NBER WORKING PAPER SERIES

HOW DO HOSPITALS RESPOND TO PAYMENT INCENTIVES?

Gautam Gowrisankaran
Keith A. Joiner
Jianjing Lin

Working Paper 26455
http://www.nber.org/papers/w26455

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2019, revised December 2023

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

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w26455

How do Hospitals Respond to Payment Incentives?
Gautam Gowrisankaran, Keith A. Joiner, and Jianjing Lin
NBER Working Paper No. 26455
November 2019, revised December 2023
JEL No. H51,I11,I13,O33

ABSTRACT

Over the past decades, Medicare has developed payment reforms that incentivize quality care, by reimbursing fixed amounts for ex ante similar patients. While these reforms may add value, they require providers to code more information on patient health conditions, which is costly. We evaluate the role of revenues and costs in coding intensity for Medicare hospitalized inpatients. We examine the role of costs by estimating hospitals' changes in coding intensity following a 2007 reform based on whether they had adopted electronic medical records (EMRs). EMR hospitals documented relatively more top billing codes after the reform with the increase occurring only for non-surgical admissions, consistent with the hypotheses that costs became an important determinant of the coding decision and EMRs lower these costs, particularly for medical admissions. We further examine whether increased reimbursements from reporting complex diagnoses led hospitals to report more of these diagnoses. We find evidence in favor of this hypothesis before the reform but not after, suggesting that increased billing complexity post-reform made coding costs a more important driver of coding decisions. Our findings suggests that recent payment innovations might add cost to providers, who may want to consider reimbursements in their technology adoption and usage decisions.

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

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

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

1 Introduction

The U.S. Medicare system—the largest health insurer in the country—aims to serve patients with high quality healthcare while keeping costs low. As part of this mission, Medicare has developed a number of payment reforms over the past several decades that seek to incentivize the provision of high quality care, rather than simply reimbursing healthcare providers for providing more care. These payment reforms—which for inpatient care involve Medicare reimbursing fixed amounts for patients who have ex ante similar health conditions—require Medicare to obtain ever greater amount of information about patient health conditions.

While there are benefits to Medicare and other insurers from learning more about patient health conditions, there are also costs. Specifically, payment innovations that rely upon this information have enormously increased the costs to providers of documenting and coding patient conditions accurately (Dunn et al., 2021; Shi, 2021), leading to massive growth in staff dedicated to these tasks. For instance, membership in the American Academy of Professional Coders has more than doubled in recent years,[1] and the U.S. Bureau of Labor Statistics lists medical billers and coders as one of the occupations likely to grow the fastest between 2016 and 2026.[2] These payment innovations and resulting costs are by no means unique to the U.S. For instance, in Australia, the demand for clinical coders has risen 62% (Healthcare Australia, 2017).

The goal of this paper is to understand the tradeoffs that hospitals face in coding more intensely, in the context of Medicare inpatient admissions for hospitals. Modern payment reforms have created ever greater incentives for hospitals to deploy coding resources selectively and strategically, where the marginal revenue gains are highest and the marginal coding costs are lowest. We evaluate the extent to which hospitals respond to changes in these incentives. We identify changes in incentives using factors that affect hospitals' revenues and likely their costs from coding intensely, notably electronic medical records (EMR) technology, which may reduce coding costs.

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

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

Understanding hospitals' coding incentives has potentially important implications for insurers, regulators, and hospital management. Specifically, private insurers and healthcare regulators may want to take coding costs into account in designing future healthcare payment models, balancing the gains from incentivizing quality care against the costs that hospitals incur from coding intensely and the potential distortions that these costs may create. Hospital managers may want to consider cost mitigation in making their EMR technology adoption decisions and in choosing where to deploy coding resources.

Focusing on payment mechanisms, since 1983, Medicare has used the inpatient prospective payment system (IPPS) to reimburse hospitals a fixed amount for each patient stay. Under the IPPS, patients are classified into a base diagnostic related group (DRG), based on the major procedure performed (for surgical admissions) or their primary condition (for other admissions, which are called medical admissions). There are up to three DRGs associated with each base DRG: without complications or comorbidities (CCs), with CCs, or with major CCs (MCCs).

Our main findings emerge from the 2007 payment reform to the IPPS. Under this reform, which was meant to further align payments with expected costs and reimburse for quality, providers needed to document a new acute manifestation of a chronic condition, provide details for an ongoing serious chronic condition, or demonstrate a new acute condition to qualify for a CC/MCC, whereas previously, documenting a chronic condition (regardless of severity or details) was sufficient. In all these cases, the reform required additional documentation to justify a CC/MCC compared to the situation prior to the reform. As a robustness check, we also consider a further reform in 2008, the penalization of complications deemed preventable when acquired during the course of hospital treatment, which are a subset of what are called "never events." Documenting these reportable adverse events further required evaluating whether secondary conditions were present on admission (POA) as a test of whether they were acquired during hospitalization. In both cases, these changes to the IPPS increased the cost of coding by increasing the complexity of the necessary documentation. Finally, we exploit changes in incentives that stem from annual variation in the extra revenues from coding

CCs/MCCs across base DRGs.

Although the 2007 payment reform affected all hospitals at the same point in time, we develop a difference-in-difference design by identifying patients and hospitals whom we believe would be more exposed to higher costs from the reforms. We exploit two related sources of variation here. First, EMRs can potentially reduce the costs of compliance with these new payment models. They are designed to elicit, record, and consolidate accurate information about a wide variety of conditions, such as those that are necessary to document CCs. They can also pinpoint missing tests that might be needed to substantiate a diagnosis, including the acuity of a condition and whether it was POA. Thus, we identify the importance of coding costs by comparing hospitals with EMRs to those without EMRs, and evaluating whether the EMR hospitals reported relatively more CCs after the reforms than before.

Second, EMRs should also disproportionately affect medical rather than surgical coding. Most EMRs are not designed for proceduralists (Kuo, 2017), who focus on the relatively narrow set of conditions relevant to the procedure they are performing. An American College of Surgeons report finds that EMRs force surgeons and their staff to laboriously document information that has little or nothing to do with surgery and that the surgeon cannot address. Surveyed surgeons further report that EMRs create unnecessary burdens by forcing them to spend effort on areas that are outside their area of expertise (Ollapally, 2018). Finally, surgeons generate revenues primarily from performing additional procedures and hence the hospitals' incentives are for the surgeon to perform more procedures rather than document CCs more completely. Thus, we compare whether EMRs have a bigger impact on medical than surgical admissions following the reforms.

We estimate a number of specifications based on the above difference-in-difference (or triple difference) design. For most specifications, the unit of observation is a hospital, base DRG, and year. We use as the main dependent variable the percent of patients with reported CCs/MCCs within this cell and include fixed effects for each base DRG at each hospital. We test whether hospitals with EMRs had relatively more reported CCs/MCCs after the 2007 payment reform than before. We find that EMR adopters had

2.0 percentage points more top codes post-reform than pre-reform, relative to non-EMR adopters. This is consistent with coding costs being an important determinant for the coding decision and EMRs lowering these costs.[3] Separating the impact of the 2007 reform on medical and surgical DRGs, we find that EMR adopters reported 2.67 percentage points more top codes for medical DRGs in the post-reform period than in the pre-reform period, relative to non-EMR adopters. In contrast, the impact on surgical DRGs is negative, with EMR adopters reporting 1.87 percentage points fewer top codes. These results provide further evidence of the importance of coding costs and providers' strategic reactions to these costs.

We also consider a number of additional robustness checks to these results. First, we analyze the impact of EMRs on the documentation of adverse reportable events, using the variation generated by the 2008 penalization of these events. Consistent with our main results, we find that EMR hospitals documented relatively more adverse hospital-acquired complications (HACs) in the post-reform period. Moreover, since greater reporting of adverse HACs *reduces* revenues to hospitals in the short-run by reducing Medicare reimbursements, these results suggest that reducing coding costs by EMR adoption is a more important determinant of how top codes are generated than are revenue gains. Second, we use propensity score weighting as an alternative way of controlling for differences across EMR and non-EMR hospitals, obtaining similar findings to our base specification. Last, we examine whether there were changes in service mix or patient characteristics at EMR hospitals following the 2007 payment reform. We do not find systematic evidence of either of these changes, lending further support to our main hypothesis.

We next turn to examining how hospitals respond to variation in revenues from reporting top codes. The extra revenues from CCs vary both by base DRG and across years for a given base DRG. We follow the literature and examine hospitals' changes in coding CCs within a base DRG when the extra revenue they earn by coding CCs changes (Silverman and Skinner, 2004; Dafny, 2005). Importantly, we allow the effect of revenues to differ before and after the payment reform. We find evidence that prior to Q4:2007,

[3] As a robustness check, we consider propensity score weighting and find similar results with this method.

increases in the extra revenue from top codes led to increases in hospitals reporting top codes, suggesting that the increased revenues affected hospitals' coding incentives. Following Q4:2007, increases in revenues no longer significantly affected reported top codes. This suggests that the extra revenue was a less important driver of coding incentives after the payment reform relative to before. The difference in results may be due to the increased complexity of billing post-reform leading coding costs to be more important relative to the revenue gains.[4]

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

Our paper builds on two main literatures. First, our paper contributes to the broad literature on how healthcare providers respond to financial incentives. Prior studies have shown revenue to be an important driver for top coding among hospitals (Silverman and Skinner, 2004; Dafny, 2005; Brown et al., 2014; Jürges and Köberlein, 2015; Januleviciute et al., 2016; Barros and Braun, 2017; Verzulli et al., 2017; Cook and Averett, 2020). Studies using more recent data have also documented similar evidence in other settings (Brunt, 2011; Bowblis and Brunt, 2014; Fang and Gong, 2017; Geruso and Layton, 2020). Our paper builds on this literature by examining to what extent hospitals respond to the cost of coding—another factor that has been understudied in the literature—in their billing practices. This is also related to Sacarny (2018) who finds that hospitals sometimes use an unspecified code for heart failure even when any specified code would yield a higher reimbursement, implying that there is a cost of complete coding.

Second, our paper also builds on a recent literature on coding and EMRs (Li, 2014; Adler-Milstein and

[4]Our research design does not allow us to distinguish whether the extra coding is justified but it does allow us to understand which incentives are the most salient for hospitals.

Jha, 2014; Ganju et al., 2021). These studies have also used a difference-in-difference design, evaluating the change in reported top codes for hospitals after they adopt EMRs. We complement this literature by showing that the increase in reported top codes is more likely to be driven by the lower cost of coding from adopting the technology—a different underlying mechanism than revenue enhancement.

