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 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 framework and testable hypotheses

3.1 Model

We develop a model to characterize a hospital's decision regarding the assignment of top bill codes 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 bill codes correspond to more specific diagnosis, which are the high severity subclass DRG or never events, depending on the context. We take as given the set of patients at each hospital and base DRG.[23] 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}. Assume that some fraction of patients α_{jdt} qualify for a top bill code. We allow α_{jdt} to vary across hospitals, base DRGs, and time, to account for variation across hospitals and base DRGs in patient severity and variation across time in coding requirements. Denote the

[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.

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

[23] Our model considers coding decisions and not treatment decisions. Hence, we do not consider hospitals shifting patients across base DRGs based on financial incentives, which might occur if, for instance, hospitals performed a different surgery if the reimbursement for that procedure increased relative to another procedure.

set of qualifying patients as $\mathcal{N}^H_{jdt}$ and the set of non-qualifying patients as $\mathcal{N}^L_{jdt}$.

We model two potential actions by hospitals: bill inflation and costly coding. We start by describing our model of bill inflation, which pertains to the coding of high severity subclass DRGs. For this model, a hospital might assign some of its $(1-\alpha_{jdt})\,N_{jdt}$ patients who *do not* qualify for a top bill code to a top bill code, in order to increase its revenues. Let u^{BI}_{ijdt} denote the net utility from coding one such patient i to the top bill code. u^{BI}_{ijdt} will be increasing in the revenue gained from obtaining the top bill code. It will also be decreasing in the cost of bill inflation, which might include penalties and stigma when caught. A hospital will report a top bill code if the net utility of this action is positive. Thus, we write

$$D^{BI}_{ijdt} = \mathbb{1}\left\{u^{BI}_{ijdt} \geq 0\right\},$$

where D^{BI}_{ijdt} is an indicator for bill inflation. We let the expectation of D^{BI}_{ijdt} take the following form:

$$E\left[D^{BI}_{ijdt}\right] = f^{BI}\left(Rev_{jdt}\right) - g^{BI}\left(Cost^{BI}_{ijdt}\right). \tag{1}$$

In (1), $f^{BI}(\cdot)$ is a function of Rev_{jdt}, which denotes the extra revenue from a top bill code for patients within the same base DRG, hospital, and time period; $g^{BI}(\cdot)$ is a function of $Cost^{BI}_{ijdt}$, which denotes the extra cost of bill inflation for a patient belonging to a given base DRG at hospital j at period t. Unlike Rev_{jdt}, $Cost^{BI}_{ijdt}$ might vary across patients. We aggregate equation (1) across patients within a hospital, base DRG, and time period:

$$Inflate_{jdt} = \sum_{i\in\mathcal{N}^L_{jdt}} \left[f^{BI}\left(Rev_{jdt}\right) - g^{BI}\left(Cost^{BI}_{ijdt}\right)\right]. \tag{2}$$

where $Inflate_{jdt}$ is the total expected number of patients who *do not* qualify for the top bill code but are assigned to the top bill code.

Much of the literature on bill inflation has focused on the revenue side (Silverman and Skinner, 2004; Dafny, 2005; Li, 2014; Geruso and Layton, 2018), examining how changes in spread lead to changes in reported top codes. A smaller set of papers that analyzes bill inflation has considered the cost side. For instance, Ganju et al. (2018) focus on Medicare's Recovery Audit Program, which increased the penalties from bill inflation. Our model captures both sides.

We now describe our model of costly coding, which pertains to both high severity subclass DRGs and never events. Under this model, a hospital might assign some of its $\alpha_{jdt}N_{jdt}$ patients who *do* qualify for the top bill code to the low bill code, due to the cost of complete coding. Let u_{ijdt}^{CC} denote the net utility from coding one such patient i to the top bill code. Similarly to $u_{i,j,d,t}^{BI}$, $u_{i,j,d,t}^{CC}$ is increasing in the revenue gained from obtaining the top bill code. Note that in the case of never events, this revenue is negative. It will also be decreasing in the cost of complete coding, which includes the time and effort that hospitals spend to document completely.[24] A hospital will completely document a patient's diagnoses and conditions for the top bill code if the net utility is positive. We write

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

where D_{ijdt}^{CC} is an indicator for complete coding. As with bill inflation, we let the expectation of D_{ijdt}^{CC} take the following form:

$$E\left[D_{ijdt}^{CC}\right] = f^{CC}\left(Rev_{jdt}\right) - g^{CC}\left(Cost_{ijdt}^{CC}\right). \tag{3}$$

The variables here are defined similarly to those in equation (1). We aggregate equation (3) across patients within a hospital, base DRG, and time period to obtain:

$$Complete_{jdt} = \sum_{i \in \mathcal{N}_{jdt}^{H}} \left[f^{CC}\left(Rev_{jdt}\right) - g^{CC}\left(Cost_{ijdt}^{CC}\right)\right], \tag{4}$$

where $Complete_{jdt}$ denotes the expected total number of patients who deserve and get assigned to the top bill code.

The costly coding model focuses on the cost side. However, even under this model, revenues can still affect the decision to report the top bill code. Thus, though the focus of

[24] Dunn et al. (2019) find both substantial monetary and time costs of coding, across a number of payors.

the costly coding model is on the cost, the two models are similar in many ways.

While we specify two potential models for generating top bill codes, our data only include the total fraction of patients with a top bill code. We define this term:

$$CodeFrac_{jdt} = \frac{1}{N_{jdt}}(Inflate_{jdt} + Complete_{jdt}) \tag{5}$$

$CodeFrac$ is composed of an aggregation of both the bill inflation and costly coding effects. Since the two explanations are similar and the data incorporate both explanations, we will need variation that has a different effect across the two models to separately identify bill inflation from costly coding. We discuss this variation below in the context of the testable implications of the model.

3.2 Testable implications of the model

Specification 1: impact of spread

We first analyze variation in *spread*—the difference in DRG weights between the top and bottom codes within a base DRG. Replacing $Inflate_{jdt}$ and $Complete_{jdt}$ in equation (5),

$$\begin{aligned} CodeFrac_{jdt} &= \frac{1}{N_{jdt}}\left[\sum_{i\in\mathcal{N}^L_{jdt}}\left(f^{BI}(Rev_{jdt}) - g^{BI}\left(Cost^{BI}_{ijdt}\right)\right) + \sum_{i\in\mathcal{N}^H_{jdt}}\left(f^{CC}\left(Rev_{jdt}\right) - g^{CC}\left(Cost^{CC}_{ijdt}\right)\right)\right] \\ &= \left[\left(1-\alpha_{jdt}\right)f^{BI}\left(Rev_{jdt}\right) + \alpha_{jdt}f^{CC}\left(Rev_{jdt}\right)\right] \\ &\quad - \frac{1}{N_{jdt}}\left[\sum_{i\in\mathcal{N}^L_{jdt}} g^{BI}\left(Cost^{BI}_{ijdt}\right) + \sum_{i\in\mathcal{N}^H_{jdt}} g^{CC}\left(Cost^{CC}_{ijdt}\right)\right]. \end{aligned} \tag{6}$$

