Upcoding under Diagnosis-based Price: Evidence from China

Yang Yue Xiaolu Zhou*

November 15, 2021

Abstract

This paper examines upcoding under the diagnosis-based prospective payment system (D-PPS) in China's public health insurance plans. Using the differential variation generated by a payment reform across diseases and hospitals, we find that: (1) hospitals tend to code patients to a more generous disease under D-PPS versus fee-for-service (FFS), and this leads to $82 excess health-care expenditures per admission on average under D-PPS versus FFS; (2) under D-PPS, hospitals' coding positively responds to disease prices; (3) there is no evidence showing that hospitals provide more resources to upcoded patients.

Keywords: Upcoding, Prospective Payment System, China
JEL Codes: H0, I0, L0

*School of Economics, and Wang Yanan Institute for Studies in Economics, Xiamen University. We thank Ginger Zhe Jin, Cheryl Long, and participants of seminars at Shanghai University of Finance and Economics and Jinan University for constructive comments. We also thank Jianhao Wu and Wenchen Xu for their excellent research assistant. This research is supported by the National Natural Science Foundation of China (grant 71803162) and the Fujian Social Science Planning Fund (grant FJ2019C058). All errors are our own. Correspondence may be addressed to authors via email at yangyue@xmu.edu.cn and xiaoluzhou@xmu.edu.cn

1 Introduction

Healthcare expenditure has seen a rapid growth in China over recent years. In 2019, the country spent $6.88 trillion—equivalent to 6.64% of its GDP—on health care, and the expenditure has grown by 23% annually since 2010.[1] To control for these expanding health-care costs, the Chinese government has started to adopt a diagnosis-based prospective payment system (D-PPS) to replace the previous fee-for-service (FFS) system. The intention for doing this is to induce hospitals to minimize costs by reimbursing health-care providers at a fixed price based on a patient's diagnosis group. However, D-PPS can also induce hospitals to code patients to diagnoses with more generous reimbursement than their service merit, which is known as upcoding. This unexpected response leads to excess expenditures to the government and thus weakens the cost-effectiveness of D-PPS.

In this paper, we examine if and how much hospitals upcode patients under D-PPS in China's public health insurance plans. To measure the intensity, we exploit a payment reform that occurred in a city (City S) during 2014—2015.[2] The reform generates two policy shocks—the introduction of D-PPS and the adjustment of pricing. In January 2014, the local government introduced D-PPS to replace FFS. Specifically, it selected 1,049 common diseases, assigned a score for each of the diseases, and began reimbursing hospitals in proportion to these scores. One year after introduction, in January 2015, the local government adjusted the score schedule by lowering the scores for 45% of diseases and increasing the scores for 5% of diseases. After the policy changes, the reimbursement may become higher or lower even for diseases with similar diagnoses. This makes some diseases more generous than others and thus provides financial incentives for upcoding.

[1] Information is from the Statistical Communiqué of the People's Republic of China on health care and development in 2019.

[2] Due to data confidentiality, we cannot disclose the city's name. City S is located at eastern China. It covers 8555 square feets, has 5.9 million residents, and recorded RMB 309.9 billion in GDP by the end of 2019.

However, it is hard to disentangle upcoding from patient selection. For example, the change in coding can be caused by the intention of hospitals to code patients towards more generous diseases, but it can also result from hospitals' selection of unhealthier patients. To overcome the challenge, we take use of the differential variation in coding incentives across diseases or hospitals. This controls for patient selection if the selection does not vary across diseases and hospitals in the same way as the upcoding incentive. Also, to check for the robustness, we exploit the within-individual variation to avoid changes in patients' composition before and after policy interventions.

The first shock—the introduction of D-PPS to replace FFS—generates differential variation in coding incentives across diseases and hospitals. After the payment change, even for diseases with the similar diagnoses, hospitals may face higher reimbursement in certain diseases but lower reimbursement in others. In response to the change in reimbursement, profit-seeking hospitals may have an incentive to code patients towards diseases subject to higher reimbursement under D-PPS versus FFS after the policy shock. Thus, we should expect an increase in the share of admissions that are coded to these diseases within a hospital.[3]

To measure the differential variation across diseases within a hospital, we use the difference-in-difference-in-difference (DDD) specification. In particular, we use a city that is located in the neighborhood (City N) as a control for City S. We compare the differential change in admission shares with respect to the change in reimbursement across diseases for City S versus that for City N. We find that the share of admissions increases for diseases subject to higher reimbursement under D-PPS versus FFS for hospitals in City S more than that in City N. Also, the increase is more pronounced for disease subject a larger increase in reimbursement under D-PPS versus FFS. Our estimates imply a 7.11% coding difference—equivalent to $82

[3]Literature argues that the objective function of a hospital is the weighted sum of its profit and volume. Thus, profit-maximization is one of the objective of hospitals.

excess expenditure per admission—under D-PPS relative to FFS on average in City S. The identification assumption is the parallel change in health status for patients in City S versus that in City N.

As argued earlier, although the above result has controlled for patient selection with the differential change in coding incentives, the aggregate result can be biased by patient selection if the magnitude of selection is consistent with the change in reimbursement under D-PPS relative to FFS. To verify the aggregate result, we also exploit the within-individual variation in coding scores. The results show a 6.3% increase in the coding score after the introduction of D-PPS for patients in City S versus that in City N, coinciding with the aggregate result. Also, we find that the upcoding intensity is more pronounced for patients generating higher treatment costs to a hospital, because hospitals face a higher opportunity cost of upcoding.

The second policy shock—the score adjustment—provides differential variation across diseases. Since some diseases have similar diagnoses but differ in scores, hospitals are likely to code patients from low-score diseases to high-score diseases. This incentive is stronger for the group of diseases with similar diagnoses but a larger dispersion in scores. The second policy shock exogenously shifts the score spread for diseases with similar diagnoses. Since score dispersion may increase or decrease across different disease groups, we should expect a higher increase in the share of admissions coded to high-score diseases within the disease group subject to a larger increase in score spread.

To quantify the differential incentive, we define a disease group for each disease according to its 3-digit ICD-10 code. Within the same disease group, we define high-score diseases as those with scores above the median score. The results indicate that the share of admissions coded to high-score diseases within the same group increases more pronouncedly for disease groups subject to a larger increase in score spread after the policy change. This suggests that hospitals tend to code patients to high-score or generous diseases, and this incentive is

strengthened if the score spread enlarges.

Similarly, the above analysis across disease groups may suffer from patient selection if the magnitude of selection exists consistently with the change in score spread. Thus, as an alternative identification, we take advantage of the differential variation in coding incentives across individuals. In particular, individuals who are subject to lower scores, and thus lower reimbursement, after the score adjustment are more likely to be upcoded. Thus, for these individuals, the probability of being coded to high-score diseases within the disease group should increase more significantly after the policy change, compared to individuals coded to the same disease group. The estimated coefficients support our conjecture. Also, consistent with the previous results, we find that patients requiring more health-care resources are more likely to be upcoded.

Lastly, we investigate whether upcoding also suggests more or better treatments to patients. This is important to understand the cost-effectiveness of D-PPS because the inefficiency of upcoding may be ameliorated if hospitals also provide more resources at the same time of upcoding. We find that the health care expenditure, length of stay, and the volume of care do not significantly increase after the policy change for patients who are upcoded more aggressively after the policy change, meaning that hospitals do not provide more resources to upcoded patients. This further confirms the inefficiency of hospitals' upcoding behavior.

Our study is related to literature about hospitals' upcoding under Medicare in the United States. Carter et al. (1990) measure the upcoding by sending a sample of patients' diagnoses to experts for coding and comparing it with hospitals' coding. Some literature relies on the ascending time trend for diagnosis-related groups (DRGs) with high weights as evidence for upcoding (Coulam and Gaumer (arch), Silverman and Skinner (2004), Nie et al. (2016)).[4]

[4]Coulam and Gaumer (arch) provides a review for literature before 1990s. Silverman and Skinner (2004) find an increasing time trend of the most generous DRG over time under PPS, and a higher increase rate for for-profit hospitals (23%) than for not-for-profit hospitals (10%). Nie et al. (2016) focus on anesthesia services for outpatient gastrointestinal procedures. They combine the time trend with a model for patients' high risk. Jürges and Köberlein (2015) also relies on the time trend, but they studies the heterogeneous

Some other literature focuses on upcoding for programs with risk adjustment, and they find a more generous coding or longer therapy hours for programs with relative to without risk adjustment (Brown et al. (2014), Bastani et al. (2019), Brunt (2011), Brunt (2014), Fang and Gong (2017)). Dafny (2005) identifies upcoding from the differential effects of a price shock in Medicare in 1989. Geruso and Layton (2019) measures upcoding in the Medicare Advantage (MA), and they rely on the market-level variation in the provision intensity of MA. We contribute to this literature in several ways. First, we study the coding responses to two policy shocks—both the introduction of D-PPS and the price change. Second, we not only estimate the treatment effect but also the differential effects of the policy shocks. Third, we measure upcoding for public health insurance plans in China, where D-PPS has become increasingly popular but no regulation has been introduced to govern upcoding.[5] Our study provides policy implications for the implementation of D-PPS in developing countries.

Our paper is also related to the empirical literature on the effectiveness of China's cost-control policies in public health insurance plans. Most of this literature focuses on outcome variables including volume of care, intensity of care, and the government's health-care expenditure (Yip and Eggleston (2001), Gao et al. (2014), Huang and Gan (2017)). Contrasting with the above, Chan and Zeng (2018) analyze hospitals' unexpected responses to the D-PPS. They find that hospitals tend to dynamically choose intensity of care over the fiscal year under the year-end fee limit, and this response leads to welfare reallocation over the year. We also analyze hospitals unexpected responses, but we focus on upcoding that reduces the effectiveness of cost-control policies.

The rest of the paper is organized as follows. Sections 2 and 3 introduce the background and data. Sections 4 and 5 present the empirics of coding responses to the first and second policy shock, respectively. Section 6 discusses whether upcoding also suggests better

upcoding in patients' health status and financial incentives.