The remainder of the paper is structured as follows. Section 2 provides institution background. Section 3 discusses our data. Section 4 discusses our empirical approach. Section 5 discusses our results and their implications. Section 6 concludes.

2 Background

2.1 The Medicare Inpatient Prospective Payment System and Coding

Following the passage of the Medicare and Medicaid Act of 1965, hospitals were reimbursed by Medicare based on submission of reports for the reasonable costs of care. This open-ended system facilitated the expansion of facilities but was unsustainable, with year-over-year reimbursement increases of nearly 20% (De Lew, 2000). Accordingly, Medicare adopted the inpatient prospective payment system (IPPS) in 1983, which reimbursed hospitals on a prospective per-case basis. The intent was multifold: to decrease length of stay through appropriate discharge planning, reduce unnecessary spending on tests and services, increase efficiencies generally, and reward delivery of care that was free of preventable complications, all of which lowered hospital expenditure growth (Dobson et al., 2013).

Under IPPS, an admission is first coded into a base Diagnostic Related Group (DRG) using the primary diagnosis code (for medical admissions) or the primary procedure code (for surgical admissions). A base DRG can be either medical or surgical. Essentially, surgical base DRGs are for patients who underwent surgery, and medical ones are for other patients. An example of a medical base DRG is "heart failure and shock" while "major chest procedures" is an example of a surgical base DRG.

Each base DRG includes one to three associated DRGs, which share the same primary diagnosis or procedure but differ in the presence of complications or comorbidities (CCs). CCs are the same across base DRGs. For instance, “spinal procedures w/o CCs” and “spinal procedures w/ CCs” belong to the same base DRG. We call each of these associated DRGs a *severity subclass*.

CMS chooses DRGs so that patients in the same DRG have similar expenditures for managing their condition in the hospital. It then assigns each DRG a numerical weight, meant to reflect the average expenditures, which include the entire cost of hospitalization. As one example, for an individual admitted for the removal of an inflamed or infected gall bladder (cholecystectomy), the DRG weight reflects the average expenses associated with the surgical procedure and post-operative management. CMS specifies different numbers of severity subclasses across base DRGs and different weights for higher subclasses, chosen with the goals of having similar expenditures within a base DRG and making reimbursements reflect expenditures.

The underlying logic of IPPS was to reward hospitals for limiting the use of resources to only those necessary for optimal care, but not penalize them for admitting/managing individuals with pre-existing comorbidities that increased the resource requirements for management. DRG weights are identical for all hospitals, regardless of other factors impacting resource use, but the ultimate reimbursement is adjusted for labor markets, geography, patient demographics, teaching program, and hospital size.

CCs indicate the presence of secondary diagnoses, all of which are *medical* conditions, and not secondary procedures. This is an important distinction, since medical physicians likely have a lower cost than surgeons of diagnosing and documenting the CCs, given their role in the healthcare production process. For example, an individual admitted for an infection in their knee joint would receive a CC DRG if they had an acute bladder infection (cystitis), a new blood clot in the legs (thrombophlebitis), or a new skin infection on their leg (cellulitis). On average, these secondary conditions increase resource usage because they require additional treatment, may confound treatment of the primary admitting diagnosis, and may increase the length of stay.

The process of coding involves selecting the appropriate base DRG and any associated CCs, to determine the exact DRG for the admission. This process is done after discharge from the hospital. It is typically performed by “coders” employed by the hospital (or outsourced to coders at a health analytics firm) and is dependent on documentation of diagnoses in the medical record by the providers.

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. 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 and imaging data, and plans for follow-up care. The operative note, specially for surgical admissions, documents in significant detail the procedure(s) that were performed. 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. Having all this information available for billing requires substantial joint efforts from both the providers and coders.

2.2 Reforms to the IPPS

Since introduction of the IPPS, multiple revisions have been made that increase the complexity of the documentation and coding process. We primarily use the October 2007 payment reform, which created Medicare Severity DRGs or MSDRGs. The 2007 reform had two main goals. The first goal was to create a system where DRG weights were substantially higher for individuals admitted with more complex conditions. This was accomplished by a relative increase in base DRG weights for those conditions, and also by creating a second tier of modifiers to the base DRG—major complicating conditions (MCCs)—reflecting more serious comorbidities requiring increased resource use. As with CCs, all MCCs were medical conditions, again making the cost of documentation lower for medical than surgical physicians.

The second goal was to eliminate less serious chronic underlying conditions from the list of CCs. This was accomplished by either limiting many CCs to acute exacerbations of those underlying chronic conditions[5] or by requiring the documentation of the features of a serious chronic condition that justify increased resource use during hospitalization.

As an example of the former, if an individual with chronic obstructive pulmonary disease (COPD – emphysema and bronchitis) was admitted for a myocardial infarction, with an accompanying acute exacerbation of COPD, this would justify a CC. In the absence of an acute exacerbation, a CC would not be justified. The acute exacerbation would be diagnosed based on clinical symptoms of increased shortness of breath, cough and sputum production and wheezing, as well as a chest x-ray, and evaluation of sputum for bacterial pathogens, with antibiotic administration if appropriate.

As an example of the latter, type I or type II diabetes qualified as a CC before the reform, without any mention of specific clinical features. After the reform, while diabetes did not qualify, diabetes with gangrene did qualify. In the hospital setting, diabetes with gangrene entails large medical costs stemming from limb elevation, dressings, antibiotic administration, and possible surgical debridement. To justify the presence of diabetes with gangrene, the hospital would need to substantiate one of the following:

- Diabetes mellitus due to underlying condition with diabetic peripheral angiopathy with gangrene
- Drug or chemical induced diabetes mellitus with diabetic peripheral angiopathy with gangrene
- Type 1 diabetes mellitus with diabetic peripheral angiopathy with gangrene
- Type 2 diabetes mellitus with diabetic peripheral angiopathy with gangrene
- Other specified diabetes mellitus with diabetic peripheral angiopathy with gangrene

[5] There are exceptions to this rule including, most prominently, for heart failure (Sacarny, 2018).

Each of these would require diagnostic evaluation to document the extent and character of the gangrene, blood flow to the affected area, involvement of pathogenic bacteria, osteomyelitis (bone infection), and more. In comparison, before the reform, the hospital would only have had to document the presence of diabetes in the discharge summary. This would be a copy and paste for paper records, and generally automatic with EMRs. Additionally, documenting MCCs was more complicated than documenting CCs, requiring more effort to justify.

The 2007 reform lowered the fraction of admissions that qualified 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.[6]

A second, smaller reform started in October 2008, when Medicare began penalizing hospitals for preventable adverse events. This reform 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. 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.

Never events include both adverse procedures—such as wrong site surgery—and adverse hospital-acquired complications (HACs)—such as a catheter-associated urinary tract infection acquired in-hospital. In 2007, CMS initially selected eight of these adverse HACs and mandated that, starting in Q4:2007, claims data designate whether diagnoses on the selected conditions were present on admission (POA) or not. Non-compliant claims were supposed to be returned to the provider for clarification.[7] In August 2008, CMS added four additional events, which together form the list of adverse HACs that could have payment implications for hospitals (Office of the Federal Register and National Archives and Records Service, 2008, p.

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

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

48,471-82). Starting in Q4:2008, CMS denied payments for the 12 HACs if they were acquired at hospitals, which is arguably the first time that Medicare withheld payments in an effort to improve the quality of care.

We focus on the subset of never events that are adverse HACs because the requirement of POA documentation imposed additional effort. The main decision-maker for determining whether a secondary diagnosis can be coded as POA is the coder. Coders designate a condition as POA if it is either a) explicitly documented as such by the provider, b) a chronic condition documented prior to admission, or c) determined at some point following admission to have been POA, based on additional information collected during the hospitalization. In situations where the above information is missing and uncertainty exists, coders can query the provider, who may then 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. Improper coding of an adverse HAC as POA when in fact it was hospital acquired can lead to *higher* reimbursement that is not justified (Saint et al., 2009).[8]

2.3 EMRs and Coding

According to the Healthcare Information and Management Systems Society (HIMSS), the following components are key for the meaningful use of EMRs: Clinical Data Repository (CDR), Clinical Decision Support Capabilities (CDS), and Computerized Physician/Provider Order Entry (CPOE) (McCullough et al., 2010). CDR is a centralized database that collects, stores, accesses, and reports health information. 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

[8] 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.

drug safety features such as checks for drug allergies, cross-drug interactions, or dosage adjustments.[9]

Both CDS and CPOE require physician training and involvement to provide real-time support. Typically, physicians interact with CDS and CPOE when making treatment decisions or ordering tests. Nonetheless, CDS and CPOE rely on CDR for the underlying databases of patient information.

The EMR system generally records the hospital course, providing templates to aid physicians in documentation. At the time of admission, if 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 provider 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 EMR 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, although this is most applicable to the ambulatory setting.[10] In addition, and returning to the diabetes example, test results from previous encounters with the medical system are more likely to be accessible by the provider with EMRs.

EMRs facilitate the task of documentation—assembling of a series of clinical and laboratory details to support the diagnosis and guide management of care—by consolidating the information in a legible, searchable, and timely fashion.[11] With EMRs, clinicians can access all clinical notes, laboratory and imaging results, medication orders and other components of the hospitalization in the patient chart, avoiding delay in acquisition or loss of the information (Giannulli, 2016). Moreover, EMRs enable remote access to and management of all this information, whereas, with paper records, all documentation and access must occur in the hospital, and specifically on the unit where the patient is admitted. While substantial time, often remotely outside of the hospital for providers, is involved in entering clinical notes, documenting the hospital course,

[9] See https://psnet.ahrq.gov/primers/primer/6/computerized-provider-order-entry.
[10] See http://www.hcca-info.org/Portals/0/PDFs/Resources/Rpt_Medicare/2016/rmc022216.pdf.
[11] See https://www.healthit.gov/faq/what-are-advantages-electronic-health-records.

and accessing clinical data associated with the case, doing all this via EMRs can ensure the information is complete and up-to-date (Congdon, 2012).

