



Published in final edited form as:

J Health Econ. 2020 September ; 73: 102319. doi:10.1016/j.jhealeco.2020.102319.

Do hospitals respond to changing incentive structures? Evidence from Medicare's 2007 DRG restructuring

Amanda Cook^{a,*}, Susan Averett^b

^aDepartment of Economics, Bowling Green State University, OH, United States

^bDepartment of Economics, Lafayette College, PA, United States

Abstract

In 2007, the Centers for Medicare and Medicaid restructured the diagnosis related group (DRG) system by expanding the number of categories within a DRG to account for complications present within certain conditions. This change allows for differential reimbursement depending on the severity of the case. We examine whether this change incentivized hospitals to upcode patients as sicker to increase their reimbursements. Using the National Inpatient Survey data from HCUP from 2005 to 2010 and three methods to detect the presence of upcoding, our most conservative estimate is an additional three percent of reimbursement is attributable to upcoding. We find evidence of upcoding in government, non-profit, and for-profit hospitals. We find spillover effects of upcoding impacting not only Medicare payers, but also private insurance companies as well.

Keywords

Upcoding; Hospital reimbursement; Health care financing; Health insurance; Diagnosis related group

1. Introduction

In the United States, rising healthcare costs have concerned policymakers and the public for decades. The aging of the population and advances in medical technology are partly to blame, but concern has also centered on provider reimbursement. Prior to 1984, hospitals who treated Medicare patients were reimbursed on a fee-for-service basis, giving hospitals the incentive to provide excess services to patients to generate revenues. Believing the fee-for-service model to be unsustainable in terms of costs, in 1984 the federal government implemented the Prospective Payment System (PPS) for reimbursing hospitals for the care of Medicare patients. Under the PPS, a hospital receives a fixed payment for each Medicare patient in a given diagnosis related group (DRG) regardless of the actual cost the hospital incurs for treating that patient. Such systems have now been adopted by many other countries in an effort to control health care costs.

While the actual calculation is a bit more complicated, the DRG system basically works as follows. Each DRG is assigned a relative weight based on the average amount of resources

*Corresponding author. accook@bgsu.edu (A. Cook), averetts@lafayette.edu (S. Averett).

required to care for a patient in that DRG. The DRG weight is then multiplied by the hospital's reimbursement rate. This rate is set by the Center for Medicare and Medicaid Services (CMS) and adjusted for labor markets, geography, patient demographics, teaching program, and hospital size.¹ This calculation determines the total compensation of the hospital for the care of a particular patient. Overtime, the CMS has made changes to DRG weights in order to better reflect the resources used in, and hence the costs of, treating certain conditions. These DRG weights are calculated and published by the CMS annually. Importantly, these weights are prospective. The hospital is only reimbursed for the particular DRG assigned to a patient upon initial diagnosis even if that patient ultimately requires significantly more care.

In 2007, the CMS made a change in the methodology for calculating DRG weights in part based on a concern that the old weights were too coarse, e.g., all patients within a given DRG were not equally ill. With Version 25 of the DRGs (effective October 2007), the CMS revised the DRG scheme. The revision consisted of expanding what was up to 540 DRGs into over 700 Medicare-severity DRGs (MS-DRGs). First, DRGs were consolidated into 335 base MS-DRGs. Of these, 106 were split into two subgroups, and 152 were split into three subgroups, to arrive at 745 total MS-DRGs. The subgroups were based on the presence of complications or comorbidities (CCs) or major CCs (MCCs). This caused a change in the weighting system. For example, patients with a particular illness who had no complications and comorbidities would now be assigned a lower MS-DRG weight than those with comorbidities, to reflect their lower level of illness severity. We provide examples of how DRGs were expanded in Appendix A.

With these changes to the DRG classification system, there was a parallel change to the reimbursement hospitals received for treating patients of different severity levels. Prior to 2008, in the DRG system, patient illness variation within a DRG was not reimbursed differentially. This meant that a hospital would benefit financially when it admitted a patient in otherwise good health who was assigned DRG X, as this patient would, in expectation, require fewer services than the average patient assigned DRG X. Similarly, the hospital would lose financially when it admitted a very ill patient-with many underlying health complications-who also was admitted and assigned DRG X, as this patient would require more than the average amount of services for treating a patient with this illness illness X, but the hospital would be compensated for treating a patient with 'average' illness.

With the implementation of the MS-DRG classification system in 2007, hospitals were compensated at a lower rate for the less severe cases within a particular group of MS-DRGs and compensated more for the more severe cases. This move to the MS-DRG system sets up the central question of our research. In particular, after this transition to MS-DRGs, have we seen a systematic shift in the distribution of comorbidities or complications recorded by the hospital? Or posed another way, now that hospitals receive more compensation for severely ill patients, are patients more likely to get coded with additional comorbidities or

¹For more information visit Payment Details: <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/index.html>.

complications? If so, this would be an example of what in the economics literature is known as upcoding.

Upcoding can take various forms. Patients are usually coded into their DRGs by specialist coders based on medical charts and using special coding software. At the coder level, as Jürges and Köberlein (2015) note, there are legal, semi-legal, and illegal types of upcoding. Legal types of upcoding would be more accurate coding of a patient's condition so that the provider can be reimbursed appropriately for the costs incurred. This is the rationale underlying the change to the MS-DRGs. A semi-legal type of upcoding would include changing the primary and secondary diagnosis regarding comorbidities. Illegal forms of upcoding include adding comorbidities that are not documented in the medical charts or false documentation, i.e. the manipulation of patient charts so that patients appear to be sicker than they really are. The semi-legal and illegal types can in principle be detected in audits but the illegal types are harder to detect, in part because they are typically not coded by hospital administrators but by medical professionals; i.e. doctors or nurses.² In our case, we define upcoding as the systematic classification of patients into higher severity levels after the 2007 DRG change which, as noted, changed the financial incentives facing hospitals. Using data from the Healthcare Cost and Utilization Project's (HCUP) National Inpatient Survey (NIS) for the years 2005–2010, we find significant evidence of upcoding but, unlike previous research, we do not find it to be more prevalent at for-profit hospitals. In what follows, we situate our work within the existent literature. We then describe our data and methods and follow with results and a conclusion.

2. Previous literature

The financial incentives created by the PPS used by hospitals and governments in many countries have generated great interest in the potential for upcoding. Here we focus our review of the literature on both some of the most influential studies in this area as well as some of the most recent work.

In one of the first studies to undertake an examination of upcoding, Silverman and Skinner (2004) use data from US Medicare claims to focus on whether hospital ownership and market structure affected upcoding behavior from 1989 to 1997. They theorize that for-profit hospitals are more likely to take risks and that the goals of the administration and the clinical staff are more aligned in a for-profit environment and hence upcoding is more likely. To measure upcoding they focus on pneumonia and respiratory infections. At the time of their writing there were four DRGs that together defined general respiratory ailments-DRGs 79, 80, 89 and 90. DRG 79 had the highest weight so they measured the ratio of DRG 79 to the other 4 and were able to show that this ratio rose over time. They rule out two competing explanations for this rise-that the patients were becoming sicker over the time frame they study or that the coding was more accurate. As they hypothesized, upcoding was particularly pronounced at for-profit hospitals.

²Bauder et al. (2017) present a comprehensive survey on the state of healthcare upcoding fraud analysis.

Because DRG weights are recalibrated annually to reflect relative increases in costs, it is difficult to identify hospital responses to changes in DRG weights. Thus more recent literature focuses on various changes to DRG weights which are exogenous to hospitals. When such weights change, prices change and hence hospitals may respond to these price changes by moving patients to DRGs with higher weights. To this end, Dafny (2005) leverages a 1988 policy change that generated large price changes for a significant number of Medicare patients. In particular, in 1987 there were 473 individual DRG codes and 40 percent of those codes belonged to a “pair” of codes that shared the same diagnosis. Within each pair, the codes were separated by age restrictions and the presence of complications. In 1988 the age restrictions were lifted in an effort to more accurately reimburse hospitals for these diagnoses. The DRG weights were thus recalibrated accordingly.

Dafny leverages this change to detect upcoding by creating a variable which is the share of admissions to pair p in year t that is assigned to the top DRG code in that pair. She regresses this created variable on the difference between the DRG weight in the top code and the DRG weight in the bottom code for each pair (she terms this the DRG spread) and finds a positive and significant effect of higher spreads on the fraction coded in the higher weighted DRG. In other words, she finds that hospitals upcoded more for DRGs in which the price increase was greater. She argues that there was no increase in intensity or quality of care to explain the upcoding. Similar to Silverman, 2004, she reports that the response of for-profit hospitals was significantly higher than that of nonprofit or government-owned hospitals, indicating a greater willingness or ability of for-profit managers to exploit the DRG weight changes.

Others have followed the lead of Dafny (2005) and focused on whether exogenous changes in DRG weights lead to upcoding. Scholars report evidence of upcoding in hospitals in Norway (Anthun et al., 2017; Januleviciute et al., 2016; Melberg et al., 2016), Portugal (Barros and Braun, 2017) and Italy (Verzulli et al., 2017).

Changing the relative profitability of DRGs in a PPS may also cause hospitals to alter their case mix and this is distinct from upcoding. In this vein, Liang (2015) report that hospital caseloads in orthopedic departments increase with the profitability of these DRGs in Taiwan. This is distinct from upcoding because as it reflects a change in who the hospital is admitting. Similar work was undertaken by Januleviciute et al. (2016) using Norwegian data. Their results indicate that hospitals substitute towards more patients treated for medical DRGs which have increased reimbursement but not for surgical DRGs. In addition, they find that as the ratio of prices between patients with and without complications increases, the proportion of patients coded with complications rises.

In related work, Jürges and Köberlein (2015) study financial incentives to upcode in German hospitals. They note that in neonatology, there are substantial differences in reimbursement rates based on infant birthweight. Being just below a particular cutoff can command a much higher reimbursement rate. They document upcoding in the form of shifting birth weights from just above DRG-relevant thresholds to just below these thresholds where reimbursement is higher. A similar finding is found by Shigeoka and Fushimi (2014) using data from Japan where, similar to the German system, reimbursement levels increase sharply at a specific infant birthweight cutoff. They report an increase in manipulation of infant

birthweights and also find longer stays in the neonatal intensive care unit, both of which are associated with greater reimbursement. Reif et al. (2018) examine if this upcoding of neonates is possibly good for the neonates in that it may lead to better care. However, they find no evidence that this is the case.