In general, Rev_{jdt} is proportional to the spread between the top and bottom codes plus some constant. Let

$$f^M\left(Rev_{jdt}\right) - g^M\left(Cost^M_{ijdt}\right) = \gamma Spread_{dt} - Cost^M_{ijdt} + \bar{c}^M_{jdt}, \tag{7}$$

where $M = \{BI, CC\}$, denoting the type of the model; γ measures the effect from spread on extra billing; $\overline{c}_{jdt}^{M}$ includes the fixed revenue, fixed costs, and the residual in model M. Plugging (7) into (6), we obtain

$$
\begin{aligned}
CodeFrac_{jdt} &= \frac{1}{N_{jdt}}\left[\sum_{i\in\mathcal{N}_{jdt}^{L}}\left(\gamma Spread_{dt} - Cost_{ijdt}^{BI} + \overline{c}_{jdt}^{BI}\right) + \sum_{i\in\mathcal{N}_{jdt}^{H}}\left(\gamma Spread_{dt} - Cost_{ijdt}^{CC} + \overline{c}_{jdt}^{CC}\right)\right] \\
&= \gamma Spread_{dt} - \frac{1}{N_{jdt}}\left(\sum_{i\in\mathcal{N}_{jdt}^{L}} Cost_{ijdt}^{BI}\right) - \frac{1}{N_{jdt}}\left(\sum_{i\in\mathcal{N}_{jdt}^{H}} Cost_{ijdt}^{CC}\right) \\
&+ \left(1-\alpha_{jdt}\right)\overline{c}_{jdt}^{BI} + \alpha_{jdt}\overline{c}_{jdt}^{CC} \qquad (8)
\end{aligned}
$$

We condition on the pre- and post-reform period, so that costs of the top bill code are relatively homogeneous across time. Thus, we parameterize the cost terms in (8), which is

$$
-\frac{1}{N_{jdt}}\left(\sum_{i\in\mathcal{N}_{jdt}^{L}} Cost_{ijdt}^{BI}\right) - \frac{1}{N_{jdt}}\left(\sum_{i\in\mathcal{N}_{jdt}^{H}} Cost_{ijdt}^{CC}\right) + \left(1-\alpha_{jdt}\right)\overline{c}_{jdt}^{BI} + \alpha_{jdt}\overline{c}_{jdt}^{CC},
$$

as $\overline{c}_{jd}^{H} + \overline{c}_{t}^{Q} + \delta X_{jt} + \varepsilon_{jdt}$. Here, we have decomposed costs into a hospital/DRG fixed effect and a time fixed effect. This results in:

$$
CodeFrac_{jdt} = \gamma Spread_{dt} + \overline{c}_{jd}^{H} + \overline{c}_{t}^{Q} + \delta X_{jt} + \varepsilon_{jdt}, \qquad (9)
$$

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 (9) 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 have reported top bill codes. 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.

Variation in revenues in our context will be given by the changes in the spread between the top and bottom bill codes. Our data include such variation, as $Spread_{dt}$ changes over time, both before and after the 2007 payment reform. Under the bill inflation model, hospitals that are inflating bills should move more patients to the top billing 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 cost of coding can be captured by the hospital/base DRG and time fixed effects.[25] 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 bill. 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 bill inflation and indirect effect on the cost of complete coding will be captured by the fixed effects for hospitals in those six states.[26]

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 inflate bills more than government or not-for-profit hospitals and that hospitals with a high debt-asset ratio exhibited a larger increase in percent top bill codes. She concluded these subsets of hospitals respond more actively to the upcoding incentive. Based on equation (9), we additionally interact $Spread_{dt}$ with measures of financial health—$Distressed$ (the 25% with the highest debt-asset ratio) and $FinHealthy$ (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.

[25] The effect on the documentation cost for percent top bill codes from the 2008 never event penalization is negligible, as the number of never event is small. On average, there are only 0.31% of discharges within a hospital and base DRG reporting never events.

[26] 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.

Importantly, while a positive γ in the estimating equation (9) would be consistent with bill inflation, it could also be consistent with costly coding. Specifically, hospitals that do not always completely code all top bill code patients will code more of these patients as top bill code patients if $Spread_{dt}$ increases. Thus, the two models have observably identical implications to variations in revenue, even though the causal mechanisms are very different.

Specification 2: impact of EMRs

To separate the bill inflation model from the costly coding model, we next consider variation in costs. Unlike for revenues, we do not directly observe hospital costs. However, we do observe two events that generate variation in costs across different types of hospitals. Specifically, 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 cost terms in equation (6) take the following form:

$$\frac{1}{N_{jdt}}\left[\sum_{i\in\mathcal{N}_{jdt}^{L}} g^{BI}\left(Cost_{ijdt}^{BI}\right)+\sum_{i\in\mathcal{N}_{jdt}^{H}} g^{CC}\left(Cost_{ijdt}^{CC}\right)\right]=\beta_t\, EMR_j+\bar{c}_t^{Q}, \tag{10}$$

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. As our main goal here is to estimate how EMR adopters respond to the change in the cost of coding from the reform, we assume that the variation in revenues can be captured by hospital/base DRG fixed effects, time fixed effects, and hospital characteristics. We then parameterize the revenue terms in equation (6), which are

$$\left(1-\alpha_{jdt}\right) f^{BI}\left(Rev_{jdt}\right)+\alpha_{jdt} f^{CC}\left(Rev_{jdt}\right),$$

as $\delta X_{jt}+\bar{c}_{jd}^{H}+\varepsilon_{jdt}$.[27] Thus, an implication of our costly coding model is that the percent top bill codes will be relatively higher in the post-reform period from in the pre-reform period

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

among EMR hospitals. We define a regression specification based on this idea:

$$CodeFrac_{jdt} = \beta_t\, EMR_j + \overline{c}_t^{Q} + \overline{c}_{jd}^{H} + \delta X_{jt} + \varepsilon_{jdt}, \tag{11}$$

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 bill codes following the reform.

Our unit of observation, dependent variable, controls X_{jt}, and clustering strategy are the same as in Specification 1. However, our analysis here only includes hospitals that had adopted EMRs prior to 2007 and hospitals that had not adopted EMRs by the end of 2009. We focus on these two sets of hospitals to avoid imprecise estimates due to the shifting control group over time (Borusyak and Jaravel, 2017; Abraham and Sun, 2018; de Chaisemartin and D'Haultfœuille, 2019).

While a finding of positive β_t values post-reform is most consistent with costly coding, we cannot rule out that positive coefficients on post-reform β_t values are driven by some combination of bill inflation and costly coding. Specifically, while researchers typically assume that the costs of bill inflation are the penalties, it is possible that the costs of bill inflation include the coding costs. Thus, by lowering the documentation costs, a positive impact on β_t values post-reform could also be consistent with bill inflation, if there are still coding costs.

Specification 3: impact of never events

As another source of variation in costs, we turn to never events. Never events change the costs of documentation in a similar way to the payment reform: they increase the specificity of documentation and hence increase the value of EMRs in generating a top code. Similar to (10), we allow for the change in the cost of coding from the 2008 penalization program by interacting EMRs with time dummies. As with Specification 2, we assume that revenues 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 (11) except that the dependent variable becomes the percent top bill codes of never events.

The difference between never events and the payment reform is that a finding of positive coefficients on β_t post-penalization is not consistent with bill inflation. Specifically, a positive coefficient indicates more top bill codes for EMR hospitals, in a circumstance where never events were penalized. Since a top bill code gives a lower revenue to a hospital, a hospital should never choose to increase reported never events following their penalization under the bill inflation model. However, if their costs from reporting are lower, EMR hospitals may report more never events under the costly coding model.