Taking the example of the infected knee joint, information from multiple sources would be required to establish the correct diagnosis, determine if potential CCs were POA, and guide management. EMRs could minimize the wait time to access such information once it is available and enable timely update of the patient's status. For instance, the results from different studies accumulate at different times after admission. An x-ray or MRI of the knee would be used to determine if the infection were limited to the joint space, or involved infected bone. An ultrasound of the leg or other imaging study would be required to substantiate a diagnosis of thrombophlebitis (blood clots). Fluid removed from the knee joint would be evaluated microscopically, and cultured to determine the causative agent. Similarly, a urine analysis and urine culture on admission would be required to establish this condition was POA. While awaiting culture results (1-3 days), a choice of antibiotics would be predicated on likely causative agents, simultaneously taking into account any known allergies of the patient or other factors dictating choice or dose of antibiotic. With paper charts, results from these tests would depend on calling the appropriate service (for example, diagnostic imaging, microbiology, pharmacy), physically going to the service site in the hospital to get the result, or awaiting the paper report to arrive at the hospital unit, which could be during the day or at night. Particularly if the latter, actions based on the results can be delayed. EMRs largely circumvent these matters, both during the admission, and in preparing the discharge summary. All of the information is available in real-time, and can be assembled on site or off site more efficiently as part of the documentation.

Finally, the extent to which EMRs complement the coding process is different between medical and surgical admissions. In general, EMRs are not designed for proceduralists (i.e., surgeons) but rather for gathering a variety of information to manage patient care, which surgeons might not find necessary (Kuo, 2017). For instance, the American College of Surgeons pointed out that EMRs frequently have pop-ups with information applicable to the patient's primary care physician or the medical physician managing the case—

such as alerts that the patient needs a colonoscopy or a screen for diabetes—that have little relevance for the surgeon (Ollapally, 2018). Moreover, surgeons generate higher margins from performing procedures than do medical physicians (Sinsky and Dugdale, 2013) and also contribute to hospital profits overall (Resnick et al., 2005). As a result, hospitals are more likely to incentivize surgeons to perform additional procedures than to focus on documenting medical conditions (completely). Accordingly, we believe that there would be more complementarities between medical admissions and EMRs in lowering coding costs.

3 Data

3.1 Data sources

Our primary dataset is the Medicare Provider Analysis and Review (MedPAR) data. For our purposes, this dataset contains information on all inpatient hospital stays for Medicare beneficiaries. Each observation in these data represents one patient stay and contains information on the hospital, the beneficiary's home zip code, age, gender, dates of service, reimbursement amount, dates of admission and discharge, DRG, and principal and secondary diagnosis and procedure codes. We drop admissions that are not paid under IPPS, such as those from Critical Access Hospitals (CAHs).[12] We construct our main dependent variable, the percent of patients with documented top codes within a particular base DRG, hospital, and quarter, from the MedPAR data. Our discharge data extend from Q1:2005 through Q4:2009. Some of our analyses use data over a shorter time period, as we discuss below.[13]

We merge our base data with information on DRGs from CMS. This information indicates whether the DRG is medical or surgical. It also indicates the weight for each DRG, which we use to define a *spread*—the difference between the top severity subclass DRG weight and the bottom severity subclass DRG weight. We

[12] CAHs receive cost-based reimbursements from Medicare (Gowrisankaran et al., 2018). Other IPPS-exempt hospitals include swing-bed short-term/acute care hospital, swing-bed long-term hospital, inpatient psychiatric hospitals, etc.

[13] We also construct other dependent variables using this dataset, including the distance traveled, length of stay, mean bottom 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.

use the spread as a measure for the incremental change in revenues from top coding. DRG weights change at the beginning of each fiscal year (FY), which corresponds to the fourth quarter of a calendar year.

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

As noted in Section 2.3, there are several components of EMRs. We use the presence of a live and operational CDS or CPOE within the organization as our measure of EMR adoption, since physicians will interact most with these components. This is roughly consistent with what has been done in the literature. For instance, Jha et al. (2009) divide EMR systems into 32 functionalities, of which they view eight (including some parts of CPOE) as necessary for "basic" EMR operation. Miller and Tucker (2009) measure EMR adoption by whether a hospital has installed an "enterprise EMR" system, which they state is a "basic" system that underlies CDR, CDS, and CPOE. Recent studies defined EMR capabilities by either enterprise EMR, CDS, or CPOE (Lee et al., 2013; Agha, 2014; Dranove et al., 2014; McCullough et al., 2016; Ganju et al., 2021). We also examine the robustness of our results to alternate definitions of EMR adoption based on different components, and the results are basically consistent.

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

3.2 Summary statistics

Table 1 provides summary statistics on overall patient samples and the samples used in the analyses. Panel 1 of Table 1 shows our overall data separated by year. On average, there were more than 13 million Medicare discharges each year in our data.[14] The mean age of a Medicare patient discharged from a hospital was 73 years in our sample and the mean DRG weight was rising over time, from 1.44 in 2005 to 1.55 in 2009.

Panel 2 shows the sample by FY for the analyses where we exploit the variation in coding costs generated from the 2007 payment reform. We use four years of data (2006-2009) and all matched base DRGs from before and after the reform with multiple severity subclasses.[15] The percent top codes for base DRGs with three severity subclasses refers to the fraction of patients with the top of the three severity subclasses within the given base DRG. The percent top codes drops substantially following the reform, from over 73% before the reform to 27% immediately after the reform. Moreover, the percent top codes is higher among medical than surgical base DRGs, both before and after the reform.

Panel 3 shows the sample for the analysis where we study the role of revenues in coding intensity. Our analysis here uses five years of data, 2005-2009. We are able to use one year earlier, because we do not consider EMR adoption in this analysis. We limit this analysis to all base DRGs with two severity subclasses pre- or post-reform for a clear definition of spread. We also observe a significant reduction in percent top codes following the reform, from over 73% before the reform to 28% immediately after the reform.[16] Note that the number of discharges is relatively small in FY2005, as we were not able to obtain

[14] This includes admissions that are not paid on IPPS.

[15] We skip the year 2005 due to a large number of missing observations on EMR adoption status for that year.

[16] The 40.3% in the background section is cited from the Federal Registry. It is calculated using the 2006 MedPAR data based on the revised CC list, whereas the 28% here is based on the sample for our analysis—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	73.1	73.1	73.0	72.9	72.9
Mean DRG weight	1.44	1.45	1.46	1.51	1.55
# discharges	13,512,853	13,223,495	12,988,668	14,238,372	14,593,058
	Panel 2: samples for studying the role of coding costs and EMR adoption				
	FY2005	FY2006	FY2007	FY2008	FY2009
% top codes	–	72.1	73.0	27.1	29.4
# discharges	–	3,065,717	2,950,601	3,146,458	3,296,201
MED					
% top codes	–	81.2	81.4	28.9	31.7
# discharges	–	1,739,824	1,696,246	1,861,234	1,961,851
SURG					
% top codes	–	60.1	61.5	24.6	26.1
# discharges	–	1,325,893	1,254,355	1,285,224	1,334,350
	Panel 3: samples for studying the role of revenues				
	FY2005	FY2006	FY2007	FY2008	FY2009
% top codes	77.4	73.1	73.8	28.1	30.8
# discharges	2,621,588	3,953,525	3,773,419	4,441,733	4,764,406

Note: Panels 2 and 3 report statistics in the fiscal year (FY), which is the accounting period for the federal government, from Q4 of the previous year to Q3 of the current year. For instance, FY2006 is Q4:2005–Q3:2006.

the data in Q4:2004.

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 outcome measures and hospital characteristics separated by adoption status. The difference in reported percent top codes is small between hospitals that adopted EMRs before 2007 and hospitals without EMRs through 2009, but EMR hospitals report significantly more adverse HACs relative to non-EMR hospitals. Also, hospitals that adopted EMRs in 2006 or earlier are on average larger and more likely to be teaching or not-for-profit hospitals. For instance, the bed size for early adopters is more than twice that of hospitals without adoption through 2009. EMR hospitals also had a slightly lower debt-to-asset ratio than non adopters.

Table 2: Summary statistics on outcomes and hospital characteristics by EMR use

	EMR adopters	EMR non-adopters
% top codes	51.4	52.5
% discharges w/ adverse HACs	0.285	0.165
Bed size	264	111
Total outpatient visits	208,992	79,201
Total admissions	12,479	4646
FTE physicians and dentists	28	7
Total number of births	1,449	486
% teaching hospital	12.6	2.34
% Medicare discharge	44.5	48.9
% Medicaid discharge	18.7	18.1
% for-profit	19.3	30.0
% not-for-profit	69.8	42.6
% public hospitals	10.9	27.4
Debt-asset ratio	0.631	0.696
Number of hospitals	1,522	256

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

Figure 1: Percent top codes by EMR adoption status

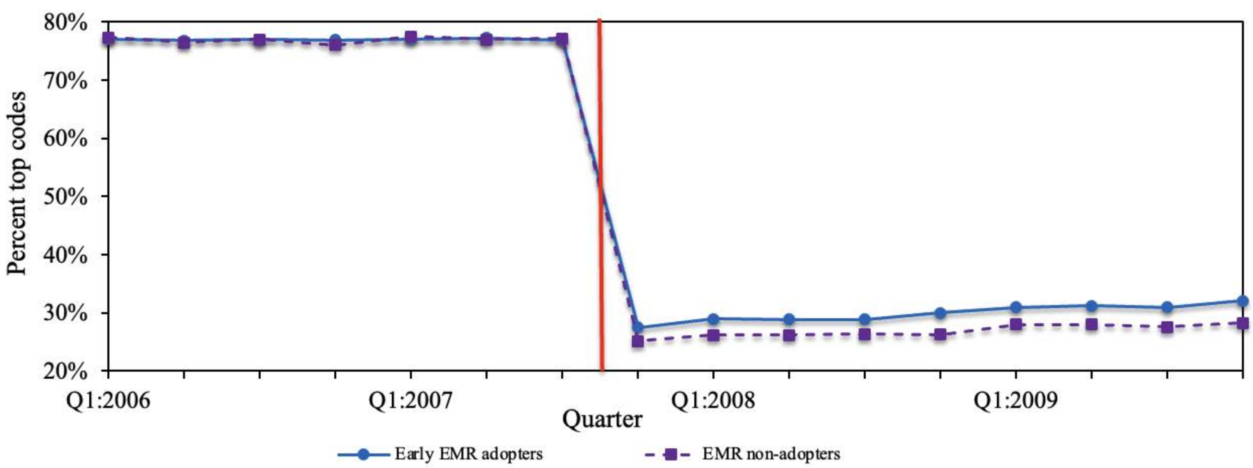


Figure 1 displays the raw trends of percent top codes separately for EMR and non-EMR hospitals based on the sample in Panel 2 of Table 1. We include a vertical line right before Q4:2007 to indicate the implementation of the 2007 payment reform. Visually, there is almost no difference in top codes between early adopters and non adopters prior to the reform. The payment reform reduced the incidence of top codes substantially, but the reduction was relatively smaller among early adopters during the post-reform period.