Upcoding can occur in settings other than hospitals. For example, Geruso and Layton (2015) identified upcoding at the market level using risk scores and variations in financial incentives for physicians. In particular, they show that enrollees in private Medicare plans generate higher diagnosis-based risk scores than they would generate under fee-for-service Medicare, where diagnoses do not affect payments. Fang and Gong (2017) examine physicians hours spent on Medicare beneficiaries. They find that about 3 percent of physicians have billed Medicare what they term to be implausibly long hours concluding this is a form of upcoding. Bowlblis and Brunt (2014) test the hypothesis that there is upcoding in the skilled nursing industry by looking at how skilled nursing facilities respond to regional variation in reimbursement rates and report evidence of upcoding. Brunt (2011) investigates office visit upcoding using Medicare Part B data. Brunt finds evidence of upcoding when the use of a higher fee code is more likely to be approved; i.e. there is a small likelihood of being caught. He concludes, using these office visit codes, that physicians have considerable latitude when choosing the code they use, and upcoding is more likely when there is a larger fee differential. Gowrisankaran et al. (2016) focus on how the adoption of electronic medical records (EMRs) may lead to upcoding. This can occur if the EMR system allows physicians to more easily add diagnoses to patient records. However, the addition of more diagnoses may not be upcoding *per se* but may reflect more accurate coding. In fact, Gowrisankaran and colleagues do not find that EMRs lead to upcoding though they do find evidence that they lead to more complete coding.

We add to this literature in several ways. To our knowledge, we are the first to exploit the 2007 CMS change in DRGs to ascertain if hospitals were engaged in upcoding as a result of this change. Second, we use data on Medicare as well as private-pay patients to see if there is a spillover from Medicare to other payers in a hospital setting. Third, we use three methods to detect upcoding. First we exploit a unique opportunity in our data to measure upcoding as described in the next section. Next, we extend Dafny (2005)'s methodology from one where the variation is in the composition of individuals within a DRG to one where the variation is different DRGs for different levels of illness severity. Finally, we take advantage of the fact that some DRGs were expanded and others were not to test for upcoding using a difference-in-differences (DiD) specification.

3. Data description

To conduct our analyses, we use the Healthcare Cost and Utilization Project's (HCUP) national inpatient sample (NIS) from the years 2005–2010. This inpatient sample consists of approximately 7 million observations of hospital discharges annually from roughly 1,000 community hospitals. Weights are provided in the data to extrapolate to a national sample, as rare conditions are over represented to facilitate study. The NIS contains information on patient's illnesses, comorbidities, procedures (ICD codes) as well as patient DRG codes. An important feature of the data for our use is that HCUP retained the 2007 DRGs along with

the current year MS-DRG allowing us to make comparisons across these two DRGs and their associated weights.

Patient demographics are also included allowing us to control for race, gender and age in our models. We also have information on the type of admission (e.g. emergency, elective) and source of admission (e.g. emergency room, transfer from acute hospital) as well as payer. The NIS also includes hospital characteristics such as hospital ownership, number of beds, geographic region and whether the hospital is a teaching hospital.

To conduct our analyses, we merge in DRG weights for each year. These weights are publicly available from CMS.³ We focus on three years after the change for two reasons. First, it may take hospitals (administrators and clinical staff) some time to adjust to the new system. Second, the Affordable Care Act was passed in 2010 and we want to avoid confounding the change in the DRG system with other changes that occurred with that legislation. We extract a sample of individuals over age 18 for our analysis.

4. Method

We test for upcoding in three different ways. The first method involves using counterfactual DRGs (from 2007) to determine the change in weights pre and post reform for the same patients. The second method, the regression on weight differentials, uses a regression approach to measure the strength of the hospitals' response to increased payments. Finally, we run a difference-in-difference estimator for DRG groups that expanded in the reform (treated) and groups that were unaffected in the reform (control group). We describe each method in detail below.

4.1. Counterfactual method

For 2008–2010, the HCUP data provides two important pieces of information about patients' illness categories: the DRG they were assigned that year, the MS-DRG, and the DRG they would have been assigned if they had been a patient in 2007, the CMS-DRG. We show these two DRGs in Table 1 in columns (2) and (4) for six hypothetical patients. In columns (3) and (5), we have the 2008 weights from MS-DRGs and the 2007 weight associated with CMS-DRGs, respectively. While all patients in 2007 were assigned the same (hypothetical) DRG of 124, this DRG has been split into two MS-DRGs: 200 and 201 to reflect severity of illness.

Finally, in the last column of Table 1 we have, for a set of six hypothetical patients, the change in weights between the old and new methodology. We expect patients to have both positive and negative changes in weights. For DRG groups that have expanded to a low and high severity, the patients in the low severity group are less sick than the average patient in the old group. These patients would have a negative weight. For example, consider a

³Changes to the DRG weights occur in the last quarter of each calendar year. For simplicity of language, when we analyze the data, we discuss the results for a calendar year. In actuality, we are using the appropriate DRG versions for each year, which includes the last quarter of the previous year and the first three quarters of the current year. As such, when the policy change occurred in the last quarter of 2007, we observe that last quarter of 2007 and the first 3 quarters of 2008 under the DRG expansion, and refer to that four quarter period as '2008'.

DRG to which 6 patients have been assigned in 2007. Under the new MS-DRG system, 3 of them have been re-classified into a refined DRG with complications, MS-DRG number 200. These patients (1, 2 and 3) are the sicker individuals. Since they are sicker, their illness weight is higher than in 2007. Patients 4, 5, and 6 were classified into MS-DRG 201, a DRG without complications. Since these patients (4, 5, and 6) are 'less sick' - they don't have complications, the average number of services they need is smaller and as such, their MS-DRG weight is lower.

These changes in illness severity are captured in the last column of Table 1. We see some patients (patient 1, 2 and 3), who are now in the refined MS-DRG with complications (200), needing more services than in the previous system. They are the sicker patients under the CMS-DRG classification (124) and hence they have a .09 increase in weight. Similarly, we see patients 4, 5, and 6, our 'less sick' patients, require fewer services and this is quantified by a change in weight of $-.09$.

This highly simplified example is illustrative. While we expect both positive and negative changes in weight for any particular DRG refinement, on average, we would not expect the aggregate service intensity to change. Specifically, if we look at the mean of *Change in Weight* for the example above, we see it is 0 which is not surprising. The DRG weights are measuring average service provision. If we look over an entire CMS-DRG, we would in fact, *expect* the change in weight to have an average of 0. If we split our old group (CMS-DRG 124) up into a group with complications (MS-DRG 200) and a group without complications (MS-DRG 201), when we look at the whole group together (MS-DRG 200 and 201), we would still expect them to need the same services as when we had no refinements (CMS-DRG 124). Therefore, if we observe a positive change in average MS-DRG weight post reform, more patients are labeled as having complications than actually had complications in previous years. Since hospitals are reimbursed more when patients are in the high severity group, this would reflect upcoding.

Expanding on this example further illustrates this counterfactual method of measuring upcoding. Imagine that Patients 1 and 2 clearly have complications, patients 5 and 6 do not, and patients 3 and 4 are on the margin between having complications or not. In the CMS-DRG system, this distinction was not important. All patients were assigned to a single CMS-DRG and the weight for the group was determined by a national average of services for patients in the same CMS-DRG in the previous year compared to the average services for an average patient. With the refinement of 'complications' or 'no complications' patients 3 and 4 are now recorded as having complications, triggering a higher reimbursement rate for the hospital. As seen in Table 2, this counterfactual example now leads to an average change in weight of $.03$. This could be indicative of upcoding.

Fig. 1 provides an illustration of upcoding. Both the left and the right hand side panels show a hypothetical distribution of patients on a continuum of illness. For ease of exposition, assume we are focusing on one CMS-DRG which has been separated into two groups. As an example, this could be digestive malignancy which was split into two groups: with complications and without complications (MS-DRGs 172 and 173). The relatively sicker patients (those with complications) are shown in the solid red diamonds and the less sick

are the spotted blue diamonds. The horizontal line divides the patients into the MS-DRGs based on their level of severity. On the left hand side, we show an example with no upcoding after the implementation of the 2008 change to the MS-DRGs and the associated MS-DRG weights for their particular diagnoses. These weights average to 1 as we would expect them to given that we are merely doing a better job classifying patients-if the distribution of patient severity overall did not change between 2007 and 2008, we have no reason to believe that the weights would not sum to one.

However, those patients whose level of illness severity is close to the cutoff may be upcoded into the higher MS-DRG with the greater weight. If this is the case (as seen by two of the blue spotted diamonds in the left panel becoming red in the right panel) the DRG weights for this group will increase (in this case to 1.05) and this is what we are terming upcoding. Importantly, the distribution of patient illness has not changed.⁴ In the next section, we describe a more formal estimation strategy for detecting upcoding.