[5]One international study is J.M.Steinbusch et al. (2007). But they do not measure upcoding intensity.

treatments, and Section 7 concludes.

2 Background

2.1 China's Public Medical Insurance System

The Chinese government has developed two universal public medical insurance systems for urban citizens: Urban Employee Basic Medical Insurance (UEBMI), and Urban Resident Basic Medical Insurance (URBMI).[6] The former was introduced in 1999 as a mandatory program for urban citizens who are employed in the public sector or state-owned enterprises. The latter was established as a voluntary program for urban citizens who are not covered by the former system, for example, children, students, and unemployment. By 2019, 95% of urban citizens had enrolled the two programs[7], and the health-care expenditure has accounted for 6.64% of GDP ($675 per capita).[8]

All of public hospitals, most of private hospitals, and some of specialized clinics are under contract with the government. However, due to the decentralization of financing, the payment method differs for treatment in hospitals within the patients' residence versus those outside their residence. Thus, we only focus on hospital treatment in the city of patients residence because only these treatment are subject to the local government's policy changes. The remaining observations still account for a majority of our data because the reimbursement rate is very low for treatment outside the patients' residence and thus only a few patients would like to receive treatment out of their city.

Also, all hospitals in China are subject to a quality system. Each hospital is granted one

[6]Since 2002, Newly Cooperative Medical Scheme (NCMS) was established for rural citizens. In 2016, NCMS was consolidated to URBMI, and the combined one is now called urban-rural residents basic medical insurance. The urban or rural citizen type is assigned by the Chinese government to residents by their place of origin, known as "hukou".

[7]Information is from the Statistical Communiqué of the People's Republic of China on health care and development in 2019.

[8]Information is from the National Bureau of Statistics of China on health care and development in 2019.

of the three quality tiers— high-, medium-, and low-tier—as an indicator for their capacity, personnel (e.g. doctors), and equipment. Upgrading and downgrading of their tiers are both possible. The above two public health insurance plans apply to hospitals in all of the three quality tiers.

Services covered by the two public medical insurance typically contain inpatient hospital care, primary and specialist care, prescription drugs, mental health care, physical therapy, emergency care, and traditional Chinese medicine. These services are subject to cost-sharing between patients and the government, and the government's payment method differs across regions, insurance plans, and services. For inpatient services, cost-based payment methods such as FFS and pay-per-capita were common before 2017, and after which the central government encouraged the adoption of D-PPS.[9] The program is enrolled gradually. Before it was implemented nationwide, several cities have experimented with D-PPS. This provides exogeneous variation for economists to understand hospitals' responses and the cost-effectiveness of D-PPS. This paper takes use of a payment reform in a city. The policy details are presented in the following section.

2.2 Policy Shocks in City S

Due to data confidentiality, we cannot disclose the city's name so we denote it as City S. City S is located at eastern China. It covers 8555 square feets, has 5.9 million residents, and recorded RMB 309.9 billion in GDP by the end of 2019. From 2014 to 2015, the city implemented payment reform generating two policy shocks for its UEBMI and URBMI enrollees.

[9]Information is from State Council No. 55, 2017. Available at `http://www.gov.cn/zhengce/content/2017-06/28/content_5206315.htm`.

The First Shock: Introduction of D-PPS

The local government of City S introduced D-PPS for inpatient services of URBMI enrollees in July 2013 and then UEBMI enrollees in January 2014. Due to data problem, we only focus on the first policy shock for UEBMI enrollees that occurred in January 2014, in the following.[10]

In City S, prior to January 2014, the health-care payment method was a combination of FFS and a global budget. Under the system, the local government estimated a year-end cap on the total actual expenditure for a hospital based on its historical records. The hospital thus was reimbursed according to the actual expenditure for all of admissions over the year below the year-end cap. In practice, the local government would pay a portion of the estimated year-end expenditure to the hospital at the beginning of a year, check the hospital's final balance, and retrieve the excess budget at the end of a year.

After January 2014, City S selected 1,049 most frequently admitted diseases and applied D-PPS to these diseases. Specifically, the government assigned a score δ_d for each disease d to reflect its relative health-care expenditure. Actually, the relative health-care expenditure was computed based on the average expenditure across all admissions coded to disease d in City S over the previous three years. The system also designed a weight w_h for hospital h to reveal the cost differences between hospitals; for example, for differences in wages. The weight on hospital h was only contingent on its quality tier $q(h)$, $w_h = w_{q(h)}$.

Given the score schedule, the government determines the actual price for a disease taking into account its total health-care revenue from the previous year. In practice, the government would decide a gross funding budget K for the city at the beginning of a year. At the end of a year, the local government would count the sum of weighted scores for all admissions in the city over the year. The ratio between these two became the actual unit price per score.

[10]We only observe UEBMI but not URBMI enrollees in the control city so that we only focus on the effect of the introduction of D-PPS under UEBMI.

Then, the actual price for an admission coded to disease d by hospital h with the quality tier $q(h)$ was:

$$P_{hd} = \frac{K}{\sum_{h'} \sum_{d'} w_{q(h')} \delta_{d'} n_{h'd'}} \times w_{q(h)} \delta_d n_{hd}, \tag{1}$$

where n_{hd} is the number of admissions in disease d at hospital h. This suggests that, for a disease d in hospital h, the actual price P_{hd} can change year over year due to the changing gross budget K and the distribution of diseases in the city.

The Second Shock: Score Adjustment

In January 2015, one year after the introduction of D-PPS, the local government revised its score schedule for both UEBMI and URBMI due to complaints from hospitals.[11] The new version of the scores lowered the relative price of 45% of the 1,049 diseases and increased the relative price of 5% of diseases.

Table 1 provides a few examples of diseases and the pre- and post-revision scores. For example, "atherosclerotic heart disease without intervention" and "atherosclerotic heart disease with intervention" both have a similar diagnosis but different therapies. In 2014, the former was assigned the score of 64, and the latter was assigned the score of 300. Some diseases are similar in both diagnoses and therapies, but they were assigned different scores. Similarities between the main diagnoses for different diseases provided opportunities for hospitals to code patients for their own benefits, when some diseases resulted in more generous value of reimbursement than others in the payment system.

[11] Hospitals argued that the 2014-version of the scores did not reflect the relative cost of care across diseases, and this would lead to hospitals losing profits. As a result, the local government bargained with hospitals and revised the 2015 schedule.

3 Data

We obtain inpatient claims data for all of URBMI and UEBMI enrollees of City S and UBEMI enrollees of City N, a city in the same province of City S, at the distance of 386.4 kilometers from it. City S and City N had similar populations (5.9 vs. 7.8 million) and GDP per capita ($8,080 vs. $14,945) in 2019. The data period is from August 2012 to December 2015. Over the sample period, City N used FFS with a global budget for its UEBMI enrollees, as did City S before January 2014.

The data contains a record for each individual-encounter, including a unique ID for the individual, the individual's birth date, gender and insurance type, the hospital name and its quality tiers, the diagnosis information, the date of admission and discharge, the total amount of actual expenditures, and reimbursement. We drop observations without insurance type and diagnostic information. To ensure all hospitals are subject to the policy shocks, we only keep hospitals located within the city for both City S and City N. We then use the data to construct the sample for event analyses.

To analyze the effect of the introduction of D-PPS, we use records of UEBMI enrollees in City S and City N from July 2012 to December 2015, which we call "Sample A". We only analyze the effect of the introduction of D-PPS for UEBMI but not URBMI because we do not have URBMI discharge records for city N as a control. Sample A includes 326,234 admissions, with 132,323 in City S and 193,911 in City N, respectively.

To analyze the effect of the 2015 score adjustment, we include all records of URBMI and UEBMI enrollees in City S from January 2014 to December 2015 and call this "Sample B". In this exercise, we do not use records in City N because its payment system differ with that of City S in 2014—D-PPS in City S and FFS in City N—and this can lead to different coding behavior in the two cities before the adjustment of scores occurred. In total, sample B includes 258,722 admissions.

We also obtain the 2014- and 2015-versions of diagnosis-specific scores from City S, and we merge the two versions of scores for samples A and B. We also match the 3-digit ICD-10 codes for samples A and B. The details about data consolidation is in Appendix 8.3.

Table 3 describes samples A and B with the diagnosis-specific scores and 3-digit ICD-10 codes included. After the merge, there are 212,158 (65%) admissions left in sample A, with 105,372 (79%) and 106,786 (55%) admissions in City S and City N, respectively, and 175,745 (68%) of admissions in Sample B.

The first two columns of Table 3 summarizes Sample A. Panel a describes the sample by individuals. There are 43,657 and 64,598 patients in City S and City N, respectively. Patients in City S have similar ages (62.75 vs. 61.02) but lower expenditure per admission ($1075 vs. $1950), relative to those in City N.

Panel b describes the sample by hospitals. There are 186 and 168 hospitals in City S and City N, respectively, with more low-tier hospitals in City S than in City N. We present the average share of admissions during the sample period (from July 2012 to December 2014) for hospitals across different quality tiers. High-tier hospitals occupy a smaller share of admissions in City S than that in City N since there are fewer high-tier hospitals. Also, for both cities, the average expenditure per admission is lower in high-tier hospitals than in medium- and low-tier ones, meaning that high-tier hospitals receive more severe patients. Comparing expenditures in the two cities, we find that each admission spends less for City S than City N.

Panel c summarizes the data by diseases. Sample A contains 821 and 902 diseases in City S and City N, respectively. Cerebral infarction and coronary heart disease are the two highest ranked disease in the number of admissions for the two cities.