Table 3: Summary statistics on DRG weights

	Panel 1: All DRGs				
Variable	FY2005	FY2006	FY2007	FY2008	FY2009
DRG weight	1.53	1.47	1.48	1.99	2.02
	(1.90)	(1.86)	(1.84)	(1.93)	(2.00)
# DRGs	518	559	579	743	744
	Panel 2: *Spread* in samples for studying the role of revenues				
Variable	FY2005	FY2006	FY2007	FY2008	FY2009
Spread	0.671	0.655	0.638	0.412	0.741
	(0.490)	(0.485)	(0.455)	(0.322)	(0.557)
Δ*Spread*	−0.029	−0.0027	–	0.328	−0.0149
	(0.240)	(0.046)	–	(0.243)	(0.093)

Note: *Spread* measures the difference between the weight in the top and bottom codes within a base DRG. Δ*Spread* for a particular FY is the difference in *Spread* between the current and following years. Standard errors in parenthesis. We do not report Δ*Spread* in FY2007 because DRGs were reclassified in FY2008.

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

4 Empirical Approach

Our empirical specifications seek to understand the incentives underneath hospitals' decisions to report CCs/MCCs within a base DRG. We focus on the tradeoffs that hospitals face between costs and revenues in assigning top codes to a greater proportion of patients. On the one hand, hospitals will be less likely to report a top code the higher the coding costs necessary to justify that code. On the other hand, hospitals will be more likely to report a top code the higher the relative revenues.[17]

4.1 The Role of Coding Costs and EMRs

We first explore how coding costs affect the incentives to report CCs/MCCs. Our main analysis uses variation in the 2007 payment reform and the differences across hospitals in EMR adoption status in a difference-in-difference design. Specifically, we regress:

$$TopFrac_{jdt} = \beta_t^1 EMR_{jt} + \alpha^1 X_{jt} + \delta_{jd}^1 + \mu_t^1 + \varepsilon_{jdt}, \tag{1}$$

where $TopFrac_{jdt}$ denotes the fraction of reported top-coded patients at hospital j with base DRG d in quarter t. EMR_{jt} indicates a hospital that was an EMR adopter before the start of our sample observed at quarter t and β_t^1 is a quarter-specific coefficient on this indicator. Equation (1) is a two-way fixed effect model, including time dummies, μ_t^1, hospital/base DRG dummies, δ_{jd}^1, and a residual ε_{jdt}. It also includes a number of controls, X_{jt}, specifically bed size, outpatient visits, admissions, number of births, number of full-time physicians and dentists, percentage of Medicare and Medicaid patients, profit status, and a teaching hospital indicator.

Our main hypothesis is that the β_t^1 coefficients will be higher in the post-reform period—starting in Q4:2007—than in the pre-reform period. If these coefficients are higher, then this implies that EMR hospi-

[17] While our approach seeks to evaluate whether the incentives are primarily based on costs or revenues, it is beyond the scope of this study to distinguish whether the reported top codes are justified.

tals are relatively more likely to top code in the post-reform period. This would then provide evidence that the costs of documentation post-reform are lower for EMR hospitals than for non-EMR hospitals, which in turn indicates that coding costs inhibit hospitals' ability to top code and that EMRs mitigate this effect.

We only include two sets of hospitals: hospitals that had adopted EMRs by 2006 (the first year of our sample)—which we call EMR hospitals—and hospitals that had not adopted EMRs by the end of 2009—which we call non-EMR hospitals or non adopters. We focus on these two sets of hospitals to avoid imprecise estimates due to the shifting control group over time (Borusyak and Jaravel, 2017; De Chaisemartin and d'Haultfoeuille, 2020; Sun and Abraham, 2021).

We estimate these regressions with ordinary least squares. We report standard errors calculated with two-way clustering at the hospital and base DRG levels (Thompson, 2011; Cameron et al., 2012). 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.

We also extend this specification to examine the difference between medical and surgical admissions in a triple difference framework. We regress:

$$\begin{aligned} TopFrac_{jdt} = {} & \beta_t^1 EMR_{jt} + \beta_t^2 EMR_{jt} \times MED_d + \alpha^1 X_{jt} + \alpha^2 X_{jt} \times MED_d \\ & + \delta_{jd}^1 + \mu_t^1 + \mu_t^2 \times MED_d + \varepsilon_{jdt}. \end{aligned} \tag{2}$$

In (2), MED_d is an indicator for base DRG d being a medical (non surgical) admission. We estimate a fully interacted specification, where we interact MED with all variables where it is feasible.[18]

Since we expect that EMRs may lower coding costs more for medical than surgical patients following the 2007 reform, our main hypothesis for this regression is that the β_t^2 dummies are positive in the post-reform

[18]We do not include MED_d as a separate regressor or as an interaction with hospital/base DRG interactions, δ_{jd}^1, because these are colinear with MED_d.

period. If these dummies are positive, this would imply that EMR hospitals are more effective at increasing top codes for medical (non-surgical) patients in the post-reform period. This would further indicate that coding costs are keeping hospitals from top coding and that EMRs mitigate this effect, particularly for non-surgical patients.

Our 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—EMR adoption prior to the start of our sample interacted with quarter dummies—is exogenous to the unobservables after controlling for a rich set of hospital characteristics and the fixed effects.

We identify the effect of coding costs from a difference-in-difference design that consider the changes in the fraction of top codes among EMR hospitals within base DRGs relative to changes in this fraction among the comparison group from the pre- to post-reform period.[19] Triple difference specifications using (2) further identify our results from the pre- to post-reform difference in medical versus surgical top codes. Given that EMR and non-EMR hospitals could be different in many aspects, we include a rich set of hospital controls and hospital/base DRG fixed effects to reduce confounding effects. We also 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.

Because the criteria necessary to justify CCs became more stringent after the reform, a threat to identification is that EMR hospitals had relatively more patients qualifying for the more stringent post-reform CC criteria than for the less stringent pre-reform criteria. To explore this possibility, we consider medical and surgical patients separately. This adds to identification because the list of CCs was the same across medical and surgical patients but the costs of coding them for EMR hospitals post-reform would have been relatively lower for medical patients.

A related threat to identification is that the patient mix at EMR hospitals relative to other hospitals

[19] 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.

changes over time. For instance, if relatively more patients with CCs seek care at EMR hospitals over time, this would make EMR hospitals report relatively more top codes following the reform, even if EMRs did not help with complete coding. It is also possible that EMRs actually caused different services to be performed post-reform, rather than only affecting coding behavior. To rule out these threats to identification, we consider a number of robustness analyses, including examining a different event from the 2007 payment reform, controlling for patient mix with propensity score methods, and examining whether there are differential changes in patient mix or service mix at EMR hospitals post-reform. We believe that examining these results together helps pinpoint if changes in coding costs following the reform is the most likely causal impact of any effect that we find.

4.2 The Role of Revenues and Changes in Payments for Top Codes

We next examine how variation in revenue affects the incentives to report CCs/MCCs. Our analysis here uses variation in revenues stemming from differences between the payment for diagnoses without CCs/MCCs and with CCs/MCCs in both the pre-reform and post-reform periods separately. Unlike in (1), where the variation is from a single policy change, the variation here derives from annual updates to the DRG weights, that affect the relative value for reporting CCs/MCCs across different base DRGs, and we are able to consider the pre- and post-reform effects separately.

We regress:

$$
\begin{aligned}
TopFrac_{jdt} = {} & \mathbb{1}\{t < \text{Q4:2007}\} \times \left[\beta^{Pre} \times Spread_{dt} + \alpha^{Pre} X_{jt} + \delta_{jd}^{Pre} + \mu_t^{Pre}\right] \\
& + \mathbb{1}\{t \geq \text{Q4:2007}\} \times \left[\beta^{Post} \times Spread_{dt} + \alpha^{Post} X_{jt} + \delta_{jd}^{Post} + \mu_t^{Post}\right] + \varepsilon_{jdt}. \qquad (3)
\end{aligned}
$$

In (3), $Spread_{dt}$ indicates the difference in DRG weights between the top and bottom codes for base DRG d. We include indicators for the pre-reform and post-reform periods respectively, $\mathbb{1}\{t < \text{Q4:2007}\}$ and

$\mathbb{1}\{t \geq \text{Q4:2007}\}$. We interact all the coefficients separately with the pre- and post-reform indicators. This allows the effect of revenues to vary before and after the reform. The controls, X_{jt}, are the same as in (1).

Our hypothesis here is that β^{Pre} and β^{Post} are positive. If they are positive, then this provides evidence that hospitals top code more when the revenues from top coding are higher, for a given base DRG. This then implies that hospitals strategically adjust their coding behavior based on the revenues that they can obtain for coding patients as having CCs/MCCs.

Following the prior literature (Silverman and Skinner, 2004; Dafny, 2005; Li, 2014), we also examine the effect from the spread across different types of hospitals. For instance, Dafny (2005) found that for-profit hospitals report more top codes than government or not-for-profit hospitals and that hospitals with a high debt-asset ratio exhibited a larger increase in percent top codes. Starting from equation (3), we additionally interact $Spread_{dt}$ with measures of financial health—*financially distressed* (the 25% with the highest debt-asset ratio) and *financially healthy* (the 25% with the lowest debt-asset ratio)—or with whether the hospital is for-profit or not-for-profit, with the omitted category being public hospitals.