4.2. Regression on weight differentials

Our regression on weight differentials method for detecting upcoding takes advantage of the presence of both the CMS and the MS DRGs in the NIS 2008–2010 data and is similar to the strategy used by Dafny (2005). To implement this procedure, we first classify DRGs which have the same illness with different severities into groups. For example the three DRGs for pneumonia with no complications, pneumonia with some complications, and pneumonia with major complications would be identified as a group. Within each ‘group’, the DRGs are ranked for the level of the severity. In this example, pneumonia with no complications would be ranked as ‘low’ severity, pneumonia with some complications would be ranked as ‘medium’ severity, and pneumonia with major complications would be ranked as ‘high’ severity. Next we create variables, *difference in weight between DRGs w/o Complications and Comorbidities and with Complications and Comorbidities*, and *difference in weight between DRGs with Complications or Comorbidities and with Major Complications or Comorbidities* that measure the difference in DRG weights for each of the adjacent severities. For example, pneumonia with no complications had a DRG weight of .8 and pneumonia with complications or comorbidities had a DRG weight of 1.3, the *difference in weight between DRGs w/o Complications and Comorbidities and with Complications and Comorbidities* variable for all patients with any kind of pneumonia is $1.3 - .8 = .5$. Similarly, if pneumonia with major complications had a weight of 1.7, then the *difference in weight between DRGs with Complications or Comorbidities and with Major Complications or Comorbidities* variable would be $1.7 - 1.3 = .4$ for all patients with pneumonia. Once we have created our groups and the corresponding difference variables, we then estimate the following equation separately for 2008, 2009, and 2010:

$$\begin{aligned}
 \text{DRG weight year}_{i,g,h} = & \beta_0 + \beta_1 \text{dif in weight between} \\
 & \text{DRGs w/o CC and with CC}_g + \beta_2 \text{dif in weight} \\
 & \text{between DRGs with CC and with MCC}_g + \beta_3 2007 \\
 & \text{weight}_{i,g,h} + \beta_4 \text{Age}_i + \beta_5 \text{Female}_i + \beta_6 \text{Race}_i + \theta_h + \varepsilon_{i,g,h}
 \end{aligned} \tag{1}$$

⁴We provide some evidence that this assumption is warranted in Table 3.

where $DRG\ weight\ year_{i,g,h}$ is the DRG weight for patient i in group g at hospital h . The variable $2007\ Weight$ is the weight the patient would have been assigned if they were admitted in 2007. θ_h is a hospital fixed effect. Race controls for whether the patient self-reports as White, Black, Hispanic, Asian or Pacific Islander, Native American, or other. Age is a fixed effect by patient age and female is a dummy variable. Our standard errors are clustered by hospital. In the absence of upcoding, we would expect the estimates of β_1 and β_2 to be zero. If the estimates of β_1 or β_2 are positive, it means that when there is a larger difference between DRG weights more patients are classified in the sicker category. We expect that the coefficient on 2007 Weight will be close to 1 because if the patient requires a considerable number of services using the 2007 classification system, we expect that same patient to use more services in 2008⁵ classification system. We estimate Equation 1 separately by year rather than pooling years together because we are concerned about confounding upcoding with technological change.

4.3. Difference-in-difference estimation

For this analysis, we capitalize on the uneven nature of the expansion. While some DRG groups expanded in 2007, some DRGs remained the same pre and post reform. We use those DRGs that were not expanded as our control group and those DRGs that were expanded as our treated group. We then estimate the following equation:

$$Y_{i,d,h,t} = \beta_0 + \beta_1 Post_t * ExpDRG_d + \beta_2 2007\ weights_{i,d,h} + \beta_3 X_i + \theta_h + \tau_t + \delta_d + \varepsilon_{i,d,h,t} \quad (2)$$

where the dependent variable is equal to the DRG weight for patient i in DRG d in hospital h in year t . The interaction $Post_t * ExpDRG_d$ is an indicator variable equal to 1 if the DRG is in a group that expanded in the DRG reform and is in the post-expansion period. The variable $2007\ weights_{i,d,h}$ is the weight a patient *would have* been assigned if they had been admitted in 2007. X is a set of patient controls, including race, gender, and age. θ_h , τ_t , δ_d are hospital, time, and current year DRG fixed effects, respectively. Standard errors are clustered by hospital.

In the next section, we present results from these three methods.

5. Results

5.1. Counterfactual: yearly results

As noted, an advantage of our data is that for each patient we have both the MS-DRG to which they were classified and the CMS-DRG to which they would have been classified *if they had been admitted in 2007*.

A DRG weight measures the average services required for a particular DRG. A patient who is quite ill might have a DRG weight of 4. A patient who is not very ill might have a DRG weight of .5. As such, a larger DRG weight corresponds to an illness which is more service intensive to treat. This increase in service intensity over that of an average patient increases

⁵2009 or 2010.

the cost of treatment and the related reimbursement to the hospital. However, after the 2007 reform, an increasing average weight could be attributed to either (1) health providers responding to the new payment incentives (upcoding), (2) patients becoming sicker over time or (3) patients being categorized correctly into the finer categories (i.e. better coding of patients into the correct category). If the higher average DRG weight is attributable to an increase in patient illness then when we apply the weights from 2007 to the 2008, 2009, and 2010 data, we should see an increase in our average DRG weight both for the current year weights *and* the 2007 weights. However, this is not what we observe in the data. Post reform, the CMS-DRG weights are reasonably constant but the MS-DRG weights increase. This is suggestive of upcoding. We discuss these results in more detail below.

Table 3 shows summary statistics for the weight in the current year (MS-DRG weight), the weight the patient would have been assigned in 2007 (CMS-DRG weight) and the difference between those weights for the years 2005–2010 (note that 2007 is the year of the change and hence the reference year). In column (3), 2008, we see in that the average patient weight in 2008 was 1.344. If we apply the 2007 weights to these same patients, their average weight is 1.302. This approximately 3% difference in average weights is not attributable to patients being ‘sicker’, as the 2007 weights would pick up an increase in illness.⁶ Applying the 2007 weights to the 2008 patients and looking at the difference in average DRG weights, allows us to determine if year-by-year increases are attributable to a patient population in worse health or upcoding. In 2008, we see evidence of about 3% upcoding.⁷

We next turn our attention to 2009 (column 4). We see an average patient weight of 1.361 and a 2007 weight for the same patients of 1.302. Here again, we see evidence of upcoding, this time about 4.3%.⁸ Looking at the 2007 weights in 2008 and 2009, we see little evidence of systematically sicker patients. Comparing 2007 DRG weights across time, in 2008 and 2009, we see an average patient weight of 1.302. There appears to be no increase in ‘true average illness’ but there is a change in DRG weight of .0589, as seen in Column (4) line (3) of Table 3.

Looking at column (5), 2010, we again see a change in weight that is substantial, .0722. Looking from 2009 to 2010, there is a small increase in average patient illness seen by comparing the 2007 weights from 2009 and 2010. We see a difference in aggregate illness of .008(1.310 – 1.302). This suggests the patients are .6% sicker⁹ in 2010 than in 2009. However, we see they are recorded as 4.9% sicker *after* accounting for the small, but measurable, increase in aggregate patient illness level.¹⁰

To strengthen the argument that this counterfactual calculation captures upcoding and is not measuring a change in the underlying distribution of illness severity, we apply the

⁶Looking only at the 2007 weights applied to 2008, 2009, and 2010 we see very limited evidence of an aging or sicker population. From 2008 to 2010, the 2007 weights in the data are 1.302, 1.302, and 1.310. We see a small increase in average illness likely attributable to a change in case mix from 2009 to 2010 (1.302 to 1.310), but this is separate from the upcoding we see in this period, when we look at the difference between the 2007 and current year weights.

⁷This calculation is the ratio of *change in weight* and *current yr weight* for 2008. Namely, $0.0418/1.344$ is .0311.

⁸ $0.0589/1.361$.

⁹ $0.008/1.31$.

¹⁰ $4.9\% = ((.0722/1.310) - .006) * 100$.

same methodology of the counterfactual calculations looking back several years prior to the reform. We look at the corresponding weights for their respective DRGs in both the current year (2005 and 2006) and 2007. The difference is the *change in weight* variable in row 3 of Table 3. What we observe is that in 2005 and 2006 the change in weight is an order of magnitude smaller than in the post-reform years. The change in weight variable in 2005 is -0.00512 , suggesting that weights were 0.00512 lower on average than in 2007. Similarly in 2006, the *change in weight* variable was -0.00548 , indicating that in 2006 the average patient received a weight 0.00548 below that of a patient in 2007. Both of these reductions (from the 2007 weights) are less than $\frac{1}{2}$ of a percent.¹¹ When we compare this to the *change in weight* variable in 2008, we see a change of nearly tenfold, 0.0418 or 3.1%.¹² The results indicate a large increase in the change in weights post reform, and a near zero change in the pre-reform *change in weight* variable. This provides additional evidence that the underlying distribution of illness appears to be fairly stable over time.¹³

5.2. Regression of weight differentials

5.2.1. Results by year—Table 4 shows the results from estimating Equation 1 for 2008–2010.¹⁴ In the first column of Table 4, we focus on patients who were classified into a MS-DRG group with an illness or procedure with three levels of severity.¹⁵ Similarly, in the second column we analyze patients who were in a MS-DRG group with two levels of severity. We would expect that the coefficient on the variable, *2007 weights*, would be very close to one, which we see in both columns (1) and (2). If the coefficient on the *2007 weight* variable was not close to one, it would suggest that there was some unobservable structural change to the way coding occurred or a change in the way hospitals coded. In other words, we would expect consistency in the old methodology in the absence of a major change in coding procedures. Estimating a coefficient on the 2007 weights variable that is close to one adds validity to our methodology, as it suggests that the pre-reform weights are a good predictor of current weights.

The *difference in weight between DRGs w/o CC and with CC* variable measures the difference in weights for the MS-DRG with no complications or comorbidities and the MS-DRG with complications and comorbidities. Since MS-DRG weights are directly tied to reimbursement, as the difference in weights between levels of severity increases, the hospital has an increased financial incentive to nudge patients who are on the margin between the two severity levels to the more severe MS-DRG. They then receive higher levels of reimbursement. If hospitals were not responding to the difference in weights between consecutive severity levels, we would expect the *difference in weight between DRGs w/o CC and with CC* coefficient to be 0. In both columns (1) and (2), we see evidence that

¹¹ $-0.00507/1.145$ and $-0.00549/1.348$.

¹² $0.0417/1.344$.

¹³We observe a change in average weight between 2005 and 2006. Further investigation revealed a lower average weight in both 2004 (unreported) and 2005. While there is a change in observed population health, we still observe the same trend in the relationship between the current year weight and the 2007 weights. Thus, this does not appear to be driving our results.

¹⁴These results are unweighted. We also estimated all of our models using sample weights. These results are available from the authors upon request and are quantitatively similar to those presented here.

¹⁵For example, one group would have 3 MS-DRGs for illness X. The ‘low’ severity MS-DRG would be illness X with no complications or comorbidities, the ‘mid’ range severity MS-DRG would be illness X with complications and comorbidities, and the ‘high’ severity MS-DRG would be illness X with major complications or comorbidities.

hospitals are responding to the increased reimbursements for the more severe MS-DRG, as the coefficients are 0.0431 and 0.127 respectively. These positive values indicate that as the difference in weights (and reimbursement) increases, patients are systematically getting recorded as sicker. While these coefficients may seem small, it is important to understand them in context.