Panel d describes the sample by 1-digit and 3-digit ICD-10 codes. There are 19 and 20 disease groups defined by 1-digit ICD-10 codes, with 43 and 47 diseases within each group, for City S and City N, respectively. The most frequently admitted disease group is I (diseases

of the circulatory system) for both cities, followed by J (diseases of the respiratory system) for City S and K (diseases of the digestive system) for City N. For disease groups defined by 3-digit ICD-10 codes, Sample A includes 386 and 399 groups, with 2.15 and 2.39 diseases per group on average, for City S and City N, respectively. The most frequently admitted group is I63 (cerebral infarction) for City S and I49 (other cardiac arryhthmias) for City N. These are followed by J44 (other chronic obstructive pulmonary disease) for City S and I10 (essential hypertension) for City N.

The last column of Table 3 summarizes Sample B. There are 175,744 admissions and 115,669 patients in the sample. From Panel a, patients are younger (53 years-old) in Sample B than in Sample A (62 years-old), because sample B include URBMI enrollees, such as teenagers and college students, who are younger than the UEBMI enrollees. From Panel b, hospitals in Sample B are similar to those in Sample A in terms of the distribution of quality tiers and market share of admissions. From panels c and d, there are 1,247 diseases, 23 disease groups defined by 1-digit ICD-10 code, and 474 disease groups defined by 3-digit ICD-10 code. The two most frequently admitted diseases are cerebral infarction and bronchitis, and the most frequently admitted disease groups remain the same as those of Sample A.

We analyze the two samples at both the aggregate level and individual level. Next, we present our model specifications and results.

4 Coding Responses to the Introduction of D-PPS

In January 2014, City S introduced D-PPS in replace of FFS. After this policy change, hospitals can receive more or less reimbursement under D-PPS relative to FFS for different diseases. The increase or decrease in reimbursement that a hospital can receive for different diseases generates incentives for hospitals' coding decisions. For a profit-seeking hospital,

there may be incentives to code patients with diseases generating lower reimbursement under D-PPS versus FFS towards diseases generating higher reimbursement after the policy change. Also, the incentive can be stronger for diseases subject to larger decreases in reimbursement under D-PPS versus FFS. The changes in admissions within a hospital across diseases associated with the changes in reimbursement under D-PPS versus FFS allow us to identify the differential effect on coding incentives. To explore this, we conduct an aggregate analysis.

However, one may challenge the aggregate result because it is difficult to control for patient selection with the aggregate data. For example, hospitals may admit more healthier or unhealthier patients under D-PPS, leading to variation in admission shares over time. This changes the patient composition coded to a disease in a hospital before and after the policy change. Given that our aggregate analysis relies on differential changes across diseases and hospitals, patient selection bias can be resolved if the selection is not systematically more or less severe in diseases subject to higher or lower reimbursement to a hospital. But it may be problematic if the assumption is violated. Thus, as a supplement, we also exploit the within-individual variation in coding scores before and after the policy change for individuals in City S relative to individuals in City N. This better controls for patient selection because it avoids the change in patients' composition before and after the policy change. This is illustrated in the individual analysis.

Below, we present the aggregate analysis and the individual analysis separately.

4.1 Aggregate Analysis

Since upcoding is more likely to happen among comparable diseases, we thus define a disease group where diseases within it are similar in diagnosis so that we can focus on upcoding

within the group. We use the 1-digit ICD-10 code as the definition for each disease group.[12] The dependent variable is the share of admissions assigned to disease d within the disease group g in hospital h during week t, $sh_{d|g,ht}$. The independent variable is the change of reimbursement for hospital h in treating disease d under D-PPS versus FFS. To approximate the reimbursement difference for a disease in a hospital under the two payment systems, we use the change of actual expenditure as a proxy. Specifically, we calculate the average of the actual expenditure for all admissions assigned to disease d in hospital h during the per-treat period t_0 (July 2012 to December 2014), $Avg.\ Expenditure_{dht_0}$, and we use it to approximate the average level of reimbursement to hospital h for disease d under FFS. Under D-PPS, the reimbursement is proportional to the diagnosis-based score, and the score is designed based on the average of expenditures for disease d across all hospitals, according to the local government. Thus, to approximate the average level of reimbursement for disease d in hospital h under D-PPS, we use the average of actual expenditures for disease d across all hospitals within the same quality tiers as hospital h during the pre-treat period t_0, $Avg.\ Expenditure_{dt_0}$. The dependent variable is the change in reimbursement under D-PPS versus FFS:

$$\Delta Reimb_{dh} = \log(Avg.\ Expenditure_{dt_0}) - \log(Avg.\ Expenditure_{dht_0}). \tag{2}$$

We use a difference-in-difference-in-difference model to estimate the differential effects of

[12] We also have tried the 3-digit ICD-10 codes as a definition for disease groups. But there is not enough diseases (1.05 to 1.11 diseases) in the same 3-digit disease group per hospital on average. This means that most of disease groups per hospital contain only 1 disease, and thus, it is not adequate for comparing how coding changes among diseases within the same disease group per hospital.

the payment system change on hospitals' coding incentives:

$$\begin{aligned}\log(sh_{d|g,ht}) &= \alpha_0 + \alpha_1 City_h \times Post_t \times \Delta Reimb_{dh} + \alpha_2 City_h \times Post_t \\ &\quad + \alpha_3 Post_t \times \Delta Reimb_{hd} + \alpha_4 City_h \times \Delta Reimb_{hd} + \alpha_5 \Delta Reimb_{hd} \\ &\quad + x_{d|g,ht} + \chi_h + \chi_d + \chi_t + \epsilon_{dht},\end{aligned} \tag{3}$$

where $City_h$ is the indicator equal to 1 if hospital h is located in City S; $Post_t$ is the dummy equal to 1 for all months after January 2014; $x_{d|g,ht}$ is the average age of patients coded to disease d, hospital h at week t; χ_h denotes the hospital-level fixed effects, capturing the time-invariant change in coding across hospitals, such as their productivity; χ_d denotes the disease-level fixed effects, capturing the time-invariant differences among diseases, such as common or rare diseases; and χ_t is the year-month-week-level fixed effects.

The parameter α_1 measures the marginal change of the differential effects of D-PPS on coding relative to FFS. Positive α_1 indicates that the admission share increases more significantly for diseases generating larger increases in reimbursement to a hospital under D-PPS versus FFS. The parameter α_2 shows the average difference in coding under D-PPS versus FFS. The underlying assumption of the model is the parallel change in the distribution of diseases over time for patients in the two cities.

Panel a1 of Table 4 summarizes the variables used in equation 3. These variables are all aggregated from individual–encounter–level Sample A to the level of disease–hospital–week by the number of admissions. Also, if there are no admissions coded to disease d in hospital h before January 2014, we drop the disease in the hospital because it has no pre-treatment period. The average change in reimbursement ($\Delta Reimb_{dh}$) is 0.083 and 0.056 for City S and City N, respectively. Unconditional on the same disease group, the average share of admissions per disease in a hospital ($sh_{d,ht}$) is 0.173 and 0.081 for City S and City N, respectively; and conditional on the same disease group, the average share ($sh_{d|g,ht}$) becomes

0.494 and 0.330 for City S and City N, respectively. The average age for a disease–hospital–week combination is similar in the two cities (61 and 59).

Before the model estimates, we provide descriptive evidence for the policy-induced up-coding incentives. In particular, for each hospital h, we compute the share of admissions for diseases that generate higher reimbursement to the hospital under D-PPS versus FFS ($\Delta Reimb_{dh} \geq 0$), and then take the average of the share across all hospitals. The average share indicates the average probability to code a patient to diseases subject to higher reimbursement after the policy intervention. We expect that, after the intervention, the share increases more for City S relative to City N.

Figure 1 shows the admission share for diseases with $\Delta Reimb_{dh} \geq 0$ in an average hospital and disease group in the two cities. The straight line indicates the starting time for the policy intervention (January 2014). In City S, the average share of admissions increases over time after the policy intervention for diseases subject to higher reimbursement under D-PPS versus FFS. However, for City N, the share almost does not change or slightly decreases after the policy intervention. This is consistent with our conjecture, suggesting a positive response to the reimbursement under D-PPS versus FFS in City S. Our empirical model exploits variation in admission shares before and after the policy intervention associated with the variation in reimbursement under D-PPS versus FFS for different diseases and hospitals.

Table 5 presents the estimates of equation 3. Columns (1) and (2) show the estimates for the full sample, with the unconditional share of admission as the dependent variable in column (1) and conditional share of admission in column (2). Both results show significant positive coefficients on the difference in reimbursement under the D-PPS versus FFS, measured by $\Delta Reimb_{dh}$. The results suggest that hospitals tend to code patients towards diseases subject to higher reimbursement under D-PPS relative to FFS.

As described before, one concern with the aggregate estimates is that the changes in admission share for all patients may indicate the incentive changes in both coding and

admissions. As a robustness check for patient selection, we restrict our sample to patients who have admitted both before and after the policy change. In particular, we re-calculate the change in reimbursement and the share of admissions for patients in the subsample, and estimate equation 3.

Columns (3) and (4) of Table 5 report the estimates of equation 3 for the subsample. Both results still show positive coefficient on the triple-difference interaction. When using the unconditional share of admission as the dependent variable in column (3), the coefficient changes little with that in column (1). But, the coefficients in columns (3) and (4) are not statistically significant, probably because there is not enough variation in the change of reimbursement. Overall, these results are consistent with our conjecture that hospitals are more likely to code patients to diseases subject to higher reimbursement under D-PPS versus FFS in City S relative to City N.