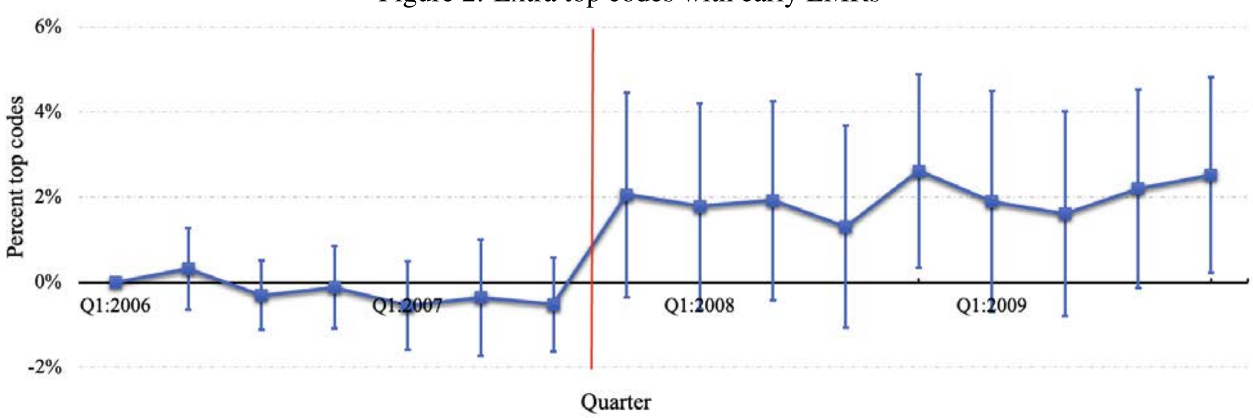
Our identification here is also based on variations within a hospital/base DRG across time. The key identifying assumption for these analyses is that the main variable of interest—$Spread_{dt}$—is exogenous to the unobservables after controlling for the hospital characteristics and fixed effects. Identification thus arises from changes in the spread across years and the resulting changes in percent top codes over time within a hospital/base DRG cell. The central threats to identification would come from correlations between changes in the spread and changes in the unobservable. For instance, a threat to identification would occur if the base DRGs where the spread increased were associated with relatively more complicating conditions for those patients due to technological change. We believe that the likelihood of such association is low 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 the Medicare Recovery Audit Program, occur more broadly across base DRGs rather than targeting to base DRGs in response to changes in spread.[20]

5 Results

5.1 The Role of Coding Costs and EMRs

We first discuss the results for the specification where we study the role of coding costs using the 2007 payment reform and the difference in EMR adoption status. Figure 2 presents the coefficients on interactions between quarter and EMR adoption, as well as their 95% confidence intervals. Appendix Table A1 provides more details on the same regression. We include a vertical line right before Q4:2007 to indicate that the reform was enacted starting in Q4:2007. All reported coefficients are relative to Q1:2006 (which we normalize to have a zero coefficient).[21]

Figure 2: Extra top codes with early EMRs



Note: The line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percentage point. The red line represents the 2007 payment reform. The vertical lines show 95% confidence intervals, based on standard errors clustered at both hospital and base DRG levels. Unit of observation is hospital/base DRG/quarter. Sample is all matched base DRGs from before and after the reform 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.

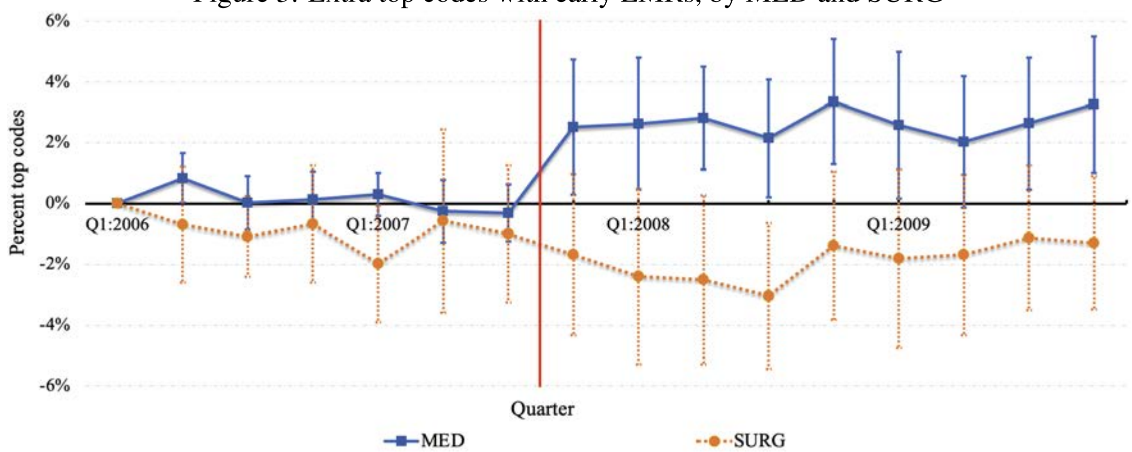
[20] The claims that were flagged to trigger review by auditors were selected based on the auditors' proprietary algorithm (Shi, 2021) or a set of claims and providers which it had determined had a high propensity for error based on the findings from the Comprehensive Error Rate Testing program (`https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/index.html?redirect=/Cert`). These sets are not likely to vary with changes in the DRG spread for a given base DRG.

[21] We need to omit one quarter because (by construction) early EMR adoption does not vary during our sample for a given hospital/base DRG.

Following the reform, we observe more top codes for EMR adopters relative to non adopters. Visually, there is no trend in either the pre-reform or post-reform period, 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.[22]

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

Figure 3: Extra top codes with early EMRs, by MED and SURG



Note: The line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percentage point. The red line represents the 2007 payment reform. The vertical lines show 95% confidence intervals, based on standard errors clustered at both hospital and base DRG levels. Unit of observation is hospital/base DRG/quarter. Sample is all matched base DRGs from before and after the reform 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.

Our result is consistent with the hypothesis that the relative increase in reported top codes for EMR hospitals post-reform is due to coding differences. A central threat to identification would come from EMR

[22]This result contrasts with Geruso and Layton (2020), 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 adoption.

hospitals having a relatively higher share of patients qualifying for the more stringent post-reform CC criteria than the less stringent pre-reform ones. In part to consider this possibility, we turn to understanding whether EMRs have differentially affected hospital coding of medical admissions versus surgical admissions. Figure 3 presents the coefficients and 95% confidence intervals for the interaction terms with the indicator for early EMR adoption. We present details of this regression in Appendix Table A2.

Starting immediately after the payment reform, EMR hospitals report relatively more top codes for medical admissions. The effect is very consistent across time. The magnitudes are large and jointly statistically significant (P=0.0044), with roughly 2.67 percentage points more top codes for early EMR adopters relative to non-EMR hospitals throughout the post-reform period, compared to the mean top coding probability of 31.6% for the medical base DRGs post reform. In contrast, the post-reform interaction coefficients for surgical base DRGs are significantly negative, with an average decrease by 1.88 percentage points (P=0.0395). The difference in the effect between medical and surgical base DRGs in the post-reform period is consistent with our proposition that there are greater complementarities between EMR adoption and medical admissions in lowering coding costs post reform to pre reform but not consistent with EMR hospitals having a relatively higher share of patients qualifying for the post-reform CC criteria.

The negative coefficients for surgical base DRGs post reform further imply that surgeons face a relatively *higher* cost of documenting diagnoses completely after the reform at EMR hospitals than at the comparison hospitals. This is likely caused by EMRs resulting in a higher fixed cost but lower marginal cost for optimal use and with the fixed costs for optimal EMR use increasing from the reform due to the stricter requirements for top codes. Surgeons are reported to face great barrier to using EMRs (Ollapally, 2018) and have less intensive interactions with EMRs (Gawande, 2018) and in particular, with documenting secondary diagnoses in EMRs. Thus, they may not find it worthwhile to bear the fixed cost of optimal use, and could even be less incentivized to do so following the reform. Substantiating this, unlike medical physicians, surgeons can increase revenues by performing more surgeries, rather than billing more top codes for existing patients.

5.2 The Role of Revenues and Changes in Payments for Top Codes

We now discuss the results of estimating equation (3) where we examine the role of revenues in the top coding decision. Table 4 reports the coefficients for the interaction terms with the spread from three separate specifications where we progressively adding more controls. Each column shows the estimates from a particular specification, with our preferred specification—as specified in equation (3)—shown in column 3.

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

	Dependent variable: Percentage of top codes within a base DRG		
	(1)	(2)	(3)
Pre × Spread	-3.68	-3.55	0.60***
	(6.97)	(6.87)	(0.090)
Post × Spread	5.47	5.19	-2.96
	(7.96)	(7.84)	(2.07)
Hospital/base DRG FEs and hospital controls	yes	yes	yes
Pre & Post interacted with hospital controls	no	yes	yes
Pre & Post interacted with hospital/base DRG FEs	no	no	yes

Note: Unit of observation is hospital/base DRG/quarter. Sample is all base DRGs with two severity subclasses, from Q1:2005-Q4:2009. Coefficients are reported in percentage point form. The three specifications progressively add more controls, as specified in the lower panel. Our preferred specification is column 3. Hospital controls 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Focusing first on the pre-reform period, the first coefficient in column 3 suggests that before Q4:2007, top coding occurs with high frequency when the financial incentive to report the top code increases. A unit increase in spread corresponds to an increase of 0.60 percentage points in reported top codes. This is consistent, though not dispositive, with the finding in prior studies that provided evidence of upcoding (Silverman and Skinner, 2004; Dafny, 2005). Documentation of top codes increasing with spread implies that revenues were an important driver of top coding in the pre-reform period.

Turning now to the post-reform period, from Q4:2007 onwards, a larger spread no longer predicts a higher fraction of top codes.[23] This difference in results may be due to a more complex reimbursement system making the cost of coding more important in the billing decision than the revenue gained. The post-reform evidence is not consistent with higher revenues leading to more top codes.

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. Column 1 in Appendix Table A3 reports the results where *Spread* is additionally interacted with hospital ownership status, besides interacted with the pre- and post-reform indicators using our preferred specification. The coefficient for *Spread* remains significantly positive for the pre-reform sample, but there is no significant difference between for-profit and not-for-profit hospitals both before and after the reform. Column 2 presents the results when we add the hospital's financial health status as a regressor. The main findings are similar except that hospitals undergoing financial distress report relatively fewer top codes post reform as the spread gets larger.

To summarize, we find that hospitals responded to revenue to a greater extent than cost in coding intensity prior to the reform. However, this significant impact on top-coding based on financial incentives does not extend to the post-reform period. It further lends supports to the relatively important role of coding costs in coding intensity, due to the increased complexity in the new payment system.

5.3 Economic Magnitudes of Complete Coding

Having found that coding costs play a more important role than revenues in the current hospital reimbursement incentives, we seek to quantify the magnitude of Medicare reimbursements affected by the change in coding costs due to the 2007 payment reform. From column 1 of Table A2, early adopters experienced a 2.67 percentage point increase in top coded medical patients in the post-reform period. On average, there

[23] The Z statistic for testing the join equality between the pre- and post-reform coefficients is 2.97, suggesting that we can reject the null hypothesis that both coefficients are the same at the 10% level of significance.

are 3.13 million patients per year with the DRGs considered in our paper, about 1.83 million of whom are medical patients, and about 1.13 million of whom are medical patients admitted to early EMR hospitals. The average spread of the DRGs on which we focus is 0.82 and the average DRG price is $6,349 for an admission with weight 1. Therefore, the in-sample extra revenue paid by CMS due to more complete coding from the 2007 reform by early EMR adopters is $157 million (=1.13 million × $6,349 × 0.82 × 2.67%) per year for the U.S.