Specification 4: variation in EMR impact between medical and surgical DRGs

Finally, to further substantiate our analysis, we also investigate whether there is any difference in cost variation for medical versus surgical base DRGs. In this case, we allow a more flexible specification for the costs terms than in (10), which led to Specification 2. Specifically, we let the cost terms from equation (6) be:

$$\mathbb{1}\{d = MED\}(\beta_{mt}\, EMR_j + \bar{c}^{Q}_{mt}) + \mathbb{1}\{d = SURG\}(\beta_{st}\, EMR_j + \bar{c}^{Q}_{st}), \tag{12}$$

where MED and $SURG$ are indicators for the base DRG being medical and surgical respectively. In equation (12), both the base costs of coding the high bill code and the incremental costs of coding the high bill code for EMR adopters post-reform vary across MED and SURG base DRGs.

We estimate a separate specification of (11) 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 bill code will be relatively greater among medical DRGs.

In sum, from Specification 1, a positive finding on spread would indicate that hospitals respond to increases in revenues by coding more high bill codes. However, this is consistent with both the costly coding and bill inflation models. From Specification 2, a finding of an increase in high bill codes for EMR hospitals post-reform would provide support for costly coding model while not ruling out bill inflation. From Specification 3, a positive finding on never events for EMR hospitals would support the costly coding model and not the bill inflation model. From Specification 4, a difference between MED and SURG DRGs that is consistent with our priors on the relative cost of coding would lend additional support to the costly coding model.

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—$Spread_{dt}$ (the difference between the high and low DRG weights within a base DRG) and EMR adoption prior to the start of our sample interacted with quarter dummies—are mean independent from the unobservable ε_{jdt}.

For Specification 1 in equation (9), identification is from changes in the spread and relative changes in percent top bill codes within a hospital/base DRG cell. The central threats to identification would have to 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.[28]

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,[29]

[28]For the Recovery Audit Program, Medicare first identified 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. Medicare then audited a random subset of the claims identified by CERT (`https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/index.html?redirect=/Cert`). The set of claims identified by CERT did not vary with changes in the DRG spread for a given base DRG.

[29]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.

from the pre- to post-reform period. One central 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 bill 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 bill 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 bill codes by MED and SURG, can mitigate this concern by adding supporting evidence to the complete 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 bill inflation 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 bill inflation for EMR hospitals. To consider the possibility of differential changes in care, we will consider the source of the never events, whether they derive from fewer diagnoses that lead to never events or from differential coding of the present on admission field.

For Specification 4, we identify costly coding for MED relative to SURG DRGs from changes in the fraction of top codes among EMR hospitals minus changes in this fraction among non-EMR hospitals for MED DRGs relative to this difference for SURG 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, Diagnostic Related Group (DRG), and principal and secondary diagnosis and procedure codes. We drop admissions to Critical Access Hospitals (CAHs) as these hospitals receive cost-based reimbursements from Medicare, instead of prospective DRG-based payments (Gowrisankaran et al., 2018). We construct our main dependent variable, the percent of patients with documented top bill 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.[30]

[30] 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.

We merge our base data with information on DRGs from the Centers for Medicare and Medicaid Services (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 of the secondary diagnoses reported in our main analysis are on the HAC list. If so, we then define a never event if the event is not present on admission (POA). According to the final rule published in 2009,[31] 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 further categorize a never event 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. We define the admission as not POA when the field takes the values of "N" or "U" as determined by Medicare rules.[32] The dependent variable for never events is then the percent of discharges with the POA field coded as not POA among discharges with at least one HAC within a hospital/base DRG/quarter cell. We also examine a related outcome, the frequency of missing values in the POA field. We define the POA field with a missing value if the field takes the value of "1" or anything other than the specified values above (Cram et al.,

[31] 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.

[32] See https://www.resdac.org/cms-data/variables/medpar-diagnosis-present-admission-indicator-code.

2014). The dependent variable for this outcome is the percent of discharges with missing POA values among discharges with at least one HAC within a hospital/base DRG/quarter cell.

Our second main dataset provides information on EMR adoption. We merge in EMR adoption data 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 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., 2018). 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: Timeline of data for different hypotheses

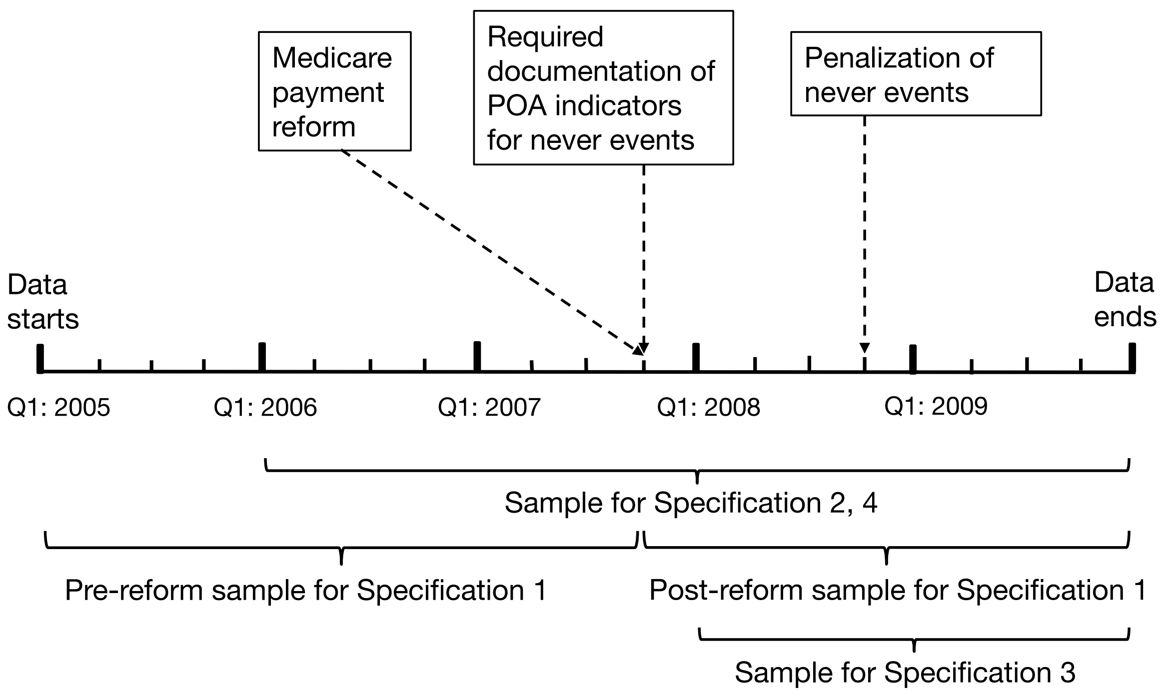


Figure 2 provides timelines for the data used in different specifications and the underlying policy changes that motivate the timelines. In Specification 1, we examine how hospitals respond to changes in revenues, ceteris paribus. This limits our data to the pre- or post-reform period separately, so as to minimize the change in the cost of coding. 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. No 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 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.[33] For base DRGs with three severity subclasses, we define $CodeFrac$ to be the fraction of patients with the top of the three severity subclasses.

For Specification 3, which considers never events, although mandatory POA reporting for the HAC list started in Q4:2007, we observe almost no POA reporting until Q1:2008. 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 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. There were more than 12 million Medicare discharges in each of the years in our data.[34] The mean age of a Medicare patient discharged from a hospital was 73 years during our sample and the mean DRG weight was rising over time, from 1.46 in 2005 to 1.57 in 2009.