From our summary statistics, we know the average patient weight in 2008 was about 1.34. The coefficient on *difference in weight between DRGs w/o CC and with CC* predicts the average patient¹⁶ is coded as 3.2% sicker¹⁷ when the difference in weights is 1. While we anticipate that the ‘marginal’ patients are being nudged as in Fig. 1, the effect is so substantial that we see an effect for all patients in expanded groups. Similarly in column (2), an average patient is recorded as being 9.4%¹⁸ sicker, when the difference in weights between the low and medium group is 1.¹⁹ Finally, the *difference in weight between DRGs with CC and with MCC* variable in column (1) measures the difference in weights between the medium and high severity cases within a MS-DRG group. Here we have evidence that the hospitals are responding to the difference in reimbursement for a ‘medium’ severity patient and a ‘high’ severity patient. This coefficient suggests that when the difference in weights between the medium and high severity MS-DRGs is 1, the average patient MS-DRG weight increases by 0.157, or roughly 11.7%.²⁰

We now turn our attention to the 2009 results found in Table 4. Column (1) displays estimation results from patients who were classified into a MS-DRG group with 3 severity levels. Column (2) shows estimation results groups with 2 severity levels. We see the variable *2007 weight* continues to be statistically significant for both specifications. This seems intuitive, as it suggests that the level of services needed for an illness in 2007 is predictive of the level of services you would need for an illness in 2009. Again, we see the *difference in weight between DRGs w/o CC and with CC* variable is positive and statistically significant in both columns (2) and (3). If hospitals were not responding to the increased incentives of the higher severity MS-DRGs, we would expect these coefficients to be close to 0. The bottom panel, 2010, in Table 4 shows a similar patterns of results. Again we see the *difference in weight between DRGs w/o CC and with CC* variable is statistically significant in both columns. Over time the coefficient for *difference in weight between DRGs w/o CC and with CC* has increased for patients who ended up in a MS-DRG group with 2 severity levels. In 2008, the *difference in weight between DRGs w/o CC and with CC* coefficient is 0.127, in 2009 it is 0.132, and in 2010 the coefficient is 0.198. For these groups with 2 severity levels we see both evidence of upcoding and that the coefficient is larger in each successive year. For patients in groups with 3 severity levels, there is consistent evidence of upcoding, as all coefficients are positive and statistically significant. However, compared to the coefficients in 2008, the coefficients in 2009 were of smaller magnitudes. In 2010, the coefficients increase again, but not to 2008 levels. While there

¹⁶In a MS-DRG group with 3 severity levels.

¹⁷0.0431/1.34.

¹⁸0.127/1.34.

¹⁹In column (2) there are only two severity levels for each MS-DRG group. Generally, these MS-DRG groups have a MS-DRG with no CC and one with CC.

²⁰.157/1.344.

is some variation from year to year, across the three years, we see a consistent pattern in the results suggesting both statistically significant and economically significant responses to changes in reimbursement rates.

Given that some scholars have found upcoding to be particularly prevalent at for-profit hospitals, we next disaggregate our results by hospital ownership and payer class.

5.2.2. Results by hospital ownership—Table 5 contains results for our estimation of Equation 1 disaggregated by hospital ownership for the years 2008, 2009, and 2010. For these three years we restrict our focus to MS-DRGs that have 3 levels of severity.²¹ In 2008, we observe that the coefficients on both the *difference in weight between DRGs w/o CC and with CC* and *difference in weight between DRGs with CC and with MCC* are quite consistent across all hospital ownership categories. *The difference in weight between DRGs w/o CC and with CC* ranges from 0.0421 to 0.0434 while the *difference in weight between DRGs with CC and with MCC* coefficient ranges from 0.153 to 0.158. Both coefficients for all hospital ownership categories show statistically and economically significant upcoding with very little variation across ownership category. This uniform response persists in 2009 and 2010. It is possible that this uniformity across ownership type is a reflection of hospitals having similar responses to the expansion of DRGs.

It is perhaps not surprising that we see upcoding at not-for-profit hospitals. These hospitals, while not maximizing shareholder profits, must remain financially solvent. As such, they may participate in upcoding. This is also consistent with the theory that not-for-profit hospitals are simply for-profit hospitals in disguise (Brickley and Van Horn, 2002).

5.2.3. Results by payer class—Table 6 shows our estimation of Equation 1 disaggregated by insurance payer class for the years 2008, 2009, and 2010. One contribution of this paper is that we look at multiple payer classes, while previous literature on upcoding in the hospital setting focuses generally on Medicare.²² For these three years we restrict our focus to DRGs that expanded to 3 levels of severity.²³ Across Medicare, Medicaid, and private pay patients,²⁴ we see economically significant upcoding in all three years. In 2009 and 2010 the privately insured patients are significantly more likely to be upcoded from low to medium severity than their government counterparts,²⁵ as seen by their higher coefficients on the difference in weight between DRGs w/o CC and with CC variable. Upcoding from CC to MCC, represented by the coefficients on the *difference in weight between DRGs with CC and with MCC* variable is fairly uniform across all payer classes and years.

5.3. Difference-in-difference estimation

Our last estimation strategy is to use a DiD estimator as described earlier. Recall that DRGs that did not expand in the reform act as our control group and DRGs that did expand are

²¹We have also analyzed the MS-DRGs that have two severity levels by hospital ownership and payer class. The results are in the Appendix in Tables 15 and 16.

²²Medicare and Medicaid both reimburse using a DRG system. Private insurance is a mix of DRG and fee-for-service payments.

²³We have also analyzed the DRGs that have two severity levels by payer type. The results are qualitatively similar and are available in Table 16 in the Appendix.

²⁴These classifications for payer class are from the variable PAY1 from the HCUP data.

²⁵We use a difference of means t-test to determine whether the coefficients are statistically different.

our treated group. Each model includes controls for race, sex, age, and has hospital, year, and current year DRG fixed effects. As seen in Equation 2, our dependent variable is the current year's DRG weight. As DiD estimation relies on parallel trends in the pre-reform period, we have plotted an event study where we have interacted the treated variable with each time period. This is shown in Fig. 2. This figure suggests that pre-trends do not appear to be driving our DiD results.

In Table 7, the effect of upcoding is seen on the coefficient for the *post*expanded DRG* variable, 0.0437. This suggests that on average *everyone* in an expanded DRG had an *increase* in DRG weight of 0.0437 after the reform. When we consider the distribution of patient illness, some patients may be close to a cutoff, for example between a DRG without CC and with CC, while others are not. This average increase in weight of 0.0437 is particularly notable when we consider that not all patients are close to a threshold where they could be upcoded.

5.3.1. Estimation of Equation 2: difference-in-difference estimation by hospital ownership category—We now estimate our DiD model by hospital ownership type. The results are in the top panel of Table 8. Our coefficient of interest is *post * expanded DRG*. We see that this coefficient varies from 0.0528 for the government hospitals to 0.0657 at the not-for-profit hospitals. For all types of hospital ownership, there is a positive and statistically significant impact on average DRG weights for patients in DRG groups which expanded in the reform. Again, a consistent pattern of upcoding emerges. Patients in groups that expanded have weights which are systematically higher than patients in groups which did not expanded in the reform. Interestingly, the most upcoding occurs in both for-profit and not-for-profit hospitals and the least in government hospitals.

5.3.2. Difference-in-difference estimation by payer class—In the bottom panel of Table 8, we present the results for the DiD models across different payer classes. Again we are interested in the *post*expanded DRG* variable. For both Medicare and Private Insurance the coefficient is positive and statistically significant. On a per-patient basis, Medicare patients experience the most upcoding with an increase in DRG weight of 0.0498 for patients in a DRG group that expanded severity level in the reform. Privately insured individuals in a DRG group that expanded saw an average increase in weight of 0.0436. Medicaid patients did not have a statistically significant increase in weights for patients in a DRG group that expanded. Given that the reform was to the Medicare payment method, it is not surprising that this payer class saw the biggest increase in upcoding. We do however, also see 'spillover' effects to private payers. These results are in line with Clemens and Gottlieb (2017), Clemens et al. (2017), and Cooper et al. (2019) who also find spillover effects across payer classes.

6. Threats to identification

We attribute the evidence presented thus far as evidence of upcoding. There are at least five other potential explanations for the results that we report. The first is that in the few years following the MS-DRG implementation patients have become sicker. Our counterfactual analysis indicates that this is not likely to be the case.

Second, the impact of technological change may be conflated with upcoding. Technology's impact is ambiguous as it could raise or reduce the services needed to treat a patient, which in turn raises or lowers the DRG weight. For example, if a new technology allows for a faster, less invasive surgery with reduced recovery time in the hospital, the procedure with the new technology may be less service intensive than previously, and thus we may see a drop in the DRG weight as the new technology becomes widely adopted. Alternatively, a new technological breakthrough might allow us to extend the life of a patient who otherwise would have died. In this case the DRG is likely to become more service intensive and command a higher DRG weight. Thus, it is not clear that technological change will bias our results in a particular direction. Additionally, in our DiD estimation, we partially control for technological innovation, as some technological improvements will also impact DRG groups that did not expand, although we would be concerned about disproportionate technological change in groups that expanded in the three years post-reform. We also try to minimize the impact of technology by examining a short time frame after the implementation of the MS-DRG system.

Third, post-reform, one might imagine patients being coded 'more intensely' to try to increase the complications and comorbidities associated with a MS-DRG. We are measuring this as upcoding. However, when we apply this patient record with the more intense coding to the 2007 methodology, this patient may be classified into a more severe CMS-DRG. If this patient had been admitted in 2007, the incentive for coding intensity may have been reduced, and so we may be overestimating the true illness of the patients in 2007. This would then underestimate the changes associated with the reform and attenuate our results.

Fourth, one might also be concerned that variation in the case mix of the hospitals over the six year sample is driving our results. To provide suggestive evidence that this is not the case, we focus on a panel of hospitals on the assumption that case mix is reasonably stable over a short period of time within hospitals. We select those hospitals who have been sampled for 4, 5, or 6 years between 2010 and 2016. Table 9 includes the full sample in column (1), a panel of hospitals sampled at least 4 of the 6 years in column (2), and a panel of hospitals which appear in at least 5 of the 6 years in column (3). In all three columns the *post * expandedDRG* coefficient is significant. In the full sample (column (1)) and the panel of 4 or more hospitals (column (2)), we see coefficients which are quite similar. In the final column, the *post * expandedDRG* coefficient is somewhat smaller, but the number of observations in the final column is only about 2% of the overall sample, and only represents 25 total hospitals, thus these results should be viewed with caution. To the extent that using a panel mitigates variation in case mix across hospitals our results do not appear to be driven by case mix changes.