To understand the implication of the model estimates, the last row of Table 5 presents the coding difference under D-PPS versus FFS. The estimates in column (2) and (4) imply that the average coding score increases by 1.83% under D-PPS versus FFS for patients in our full sample, and at least 7.11% for patients with admissions both before and after the policy intervention.[13] Our result for the subsample is similar to the 5% to 7% upcoding intensity under Medicare Advantages in the United States (Geruso and Layton (2019)). To give a monetary representation, the 7.11% upcoding intensity implies $82 excess expenditure per admission (or $3 million) for health care and $2.5 million excess total expenditure to the government for UEBMI in 2014.[14]

[13]The result is calculated as follows. We calculate the share of admission sh'_{dht} without D-PPS: $\log(sh'_{dht}) = \log(sh_{dht}) - (\hat{\alpha}_1 \Delta Reimb_{dh} + \hat{\alpha}_2)$ for all hospitals h and diseases d in City S, week t in 2014. We then calculate the average weighted scores with versus without D-PPS: $s^{pps}_{ht} = \sum_d \delta^{2014}_d \times sh_{dht}$, and $s^{ffs}_{ht} = \sum_d \delta^{2014}_d \times sh'_{dht}$ for all hospital h and week t. We then aggregate across hospitals and weeks by the number of admissions: $s^{pps} = \sum_{h,t} s^{pps}_{ht} n_{ht} / (\sum_{h,t} n_{ht})$, and $s^{ffs} = \sum_{h,t} s^{ffs}_{ht} n_{ht} / (\sum_{h,t} n_{ht})$.

[14]The result is from the average expenditure of $1242 per admission for City S in 2014. Given the upcoding intensity of 7.11%, the expenditure without the upcoding is thus $1242/1.0711×0.0711 = $82. As the total number of admissions is 36,711 for City S in 2014 in Sample A, the total excess health care expenditure is thus $82 × 36,711= $3 million. Also, given the 16% average out-of-pocket ratio, the excess health-care

4.2 Individual Analysis

As argued above, hospitals are likely to code patients to the disease with a higher score under D-PPS, but this incentive is weak under FFS because hospitals can be reimbursed according to their actual expenditure. Thus, we should expect that, conditional on the same patient, the coding score increases by a greater magnitude after the policy intervention for patients in City S versus City N.

For patient i's admissons in week t, we calculate her average coding score $\bar{\delta}_{it}$. The log of the average score is the dependent variable. The model specification is:

$$\log(\bar{\delta}_{it}) = \beta_0 + \beta_1 Post_t \times CityS_i + \xi_i + \xi_t + \epsilon_{it}, \tag{4}$$

where $Post_t$ is the indicator equal to 1 for all weeks t after the policy intervention; $CityS_i$ is the indicator equal to 1 if individual i is covered by the insurance system in City S; ξ_t is the full set of year–month–week–level fixed effects controlling for admission time; ξ_i is the fixed effects at the individual–level. The parameter β_1 presents the treatment effect of the policy intervention on the coding score conditional on the same individual. We expect positive β_1, meaning that hospitals tend to code patients to a more generous disease under D-PPS versus FFS.

Panel a2 of table 4 describes the variables. It shows that the average coding score is 115 and 131 for City S and City N, respectively. Patients are very similar in the two cities, with average age of 67.6 and 67.5 years for City S and City N, respectively, and the standard deviation of 13.9 for both cities.

Table 6 presents the estimates of equation 4. Column (1) displays the OLS estimates with no fixed effects. The result indicates a 6.8% increase in a patient's coding score under D-PPS relative to FFS for individuals in City S relative to individuals in City N. The

expenditure to the government is $\$3\times(1-0.16) = \$2.5$ million for UEBMI in City S in 2014.

coefficient on the dummy of City S is significantly negative, meaning that patients in City S are systematically healthier than those in City N. Column (2) controls for individual–level fixed effects, and this slightly reduces the magnitude of the upcoding intensity to 6.3%. This suggests that without controlling for the individual–level heterogeneity can over-estimate the upcoding intensity, but the bias is not sizable.

We then check the parallel trend in coding scores for individuals in City S and City N. Specifically, we estimate equation 4 while interacting $Post_t \times CityS_i$ with monthly dummies. Figure 2 displays the estimates and 95% confidence intervals for the interactions. The figure indicates that the coefficients on the interactions are all insignificantly different from zero before the policy change under FFS; however, after the policy change, the coefficients are all positive and become significant six months after the policy change. This illustrates that the model has controlled the differences for individuals in City S and City N.

Next we investigate the intensive margin versus extensive margin for upcoding. After the introduction of D-PPS, patients are likely to be coded to high-score diseases within a hospital, which is the "intensive margin." However, it is also possible that patients are transferred from low-tier hospitals treating less severe diseases to high-tier hospitals treating more severe diseases, which is the "extensive margin." The result in column (2) shows both intensive and extensive margins. To see the intensive margin only, we need to aggregate the individual-encounter level data to the unit of individual–disease–hospital–week. There are 63,595 number of observations after the aggregation.[15]

Column (3) presents the estimates of equation 4 using the individual–disease–hospital–weekly level data. By controlling for the individual-hospital level fixed effects, column (3) only indicates the intensive margin. The estimate coefficient is slightly lower (6%) than that in column (2) (6.3%), reflecting that upcoding mainly occurs in the intensive margin.

[15]In particular, we keep an observation for each combination of individual–disease–hospital–week δ_{idht}, and use the specification as equation 4 with various sets of fixed effects for columns (3) to (7) of Table 6.

We then provide robustness checks for the aggregate results. The aggregate analysis investigates coding across diseases within the same disease group. To verify for this, column (4) of Table 6 includes individual–hospital–disease group level fixed effects. The estimate result indicates a 5.3% increase in coding score, similar to the aggregate results (7%).

Also, the aggregate result exploit the differential changes in coding across diseases. To check for this, columns (6) and (7) of Table 6 estimate equation 4 for two subsamples of disease—diseases subject to increasing/decreasing reimbursement under D-PPS versus FFS ($\Delta Reimb_{dh} \geq, < 0$). Coinciding with the aggregate result, we find a higher upcoding intensity for diseases subject to lower reimbursement under D-PPS versus FFS (8.6%) relative to diseases subject to non-lower reimbursement under D-PPS versus FFS (3.5%), even controlling for the fixed effects for the combination of individuals and hospitals.

4.3 Heterogeneity among Patients and Hospitals

Due to the heterogeneity in hospitals' benefit and costs of upcoding, the upcoding intensity may vary across patients and hospitals. This exercise can provide policy implications for the design of D-PPS. In particular, we use the data aggregated to the level of individual–hospital–week, and we estimate equation 4 including both the individual–level and hospital–level fixed effects.[16]

Columns (1) and (2) of Table 7 investigate how the upcoding intensity varies for old and young patients, respectively. Upcoding can be more significant for old patients because they are less healthier and thus require more treatment resources. However, it is also possible that it is easier to upcode young patients because there is a lower risk of upcoding this group.

[16] In particular, we calculate the average coding score for a patient i's admissions in hospital h during week t, $\bar{\delta}_{iht}$, and run the regression as follows:

$$\log(\bar{\delta}_{iht}) = \beta_0 + \beta_1 Post_t \times CityS_i + \xi_i + \xi_t + \xi_h + \epsilon_{it},$$

where ξ_h denotes the set of hospital-level fixed effects.

Empirical results indicate upcoding for both old and young patients, and the upcoding intensity is similar between them (5.1% and 5.3%).

Columns (3) and (4) of Table 7 study the upcoding for high-cost versus low-cost patients. Since high-cost patients require more resources from hospitals, we should expect a stronger upcoding intensity for high-cost patients than low-cost patients. To define a high-cost patient, we use her average expenditure for all admissions during the pre-treatment period. If a patient's pre-treatment average expenditure is above the median across all patients, she is defined as being a high-cost patient. As predicted, the results show a higher upcoding intensity for high-cost patients (8.6%) than low-cost patients (5.3%).

Columns (5) (6) (7) of Table 7 examine upcoding in high-, medium-, and low-tier hospitals. High-tier hospitals admit more severe patients and may upcode them more aggressively, but high-tier hospitals may also concern about their reputation and thus upcode patients less aggressively. Results show a higher upcoding intensity for high- (6.4%) and medium-tier (7.7%) hospitals than low-tier ones (5.3%), probably because high- and medium-tier hospitals receive more severe patients. Also, the upcoding intensity is slightly higher for medium-tier hospitals than high-tier ones, probably because medium-tier hospitals are less concerned about their reputation.

5 Coding Responses to the 2015 Score Adjustment

The adjustment on scores provides exogenous variation in coding across diseases. Since the reimbursement is contingent on diagnoses, when some diseases are similar in diagnoses but differ in scores, hospitals are likely to code patients towards the disease with a higher score. This incentive is stronger for the group of diseases subject to a larger spread in scores. Thus, using the exogeneous variation in scores for the group of diseases with similar diagnoses, we can investigate how the coding behavior within the group changes as the score spread among

diseases within the group changes. The identification strategy is similar to Dafny (2005), and it is illustrated in the aggregate analysis.

Apart from the variation across disease groups, the adjustment on scores also generates differential coding incentives across individuals. Patients who are frequently coded to diseases subject to a lower score under the new version (2015–version) before 2015 will be reimbursed less after 2015 if their health status remain stable over time and they are "correctly" coded to the same diseases.[17] Therefore, due to the reduced reimbursement, these patients are more likely to be upcoded after the policy change, leading to an increase in the probability of being coded to high-score diseases relative to other patients within the same disease group. This provides us variation across individuals conditional on the same disease group. Thus, as a supplementary for the aggregate analysis, we also conduct the individual analysis.

We illustrate the two identification strategies in the following sections.

5.1 Aggregate Analysis

We define the group of disease with similar diagnoses by the 3-digit ICD-10 codes. We do not use the 1-digit ICD-10 codes as a definition as Section 4.1 because we now use the variation in a disease group rather than the variation in a disease group per hospital, and there are adequate number of diseases in each 3-digit ICD-10 codes.[18] Given the disease group g, we compute the median score for all admissions assigned to the disease group over our sample period. The dependent variable is the share of admissions that are assigned to diseases with a score above the median score within the group g in week t, City S, $frac_{gt}$.[19] Also, to measure

[17]The "correctly" means no coding incentives induced by the score adjustment.