Panel 2 shows the sample by fiscal year 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 bill codes drops substantially following the reform, from over 73% before the reform to 30%[35] immediately after the reform. 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

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

[34] This excludes admissions to CAH hospitals.

[35] The 40.3% in the Introduction is cited from the Federal Registry, specifying that the number is calculated using the 2006 MedPAR data based on the revised CC list, whereas the 30% here is based on the sample for Specification 2—the MedPAR data for a subset of base DRGs with two subclasses in FY2008.

Table 1: Summary statistics on patient sample

Panel 1: universe of Medicare patients in sample					
	2005	2006	2007	2008	2009
Mean age	73.0	72.9	72.8	73.0	72.8
Mean DRG weight	1.46	1.47	1.48	1.52	1.57
# discharges	13,095,851	12,821,616	12,535,313	13,793,662	14,186,766
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	77.5	73.3	73.9	30.4	32.9
# discharges	2,703,129	4,072,710	3,887,605	4,834,427	5,171,755
Panel 3: samples for Specifications 2 and 4 (matched base DRGs with multiple severity subclasses)					
	2005	2006	2007	2008	2009
% top codes	–	72.3	62.0	27.9	30.0
# discharges	–	3,135,339	2,983,042	3,352,411	3,365,488
MED					
% top codes	–	81.4	68.6	29.8	32.4
# discharges	–	1,785,891	1,735,792	1,988,372	2,011,390
SURG					
% top codes	–	60.3	52.9	25.1	26.4
# discharges	–	1,349,448	1,247,250	1,364,039	1,354,098
Panel 4: sample for Specification 3 (all base DRGs)					
	2005	2006	2007	2008	2009
% discharges w/ HACs	4.12	4.07	4.46	5.18	4.52
% never events given HAC	–	–	–	4.46	8.38
% missing POA given HAC	–	–	–	59.4	8.18

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 Specification 2. "POA" indicates present on admission and "HAC" indicates hospital-acquired conditions.

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 bill codes in 2007 is a bit lower than that in 2006, as it is a mix of the percent top bill codes before and after the reform.[36] Specification 4 also decomposes this sample by whether the base DRG is medical or surgical. The percent top bill 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 reported never events was much higher in 2009 than that in 2008, whereas the number of missing values in the POA field dropped substantially from 2008 to 2009.

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.

Table 2: Summary statistics on hospital characteristics by EMR use

	EMR adopters $\leq$2006	EMR non-adopters through 2009
Bed size	253	113
Total outpatient visits	199,292	71,498
Total admissions	11,860	4,368
FTE physicians and dentists	28	7
Total number of births	1,335	405
% teaching hospital	12.9	2.46
% Medicare discharge	44.1	48.3
% Medicaid discharge	18.9	17.8
% for-profit	21.2	40.0
% not-for-profit	68.5	38.5
% public hospitals	10.3	21.5
Debt-asset ratio	0.651	0.772
Number of hospitals	1,696	325

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

Table 3 provides summary statistics on DRG weights with the mean DRG weights in the

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

Table 3: Summary statistics on DRG weights

Variable	Obs	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
$\Delta Spread$, FY2005 to FY2006	101	-0.0173	0.182
$\Delta Spread$, FY2006 to FY2007	106	-0.0053	0.0726
$\Delta Spread$, FY2008 to FY2009	110	0.42	0.486
$\Delta Spread$, FY2009 to FY2010	110	0.0053	0.242

Note: *Spread* measures the difference between the weight in the top and bottom codes.

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 considered in Specification 1. The fiscal years used by CMS start one quarter before the calendar year. There were large changes in spread across fiscal years, with standard deviations in the change in spread of 0.0726 to 0.486 depending on the year. Thus, there is substantial variation to identify this specification.

5 Results

5.1 Specification 1: impact of spread

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 between bottom and top bill codes on the probability of top coding within a base DRG. We estimate this specification separately before and after the reform, with the percent of the top bill code 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 bill code.

Table 4 presents the coefficient for the key variable of interest: *Spread*, for both the pre-

Table 4: extra top bill codes with spread (Specification 1)

	Dependent variable: Percent top bill codes within a base DRG			
	(1)	(2)	(3)	(4)
Pre-reform	−2.59	.637***	.613***	.614***
	(4.77)	(.102)	(.102)	(.102)
Post-reform	−13.7	−2.31	−2.35	−2.36
	(8.47)	(1.58)	(1.58)	(1.58)
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and post-reform samples. Thus, 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 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 a high code increases. A unit increase in spread corresponds to an increase of 0.614 percentage points in the reporting of a top bill code. This is consistent, though not dispositive, with the finding in prior studies that provided evidence of bill inflation (Silverman and Skinner, 2004; Dafny, 2005). While documentation of top bill codes increasing with spread may imply that changes in revenues could be an important driver of coding, it cannot distinguish between the two models we proposed. Hospitals will report more top bill codes when the financial returns are greater, under both the bill inflation and costly coding model.

Turning now to the post-reform period, from Q4:2007 onwards, a larger spread no longer predicts a higher fraction of top bill codes. For this time period, the column (4) point estimate on spread is negative and not statistically significant. 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. Regardless of the cause of the change in coefficient from the pre- to the post-reform period, the post-reform evidence is inconsistent with a bill inflation story where higher revenues lead to more unjustified top codes. While reducing bill inflation is not one of the stated policy goals of the 2007 payment reform, our findings show that it is an important side effect.

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 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. 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 are not statistically significant.

To summarize, we find that hospitals responded to financial incentives in coding before the reform. This could be due to bill inflation or costly coding. Prior to the reform, there is little evidence that the impact varied by hospital type. The significant impact of 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.

Table 5: extra top bill 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	.696***	(.156)	−1.17	(1.6)
ForProfit×Spread	−.11	(.15)	−.408	(.942)
NotForProfit×Spread	−.0896	(.131)	−1.58**	(.665)
By financial health status				
Spread	.654***	(.105)	−2.57*	(1.51)
FinanciallyDistressed×Spread	−.0367	(.101)	−.186	(.511)
FinanciallyHealthy×Spread	−.066	(.0907)	1.04	(.84)

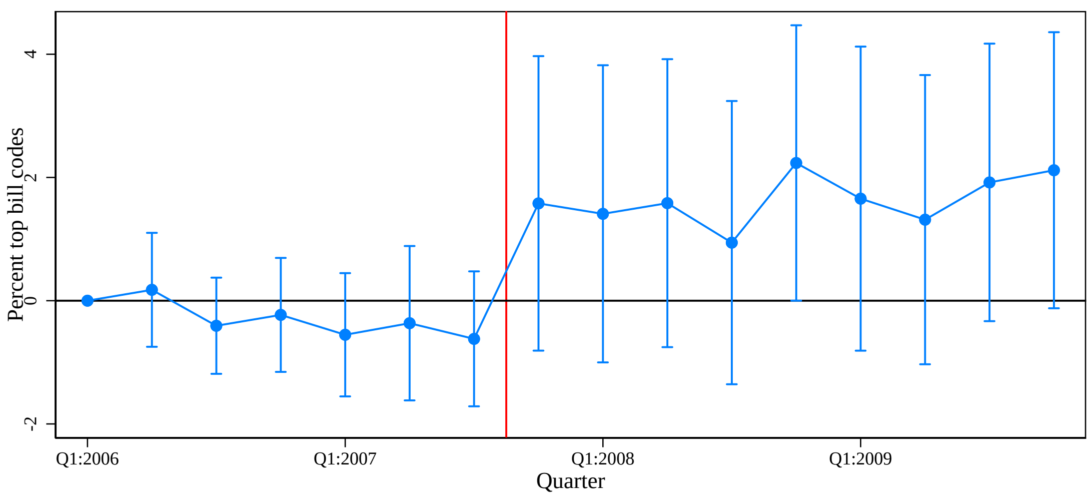
Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percentage point form. Pre-reform sample is base DRGs with two severity subclasses from Q1:2005-Q3:2007, and post-reform sample is from Q4:2007-Q4:2009. For the analysis by profit status, the omitted category is public 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, EMRs interacting with quarter dummies, 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Specification 2: impact of EMRs

Specification 2 evaluates the impact of EMRs on reported high bill 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 3 presents the coefficients on interactions between quarter and EMR adoption, as well as their 95% confidence intervals. Table A1 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.