Lastly, one may be concerned about confounding reforms that also occurred between 2005 and 2010. For example, in 2008, Medicare announced that it would stop reimbursing hospitals for 8 conditions which were associated with poor hospital care-Hospital Acquired Conditions (HAC) (CMS, 2008).²⁶ Hospitals may increase the frequency or intensity of upcoding to cover the financial losses the hospital now faces if a patient has a HAC. Table

²⁶A list of these HACs is available at [dphh.nv.gov/Programs/SER/Docs/2008-10_hospital-acquired_conditions_\(HACs\)](http://dphh.nv.gov/Programs/SER/Docs/2008-10_hospital-acquired_conditions_(HACs)).

10 estimates Equation 2 for both the full sample of patients (column 1) and the sub sample of patients who did not have a HAC (column 2). We observe that the full sample has a coefficient for *post * expandedDRG* of 0.0437, which would suggest that the impact of being in a DRG that expanded in the post expansion period is an increase to the current year DRG weight of 0.0437. If we exclude those cases with HAC (column 2), we estimate a very similar result for the *post * expandedDRG* coefficient, 0.0427. Comparing columns (1) and (2), these results suggest that the HACs – and the differential payments for them – are not driving the upcoding we see.

7. Robustness checks

While the majority of our analysis occurs with a patient record being the level of observation, in Table 11 we consider alternative levels of observation. First we re-estimate Equation (2), after collapsing the data so that the level of observation is the MS-DRG-Hospital-Year. A MS-DRG treated in hospital *h* in year *y* becomes the level of observation, regardless of how many times that illness was treated. This functionally acts as a weighting mechanism for the data. In previous estimations, where the level of observation was at the patient level, each observation represented an occurrence of an illness in a hospital. In the MS-DRG-Year-Hospital level of observation, each MS-DRG is weighted equally. If one patient was classified into a DRG or 1000 patients were classified into a DRG, both would constitute one observation. This specification ignores relative illness prevalence. However, we still see a positive and significant coefficient on *post * expandedDRG*. Second, in column (2), the level of observation is now averaged across hospitals. This weights each MS-DRG in each year equally. This specification also does not reflecting illness prevalence or relative size of hospitals. Here *post * expandedDRG* is not significant.

In Table 12, we illustrate the robustness of our results when we include alternative and additional fixed effects to those in Equation 2. While Equation 2 contains fixed effects for hospital, year, and MS-DRG, in columns (2)(4), we include interacted hospital-by-year fixed effects, interacted MS-DRG-by-year fixed effects, and interacted hospital-by-MS-DRG fixed effects. We find that in all specifications that there is minimal variation on our *post * expanded DRG* coefficient. Each specification predicts between a 0.043 to 0.048 increase in DRG weight for patients in groups in an expanded DRG post reform.

To alleviate concerns that our results arose from variation between hospitals selected as part of the NIS sampling procedures,²⁷ we replicate Table 3, our counterfactual analysis, using a subsample of hospitals which appear in our panel at least 4 of the 6 years. These results are in Table 13. Looking at the mean of 2007 *weights* across years, we see very little variation between 2006 and 2010. This consistency, using the old weighting system, is suggestive evidence that hospitals did not change their admitting behavior substantially post-reform. If they had, their average case mix (represented by the 2007 *weights* variable) would have changed. The 2007 *weights* variable allows us to isolate the true case mix from

²⁷We include hospital-specific fixed effects to mitigate concerns that our results are due to favorable NIS sampling variation in the post-reform years.

the reported case mix in the MS-DRG system and quantify the financial incentives provided from upcoding.

In the time frame of this study, there was an increase in the frequency of outpatient surgical cases. Previously, many surgical patients would have been admitted to the hospital but now many are having surgery on an outpatient basis. As DRG weights are the ratio of average service intensity for the particular DRG to the average service intensity over all DRGs, one might worry that upcoding results are being driven by an increase in the average service intensity for surgical patients. Some DRGs are medical (treating illnesses) and some are surgical (involving surgical procedures). An increase in outpatient surgery would increase the relative service intensity for procedures, and would reduce weights for treating illnesses. We would anticipate non-surgical medical procedures being assigned a lower DRG weight as the shift to outpatient surgery becomes more pronounced. The service provision for medical DRGs would remain constant, and the average service provision for surgical DRG patients would increase. In Table 14, for the medical cases, we have a positive and statistically significant increase in our DiD coefficient, *Post * Expanded DRG*. This is suggestive evidence that our results are not from a change in case mix associated with a shift of healthier surgical patients to an outpatient setting. In fact, one might imagine that this shifting of surgical patients to the outpatient setting would attenuate the results for medical patients, as the average service intensity of surgical patients (and thus the average for all patients) would increase.

8. Conclusions

Upcoding has attracted considerable attention by policymakers and economists. Here we add to the literature on upcoding by leveraging a CMS change in 2007 that implemented the MS-DRG system. This change allowed for a refinement of patient illness categories and their associated reimbursements. We use three methods to determine if hospitals responded to the new MS-DRG weights by systematically upcoding patients. Our results indicate evidence of upcoding, and are in line with others who have documented upcoding by hospitals. To the best of our knowledge, we are the first to use the 2008 change to the MS-DRG system as a way to identify upcoding.

While all of our estimates suggest upcoding is occurring, each estimation method measures upcoding in a different way. This naturally leads to some variation in our estimation of the magnitude of upcoding. In the counterfactual estimation, we see aggregate estimates of weights increase between 0.0418 and 0.0722 for the three post-reform years. Our estimates in the regression on weight differentials section focus on patients in DRG groups that had 3 severity levels. There we find that when there was a difference of 1 in the weights between severity levels, we estimate an increase in the patients' DRG weight between 0.0214 and 0.198. Finally when we do the DiD estimation, we see that patients in groups which expanded in the post-reform period were coded with weights 0.0437 higher. In other words, patients in a DRG group that was expanded have an average DRG weight that is 0.0437 greater than their non-expanded counterparts.

Notably, we use data from community hospitals in our analysis. In general, we think of these hospitals as seeing the easier cases with the more complex cases being promoted up the referral pyramid. Hence, we view our findings as a lower bound for upcoding. Teaching hospitals and academic medical centers treat patients with more complex case profiles, where the ability to upcode by nudging a patient to a higher severity DRG might be greater given the generally higher case mix at these hospitals.

Our estimates of upcoding are not only statistically significant, they are also economically meaningful as they correspond to a sizable dollar amount of excess reimbursement. Using the lower estimate for upcoding, we observe a 3% markup as the result of the MS-DRG system. In 2008, expenditures on hospital care in the United States were approximately 730 billion dollars.²⁸ Using our lower bound estimate of 3% suggests that about 20 billion dollars in excess payments were attributable to upcoding.²⁹ It is unclear if any of these additional expenditures lead to better health outcomes. It is clear patients are receiving more services, what is unclear is if patients are truly sicker or sicker on paper.

Given the potential costs of upcoding, it is not surprising that in the United States legislation has been introduced to curb such behavior (see e.g. Angeles and Park, 2009). Indeed, the costs of policing Medicare fraud are estimated to be 2.5 million dollars annually (Perez and Wing, 2018). However, increased spending to detect fraud in Medicare could have both desirable deterrence effects and undesirable defensive medicine effects.

Further research can go in many directions. For example, future work should investigate further the potential role of technological change in estimating the extent of upcoding. Second, while one might hypothesize that those patients who are upcoded might benefit if they indeed received more intensive treatment, it is possible that additional treatment could be harmful to a patient. Thus, an examination of the effect of upcoding on patient outcomes would be useful. Third, we provide some evidence that payer type is a predictor of upcoding, but what is still unknown is how, conditional on a particular illness, payer type influences upcoding. Finally, an avenue for further research would be to examine which patients are most likely to be upcoded. In particular, future work should focus on identifying the ‘marginal’ patients who are getting upcoded. Understanding which patients are more likely to be upcoded will also be useful in forming future regulatory frameworks that are less susceptible to such manipulation.

Acknowledgments

The authors gratefully acknowledge the suggestions and contributions from Eric Roberts, Hoda Nouri Khajavi, and Hau Nguyen as well as the University of Toledo and Xavier University’s economics seminar series and participants in sessions at the 2019 MEA and 2018 ASHE meetings.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

²⁸Hospital spending in 2008 obtained by [Statisa.com](https://www.statista.com).

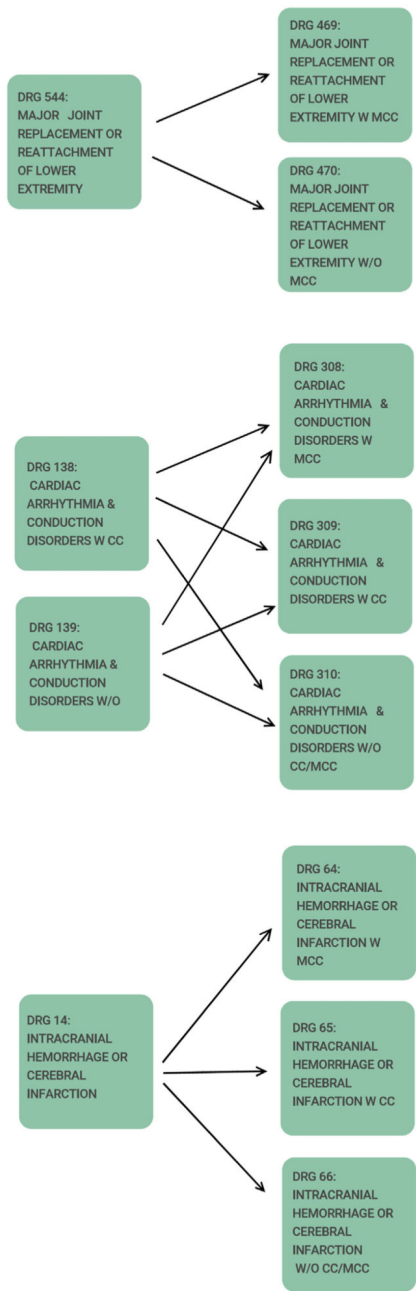
²⁹730 billion divided by 1.03 = 708 billion.