Our sample accounts for about 22% of the Medicare inpatient population. Considering the fact that almost 89% of patients are in DRGs with multiple subclasses, we expect that the costs for early EMR adopters would amount to $634 million[24] when extrapolating to all DRGs. Moreover, given that almost all hospitals have adopted EMRs by 2020, when considering the full Medicare sample, the costs for all hospitals would amount to $1.03 billion[25] per year, which is 0.84 percent of total Medicare hospital claims costs.[26]

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 the DRG reclassification on extra charges due to EMRs facilitating complete coding would translate into approximately $3.4 billion in annual billed costs.[27]

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

[24] [$157 million / 22% (percent of Medicare inpatient population in sample)] × 89%.

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

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

[27] The economic magnitudes we estimate here do not account for the penalty from documenting adverse HACs (a second policy reform we consider as a robustness check on which we provide more details in Section 5.4). However, we believe that the direct dollar amount associated with the second reform is likely to be small, because, on average, only 0.256% of discharges within a hospital and base DRG reported adverse HACs, and EMR hospitals reported 0.0617 percentage points more adverse HACs than non-EMR hospitals based on our findings in Section 5.4 (shown in Appendix Table A4).

systems across many countries, the worldwide impact is likely much larger.

Overall, our takeaway is that the extra revenues hospitals receive 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 economic effects caused by the costs of complete coding may also be substantial.

5.4 Robustness

Testing the role of coding costs using adverse HAC penalization. We examine the penalization of adverse HACs using similar specifications to equation (1) but with a different dependent variable—the percent of adverse HACs within a hospital/base DRG/quarter cell. We define a reported secondary diagnose as an adverse HAC, if it is on the HAC list specified by CMS *and* if it is coded as "acquired at hospital" in the code "present on admission (POA) indicator." According to the final rule published in 2009 (Office of the Federal Register and National Archives and Records Service, 2008, p. 48,471-82), the HAC list broadly includes 12 conditions, each of which is associated with a series of ICD-9 codes and some with procedures. Our definition of each HAC follows the codes published by CMS (Office of the Federal Register and National Archives and Records Service, 2008, p. 48,471-82). CMS mandated the code for the POA indicator in FY 2008, requiring hospitals to document whether each reported diagnosis occurred before or after hospital admission (Office of the Federal Register and National Archives and Records Service, 2008, p. 48,486).

Reporting adverse HACs results in lower revenue in the short run—an opposite financial consequence to that of top-coding patients. Our main hypothesis is that the β^1 coefficients are negative in the period following the penalization of adverse HACs (Q4:2008). If this were to be the case, this would indicate that EMR hospitals are more effective at helping hospitals increase short-term revenues by allowing them to document fewer adverse HACs.

Appendix Figure A1 shows the interaction coefficients and their 95% confidence intervals, with more details in Appendix Table A4. In this figure, we normalize the coefficients by the 2008-09 sample mean of the adverse HAC probability, which is 0.256%. We include a vertical line right before Q4:2008 to indicate the start of penalization for adverse HACs. All reported coefficients are relative to Q1:2008.

All coefficients for the post-penalization periods are positive, implying that early EMR adopters had more adverse HACs—0.0617 percentage points on average—in every quarter of the post-penalization period, in the difference-in-differences from their Q1:2008 values and non-EMR adopters. The relative magnitudes, compared to the sample mean, are large: the post-penalization coefficients show that early EMR adoption predicts between 16 and 34 percent more reported adverse HACs than the comparison group. The coefficients during the penalty phase are jointly significantly positive (P=0.0385). In fact, if coding costs were the main driver for coding intensity, EMR hospitals would be expected to report relatively more adverse HACs even *before* the penalty provision. Appendix Figure A1 shows that relatively more adverse HACs were documented among early adopters in Q3:2008, one quarter before the penalization started.[28]

Propensity score weighting. As a further robustness check, we apply propensity score weighting in our difference-in-differences estimation where we examine the impact of coding costs on coding intensity. We use propensity score weighting to create sets of EMR and non-EMR hospitals that are similar based on observables regressors other than EMR adoption status.[29] We implement this alternative estimator in two steps. First, we estimate the probability—the propensity score—of being an early EMR adopter given a hospital's characteristics, using 2006 data. Second, we rerun the fixed effects model, using the predicted probability in the first step to derive the weight for each observation.[30] Specifically, we use the inverse probability of being an early adopter for EMR hospitals and the inverse of one minus the probability for non-EMR hospitals.[31] Such a weighting scheme allows the hospitals with relatively low (high) probability

[28]The observed increase prior to the reform could also be driven by anticipatory effects. The reimbursement changes due to adverse HACs were mandated by section 5001(c) of the Deficit Reduction Act of 2005, and the HAC list was announced in October 2007, both of which were before the implementation of the penalty provision (Mattie and Webster, 2008). However, it is hard to separate the anticipatory effect from the treatment effect, because the reporting of POA is not available before January 2008.

[29]This method was proposed by Imbens (2000) and is similar to Robins and Rotnitzky (1995) and Robins et al. (2000).

[30]The weight used in the regression is a product of the probability and the number of patients within the hospital/base DRG cell.

[31]To avoid the presence of extreme weights, we drop the hospitals with a weight of less than 1% in the second-step estimation (Chesnaye et al., 2022).

of being an early adopter to contribute more (less) to the estimation so as to create a balanced sample where hospitals can be considered exchangeable conditional on the probability of being an early adopter.[32]

Appendix Figure A2 presents the results for equation (1) from this estimator, and Appendix Figure A3 shows the estimates for equation (2) using this weighting scheme. The estimated effects are in a similar pattern to those from the conventional two-way fixed effects model (shown in Figures 2 and 3), which provides further evidence that the bias due to the potential confounding factors could be limited.

Testing for changes in patient severity of illness. We next test whether the changes following the payment reform could be instead due to differences in patient severity of illness. Appendix Figure A4 considers three patient characteristics: distance traveled to the hospital, the number of reported diagnoses, and the mean DRG weight at the hospital.[33] For the first two measures, we calculate the mean within a hospital/base DRG/quarter cell and re-estimate equation (1), replacing the dependent variable with the appropriate measure. For the last measure, the unit of observation is a hospital/quarter rather than a hospital/base DRG/quarter and we calculate mean DRG weight by averaging across base DRGs within a hospital/quarter cell. This specification is also similar to equation (1), except that we include hospital fixed effects rather than hospital/base DRG FEs, given the unit of observation. 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=0.657) and for number of diagnoses (P=0.379). For the mean DRG weight specification, the post-reform coefficients are statistically significant (P=0.0605), but this reflects both positive and negative values in different quarters.[34] Hence, it does not appear that the increase in reported top codes post-reform is due to changes in the patient mix at EMR hospitals relative to others.[35]

[32]There are other propensity score approaches, notably matching and stratification. We apply propensity score weighting because (1) we can use all the observations in the estimation, unlike matching where non-matched observations will be excluded, and (2) the weights can be directly incorporated into our main regression.

[33]We use the mean weight of the *lowest* severity subclass within a base DRG to allow the effect to be robust to misreporting of severity subclasses.

[34]To test for the presence of monotonic effects, we re-estimate each of these regressions by replacing the interactions involving EMR indicators with a single post-reform / early EMR interaction. The p value of the interaction coefficient is 0.095 for distance, 0.437 for number of diagnoses, and 0.765 for mean bottom DRG weights, respectively.

[35]Assuming there is no selection of patients in response to the payment reform, the mean DRG weight specification further shows that hospitals do not appear to change the coding of the base DRG of the patient in response to the payment reform.

Testing for changes in services performed. Finally, we consider whether the changes following the payment reform could be due to changes in services performed. Appendix Figure A5 considers two measures of services performed: length-of-stay and number of procedures performed. To obtain these estimates, we calculate the mean for these two measures at the hospital/base DRG/quarter level and re-estimate equation (1) using them as the dependent variable. Again, neither measure shows a sharp change post-reform relative to pre-reform. The post-reform indicators for number of procedures are not jointly statistically significant (P=0.583). The post-reform indicators for length-of-stay are significantly negative (P=0.00133), which could not be caused by sicker patients at EMR hospitals post-reform but may reflect better treatments and hence shorter length of stay at these hospitals over time.[36]

Summary of robustness results. The robustness results provide further evidence that revenue is not a main driver for coding intensity. EMR hospitals would not report more adverse HACs in the post-penalization period relative to the comparison group if short-run revenue enhancement were the predominant incentive. In fact, it is consistent with the proposition that EMRs help with the specificity of necessary documentation.

The similar findings after we apply propensity score weighting as well as the lack of substantial differences in the alternative outcome measures between EMR and non-EMR hospitals lend support to our causal interpretation of the results.

6 Conclusion

Over the past decades, the U.S. healthcare systems have devised a number of payment innovations, seeking to reward quality over quantity. These payment mechanisms require providers to obtain more information on patient health conditions. While it is beneficial for providers and insurers to learn more about patients' health conditions, gathering the information imposes a significant amount of costs in documentation and coding

[36] Here again, we re-estimate each of these regressions using a single post-reform / early EMR interaction as the key variable of interest. The p values of the interaction coefficient for length of stay and number of procedures are 0.380 and 0.214, respectively.

for providers. Complete documentation of patient health information could result in more reimbursements but also involves substantial time and efforts from providers. We seek to understand the tradeoff facing hospitals in coding intensity in the context of Medicare inpatient hospitalization care.

We examine the importance of coding costs in coding intensity using a difference-in-difference framework, where we exploit the variations in coding costs from a national payment reform and the difference in EMR adoption status across hospitals. We find that EMR hospitals report more top codes following the reform, relative to non adopters. We also find that the relative increase in top coding among EMR hospitals mainly occur to medical admissions, which is consistent with our proposition that EMRs are generally more helpful in documenting care for medical physicians than surgeons.

We then estimate how financial incentives to report more complex secondary diagnoses affect the fraction of patients reported as having these diagnoses. We find that hospitals report more top codes when the additional revenues from reporting them increase prior to the reform but the effect is no longer present after the reform. The change may be due to the payment reform making the coding costs more important relative to the revenues from reporting secondary diagnoses.