Figure 3: extra top bill codes with early EMRs

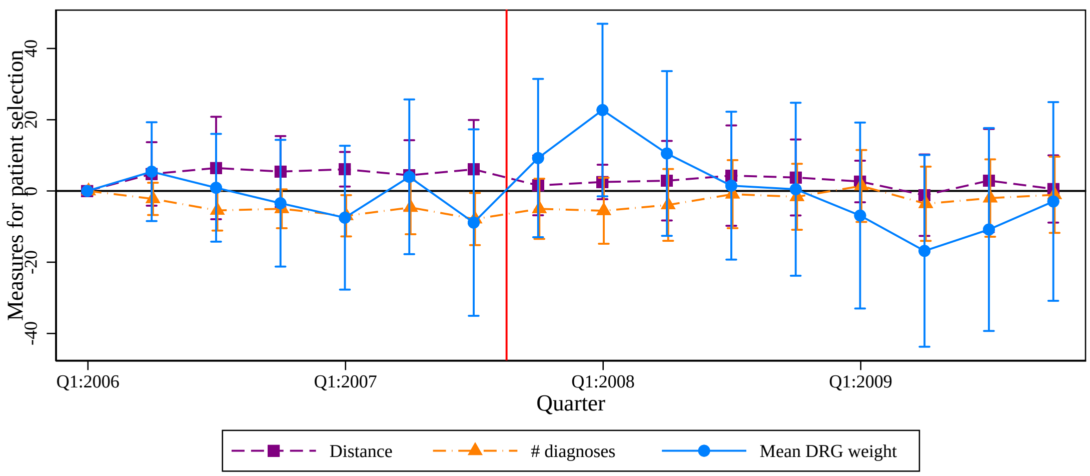


Note: The line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percent of the dependent variable. The red line represents the 2007 payment reform. The vertical lines show 95% confidence intervals, based on standard 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.

Following the reform, we observe more top bill 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.[37]

The coefficients for the post-reform interaction terms show a mean increase in top coding of 1.64 percentage points. A joint significance test reveals that these coefficients are together significantly non-zero (P=0.0353). Thus, the results show that early EMR adopters had a significant comparative advantage in reporting top billing codes post-reform to pre-reform.

Figure 4: 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, and hospital/base DRG fixed effects.

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

[37]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.

be instead due to increases in the severity of illness of patients at those hospitals post-reform rather than from coding differences.

Figure 4 considers three patient characteristics: distance traveled to the hospital, the number of reported diagnoses, and the mean DRG weight based on the base DRG.[38] In 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=0.47) and for number of diagnoses (P=0.425). For mean DRG weight of the base DRG, the coefficients are statistically significant (P=0.0216) but this reflects both positive and negative values in different quarters.[39] 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.

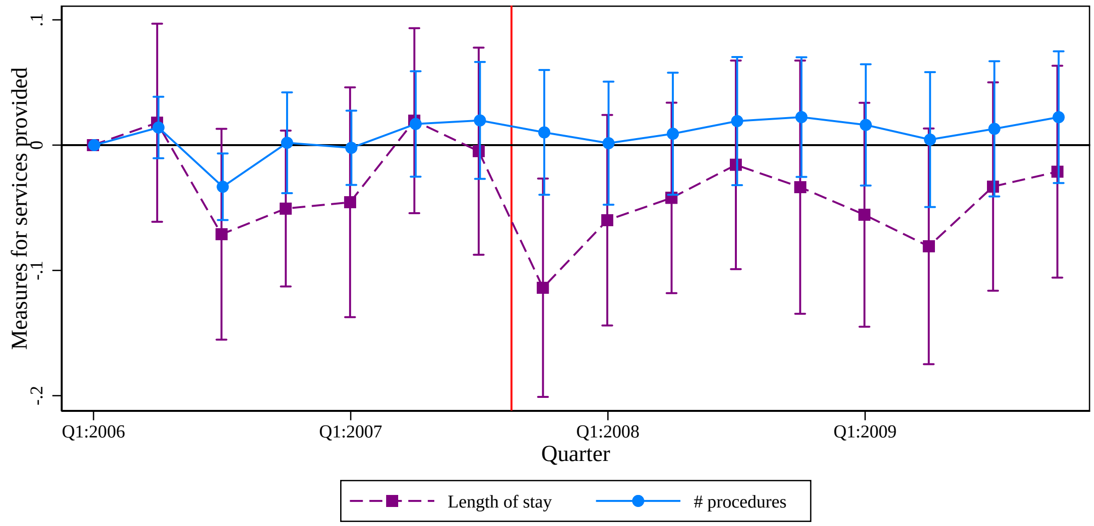
As a further test, Figure 5 considers two measures of patient services: length of stay and number of procedures performed. Again, neither measures shows a sharp change post-reform relative to pre-reform. The post-reform indicators for number of procedures are not jointly statistically significant (P=0.584. The post-reform indicators for length-of-stay are significantly negative (P=0.0187), which could not be caused by sicker patients at EMR hospitals post-reform but may reflect better treatments and hence shorter care at these hospitals over time.[40]

[38] 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.

[39] 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=0.040, P=0.558, and P=0.743, respectively.

[40] Here again, we estimate each of these models with a single post-reform / early EMR interaction coefficient, finding test statistics of P=0.491 and P=0.238, respectively.

Figure 5: Services provided with early EMRs



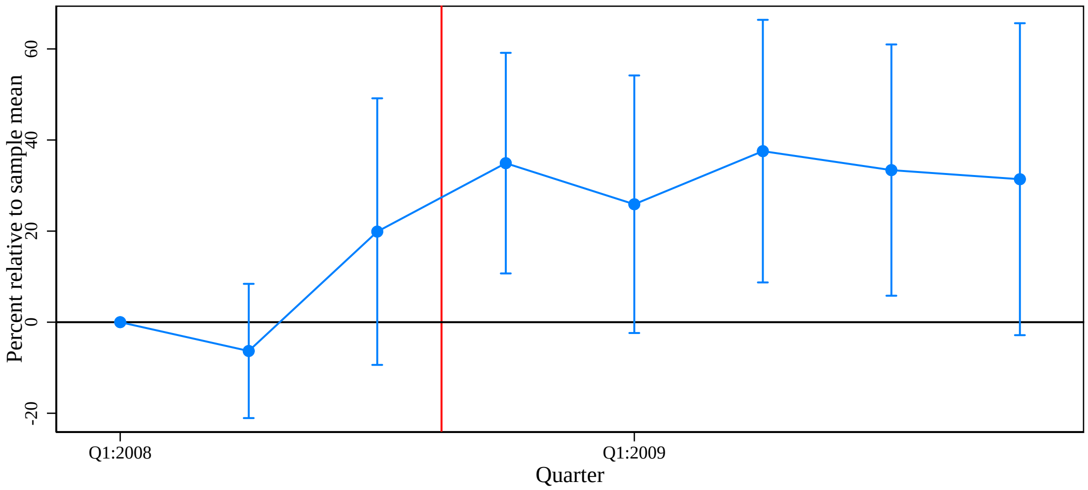
Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percent point form. Sample is all matched base DRGs from before and after the reform with multiple severity subclasses, from Q1:2006-Q4:2009. Standard errors are clustered at both hospital and base DRG levels. Other 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, and hospital/base DRG fixed effects.