APPENDIX

Appendix

Appendix A. Illustration of DRG groups pre- and post-reform

2007 Group Post Group



Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Appendix B. Replication of Table 5's hospital ownership analysis for DRG groups that expanded from 1 to 2 levels of severity

Table 15

Estimation of Eq. 1 for 2008, 2009, 2010 by hospital characteristics for DRG groups that expanded from 1 to 2 levels of severity: $Y = \text{MS-DRG weight in the current year}$.

	Government	Not For Profit	For Profit
2007 weights	0.954 ^{***} (0.0171)	0.916 ^{***} (0.00648)	0.894 ^{***} (0.00739)
Difference in weight between DRGs w/o CC and with CC	0.0936 ^{***} (0.00849)	0.133 ^{***} (0.00728)	0.140 ^{***} (0.0142)
	Government	Not For Profit	For Profit
2007 weights	0.951 ^{***} (0.0174)	0.914 ^{***} (0.00617)	0.921 ^{***} (0.00889)
Difference in weight between DRGs w/o CC and with CC	0.0993 ^{***} (0.0262)	0.142 ^{***} (0.0114)	0.107 ^{***} (0.0173)
	Government	Not For Profit	For Profit
2007 weights	0.910 ^{***} (0.0142)	0.890 ^{***} (0.00701)	0.901 ^{***} (0.00896)
Difference in weight between DRGs w/o CC and with CC	0.166 ^{***} (0.0310)	0.212 ^{***} (0.0149)	0.154 ^{***} (0.0158)

Notes: All models control for age, race, and gender. CC refers to complications and comorbidities, MCC refers to major complications and comorbidities. We base these classifications for hospital types on the variable HOSP CONTROL from the HCUP data. Standard errors are in parentheses and clustered at the hospital level.

*
p<0.05

**
p<0.01

p<0.001.

Appendix C. Replication of Table 6's Payer Class analysis for DRG groups that expanded from 1 to 2 levels of severity

Table 16

Estimation of Equation 1 for 2008, 2009, 2010 by Payer Class for DRG groups that expanded from 1 to 2 levels of severity: $Y = \text{MS-DRG weight in the current year}$.

2008	Medicare	Medicaid	Private
2007 weights	1.004 ^{***} (0.00406)	0.977 ^{***} (0.00496)	0.917 ^{***} (0.00520)
Difference in weight between DRGs w/o CC and with CC	0.00280 [*] (0.00123)	0.0109 ^{***} (0.00238)	0.0590 ^{***} (0.00439)
Observations	2167496	646608	1599540
2009	Medicare	Medicaid	Private

2008	Medicare	Medicaid	Private
2007 weights	1.003 *** (0.00344)	0.972 *** (0.00552)	0.922 *** (0.00423)
Difference in weight between DRGs w/o CC and with CC	0.00206 (0.00142)	0.0132 *** (0.00312)	0.0696 *** (0.00488)
Observations	2170762	715228	1515305
2010	Medicare	Medicaid	Private
2007 weights	1.004 *** (0.00449)	0.969 *** (0.00687)	0.918 *** (0.00568)
Difference in weight between DRGs w/o CC and with CC	0.00457 ** (0.00158)	0.0155 *** (0.00314)	0.0712 *** (0.00543)
Observations	2290632	783486	1530095

Notes: All models control for age, race, and gender. CC refers to complications and comorbidities, MCC refers to major complications and comorbidities. We base these classifications for payer class on the variable PAY1 from the HCUP data. Standard errors are in parentheses and clustered at the hospital level.

*
p<0.05
**
p<0.01

p<0.001.

References

- Angeles J, Park E, 2009. Upcoding problem exacerbates overpayments to medicare advantage plans. Center on Budget and Policy Priorities <http://www.cbpp.org/cms/index.cfm>.
- Anthun KS, Bjørngaard JH, Magnussen J, 2017. Economic incentives and diagnostic coding in a public health care system. *Int. J. Health Econ. Manage* 17 (1), 83–101.
- Barros P, Braun G, 2017. Upcoding in a national health service: The evidence from Portugal. *Health Econ.* 26 (5), 600–618. [PubMed: 26988634]
- Bauder R, Khoshgoftaar TM, Seliya N, 2017. A survey on the state of healthcare upcoding fraud analysis and detection. *Health Serv. Outcomes Res. Methodol* 17 (1), 31–55.
- Bowblis JR, Brunt CS, 2014. Medicare skilled nursing facility reimbursement and upcoding. *Health Econ.* 23 (7), 821–840. [PubMed: 23775721]
- Brickley JA, Van Horn RL, 2002. Managerial incentives in nonprofit organizations: Evidence from hospitals. *J. Law Econ* 45 (1), 227–249.
- Brunt CS, 2011. CPT fee differentials and visit upcoding under Medicare Part B. *Health Econ.* 20 (7), 831–841. [PubMed: 20681033]
- Clemens J, Gottlieb JD, 2017. In the shadow of a giant: Medicare's influence on private physician payments. *J. Polit. Econ* 125 (1), 1–39. [PubMed: 28713176]
- Clemens J, Gottlieb JD, Molnár TL, 2017. Do health insurers innovate? evidence from the anatomy of physician payments. *J. Health Econ* 55, 153–167. [PubMed: 28784289]
- CMS, 2008. Hospital acquired conditions in acute inpatient prospective payment system hospitals. [dphh.nv.gov/Programs/SER/Docs/200810_hospital-acquired_conditions_\(HACs\)](http://dphh.nv.gov/Programs/SER/Docs/200810_hospital-acquired_conditions_(HACs)).
- Cooper Z, Craig SV, Gaynor M, Van Reenen J, 2019. The price aint right? hospital prices and health spending on the privately insured. *Quart. J. Econ* 134 (1), 51–107.
- Dafny LS, 2005. How do hospitals respond to price changes? *Am. Econ. Rev.* 1525–1547. [PubMed: 29125726]

- Fang H, Gong Q, 2017. Detecting potential overbilling in Medicare reimbursement via hours worked. *Am. Econ. Rev* 107 (2), 562–591. [PubMed: 29553228]
- Geruso M, Layton T, 2015. Upcoding: evidence from Medicare on squishy risk adjustment. National Bureau of Economic Research, Technical report.
- Gowrisankaran G, Joiner KA, Lin J, 2016. Does health it adoption lead to better information or worse incentives? National Bureau of Economic Research, Technical report.
- Januleviciute J, Askildsen JE, Kaarboe O, Siciliani L, Sutton M, 2016. How do hospitals respond to price changes? Evidence from Norway. *Health Econ.* 25 (5), 620–636. [PubMed: 25929559]
- Jürges H, Köberlein J, 2015. What explains DRG upcoding in neonatology? the roles of financial incentives and infant health. *J. Health Econ* 43, 13–26. [PubMed: 26114589]
- Liang L-L, 2015. Do diagnosis-related group-based payments incentivise hospitals to adjust output mix? *Health Econ.* 24 (4), 454–469. [PubMed: 24519749]
- Melberg HO, Olsen CB, Pedersen K, 2016. Did hospitals respond to changes in weights of Diagnosis Related Groups in Norway between 2006 and 2013? *Health Policy* 120 (9), 992–1000. [PubMed: 27567179]
- Perez V, Wing C, 2018. Should we do more to police Medicaid fraud? Evidence on intended and unintended consequences of expanded enforcement. *Am. J. Health Econ*
- Reif S, Wichert S, Wuppermann A, 2018. Is it good to be too light? birth weight thresholds in hospital reimbursement systems. *J. Health Econ* 59, 1–25. [PubMed: 29627674]
- Shigeoka H, Fushimi K, 2014. Supplier-induced demand for newborn treatment: Evidence from Japan. *J. Health Econ* 35, 162–178. [PubMed: 24709038]
- Silverman E, Skinner J, 2004. Medicare upcoding and hospital ownership. *J. Health Econ* 23 (2), 369–389. [PubMed: 15019762]
- Verzulli R, Fiorentini G, Lippi Bruni M, Ugolini C, 2017. Price changes in regulated healthcare markets: Do public hospitals respond and how? *Health Econ.* 26 (11), 1429–1446. [PubMed: 27785849]

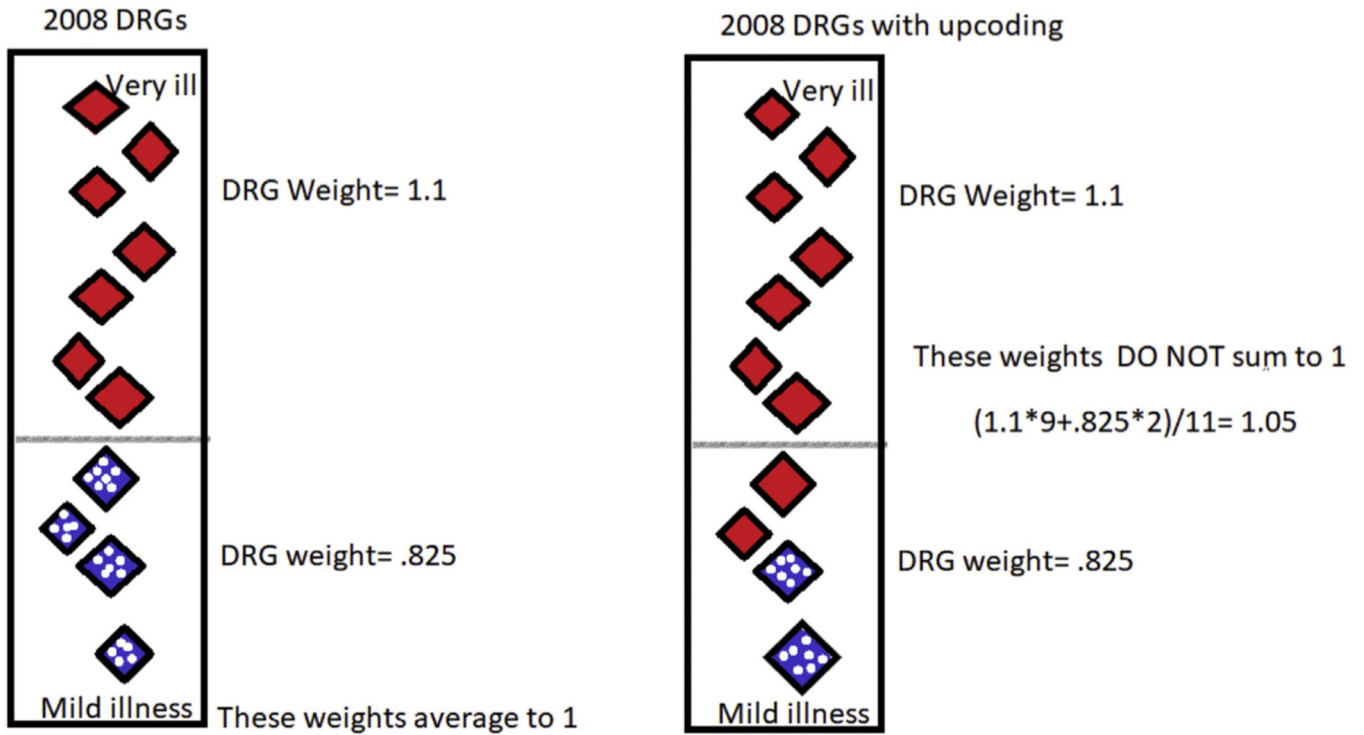


Fig. 1. Upcoding Illustration.