[18]Sepcifically, there are 2.29 diseases in a 3-digit disease group, but 1.34 diseases in a 3-digit disease group per hospital. If we define a disease group with 3-digit ICD-10, there may be only one disease in the group per hospital so that we are not able to compare the coding behaviors among diseases within the disease group per hospital in Section 4.1. Therefore, we define a disease group with a 1-digit ICD-10 code in Section 4.1. But, in Section 5.1, we rely on the variation at the level of disease group rather than the combination of disease group and hospital. Given the 2.29 diseases within a 3-digit disease group, we are able to do analysis.

[19]The median is calculated from the data where each disease is weighted by the number of admissions.

the score spread among diseases within the disease group, we use the variance of scores across diseases within the disease group g in year $y = 2014, 2015$, $\sigma^2_{g,y}$. The independent variable is the change in the log of variance in 2015 and 2014 for disease group g:

$$\Delta Spread_g = \log(\sigma^2_{g,2015}) - \log(\sigma^2_{g,2014}). \tag{5}$$

The model specification we use is:

$$frac_{gt} \quad = \quad \beta_0 + \beta_1 \Delta Spread_g \times Post_t + x_{gt} + \chi_t + \chi_g + \epsilon_{gt} \tag{6}$$

where $Post_t$ is the indicator equal to 1 for all months after January 2015; x_{gt} is the average age for patients in disease group g during week t; χ_t and χ_g are the year-month-week–level and the disease group–level fixed effects, respectively. The parameter β_1 shows the marginal effect of the differential upcoding incentive. Positive β_1 means that a larger increase in score spread across diseases within a disease group leads to a higher increase in the share of admissions assigned to high-score diseases within the group.

The identification assumption is the parallel change in the health status across disease groups. In this exercise, we only use the variation within City S rather than comparing the variation in City S with that in City N, although we obtain the discharge records for City N from January 2014 to December 2015. It is because the payment systems in the two cities differ in 2014, and this can lead to different coding behaviors in the two cities during 2014, namely non-parallel trend in coding before the policy intervention.

Panel b1 of Table 4 displays the variables in equation 6. The sample is aggregated from the individual-encounter level Sample B to disease group-week level by the number of admissions. We drop disease groups with no admissions before January 2015 since these groups have no pre-treatment period. This gives us 151 disease groups, with 2.29 diseases in each group on average, and 10,037 disease group-week-level observations. The average

change in score spread ($\Delta Spread_g$) is -0.174, and the standard deviation is 1.816. The fraction of admissions coded to high-score diseases in a disease group ($frac_{gt}$) is 0.184 on average. The correlation in the change of these two variables before and after the policy intervention helps us to identify the policy-induced coding incentives.

Table 8 presents the results. The first three columns use all patients in our sample, and the last three columns use the subsample of patients who have admitted both before and after the policy change in City S.

Column (1) of Table 8 reports the estimates for equation 5. The results indicate a positive effect of the change in score spread on the admission share for high-score diseases within the group, but the coefficient is not significant. One concern for the result is heterogeneous time trends for patients' health status across disease groups. Thus, columns (2) and (3) control for a linear time trend for each of the 127 number of 3-digit disease groups and the 18 number of 1-digit disease groups in equation 6, respectively. Columns (2) and (3) show that the coefficient estimates on $\Delta Spread_g \times Post_t$ have little changes with column (2), and the coefficient becomes significant when we control for the heterogeneous time trends at the disease group level in column (3). This indicates that hospitals are more likely to code patients to the high-score disease within the group when the score become more dispersed within the group after the policy change. Also, when we restrict our sample to patients who have been admitted both before and after the policy change, the results are still robust, as columns (4) (5) (6) show.

5.2 Individual Analysis

Same as above, we also define disease groups according to the 3-digit ICD-10 codes.[20] For disease group g and patient i coded to the group, we calculate the median score across all

[20]The reason for using the 3-digit ICD-10 codes rather than 1-digit ICD-10 codes here is similar to that in section 5.1. There are adequate number of diseases within one 3-digit ICD-10 for an average patient, so that we are able to do the comparison across diseases within the 3-digit ICD-10 codes for the same patient.

of patient i's admissions coded to disease group g during our sample period. This median score denotes patient i's median level health status within the group. Given this, for patient i's admissions coded to group g during week t, we then calculate the fraction for diseases with a score above the median score, $frac_{igt}$. This indicates the probability that patient i is coded to high-score diseases within the same disease group. We expect the fraction to increase more pronouncedly for patients who are subject to lower reimbursement after the policy change. To measure the change in reimbursement for a patient before and after the policy change, we assume the patient's health status remain stable over time so that we focus on her diseases during the pre-treatment period t_0. For patient i's pre-treatment diseases in disease group g, we calculate her average coding score according to the 2014- and 2015-version score schedules, respectively: $\overline{\delta^{y}_{igt_0}}$, for $y = 2014, 2015$. The difference between the two average coding scores is our independent variable:

$$\Delta\delta_{ig} = \log(\overline{\delta^{2015}_{igt_0}}) - \log(\overline{\delta^{2014}_{igt_0}}). \tag{7}$$

The specification we use is:

$$frac_{igt} = \beta_0 + \beta_1 \Delta\delta_{ig} \times Post_t + \chi_{ig} + \chi_t + \epsilon_{igt}, \tag{8}$$

where χ_{ig} is the set of patient-disease group-level fixed effects, capturing the time-invariant change in the scores; χ_t is the set of year-month-week-level fixed effects, controlling for the over time variation irrelevant to the policy shock. The coefficient β_1 tells the marginal effect of the differential change in scores on coding incentives. We expect negative β_1, meaning that patients with diseases subject to lower scores (or, lower reimbursement) under the new version are more likely to be coded to high-score diseases within the disease group after the policy change. The underlying identification assumption is the parallel change in patients' health status for those assigned to the same disease group.

The advantage of the individual analysis is that it compares coding for individuals conditional on the same disease group, while, the aggregate analysis uses the variation across disease groups. If the coding behavior intrinsically vary differently over time across different disease groups, the aggregate analysis may over- or under-estimate hospitals' coding responses to price changes. In this sense, the individual analysis better controls for the heterogeneous time trends in coding across disease groups.

Panel b2 of Table 4 describes the variables used in equation 8. The sample is aggregated from Sample B to the level of individual-disease group-week by the number of admissions. We focus on individuals with admissions both before and after the policy intervention in January 2015 because we use the within-individual variation. There are 340 disease groups and 14,344 patients after the aggregation. The change in a patient's pre-treatment coding score under the two versions of scores ($\Delta\delta_{ig}$) has the mean of -0.283 and the standard deviation of 0.239. The average share of admissions coded to high-score diseases within a disease group ($frac_{igt}$) is 0.294 on average, with a standard deviation of 0.455.

Column (1) of Table 9 presents the estimates for equation 8. The result is consistent with our conjecture: the coefficient on $\Delta\delta_{ig} \times Post_t$ is negative (-0.014). But the coefficient is not statistically significant, probably because there is not enough variation in the change of the average coding score for a patient's pre-treatment diseases. Column (2) of Table 9 thus uses the discrete value for the change in coding scores: $1\{\Delta\delta_{ig} \geq 0\}$. The result shows a significant negative coefficient (-0.042) on the interaction. This means that the probability of being coded to high-score diseases within a disease group increases to a larger magnitude for patients whose pre-treatment coding diseases are subject to lower scores after 2015.

Similar to before, the above estimates may be challenged by the heterogeneous time trends across patients and disease groups. Ideally, to check for the robustness, we want to add a linear time trend for each combination of patients and disease groups. But due to the large number of the time trends (18,761), we group patients into two types—patients

with $\Delta\delta_{ig} \geq 0$, and patients with negative $\Delta\delta_{ig} < 0$—and control for a linear time trend for each combination of patient types and disease groups, namely 680 linear time trends. Columns (3) and (4) show the result: both coefficients change little with those in columns (1) and (2), and the coefficient in column (4) stays significant at the p-value<0.05. Above all, our individual-level result supports the aggregate-level result: there is significant coding response to price changes under D-PPS.

The estimate in column (1) of Table 9 implies a 0.07% increase in the coding score for patients with admissions both before and after the policy intervention in our sample.[21] The 0.07% increase in coding score is equivalent to $0.75 excess expenditure per admission, or $70,448 excess expenditure for City S in 2015. [22]

5.3 Heterogeneity

As the benefit and cost of the coding response differ across patients and hospitals, we also investigate the heterogeneous coding responses across them. Table 10 presents the results.

Columns (1) and (2) report the estimates of equation 8 for old versus young patients. As before, we define a patient as old if her pre-treatment age is above the median across all patients in our sample. The result shows a significantly lower probability of being coded to high-score diseases for patients subject to higher scores after the score adjustment for young patients, but this incentive is insignificant among old patients. This is probably because the

[21]This is calculated as follows. Using the estimate in column (1) as an example, given the average change in a patient's average coding score -0.291 in 2015, the policy change induces a -0.291×(-0.014) = 0.004 increase in the fraction that a patient is coded to high-score diseases within one disease group. Given the average fraction under the current environment 0.317 in 2015, the average fraction without the policy is 0.317 - 0.004 = 0.313. Also, since the average coding score in high- and low-score diseases are 96 and 113 in 2015 for an average disease group, the average coding score is thus 0.317×113 + (1-0.317)×96 = 101.32 with the policy intervention and 0.313×113+(1-0.313)×96 = 101.25 without the policy intervention, namely (101.32 - 101.25)/101.25 =0.07%.

[22]This calculation is from the average expenditure per admission of $1082 in our sample during 2015 and 93,079 admissions in City S in 2015. The average expenditure per admission without the policy is $1082/(1+0.07%) = $1081.24; namely, $0.75 expenditure per admission, or $0.75 *93079= $70,448 health-care expenditure for City S, or $70448*(1-0.16) = $59,176 health-care expenditure to the government of City S in 2015.

risk of upcoding is smaller for young patients.