5.3 Specification 3: impact of never events

Our findings in Specification 2 show that the cost of coding plays a role in reported top codes post-reform relative to pre-reform. It suggests that the costly coding model has more support in the data than the bill inflation model. However, Specification 2 does not allow us to rule out the bill inflation model since it is possible that bill inflation is driven by the costs of complete coding rather than the revenues gained.

Specification 3 seeks to further test the two models against each other by examining never events. We base our analysis on the 2008 penalization of never events, which creates variation in costs. In particular, it increases the specificity of coding by requiring the present on admission field. For this specification, the dependent variable is the percent of discharges which have at least one listed HAC within a hospital/base DRG/quarter cell for which the HAC is coded as not present on admission. Thus, a higher value of the dependent variable indicates a never event and lower reimbursements in the post-penalization period. We use the same regressors as in Specification 2.

Figure 6: 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.

Similar to Figure 3, Figure 6 shows the interaction coefficients and their 95% confidence intervals, with more details in Table A2 in Appendix A. In Figure 6, we normalize the coefficients by the 2008-09 sample mean of the never event probability, which is 6.45%, 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.

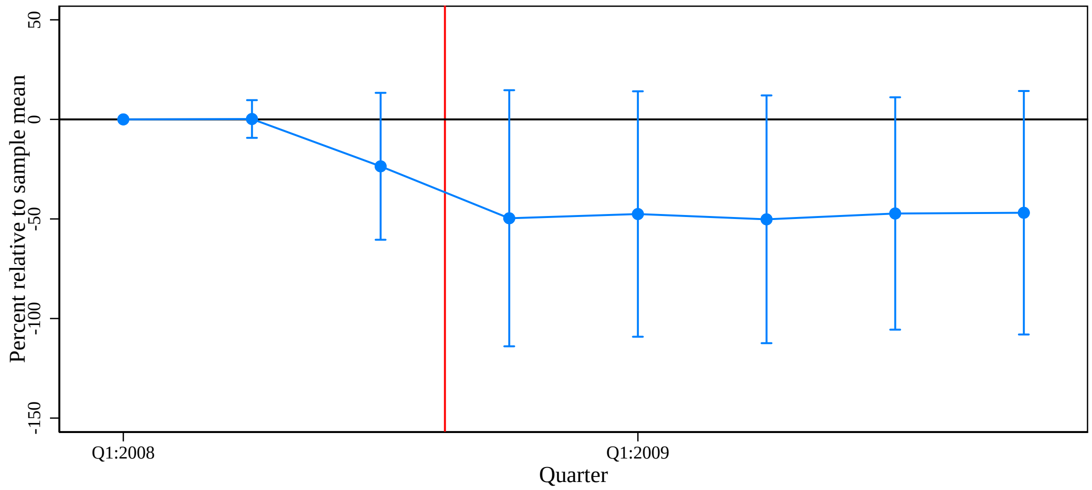
All coefficients for the post-penalization periods are positive, implying that early EMR adopters had more never events—2.1 percentage points on average—in every quarter of the post-penalization period, in the difference-in-difference from their Q1:2008 values and non-EMR adopters. The relative magnitudes, compared to the sample mean, are large: the post-penalization coefficients show that early EMR adoption predicts between 25 and 37 percent more reported never events than the comparison group. The coefficients during the penalty phase are jointly significantly positive (P=0.000345).

Unlike with Specification 2, the pre-trends are positive (though not significant): EMR hospitals started improved documentation of never events one quarter before the penalization started, in Q3:2008. This is likely because EMR hospitals updated their systems to better capture the never event reporting requirements, and did this before the start of Q4:2008 in response to the new rules.

As a robustness check, Figure A1 in the Appendix A provides a similar specification to Figure 6, but where the dependent variable is the percent of discharges for which there is an HAC coded as not present on admission (with the difference being that the denominator now includes discharges without any listed HAC). Our findings are similar to Figure 6: early EMR adopters report relatively more never events *overall* in the post-penalization period, in addition to reporting that more of the conditions on the HAC list were acquired in-hospital.

Finally, we further decompose why EMR hospitals report more never events in the post-penalization period. Figure 7 presents the interaction coefficients from a specification where the dependent variable is the frequency of missing values in the POA field, with more details in Table A3. We again normalize the reported coefficients by the mean probability of missing POA values for the 2008-09 sample, which is 34.7%. The post-penalization coefficients here are negative and the relative magnitudes are large: early EMR adopters report between 46 and 50 percent fewer missing values than non-EMR adopters. The post-penalization period coefficients are jointly significantly negative (P=0.0476). Thus, early EMR adopters report more never events because they decrease their missing POA codes in the post-penalization period more than the comparison group.

Figure 7: extra missing values in POA field 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.

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 are not consistent with EMRs leading to bill inflation for never events. EMR hospitals would not increase reporting never events in the post-penalization period relative to the comparison group if bill inflation were the predominant incentive for reporting these events and EMRs helped with the specificity of coding. However, under the costly coding model, 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 bill inflation in the post-reform period.

5.4 Specification 4: variation between medical and surgical DRGs

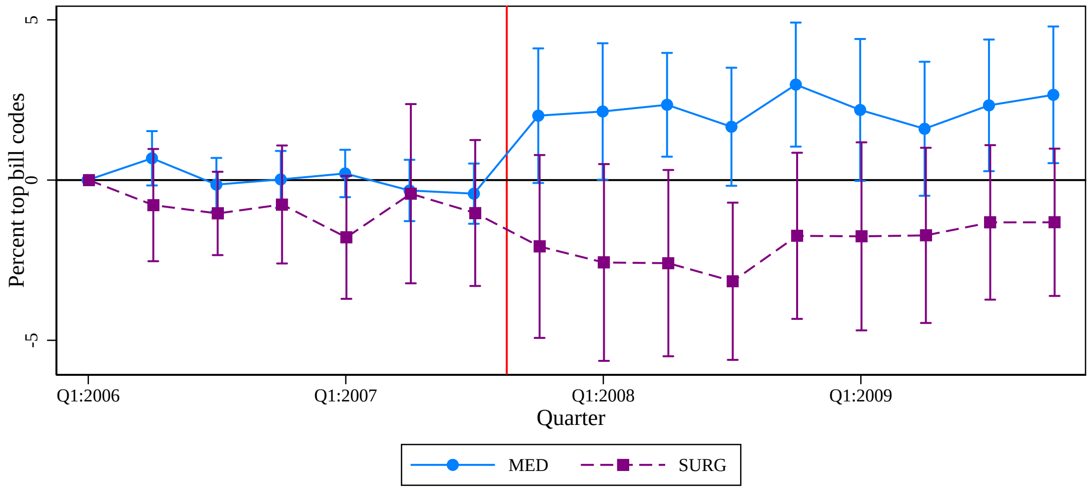
Specification 4 evaluates how EMRs may have differentially affected hospital coding across different types of base DRGs, in order to corroborate our finding that costly coding is the current dominant incentive. Similarly to Specification 2, we use the variation induced by the

2007 payment reform. As we discussed in Section 3, EMRs help lower documentation costs more for medical physicians than surgeons. If hospitals respond to the increased complexity of billing as predicted by our costly coding model, we expect an increase in reporting top bill codes where EMRs are more helpful in lowering the cost of coding. Thus, 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.