Notes: In this figure, we are considering that within each CMS-DRG there is a continuum of illness. The patients who are the most ill –i.e. those who require the most services to treat their illness– are at the top of the column; those that are the least ill are at the bottom. Here we imagine that there is a cutoff, depicted by the horizontal gray line partway up the column, which delineates those patients who are appropriately classified into the high severity MS-DRG (at the top in red) compared to those who are appropriately categorized in the less severe MS-DRG. Theoretically, we are hypothesizing that the patients who are being upcoded are those who are close to the true threshold.

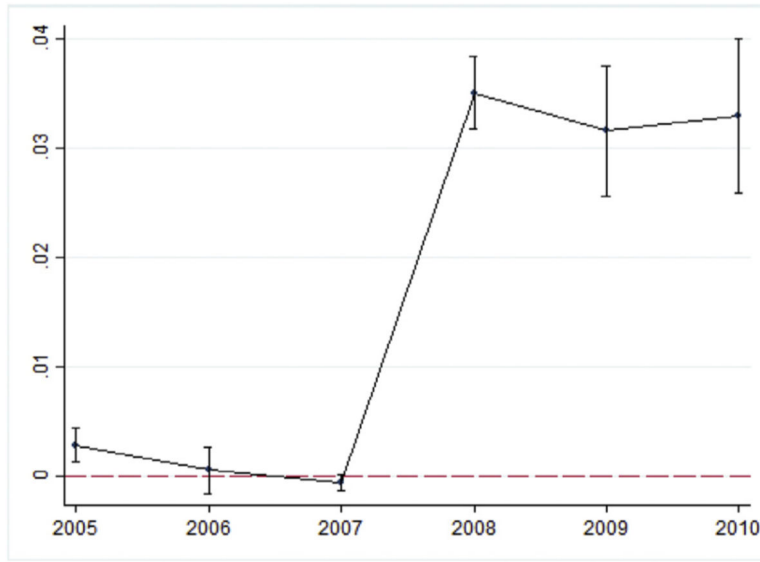


Fig. 2. Event Study Plot of coefficients on Expanded DRG Dummy by Year.

Notes: On the Y axis we have the magnitude of the coefficients from the event study estimation. $Y_{i,d,h} = \beta_0 + \beta_1 Year_t * ExpDRG_d + \beta_2 2007\ weights_{i,d,h} + \beta_3 X_i + \theta_h + \tau_t + \epsilon_{i,d,h}$ where $Y = DRG\ weight\ year_{i,g,h}$; the DRG weight for patient i in group g at hospital h in year 2005–2010. X contains patient specific characteristics: Age, race, and sex. θ_h , τ_t are fixed effects for hospitals and years, respectively. The $Year_t * ExpDRG_d$ variable is our event study coefficient, representing the change in DRG weight in the current year for patients in an expanded DRG group pre and post reform.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Hypothetical patients and counterfactual example.

2008	MS-DRG (Post 2007)	MS-DRG Weight	CMS-DRG (2007 and before)	CMS-DRG Weight	Change in Weight
Patient 1	200	1.3	124	1.21	.09
Patient 2	200	1.3	124	1.21	.09
Patient 3	200	1.3	124	1.21	.09
Patient 4	201	1.12	124	1.21	-.09
Patient 5	201	1.12	124	1.21	-.09
Patient 6	201	1.12	124	1.21	-.09
Average Change in weight:					0

Notes: This table illustrates for 6 hypothetical patients how, in the absence of upcoding, we would not expect a change in average DRG weight. The column titled MS-DRG weight shows the weight associated with the new DRG post reform. The CMS-DRG weight column shows the DRG the patient would have been assigned if they had been admitted prior to the DRG expansion. The CMS-DRG Weight column shows the weight associated with the CMS-DRG. The column Change in Weight is the difference between MS-DRG Weights and CMS-DRG Weights. The Change in Weight column highlights that in the refinement of DRGs, some patients would be less sick than the average patient in the CMS-DRG, and thus would have negative weights post-reform. Other patients would be sicker than the average patient in the CMS-DRG, and would have a higher weight post-reform.

Table 2

Hypothetical patients and counterfactual example with upcoding.

2008	MS-DRG (Post 2007)	MS-DRG Weight	CMS-DRG (2007 and before)	CMS-DRG Weight	Change in Weight
Patient 1	200	1.3	124	1.21	.09
Patient 2	200	1.3	124	1.21	.09
Patient 3	200	1.3	124	1.21	.09
Patient 4	200	1.3	124	1.21	.09
Patient 5	201	1.12	124	1.21	-.09
Patient 6	201	1.12	124	1.21	-.09
Average Change in weight:					0.03

Notes: See notes to Table 1. Change in Weights is the difference between MS-DRG Weights and CMS-DRG Weights. In this example, Patient 4 was incorrectly classified as belonging to the more severe MS-DRG. As a result of upcoding patient 4, the average change in weight is positive (0.03).

Table 3

Weights and counterfactual weights from 2005 to 2010.

	2005	2006	2008	2009	2010	Total
2007 weights	1.149 (0.000658)	1.355 (0.000636)	1.302 (0.000501)	1.302 (0.000509)	1.310 (0.000514)	1.291 (0.000248)
current yr weight	1.144 (0.000676)	1.349 (0.000650)	1.344 (0.000502)	1.361 (0.000522)	1.382 (0.000527)	1.328 (0.000253)
change in weight	-0.00512 (0.0000393)	-0.00548 (0.000103)	0.0418 (0.000123)	0.0589 (0.000172)	0.0722 (0.000177)	0.0375 (0.0000653)

Notes: The variable 2007 weight is the CMS-DRG weight the patient would have been assigned, if the patient had been admitted in 2007. The variable current year weight represents the DRG weight of the patient in the year they were admitted. The change in weight variable is row 2 minus row 1, and this measures the difference between the DRG weight which was assigned to the patient and the DRG weight that would have been assigned if the patient had been admitted in 2007. In parentheses are the standard errors of the mean; standard deviation / square root of sample size.

Table 4

Estimation of Eq 1 for 2008, 2009, 2010: Y = MS-DRG weight in the current year.

	3 groups	2 groups
2008		
2007 weights	0.878*** (0.00225)	0.919*** (0.00563)
Difference in weight between DRGs w/o CC and with CC	0.0431*** (0.000863)	0.127*** (0.00563)
Difference in weight between DRGs with CC and with MCC	0.157*** (0.00245)	
2009		
2007 weights	0.911*** (0.00296)	0.918*** (0.00496)
Difference in weight between DRGs w/o CC and with CC	0.0214*** (0.000754)	0.132*** (0.00938)
Difference in weight between DRGs with CC and with MCC	0.104*** (0.00252)	
2010		
2007 weights	0.899*** (0.00369)	0.895*** (0.00542)
Difference in weight between DRGs w/o CC and with CC	0.0323*** (0.000958)	0.198*** (0.0122)
Difference in weight between DRGs with CC and with MCC	0.114*** (0.00267)	
Observations	2601342	2484981

Note: Estimation of Equation 1. All models control for age, race, and gender. The dependent variable is MS-DRG weight in the current year. CC refers to complications and comorbidities, MCC refers to major complications and comorbidities. Standard errors are in parentheses and clustered at the hospital level. Column (1), 3 groups, restricts the expanded MS-DRGs to illnesses that have 3 levels of severity (and 3 associated MS-DRGs) after 2007. (i.e. there are 3 MS-DRGs for the same illness, one which has no complications or comorbidities, the second which has complications and comorbidities, and the third which has major complications and comorbidities.) Column (2), 2 groups, restricts the expanded MS-DRGs to illnesses that have 2 levels of severity.

* p<0.05

.1000>d

10<0>d
**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Estimation of Eq 1 for 2008, 2009, 2010 by hospital ownership: Y = MS-DRG weight in the current year.

	Government	Not for Profit	For Profit	All
2008				
2007 weights	0.873 *** (0.00623)	0.878 *** (0.00270)	0.885 *** (0.00490)	0.878 *** (0.00225)
Difference in weight between DRGs w/o CC and with CC	0.0430 *** (0.00325)	0.0434 *** (0.000988)	0.0421 *** (0.00151)	0.0431 *** (0.000863)
Difference in weight between DRGs with CC and with MCC	0.153 *** (0.00529)	0.158 *** (0.00309)	0.155 *** (0.00459)	0.157 *** (0.00245)
2009				
2007 weights	0.889 *** (0.0131)	0.916 *** (0.00329)	0.902 *** (0.00665)	0.911 *** (0.00296)
Difference in weight between DRGs w/o CC and with CC	0.0236 *** (0.00319)	0.0206 *** (0.000821)	0.0231 *** (0.00204)	0.0214 *** (0.000754)
Difference in weight between DRGs with CC and with MCC	0.108 *** (0.00550)	0.101 *** (0.00296)	0.116 *** (0.00692)	0.104 *** (0.00252)
2010				
2007 weights	0.887 *** (0.0132)	0.904 *** (0.00396)	0.887 *** (0.0113)	0.899 *** (0.00369)
Difference in weight between DRGs w/o CC and with CC	0.0341 *** (0.00319)	0.0319 *** (0.00104)	0.0329 *** (0.00317)	0.0323 *** (0.000958)
Difference in weight between DRGs with CC and with MCC	0.120 *** (0.00594)	0.108 *** (0.00288)	0.145 *** (0.00987)	0.114 *** (0.00267)
Observations	379507	1827324	360268	2601342

Notes: All models control for age, race, and gender. CC refers to complications and comorbidities, MCC refers to major complications and comorbidities. We base these classifications for hospital ownership on the variable HOSP CONTROL from the HCUP data. The last column "all" estimates the same coefficients using the entire data set. Standard errors are in parentheses and clustered at the hospital level.