Columns (3) and (4) show the estimates of equation 8 for high- and low-cost patients. We define a patient as high-cost if her average expenditure per admission during the pre-treatment period is above the median. The result shows the significant and sizable heterogeneity: the coefficient on $\Delta\delta_i \times Post_t$ is significantly negative (-0.096) for high-cost patients but insignificant (0.006) for low-cost patients. This suggests that hospitals tend to mainly upcode high-cost, severe patients.

Columns (5) to (7) present the heterogeneity of hospitals in different quality tiers. In this exercise, we aggregate the individual-encounter-level data to the unit of individual-hospital-disease group. Also, to control for the time-invariant differences across hospitals, we add the hospital-level fixed effects to equation 8.[23] The results show a significant negative coefficient on the interaction for high-tier hospitals (-0.093), but insignificant coefficients for medium- and low-tier hospitals. This means that the upcoding response to the score adjustment mainly occurs in high-tier hospitals, probably because there are more severe patients.

6 Whether Upcoding Suggests Better Treatments?

The above results suggest upcoding behavior under D-PPS, and this generates excess health-care expenditure. But if upcoding also indicates better treatments, the inefficiency caused by upcoding is ameliorated. Thus, in this section, we examine whether hospitals also invest more health-care resources to upcoded patients. Our identification strategy is similar to what have been used previously. If it is the case, we should expect that hospitals' intensity

[23]Thus, the dependent variable is the fraction of admissions that patient i is coded to high-score disases within one disease group g, hospital h in week t, $frac_{ight}$. The independent variable is the change in average pre-treatment coding score for patient i's admissions in disease group g, hospital h under the two version scores: $\Delta\delta_{igh}$. The model specification becomes:

$$frac_{ight} = \beta_0 + \beta_1 \Delta\delta_{igh} \times Post_t + \chi_{ig} + \chi_h + \chi_t + \epsilon_{ight}. \quad (9)$$

of care responds to the two policy shocks in the same way as their coding behavior.

In response to the first policy shock, hospitals are more likely to code patients to diseases subject to higher reimbursement under D-PPS versus FFS. Similarly, if hospitals also provide better services to upcoded patients, we should expect the intensity of care to also increase more significantly for diseases generating higher reimbursement to a hospital. Thus, we estimate equation 3 for 3 dependent variables of the intensity of care—average health-care expenditure, average length of stay, and total volume. A higher average expenditure, length of stay, or a larger volume suggest more resources that are invested on patients. The key parameter we care about is the coefficient on the triple interaction $City_s \times Post_t \times \Delta Reimb_{dh}$, and it is expected to be positive.

Table 11 presents the estimates. The sample used in the estimation is aggregated from Sample A to the level of disease-hospital-week by the number of admissions. In particular, the first three columns are based on an aggregation for the full sample, and the last three columns are conditional on patients who have been admitted both before and after the policy change. All specifications control for the disease-level and the hospital-level fixed effects and a full set of time fixed effects. The coefficients on the triple interaction are all insignificant except for column (1), and some of the coefficients are even negative. This indicates that hospitals do not significantly invest more resources on diseases with stronger upcoding incentives. The coefficient on the interaction $City_s \times Post_t$ indicates the average change in the intensity of care under D-PPS versus FFS. The estimates are all negative, meaning that hospitals tend to reduce the cost of care under D-PPS versus FFS. This is consistent with previous literature.

In response to the second policy shock, hospitals have more incentives to code patients to high-score diseases within a disease group if the disease group are subject to a larger score spread after the policy change. Thus, similarly, we expect the intensity of care to increase for disease groups subject to a higher score spread after the policy change, and the increase

is more pronounced for the disease group with a larger increase in score spread. Thus, we estimate equation 6 for the 3 dependent variables—average health care expenditure, average length of stay, and total volume.

Table 12 reports the results. All specifications have controlled for the disease group-level fixed effects, linear time trends for disease groups (defined by 1-digit ICD-10 codes), and a full set of time fixed effects. The estimated coefficients on the interaction $\Delta Spread_g \times Post_t$ are all insignificant, meaning that hospitals do not significantly provide more resources for the disease group subject to a larger score spread after the 2015 score adjustment, or namely a higher upcoding intensity.

Above all, we do not find significant evidence that upcoded patients also receive better treatment. In other words, hospitals' upcoding behavior only generates excess health-care expenditure and it does not significantly improve the intensity of care on patients.

7 Conclusion

This paper studies hospitals' upcoding behavior in China's UEBMI. We investigate upcoding responses to a payment reform that has, first, replaced FFS to D-PPS and, second, adjusted the score one year after the introduction. We find that hospitals' coding are more likely to positively respond to diseases with the largest increases in reimbursement under D-PPS versus FFS. Also, hospitals tend to code patients to high-score diseases in response to an increase in diseases' prices. The upcoding responses translate to a 7.11% (or \$82 per admission) increase in health-care expenditure under D-PPS versus FFS in 2014 for City S, and a 0.07% (or \$0.75 per admission) increase in health-care expenditure after score adjustment in 2015 for City S. Also, we do not find significant evidence that hospitals also improve the intensity of care for patients who are upcoded. Our results suggest that upcoding exists in D-PPS, and this weakens its efficiency in cost-control. Policy-makers need to regulate

hospitals' upcoding behavior when applying D-PPS.

References

Bastani, H., J. Goh, and M. Bayati (March 2019). Evidence of upcoding in pay-for-performance programs. *Management Science 65*(3).

Brown, J., M. Duggan, I. Kuziemko, and W. Woolston (OCTOBER 2014). How does risk selection respond to risk adjustment? new evidence from the medicare advantage program. *AMERICAN ECONOMIC REVIEW 104*(10).

Brunt, C. S. (July 2011). Cpt fee differentials and visit upcoding under medicare part b. *Health Economics 20*.

Brunt, J. R. B. C. S. (July 2014). Medicare skilled nursing facility reimbursement and upcoding. *Health Economics 23*.

Carter, G. M., J. P. Newhouse, and D. A. Relles (1990). How much change in the case mix index is drg creep? *Journal of Health Economics 9*(4), 411–428.

Chan, M. K. and G. Zeng (2018). Unintended consequences of supply-side cost control? evidence from china's new cooperative medical scheme. *Journal of Health Economics 61*, 27–46.

Coulam, R. F. and G. L. Gaumer (1992 March). Medicare's prospective payment system: A critical appraisal.

Dafny, L. S. (Dec., 2005). How do hospitals respond to price changes? *The American Economic Review 95*(5), 1525–1547.

Fang, H. and Q. Gong (FEBRUARY 2017). Detecting potential overbilling in medicare reimbursement via hours worked. *AMERICAN ECONOMIC REVIEW 107*(2).

Gao, C., F. Xu, and G. G. Liu (2014). Payment reform and changes in health care in china. *Social Science & Medicine 111*, 10–16.

Geruso, M. and T. Layton (June 2019). Upcoding: Evidence from medicare on squishy risk adjustment. *Journal of Political Economics forthcoming*.

Huang, F. and L. Gan (2017). The impacts of china's urban employee basic medical insurance on healthcare expenditures and health outcomes. *Health Economics 26*, 149–163.

J.M.Steinbusch, P., J. B.Oostenbrink, J. J.Zuurbier, and F. J.M.Schaepkens (May 2007). The risk of upcoding in casemix systems: A comparative study. *Helath Policy 81*.

Jürges, H. and J. Köberlein (September 2015). What explains drg upcoding in neonatology? the roles of financial incentives and infant health. *Journal of Health Economics 43*.

Nie, X., S. Mattke, Z. Predmore, and H. Liu (2016). Upcoding and anesthesia risk in outpatient gastrointestinal endoscopy procedures. *JAMA Internal Medicine 176*(6).

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

Yip, W. and K. Eggleston (2001). Provider payment reform in china: the case of hospital reimbursement in hainan province. *Health Economics 10*(4), 325–339.

8 Appendix

8.1 Tables

Table 1: An Example for Diagnosis Scores in City S

3-digit ICD	Disease	2014 Score	2015 Score
B05	Melease	38	30
	Melease with pneumonia	68	68
J18	pneumonia	90	45
	Fungal pneumonia	300	200
	Bronchitis	68	60
	carcinoma of gallbladder	400	400
	gallstone	110	88
	gallstone with other cholecystitis	122	92
	gallstone with cholecystitis	118	92
	gallstone with Gangrenous cholecystitis	146	138
	gallstone with acute cholecystitis	130	92
	gallstone with chronic cholecystitis	112	92

Notes: This table shows the distribution of the two versions of diagnosis scores. We also provide examples for diagnoses and the scores.

Table 2: Sample Description

	Sample A	Sample B
Sample Period	Aug., 2015 - Dec., 2015	Jan., 2014 - Dec., 2015
Event	The First Policy Shock The Introduction of D-PPS in replace of FFS	The Second Policy Shock The Adjustment in Diagnosis-specific Scores
Policy Intervention Time	January 2014	January 2015
Treatment	UEBMI Enrollees in City S	URBMI and UEBMI Enrollees in City S
Control	UEBMI Enrollees in City N	

Notes: This table introduces the two sample for event analyses of the two policy shocks.