We present coefficients and 95% confidence intervals for the early EMR adoption in Figure 8, with details in Table A4 in the Appendix A. Starting immediately after the payment reform, EMR hospitals report relatively more top bill codes for MED. The effect is very consistent across time. The magnitudes are large and jointly statistically significant (P=0.0075), with roughly 2.21 percentage points more top bill codes for early EMR adopters relative to non-EMR hospitals throughout the post-reform period, compared to the mean top coding probability of 31% for the post-reform MED sample. In contrast, the post-reform interaction coefficients for surgical DRGs are negative and jointly statistically significant (P=0.000487), with roughly 2.03 percentage points *fewer* top bill codes for early adopting hospitals in the SURG sample post reform.

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 DRGs compared to SURG DRGs. A plausible cause is that the complete documentation of secondary conditions, which generates the top bill 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.

Figure 8: extra top bill codes with early EMRs, by MED and SURG



Note: Each line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as one percent of the dependent variable. 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.

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. 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 DRGs.

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 bill inflation, we seek to quantify the economic magnitude of this effect. From column 1 of Table A4, early adopters experienced a 2.21 percentage point increase in top coding medical patients in the post-reform period. On average, there are 3.34 million patients per year with the DRGs considered in our paper, about 1.97 million of whom are medical patients, and about 1.28 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 1.28 million $\times$ $6,349 $\times$ 0.82 $\times$ 2.21% = $147 million per year for the U.S.

Our sample accounts for about one fifth 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 the early EMR adopters would amount to $676 million[41] when extrapolating to all DRGs. Moreover, given that almost all hospitals have adopted EMRs by 2018, when considering the full Medicare sample, the costs for all hospitals would amount to $1.04 billion[42] per year, which is 0.85 percent of total Medicare hospital claims costs.

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 $3.47 billion in annual billed costs.

The $3.47 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

[41] [$147 million / 19.4% (percent of Medicare population in sample)] $\times$ 89%.

[42] Assuming that all hospitals behave like early adopters in 2018, we extrapolate from the medical patients admitted to early adopters in our sample to medical patients admitted to all hospitals: ($676 million / 1.28 million) $\times$ $1.97 million.

for MED patients, it does not incorporate the revenues from complete coding for non-EMR hospitals or for SURG patients. Finally, given the move to complex DRG-based payment systems across many countries, the worldwide impact of complete coding costs is likely much larger still.

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 with financial incentives to do so, the causes of this are less understood.

We develop a simple model of bill inflation and costly coding. Our model specifies that hospitals maximize revenues from coding, net 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 bill codes when the additional revenues from reporting them increase. This finding is consistent with both bill inflation and costly coding. Post-reform, this effect is no longer present. The change may be due to the payment reform making the cost of coding more important relative to the revenues from reporting secondary diagnoses.

We test the costly coding model using variation generated by the 2007 payment reform across hospitals. We find that EMR hospitals report more top bill codes following the reform, relative to non-EMR adopters. This finding is suggestive that coding is costly post-reform and that EMRs lower coding costs. However, this result is consistent with EMRs also leading to more bill inflation by lowering coding costs.

Using the 2008 penalization of never events, we further examine whether our results are consistent with bill inflation. 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 bill inflation but is consistent with EMRs reducing the costs of documentation.

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 bill code—is more central to medical than surgical admissions. We find that EMR hospitals reported more top bill 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 \$1.04 billion more to CMS and \$3.47 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 bill inflation after 2007 suggests that there is a limited ability to reduce Medicare hospital expenditures through increased enforcement. It also suggests that other countries with ongoing widespread bill inflation (Jürges and Köberlein, 2015) may be able to reduce bill inflation with greater 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

Abraham, S. and Sun, L. (2018). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Available at SSRN 3158747*.

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.

Cram, P., Bozic, K. J., Callaghan, J. J., Lu, X., and Li, Y. (2014). Use of present-on-admission indicators for complications after total knee arthroplasty: an analysis of Medicare administrative data. *The Journal of Arthroplasty*, 29(5):923–928.

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

de Chaisemartin, C. and D'Haultfœuille, X. (2019). Two-way fixed effects estimators with heterogeneous treatment effects.

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. H., and Tebaldi, P. (2019). The costs of payment uncertainty in health care markets. *Working paper.*

Edwardson, N., Kash, B. A., and Janakiraman, R. (2017). Measuring the impact of electronic health record adoption on charge capture. *Medical Care Research and Review*, 74(5):582–594.

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).

Freedman, S., Lin, H., and Prince, J. (2018). Information technology and patient health: Analyzing outcomes, populations, and mechanisms. *American Journal of Health Economics*, 4(1):51–79.

Ganju, K., Atasoy, H., and Pavlou, P. (2018). Does the adoption of electronic medical record systems inflate Medicare reimbursements? *Working paper.*

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.

Healthcare Australia (2017). Myth buster! clinical coding and health information management in high demand and pay is on the rise in 2017. Available at `http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm`.

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.

Miller, A. R. and Tucker, C. (2011). Can health care information technology save babies? *Journal of Political Economy*, 119(2):289–324.

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.

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.

Parente, S. T. and McCullough, J. S. (2009). Health information technology and patient safety: evidence from panel data. *Health Affairs*, 28(2):357–360.

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.

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

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

Appendix A

Table A1: extra top bill codes with early EMRs (Specification 2)

	Dependent variable: Percent top bill codes within a base DRG	
	Coefficient	S.E.
EMR×Quarter 1	.176	(.465)
EMR×Quarter 2	−.407	(.393)
EMR×Quarter 3	−.231	(.465)
EMR×Quarter 4 (Q1:2007)	−.553	(.503)
EMR×Quarter 5	−.365	(.63)
EMR×Quarter 6	−.619	(.551)
EMR×Quarter 7	1.58	(1.2)
EMR×Quarter 8 (Q1:2008)	1.41	(1.21)
EMR×Quarter 9	1.58	(1.18)
EMR×Quarter 10	.942	(1.16)
EMR×Quarter 11	2.24**	(1.12)
EMR×Quarter 12 (Q1:2009)	1.66	(1.24)
EMR×Quarter 13	1.31	(1.18)
EMR×Quarter 14	1.92*	(1.13)
EMR×Quarter 15	2.12*	(1.13)
Quarter 1	−.531	(.405)
Quarter 2	.152	(.403)
Quarter 3	−.0356	(.45)
Quarter 4 (Q1:2007)	.553	(.377)
Quarter 5	.353	(.539)
Quarter 6	.283	(.51)
Quarter 7	−48.7***	(3.48)
Quarter 8 (Q1:2008)	−46.9***	(3.49)
Quarter 9	−47.4***	(3.49)
Quarter 10	−46.9***	(3.52)
Quarter 11	−46.7***	(3.5)
Quarter 12 (Q1:2009)	−45.2***	(3.57)
Quarter 13	−44.8***	(3.57)
Quarter 14	−45.5***	(3.44)
Quarter 15	−44.6***	(3.47)
N	1,311,245	
p-value for joint significance of post-reform EMR coefficients	.0353	