* p<0.05

.1000>d

10<0>d
**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6

Estimation of Eq 1 by Year and Payer Class: $Y = MS-DRG$ weight in the current year.

	Medicare	Medicaid	Private
2008			
2007 weights	0.937 *** (0.00279)	0.915 *** (0.00424)	0.845 *** (0.00477)
Difference in weight between DRGs w/o CC and with CC	0.0260 *** (0.000901)	0.0272 *** (0.00254)	0.0371 *** (0.00305)
Difference in weight between DRGs with CC and with MCC	0.135 *** (0.00447)	0.130 *** (0.00512)	0.134 *** (0.00469)
Observations	1349113	212371	669659
2009			
2007 weights	0.932 *** (0.00272)	0.903 *** (0.00488)	0.850 *** (0.00423)
Difference in weight between DRGs w/o CC and with CC	0.0246 *** (0.000772)	0.0287 *** (0.00189)	0.0416 *** (0.00185)
Difference in weight between DRGs with CC and with MCC	0.124 *** (0.00380)	0.119 *** (0.00449)	0.108 *** (0.00299)
Observations	1345626	231847	634182
2010			
2007 weights	0.932 *** (0.00292)	0.900 *** (0.00765)	0.844 *** (0.00542)
Difference in weight between DRGs w/o CC and with CC	0.0249 *** (0.000794)	0.0285 *** (0.00183)	0.0425 *** (0.00198)
Difference in weight between DRGs with CC and with MCC	0.117 *** (0.00319)	0.117 *** (0.00487)	0.112 *** (0.00359)
Observations	1424377	268965	658716

Notes: All models control for age, race, and gender. CC refers to complications and comorbidities. MCC refers to major complications and comorbidities. We base these classifications for payer class on the variable PAY1 from the HCUP data. Standard errors are in parentheses and clustered at the hospital level.

p>0.0001

p<0.01
**
p>0.05
*

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7

Estimation of Eq 2: difference-in-difference estimation: $Y = \text{MS-DRG weight in the current year}$.

	Full Sample
Post * expanded DRG	0.0437 *** (0.00242)
2007 weights	0.939 *** (0.00363)
Observations	26904427

Notes: This model controls for age, race, and gender, and year, DRG and hospital fixed effects. *Post * ExpandedDRG* is an indicator variable equal to 1 if the DRG group expanded in the DRG reform and is in the post expansion period. Standard errors are in parentheses and clustered at the hospital level.

*
p<0.05

**
p<0.01

p<0.001.

Table 8

Est. of Eq 2: difference-in-difference estimation by hospital & payer class: $Y = \text{MS-DRG weight in the current year.}$

Hospital Ownership			
	Government	Not For Profit	For Profit
Post * expanded DRG	0.0528 *** (0.0133)	0.0647 *** (0.00565)	0.0657 *** (0.00685)
2007 weights	0.870 *** (0.0168)	0.864 *** (0.0127)	0.864 *** (0.0122)
Observations	1353177	4000567	2304196
Payer Class			
	Medicare	Medicaid	Private Ins.
Post * expanded DRG	0.0498 *** (0.00252)	0.00572 (0.00340)	0.0436 *** (0.00305)
2007 weights	0.937 *** (0.00328)	0.944 *** (0.00494)	0.933 *** (0.00551)
Observations	11792191	4112707	8510458

Note: This model controls for age, race, and gender, and year, DRG, and hospital fixed effects. *Post * ExpandedDRG* is an indicator variable equal to 1 if the DRG group expanded in the DRG reform and is in the post expansion period. We base these classifications for hospital type and payer class on the variables HOSP CONTROL and PAY1 from the HCUP data, respectively. Standard errors are in parentheses and clustered at the hospital level.

*
p<0.05

**
p<0.01

p<0.001.

Table 9

Difference-in-difference estimation using a panel of hospitals. Y = DRG weight in current year.

	Full sample	4,5,6 years	5 or 6 years
Post * expanded DRG	0.0437 *** (0.00242)	0.0413 *** (0.00887)	0.0298 *** (0.00563)
2007 weights	0.939 *** (0.00363)	0.955 *** (0.0131)	0.918 *** (0.00368)
Observations	26904427	2627929	368379

Notes: This table estimates Equation 2 but only on those hospitals that appear in the sample for at least four or more years. This model controls for age, race, and gender, and year, DRG, and hospital fixed effects. *Post * ExpandedDRG* is an indicator variable equal to 1 if the DRG group expanded in the DRG reform and is in the post expansion period. Column (1) represents the full sample of all hospitals. Column (2) restricts the sample to hospitals which are sampled at least 4 of the 6 years in the sample. Column (3) restricts the sample to Hospitals which appear in 5 or more years in the sample. We do not estimate a panel of hospitals in the sample all 6 years, as there are only 5 of them. Standard errors are in parentheses and clustered at the hospital level.

*
p<0.05

**
p<0.01

p<0.001.

Table 10Sensitivity of results to hospital acquired conditions: $Y = \text{DRG weight in current year}$.

	Full Sample	No HAC
Post * expanded DRG	0.0437 *** (0.00242)	0.0427 *** (0.00243)
Observations	26904427	26587472

Notes: This table estimates equation 2. This model controls for age, race, and gender, and year, DRG, and hospital fixed effects. *Post * ExpandedDRG* is an indicator variable equal to 1 if the DRG group expanded in the DRG reform and is in the post expansion period. In column (1) there are results for the full sample of patients. In column (2) there are results for a subsample of patients who did not have a Hospital Acquired Condition (HAC). Standard errors are in parentheses and clustered at the hospital level.

*
p<0.05

**
p<0.01

p<0.001.

Table 11

Difference-in-difference estimation with different levels of observation. Y = DRG weight in current year.

	DiD collapsed to DRG-Year-Hospital	DiD collapsed to DRG-Year
Post * expanded DRG	0.0876 ** (0.0292)	0.0755 (0.0650)
2007 weights	0.997 *** (0.0222)	0.986 *** (0.0506)
Observations	1716919	3637

Notes: This table estimates Equation 2 without the individual characteristics, X_i . Column (1) represents the data collapsed to the DRG-Year-Hospital level of observation. Column (2) collapses to the DRG-Year level of observation. Standard errors are in parentheses and clustered at the DRG level.

*
p<0.05

**
p<0.01

p<0.001.

Table 12

Difference-in-difference estimation with additional FE. $Y = DRG$ weight in current year.

	Eq 2	+Hospital [*] year FE	+DRG [*] year FE	+DRG [*] Hospital FE
Post [*] expanded DRG	0.0437 ^{***} (0.00242)	0.0436 ^{***} (0.00242)	0.0482 ^{***} (0.00864)	0.0484 ^{***} (0.000308)
2007 weights	0.939 ^{***} (0.00363)	0.939 ^{***} (0.00363)	0.547 ^{***} (0.0199)	0.935 ^{***} (0.0000827)
Observations	26904427	26904427	26904427	26904427

Notes: In column (1), we estimate Equation 2, a DiD model with hospital, time, and DRG fixed effects. In column (2), we estimate Equation 2 with an additional interaction term: Hospital by year fixed effects. In column (3), we estimate Equation 2 plus an additional interaction term: DRG by year fixed effects. In column (4), we estimate Equation 2 with an additional fixed effect term: DRG by hospital fixed effects. Standard errors are in parentheses and clustered at the hospital level for columns (1), (2) and (3). Errors are clustered at the Hospital DRG level in Column (4).

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$.

Table 13

Counterfactual estimation on panel of hospitals that appear at least 4 out of 6 years in the sample.

	2005	2006	2008	2009	2010
2007 weights	1.187 (0.00211)	1.409 (0.00194)	1.331 (0.00193)	1.340 (0.00172)	1.314 (0.00170)
Current yr weight	1.182 (0.00217)	1.406 (0.00199)	1.373 (0.00194)	1.396 (0.00143)	1.388 (0.00178)
Change in weight	-0.00503 (0.000126)	-0.00234 (0.000276)	0.0419 (0.000473)	0.0559 (0.000570)	0.0739 (0.000591)
Observations	442671	608048	501170	640686	610510

Notes: This table replicates the counterfactual presented in Table 3 for the sample of hospitals which appear in the panel at least 4 out of the 6 years of the sample. The variable 2007 weight is the CMS-DRG weight the patient would have been assigned, if the patient had been admitted in 2007. The variable current year weight represents the DRG weight of the patient in the year they were admitted. The change in weight variable is row 2 minus row 1, and this measures the difference between the DRG weight which was assigned to the patient and the DRG weight that would have been assigned if the patient had been admitted in 2007. In parentheses are the standard errors of the means: standard deviation / square root of sample size.

Table 14

Estimation of Eq. (2) on medical and surgical patients: $Y = \text{DRG weight in current year}$.

	All	Medical	Surgical
Post [*] expand	0.0437 ^{***} (0.00242)	0.212 ^{***} (0.00441)	0.0258 ^{***} (0.00530)
2007 weights	0.939 ^{***} (0.00363)	0.708 ^{***} (0.00433)	0.908 ^{***} (0.00755)
Observations	26904427	23434685	3469707

Notes: This table presents results from separately estimating Equation 2 on two distinct subsets of the inpatients in our sample. In column (1), we have patients whose DRG was classified as 'medical'. In column (2), we have patients whose DRG was classified as 'surgical'. Classifications are provided by CMS in the DRG weight files. Standard errors are in parentheses and clustered at the DRG level.

*
p<0.05

**
p<0.01

p<0.001.