Our paper provides general evidence on incentives in the health sector. Our finding of the relatively more importance of coding costs than revenue in coding intensity suggests that the recently-implemented payment mechanisms that reward high-quality care could imply a hidden cost to healthcare providers. Private insurers and CMS may want to account for the costs of coding in designing future reimbursement models to encourage proper documentation among providers. Our finding that EMRs help capture the stricter requirements in reporting particular medical conditions suggests that the technology has the potential to combat the increasing complexity in the reimbursement system, but the effect could vary across different types of patients. Managers of healthcare organizations may want to consider how to take the full advantage of the technology to optimize reimbursements.

References

Adler-Milstein, J. and Jha, A. K. (2014). No evidence found that hospitals are using new electronic health records to increase Medicare reimbursements. *Health Affairs*, 33(7):1271–1277.

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

Barros, P. and Braun, G. (2017). Upcoding in a national health service: The evidence from portugal. *Health economics*, 26(5):600–618.

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

Bowblis, J. R. and Brunt, C. S. (2014). Medicare skilled nursing facility reimbursement and upcoding. *Health economics*, 23(7):821–840.

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.

Brunt, C. S. (2011). Cpt fee differentials and visit upcoding under medicare part b. *Health economics*, 20(7):831–841.

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

Chesnaye, N. C., Stel, V. S., Tripepi, G., Dekker, F. W., Fu, E. L., Zoccali, C., and Jager, K. J. (2022). An introduction to inverse probability of treatment weighting in observational research. *Clinical Kidney Journal*, 15(1):14–20.

Congdon, K. (2012). The Reality Of EHR Use In The Physicians Practice. Available at https://www.healthitoutcomes.com/doc/the-reality-of-ehr-use-in-the-physicians-practice-0001.

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

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

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

De Lew, N. (2000). Medicare: 35 years of service. *Health Care Financing Review*, 22(1):75.

Dobson, A., DaVanzo, J. E., Berger, G., and El-gamil, A. (2013). The Purpose, Promise, and Impact of the Original 1983 Medicare Hospital Inpatient Prospective Payment System (IPPS). Available at https://www.dobsondavanzo.com/clientuploads/Publications/Hospital%20Finance/The_Purpose,_Promise,_and_Impact_of_the_Original_1983_(White_Paper).pdf.

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

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

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

Fang, H. and Gong, Q. (2017). Detecting potential overbilling in medicare reimbursement via hours worked. *American Economic Review*, 107(2):562–591.

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

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

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

Giannulli, T. (2016). 7 Time-Saving EHR Use Tips to Boost Physician Productivity. Available at https://ehrintelligence.com/news/7-time-saving-emr-use-tips-to-boost-physician-productivity.

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.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Januleviciute, J., Askildsen, J. E., Kaarboe, O., Siciliani, L., and Sutton, M. (2016). How do hospitals respond to price changes? evidence from norway. *Health economics*, 25(5):620–636.

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.

Kuo, P. (2017). Anxious over a transition to electronic records? Fear not. Available at https://www.aorn.org/outpatient-surgery/article/2017-May-business-advisor-clearing-8-common-ehr-hurdles.

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.

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

Mattie, A. S. and Webster, B. L. (2008). Centers for Medicare and Medicaid Services "never events": An analysis and recommendations to hospitals. *Health Care Management*, 27(4):338–349.

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.

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 Reg-*

ister (Wednesday, August 22, 2007), 72(162):47129–48175. From the Federal Register Online via the Government Publishing Office [www.gpo.gov], FR Doc No: 07-3820.

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

Ollapally, V. (2018). EHRs add to surgeons' administrative burdens: The ACS responds. Available at https://bulletin.facs.org/2018/09/ehrs-add-to-surgeons-administrative-burdens-the-acs-responds/.

Resnick, A. S., Corrigan, D., Mullen, J. L., and Kaiser, L. R. (2005). Surgeon contribution to hospital bottom line: not all are created equal. *Annals of surgery*, 242(4):530.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, pages 550–560.

Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429):122–129.

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

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

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

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

Sinsky, C. A. and Dugdale, D. C. (2013). Medicare payment for cognitive vs procedural care: minding the gap. *JAMA internal medicine*, 173(18):1733–1737.

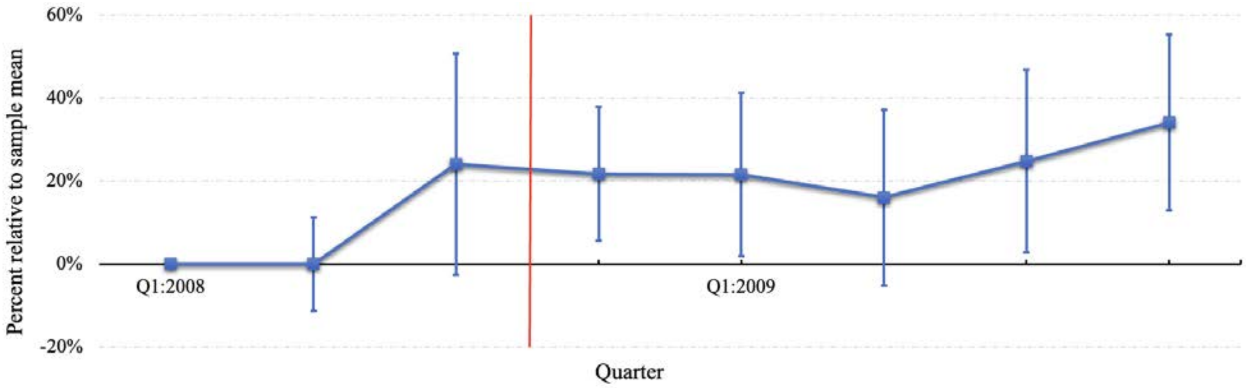
Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

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

Verzulli, R., Fiorentini, G., Lippi Bruni, M., and Ugolini, C. (2017). Price changes in regulated healthcare markets: do public hospitals respond and how? *Health economics*, 26(11):1429–1446.

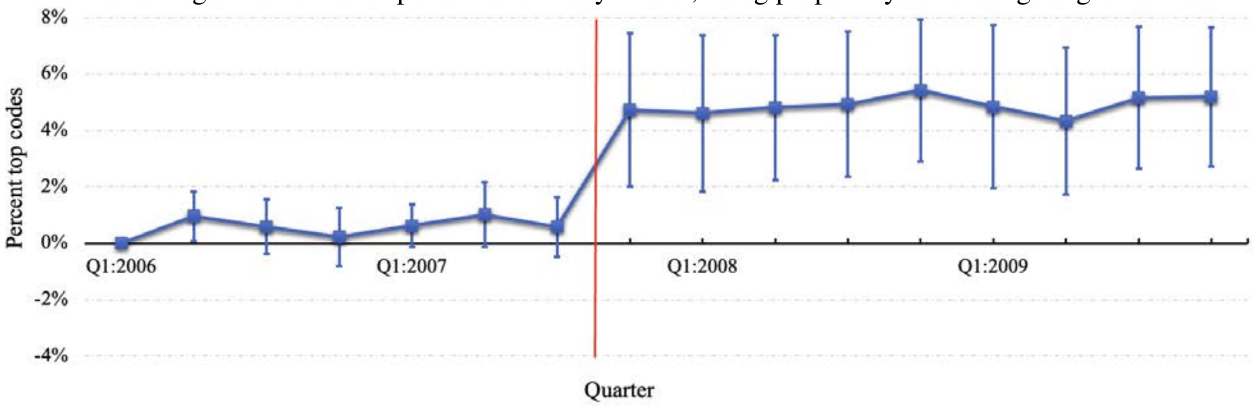
Appendix

Figure A1: Extra reported adverse HACs with early EMRs



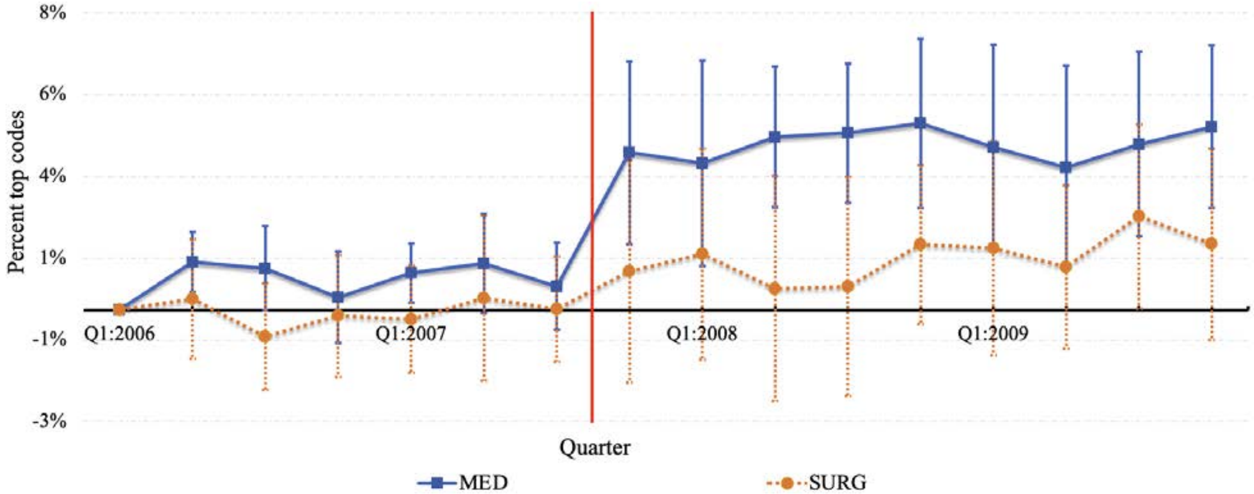
Note: Each line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percent relative to the sample mean. The red line represents the 2008 penalization of adverse HACs. The vertical lines show 95% confidence intervals, 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.