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percent 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 A2: extra reported never events with early EMRs (Specification 3)

	Dependent variable: Percent reported never events within a base DRG	
	Coefficient	S.E.
EMR×Quarter 1	−.408	(.483)
EMR×Quarter 2	1.28	(.959)
EMR×Quarter 3	2.25***	(.794)
EMR×Quarter 4 (Q1:2009)	1.67*	(.926)
EMR×Quarter 5	2.42**	(.945)
EMR×Quarter 6	2.15**	(.904)
EMR×Quarter 7	2.02*	(1.12)
Quarter 1	.817	(.538)
Quarter 2	4.53***	(1.59)
Quarter 3	4.65**	(2.28)
Quarter 4 (Q1:2009)	5.37**	(2.14)
Quarter 5	4.56**	(1.81)
Quarter 6	4.64***	(1.69)
Quarter 7	5.8***	(2.13)
N	373,910	
p-value for joint significance of post-reform EMR coefficients	.000345	

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are expressed as a percent relative to the sample mean. 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 A3: extra missing POA values with early EMRs (Specification 3)

	Dependent variable: Percent missing POA values within a base DRG	
	Coefficient	S.E.
EMR×Quarter 1	.0664	(1.67)
EMR×Quarter 2	−8.18	(6.51)
EMR×Quarter 3	−17.2	(11.3)
EMR×Quarter 4 (Q1:2009)	−16.5	(10.9)
EMR×Quarter 5	−17.4	(11)
EMR×Quarter 6	−16.4	(10.3)
EMR×Quarter 7	−16.3	(10.8)
Quarter 1	−5.18***	(1.65)
Quarter 2	−39.1***	(7.4)
Quarter 3	−67.3***	(13.1)
Quarter 4 (Q1:2009)	−67.3***	(12.7)
Quarter 5	−66.2***	(13)
Quarter 6	−63.7***	(13.3)
Quarter 7	−66.2***	(12.5)
N	373,910	
p-value for joint significance of post-reform EMR coefficients	.0476	

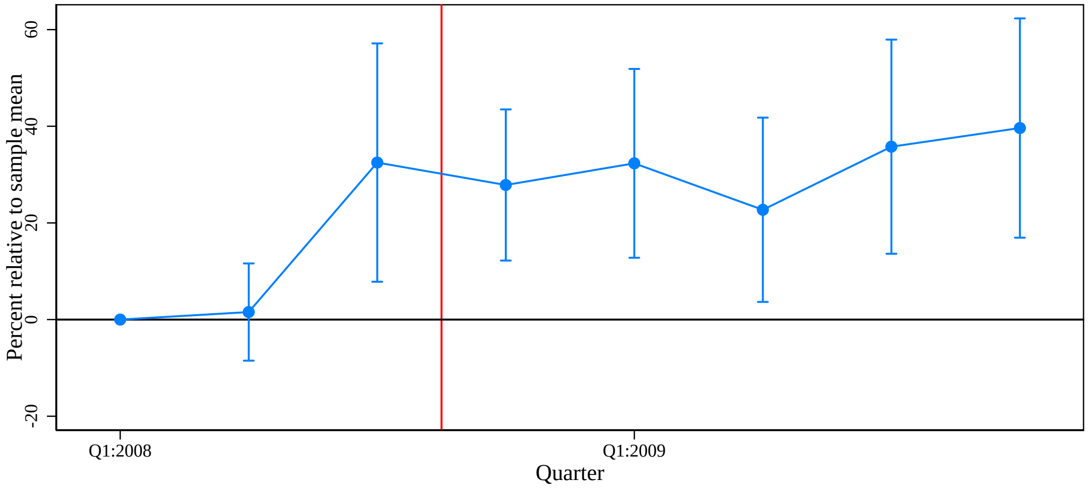
Note: Unit of observation is hospital/base DRG/quarter. Coefficients are expressed as a percent relative to the sample mean.. Sample is all base DRGs, from Q1:2008-Q4:2009. Standard errors are clustered at both hospital and base DRG levels. Other regressors are bed size, total outpatient visits, teaching 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 bill codes with early EMRs, by MED and SURG (Specification 4)

	Dependent variable: Percent top code within a base DRG			
	MED		SURG	
	Coefficient	S.E.	Coefficient	S.E.
EMR×Quarter 1	.68	(.418)	−.781	(.871)
EMR×Quarter 2	−.14	(.41)	−1.04	(.646)
EMR×Quarter 3	.0203	(.439)	−.762	(.916)
EMR×Quarter 4 (Q1:2007)	.206	(.365)	−1.79*	(.953)
EMR×Quarter 5	−.324	(.472)	−.425	(1.39)
EMR×Quarter 6	−.423	(.463)	−1.03	(1.13)
EMR×Quarter 7	2.01*	(1.04)	−2.07	(1.42)
EMR×Quarter 8 (Q1:2008)	2.14**	(1.05)	−2.57*	(1.53)
EMR×Quarter 9	2.35***	(.799)	−2.59*	(1.45)
EMR×Quarter 10	1.66*	(.909)	−3.16**	(1.22)
EMR×Quarter 11	2.98***	(.955)	−1.74	(1.29)
EMR×Quarter 12 (Q1:2009)	2.19*	(1.09)	−1.76	(1.46)
EMR×Quarter 13	1.6	(1.03)	−1.73	(1.36)
EMR×Quarter 14	2.33**	(1.01)	−1.32	(1.2)
EMR×Quarter 15	2.66**	(1.05)	−1.32	(1.14)
Quarter 1	−.943**	(.403)	.303	(.718)
Quarter 2	−.15	(.48)	.827	(.578)
Quarter 3	−.264	(.423)	.475	(.932)
Quarter 4 (Q1:2007)	.104	(.328)	1.43*	(.748)
Quarter 5	.442	(.457)	.302	(1.16)
Quarter 6	.215	(.467)	.572	(1.04)
Quarter 7	−55.4***	(3.84)	−36.4***	(3.54)
Quarter 8 (Q1:2008)	−53.6***	(3.91)	−34.7***	(3.75)
Quarter 9	−54.3***	(3.72)	−34.7***	(3.79)
Quarter 10	−53.9***	(3.74)	−34.1***	(3.77)
Quarter 11	−53.3***	(3.84)	−34.6***	(3.89)
Quarter 12 (Q1:2009)	−51.3***	(4.18)	−34.1***	(3.7)
Quarter 13	−50.8***	(4.14)	−33.9***	(3.89)
Quarter 14	−51.5***	(3.96)	−34.4***	(3.72)
Quarter 15	−50.6***	(3.97)	−33.6***	(3.8)
N	670,687		640,558	
p-value for joint significance of post-reform EMR coefficients	.0075		.000487	

Note: Unit of observation is hospital/base DRG/quarter. Coefficients are reported in percent 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A1: extra reported never events with early EMRs (with different denominator in the dependent variable)



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.