Table 3: Summary Statistics for Samples

	Sample A		Sample B
	City S	City N	City S
No. of Admissions	105,372	106,786	175,745
Panel a: Individuals			
No. of Individuals	43,657	64,598	115,669
Individuals' Age	62.75	61.02	53.61
Avg. Expenditure per admission ($)	1,075	1,950	926
Panel b: Hospitals			
No. of Hospitals	186	168	175
#High-tier	4	8	5
#Medium-tier	29	36	30
#Low-tier	153	124	145
Share of Admissions			
High-tier	0.301	0.731	0.304
Medium-tier	0.525	0.204	0.520
Low-tier	0.174	0.065	0.176
Avg. Expenditure per admission ($)			
High-tier	1,642	2,205	1,303
Medium-tier	971	1,446	872
Low-tier	404	662	435
Panel c: Diseases			
#Diseases	821	902	992
Top 1 Disease	Cerebral	Coronary	Cerebral
(%Admissions)	Infarction	Heart Disease	Infarction
	(0.156)	(0.066)	(0.115)
Top 2 Disease	Coronary	Cerebral	Bronchitis
(%Admissions)	Heart Disease	Infarction	
	(0.053)	(0.059)	(0.041)
Panel d: Disease Groups			
#1-digit Disease Groups	19	20	20
Avg. #Diseases per group	43.473	47.650	49.650
Top 1 Group (%Admissions)	I (0.422)	I (0.298)	I (0.309)
Top 2 Group (%Admissions)	J (0.141)	K (0.150)	J (0.183)
#3-digit Disease Groups	386	399	432
Avg. #Diseases per group	2.147	2.388	2.299
Top 1 Group (%Admissions)	I63 (0.159)	I49 (0.080)	I63 (0.118)
Top 2 Group (%Admissions)	J44 (0.066)	I10 (0.064)	J44 (0.072)

Notes: This table presents the summary statistics for sample A and B (see Table 2). Sample A is used for the event analysis of the introduction of D-PPS, and sample B is used for the analysis for the 2015 score adjustment.

Table 4: Summary Statistics for Variables

	Observations	Mean	S.D.	Min	Max
Panel a: The Effect of The Introduction of D-PPS					
Panel a1: Aggregate Analysis (Equation 3)					
$sh_{d,ht}$	45,345	.173	.272	.005	1
	49,624	.081	.174	.003	1
$sh_{d\|g,ht}$	45,345	.494	.366	.016	1
	49,624	.330	.333	.009	1
$\Delta Reimb_{dh}$	45,345	.083	.393	-2.547	4.295
	49,624	.056	.348	-2.568	12.389
Average Age	45,345	61.173	16.360	0	115
	49,624	59.654	16.056	19	112
Panel a2: Individual Analysis (Equation 4)					
$\bar{\delta}_{it}$	30,073	115.467	69.419	25	500
	33,015	131.405	91.730	25	500
Average Age	30,073	67.656	13.879	0	115
	33,015	67.478	13.895	19	113
Panel b: The Effect of The Score Adjustment					
Panel b1: Aggregate Analysis (Equation 6)					
$frac_{gt}$	10,037	.184	.306	0	1
$\Delta Spread_g$	10,037	-.174	1.816	-9.919	6.271
Average Age	10,037	50.612	19.293	0	116
Panel b2: Individual Analysis (Equation 8)					
$frac_{igt}$	37,530	.294	.455	0	1
$\Delta\delta_{ig}$	37,530	-.283	.239	-1.793	.602

Notes: This table describes the variables used for estimating the coding responsese to the two policy shocks. Panel a1 and a2 are both aggregated from Sample A by the number of admissions; and Panel b1 and b2 are aggregated from Sample B by the number of admissions. The unit of observation is a hospital-disease-week in panel a1 and a patient-week in panel a2. For each variable, we show the statistics for City S and City N separately, with the first row for City S and the second row for City N. The unit of observation is a (3-digit) disease group-week in panel b1, and a patient-(3-digit) disease group-week in panel b2.

Table 5: Coding Responses to the First Policy Shock: Aggregate Results

Sample	Full Sample		Patients with Admissions in Pre&Post-treat Periods	
Dependent Variable	$\log(sh_{d,h,t})$	$\log(sh_{d\|g,h,t})$	$\log(sh_{d,h,t})$	$\log(sh_{d\|g,h,t})$
	(1)	(2)	(3)	(4)
$City_s \times Post_t \times \Delta Reimb_{dh}$	0.066***	0.056*	0.044	0.012
	(0.019)	(0.023)	(0.029)	(0.032)
$City_s \times Post_t$	0.012	0.014	0.106***	0.070***
	(0.008)	(0.008)	(0.011)	(0.012)
$City_s \times \Delta Reimb_{dh}$	-0.071***	-0.028	-0.064**	-0.020
	(0.014)	(0.017)	(0.021)	(0.024)
$Post_t \times \Delta Reimb_{dh}$	-0.035**	-0.046**	-0.033	0.018
	(0.013)	(0.017)	(0.022)	(0.025)
$\Delta Reimb_{dh}$	0.033***	0.028*	0.057***	0.008
	(0.009)	(0.012)	(0.015)	(0.018)
Average Age	0.000	0.001***	-0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Hospital F.E.	Y	Y	Y	Y
Disease F.E.	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y
Observations	94,969	94,969	38,573	38,573
Implied %ΔScore	1.69%	1.83%	10.83%	7.11%

Notes: This table presents the estimates of equation 3: effects of the change in average reimbursement under the D-PPS versus FFS, measured by $\Delta Reimb_{hd}$ (defined in equation 1), on the log of share of admissions. The sample is aggregated from Sample A to the level of disease/hospital/week by the number of admissions. The dependent variables in columns (1) and (3) are both the log of share of admissions that are coded to disease d in hospital h at week t among all diseases, log of $sh_{d,h,t}$; and the dependent variables in columns (2) and (4) are both the log of share of admissions that are coded to disease d among the same disease group (defined by 1-digit ICD-10) g in hospital h at week t, log of $sh_{d|g,h,t}$. Standard errors in parentheses are robust, * p<0.05,** p<0.01, *** p<0.001. The implied changes in score, or equivalently the implied change in health-care expenditure, displayed in the last row are calculated from the estimates.

Table 6: Coding Responses to the First Policy Shock on Coding: Individual Results

Dependent Variable	Log of 2014 Diagnosis Score					
Sample	Patients w. Admissions in Pre&Post-treat Periods				Diseases with Higher Reimb.	Diseases with Lower Reimb.
	(1)	(2)	(3)	(4)	(5)	(6)
$CityS_i \times Post_t$	0.068***	0.063***	0.060***	0.052***	0.035**	0.086***
	(0.012)	(0.008)	(0.009)	(0.008)	(0.011)	(0.012)
$CityS_i$	-0.101***					
	(0.013)					
Individual F.E.		Y				
Ind×Hos F.E.			Y		Y	Y
Ind×Hos×1-d ICD F.E.				Y		
Time F.E.	Y	Y	Y	Y	Y	Y
Observations	63,088	63,088	52,869	52,869	31,843	21,026

Notes: This table presents the estimates of equation 4: the effect of the introduction of D-PPS in replace of FFS on coding. All columns are estimated conditional on patients who have admissions both before and after January 2014. Conditional on these patients, columns (5) and (6) use the subsample of admissions in diseases with higher or lower reimbursement under the D-PPS versus FFS, measured by non-negative/negative $\Delta Reimb_{hd}$ (defined in equation 1). The unit of observation is an individual/week in columns (1) and (2), and the unit of observation is an individual/disease/hospital/week for the last four columns, all of which are aggregated from Sample A by the number of admissions. Standard errors in parentheses are robust and clustered at the individual level in columns (1) (2), at the individual-hospital level in columns (3) (5) (6), and at the level of individual/hospital/disease group (defined by 1-digit-ICD) in column (4), * p<0.05,** p<0.01, *** p<0.001.

Table 7: Heterogeneous Coding Responses to the First Policy Shock: Individual Results

Dependent Variable	Log of 2014 Diagnosis Score						
Sample	By Patients' Age		By Patients' Avg. Expenditure		By Hospitals' Quality Tier		
	≥Median (69)	<Median (69)	≥Median ($1170)	<Median ($1170)	High Tier	Med. Tier	Low Tier
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$CityS_i \times Post_t$	0.051***	0.053***	0.086***	0.053***	0.064***	0.077***	0.053*
	(0.011)	(0.011)	(0.013)	(0.012)	(0.015)	(0.015)	(0.023)
Individual F.E.	Y	Y	Y	Y	Y	Y	Y
Hospital F.E.	Y	Y	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y	Y	Y
Obervations	33,127	30,209	33,633	29,703	26,025	20,000	6844

Notes: This table presents the estimates of equation 4 for different subsamples of patients and hospitals. The sample is aggregated from Sample A to the level of individual/hospital/week by the number of admissions. Thus, each column has controlled for both individual- and hospital-level fixed effects. Standard errors in parentheses are robust and clustered at the individual level, * p<0.05,** p<0.01, *** p<0.001.

Table 8: Coding Responses to the Second Policy Shock: Aggregate Results

Sample	Full Sample			Patients w. Admissions in Pre&Post-treat Periods		
Dependent Varibale	$frac_{gt}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta Spread_g \times Post_t$	0.020	0.023	0.030*	0.005	0.006	0.010
	(0.011)	(0.012)	(0.012)	(0.010)	(0.009)	(0.012)
disease group Time Trend		Y			Y	
One-digit ICD Time Trend			Y			Y
Observations	10,037	10,037	10,037	4,910	4,910	4,910

Notes: This table reports the estimates of equation 6: the effect of the percentage change in the standard deivation of scores for diseases within one disease group ($\Delta Spread_g$) on the share of admissions that are assigned to diseases subject to non-lower scores within the disease group ($frac_{gt}$). The disease group is defined as the three-digit ICD-10 code. The sample is aggregated from Sample B to the disease group-week level by the number of admissions. Columns (1) to (3) include admissions for all patients; and columns (5) to (7) only include admissions for patients who have admissions both before and after the policy intervention in January 2015. The unit of observation is disease group-week, and the observations where disease group g has no admissions during the pre-treat period (before January 2015) have been dropped. All specifications have controlled for year-month-weekly fixed effects and disease group-level fixed effects. Standard errors are robust and clustered at the disease group-level, * p<0.05,** p<0.01, *** p<0.001.