Figure A2: Extra top codes with early EMRs, using propensity score weighting



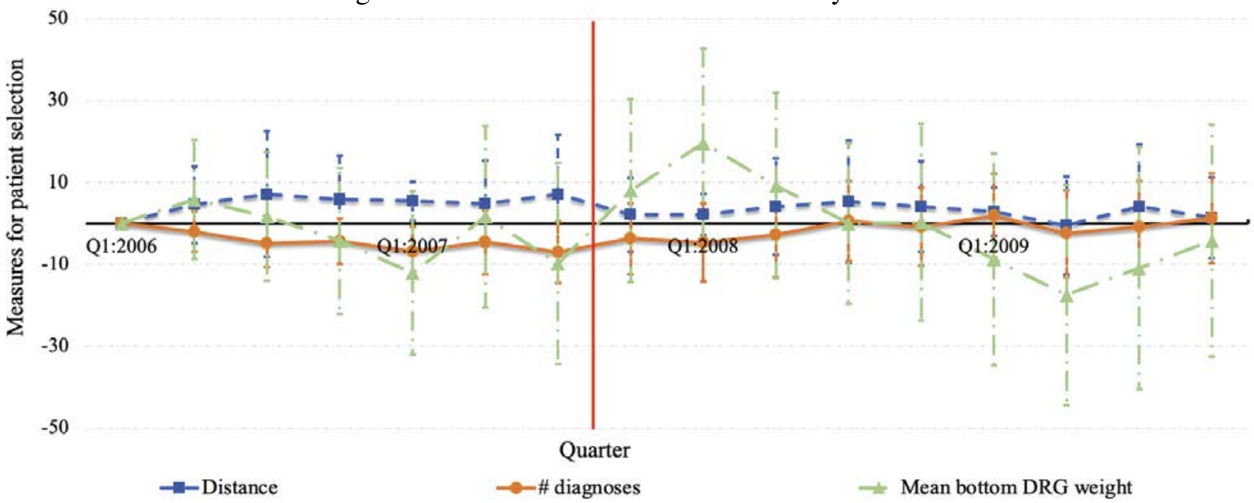
Note: The line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percentage point. The red line represents the 2007 payment reform. The vertical lines show 95% confidence intervals, based on standard errors clustered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

Figure A3: Extra top codes with early EMRs by MED, using propensity score weighting



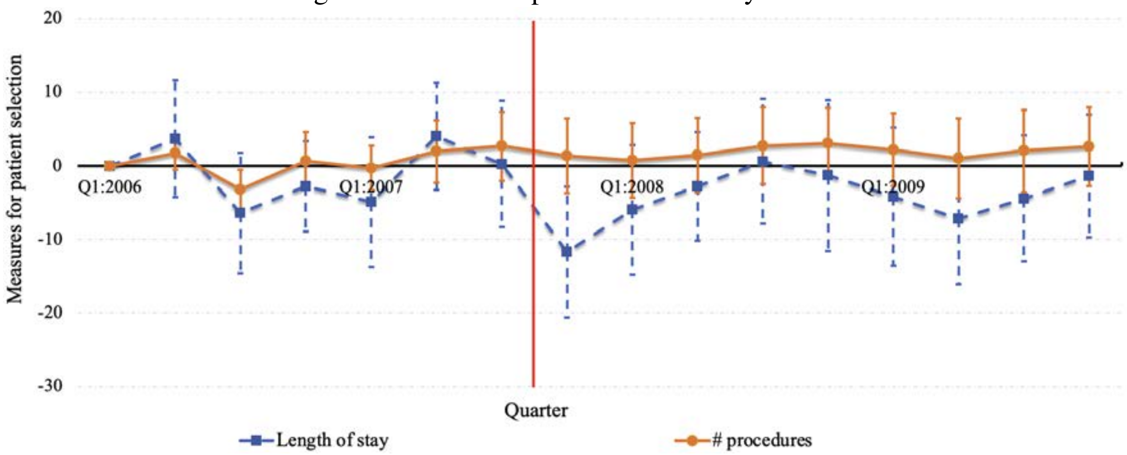
Note: The line reports the coefficients for early EMR adoption interacting with quarter dummies and each dot is a regression coefficient expressed as a percentage point. The red line represents the 2007 payment reform. The vertical lines show 95% confidence intervals, based on standard errors clustered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

Figure A4: 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 bottom DRG weight. The mean bottom DRG weight is calculated using the lowest weight for any base DRG. Coefficients are rescaled by 100 times for number of diagnoses and by 1000 times for mean bottom 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 bottom DRG weight. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects. We use hospital DRG fixed effects instead of hospital/base DRG fixed effects for mean bottom DRG weight.

Figure A5: Services provided with early EMRs



Note: Unit of observation is hospital/base DRG/quarter. Coefficients are rescaled by 100 times. Sample is all matched base DRGs from before and after the reform with multiple severity subclasses, from Q1:2006-Q4:2009. Standard errors are clustered at both hospital and base DRG levels. Other regressors are bed size, total outpatient visits, teaching hospital, total admissions, full-time physicians and dentists, percentage Medicare, percentage Medicaid, profit status, total number of births, quarter dummies and hospital/base DRG fixed effects.

Table A1: Extra top codes with early EMRs

	Dependent variable: Percentage of top codes within a base DRG	
	Coefficient	S.E.
	(1)	(2)
EMR×Quarter 1	.32	(.479)
EMR×Quarter 2	-.301	(.41)
EMR×Quarter 3	-.117	(.484)
EMR×Quarter 4 (Q1:2007)	-.547	(.522)
EMR×Quarter 5	-.352	(.685)
EMR×Quarter 6	-.517	(.551)
EMR×Quarter 7	2.06*	(1.21)
EMR×Quarter 8 (Q1:2008)	1.79	(1.21)
EMR×Quarter 9	1.92	(1.18)
EMR×Quarter 10	1.31	(1.19)
EMR×Quarter 11	2.62**	(1.14)
EMR×Quarter 12 (Q1:2009)	1.9	(1.3)
EMR×Quarter 13	1.62	(1.21)
EMR×Quarter 14	2.2*	(1.17)
EMR×Quarter 15	2.52**	(1.15)
Quarter 1	-.656	(.429)
Quarter 2	.0562	(.423)
Quarter 3	-.133	(.467)
Quarter 4 (Q1:2007)	.574	(.412)
Quarter 5	.373	(.604)
Quarter 6	.206	(.52)
Quarter 7	-49.1***	(3.48)
Quarter 8 (Q1:2008)	-47.3***	(3.49)
Quarter 9	-47.7***	(3.49)
Quarter 10	-47.1***	(3.54)
Quarter 11	-47***	(3.5)
Quarter 12 (Q1:2009)	-45.3***	(3.59)
Quarter 13	-45***	(3.62)
Quarter 14	-45.6***	(3.48)
Quarter 15	-44.9***	(3.49)
N	1,271,517	
p-value for joint significance of post-reform EMR coefficients	.0039	

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

Table A2: Extra top codes with early EMRs, by MED and SURG

	Dependent variable: Percentage of top codes within a base DRG			
	MED		SURG	
	Coefficient	S.E.	Coefficient	S.E.
	(1)	(2)	(3)	(4)
EMR×Quarter 1	.843**	(.407)	-.696	(.959)
EMR×Quarter 2	.0282	(.443)	-1.08	(.670)
EMR×Quarter 3	.139	(.470)	-.673	(.973)
EMR×Quarter 4 (Q1:2007)	.317	(.360)	-1.98**	(.967)
EMR×Quarter 5	-.245	(.518)	-.556	(1.52)
EMR×Quarter 6	-.301	(.469)	-1	(1.13)
EMR×Quarter 7	2.52**	(1.12)	-1.67	(1.34)
EMR×Quarter 8 (Q1:2008)	2.63**	(1.09)	-2.4	(1.45)
EMR×Quarter 9	2.82***	(.851)	-2.5*	(1.40)
EMR×Quarter 10	2.16**	(.978)	-3.04**	(1.21)
EMR×Quarter 11	3.35***	(1.03)	-1.37	(1.23)
EMR×Quarter 12 (Q1:2009)	2.58**	(1.22)	-1.81	(1.48)
EMR×Quarter 13	2.03*	(1.09)	-1.68	(1.33)
EMR×Quarter 14	2.64**	(1.09)	-1.12	(1.19)
EMR×Quarter 15	3.27***	(1.13)	-1.29	(1.10)
Quarter 1	-1.09**	(.429)	.24	(.801)
Quarter 2	-.306	(.516)	.868	(.622)
Quarter 3	-.369	(.451)	.405	(.970)
Quarter 4 (Q1:2007)	.00965	(.350)	1.66**	(.832)
Quarter 5	.39	(.517)	.479	(1.31)
Quarter 6	.112	(.476)	.58	(1.06)
Quarter 7	-55.8***	(3.90)	-36.7***	(3.44)
Quarter 8 (Q1:2008)	-54.1***	(3.94)	-34.9***	(3.65)
Quarter 9	-54.8***	(3.74)	-34.7***	(3.73)
Quarter 10	-54.3***	(3.79)	-34.1***	(3.67)
Quarter 11	-53.6***	(3.91)	-34.9***	(3.80)
Quarter 12 (Q1:2009)	-51.6***	(4.23)	-33.9***	(3.62)
Quarter 13	-51.2***	(4.21)	-33.8***	(3.87)
Quarter 14	-51.7***	(4.05)	-34.5***	(3.66)
Quarter 15	-51.1***	(4.00)	-33.5***	(3.73)
N	1,271,517			
p-value for joint significance of post-reform EMR coefficients	.0044		.0395	

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

Table A3: Extra top codes with spread, by hospital type

	Dependent variable: Percentage of top codes within a base DRG	
	Profit status	Financial health status
	(1)	(2)
Spread×Pre	0.660***	0.611***
	(0.107)	(0.091)
Spread×Post	-2.934	-2.924
	(2.007)	(2.124)
ForProfit×Spread×Pre	-0.046	
	(0.072)	
ForProfit×Spread×Post	-0.503	
	(0.642)	
NotForProfit×Spread×Pre	-0.073	
	(0.086)	
NotForProfit×Spread×Post	0.048)	
	(0.633)	
FinanciallyDistressed×Spread×Pre		-0.020
		(0.055)
FinanciallyDistressed×Spread×Post		-0.823*
		(0.426)
FinanciallyHealthy×Spread×Pre		0.047
		(0.070)
FinanciallyHealthy×Spread×Post		0.274
		(0.360)

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

Table A4: Extra adverse HACs with early EMRs

	Dependent variable: Percentage of patients w/ adverse HACs within a base DRG	
	Coefficient	S.E.
	(1)	(2)
EMR×Quarter 1	.0000106	(.0147)
EMR×Quarter 2	.0616*	(.0348)
EMR×Quarter 3	.0556***	(.0209)
EMR×Quarter 4 (Q1:2009)	.0551**	(.0256)
EMR×Quarter 5	.041	(.0276)
EMR×Quarter 6	.0633**	(.0286)
EMR×Quarter 7	.0873***	(.0275)
Quarter 1	.0296**	(.0148)
Quarter 2	.226***	(.0406)
Quarter 3	.146***	(.0299)
Quarter 4 (Q1:2009)	.194***	(.0351)
Quarter 5	.184***	(.0326)
Quarter 6	.214***	(.0378)
Quarter 7	.228***	(.0391)
N	1,928,450	
Mean DV	0.00256	
p-value for joint significance of post-reform EMR coefficients	.0385	

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