Table 9: Coding Responses to the Second Policy Shock: Individual Results

Sample	Patients w. Admissions in Pre&Post-treat Periods			
Dependent Variable	$frac_{igt}$			
	(1)	(2)	(3)	(4)
$\Delta\delta_{ig} \times Post_t$	-0.014		-0.012	
	(0.012)		(0.017)	
$1\{\Delta\delta_{ig} \geq 0\} \times Post_t$		-0.042***		-0.042*
		(0.011)		(0.017)
Ind×Disease-group F.E.	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y
Ind-type×Disease-group Linear Time Trends			Y	Y
Observations	37,530	37,530	37,530	37,530
Implied %Coding Score	0.07%			

Notes: This table reports the estimates of equation 8: the effect of the change in average coding scores for a patient's pre-treat diseases on the probability of being coded to the high-score diseases within the same disease group. The sample is aggregated from Sample B to the level of patient-disease group-week by the number of admissions. Standard errors are robust and clustered at the level of patient-disease groups, * p<0.05,** p<0.01, *** p<0.001.

Table 10: Heterogeneous Coding Responses to the Second Policy Shock: Individual Results

Sample	Patients w. Admissions in Pre&Post-treat Periods						
	By Patients' Age		By Patients' Avg. Expenditure		By Hospitals' Quality Tier		
	≥Med. (70)	<Med. (70)	≥Med. ($787)	<Med. ($787)	High	Med.	Low
Dependent Variable	$frac_{igt}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$1\{\Delta\delta_{ig} \geq 0\} \times Post_t$	-0.008	-0.064**	-0.051**	0.006	-0.093**	-0.016	-0.034
	(0.027)	(0.022)	(0.019)	(0.034)	(0.033)	(0.011)	(0.028)
Ind×Disease-group F.E.	Y	Y	Y	Y	Y	Y	Y
Hospital F.E.					Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y	Y	Y
Observations	20,192	17,338	20,363	17,167	8,853	18,259	8,329

Notes: This table reports the estimates of equation 8 for different subsample of patients. The unit of observation is a patient-disease group-week for columns (1) to (4) and a patient-hospital-disease group-week for columns (5) to (7), all of which are aggregated from sample B by the number of observations. Standard errors are robust and clustered at the level of patient-disease groups, * p<0.05,** p<0.01, *** p<0.001.

Table 11: Intensity of Care Responses to the First Policy Shock

Sample	Full Sample			Patients w. Admissions in Pre&Post-treat Periods		
Dependent Variable	log(Expense)	log(LOS)	log(Volume)	log(Expense)	log(LOS)	log(Volume)
	(1)	(2)	(3)	(4)	(5)	(6)
$City_s \times Post_t \times \Delta Reimb_{dh}$	0.083*	-0.023	-0.012	0.081	-0.002	0.004
	(0.033)	(0.027)	(0.016)	(0.045)	(0.043)	(0.022)
$City_s \times Post_t$	-0.104***	-0.163***	-0.187***	-0.122***	-0.159***	-0.209***
	(0.008)	(0.008)	(0.007)	(0.013)	(0.013)	(0.010)
$City_s \times \Delta Reimb_{dh}$	0.112***	0.084***	-0.039***	0.128***	0.071*	-0.041*
	(0.021)	(0.020)	(0.011)	(0.035)	(0.032)	(0.016)
$Post_t \times \Delta Reimb_{dh}$	0.455***	0.335***	-0.023*	0.483***	0.336***	-0.041*
	(0.028)	(0.022)	(0.011)	(0.035)	(0.033)	(0.016)
$\Delta Reimb_{dh}$	-0.775***	-0.461***	0.036***	-0.779***	-0.431***	0.058***
	(0.017)	(0.016)	(0.008)	(0.028)	(0.025)	(0.011)
Average Age	0.005***	0.003***	-0.000	0.003***	0.002***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	94,968	94,841	94,969	38,573	38,526	38,573

Notes: This table reports the estimates of equation 3 for the dependent variable of three measures of the intensity of care. The unit of observation is a disease-hospital-week that are aggregated from sample A by the number of observations. All specifications have included disease-level fixed effects, hospital-level fixed effects and the full set of time fixed effects. Standard errors are robust, * p<0.05,** p<0.01, *** p<0.001.

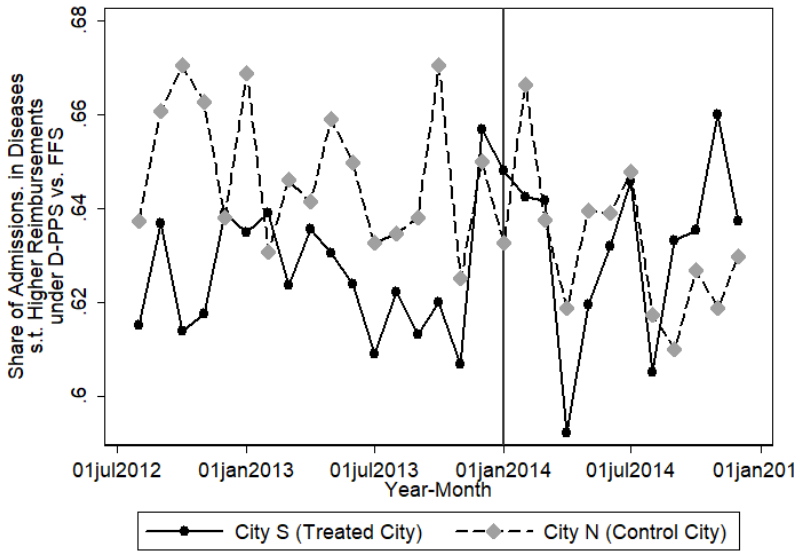
Table 12: Intensity of Care Responses to the Second Policy Shock

Sample	Full Sample			Patients w. Admissions in Pre&Post-treat Periods		
Dependent Variable	log(Expense)	log(LOS)	log(Volume)	log(Expense)	log(LOS)	log(Volume)
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta Spread_g \times Post_t$	0.005	0.008	-0.003	0.005	0.009	0.013
	(0.008)	(0.005)	(0.017)	(0.012)	(0.009)	(0.011)
Average Age	0.003***	0.002***	-0.001	0.003**	0.002*	-0.000
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	10,037	10,033	10,037	4,910	4,902	4,910

Notes: This table reports the estimates of equation 6 for the dependent variable of three measures of the intensity of care. The unit of observation is a disease group-week that are aggregated from sample B by the number of observations. All specifications have included the disease group-level fixed effects, linear time trends at the level of 1-digit ICD group, and the full set of time fixed effects. Standard errors are robust, * p<0.05,** p<0.01, *** p<0.001.

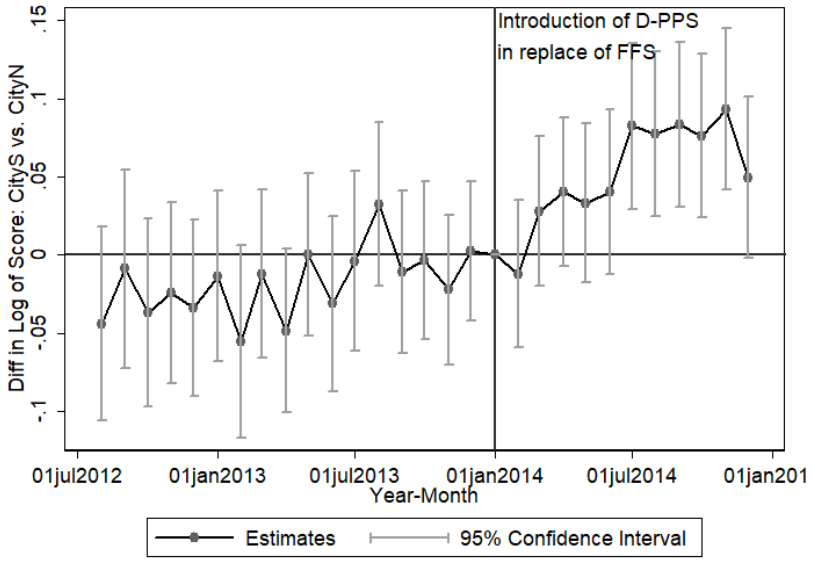
8.2 Figures

Figure 1: The Share of Admissions Assigned to Diseases subject to Higher reimbursement under D-PPS vs. FFS



Notes: This figure displays the average share of admissions that are assigned to diseases with $\Delta Reimb_{hd} >= 0$ within the same disease group (defined by 1-digit ICD-10) for a hospital in City S versus City N. In particular, we compute the fraction of admissions that are assigned to diseases with $\Delta Reimb \geq 0$ within the same disease group g in hospital h during month t; and then we compute the average of the fraction across disease groups and hospitals over time. The straight line indicates the starting time of the policy intervention in January 2015.

Figure 2: Differences in Log of Coding Score for City S vs. City N



Notes: This figure displays the estimate coefficients and confidence intervals for equation 4, while allowing the difference-in-difference parameter β_1 varies over months. The month dummy for the starting time of the policy change (January 2014) is omitted.

8.3 Data Consolidation

Before merging the diagnosis-specific scores to samples A and B, we clean up the diagnosis information contained in samples A and B as follows. First, we clean the format and abbreviations of the diagnosis information to be consistent with that in the score schedules. Second, in records with multiple diseases, we separate these to multiple observations with one disease per observation. For example, for a record "herpes zoster and hypertension," we separate this into two observations: one with "herpes zoster" as an observation and "hypertension" as the other, since these two diseases differ in their scores. Third, we also drop records with insufficient information to identify a diagnosis. For example, a record describing "malignant tumor" can be either "malignant tumor of the kidney" or "malignant tumor of the liver," with different scores assigned. After the above, we use the diagnosis information to merge scores to samples A and B, and we drop records without scores.

We then explain procedures to match the ICD-10 codes for samples A and B. Since we observe the ICD-10 codes for all of the discharge records after January 2015 for City S and some of the records before that date, we impute the 3-digit ICD-10 codes for the records with missing values based on the records with non-missing values. We observe records with the same diagnosis but different ICD-10 codes, and these usually differ after the 3 digits. For these records with multiple ICD-10 codes observed in the data, we calculate the average expenditure for each ICD-10 code and assign the ICD-10 code with the closest level of average expenditure to the record. For the remaining records without ICD-10 codes, we look for their ICD-10 code (2015 version) through online searching.[24]

[24]The website we use is `https://icd.who.int/browse10/2015/en`.