OXFORD

Hospital response to a new case-based payment system in China: the patient selection effect

Xinyu Zhang [1,2], **Shenglan Tang** [2,3,4], **Ruixin Wang** [1], **Mengcen Qian** [1,5], **Xiaohua Ying** [1,5,*] and **Matthew L Maciejewski** [6,7,8,9]

[1]School of Public Health, Fudan University, 130 Dong'an Road, Shanghai 200030, China
[2]Duke Global Health Institute, Duke University, 310 Trent Drive, Durham, NC 27710, United States
[3]Global Health Research Center, Duke Kunshan University, No. 8 Duke Avenue, Kunshan, Jiangsu 215316, China
[4]SingHealth Duke-NUS Global Health Institute, Duke-NUS, 8 College Road, Singapore 169857, Singapore
[5]Key Laboratory of Health Technology Assessment, National Health Commission (Fudan University), 130 Dong'an Road, Shanghai 200030, China
[6]Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT), Durham Veterans Affairs Health Care System, 508 Fulton Street, Durham, NC 27705, United States
[7]Department of Population Health Sciences, Duke University, 215 Morris Street, Durham, NC 27701, United States
[8]Division of General Internal Medicine, Department of Medicine, Duke University, 40 Duke Medicine Circle, Durham, NC 27710, United States
[9]Duke-Margolis Health Policy Center, Duke University, 230 Science Drive, Durham, NC 27708, United States

*Corresponding author. School of Public Health, Fudan University, 130 Dong'an Road, Shanghai 200030, China. E-mail: xhying@fudan.edu.cn

Abstract

Providers have intended and unintended responses to payment reforms, such as China's new case-based payment system, i.e. Diagnosis-Intervention Packet (DIP) under global budget, that classified patients based on the combination of principal diagnosis and procedures. Our study explores the impact of DIP payment reform on hospital selection of patients undergoing total hip/knee arthroplasty (THA/TKA) or with arteriosclerotic heart disease (AHD) from July 2017 to June 2021 in a large city. We used a difference-in-differences approach to compare the changes in patient age, severity reflected by the Charlson Comorbidity Index (CCI), and a measure of treatment intensity [relative weight (RW)] in hospitals that were and were not subject to DIP incentives before and after the DIP payment reform in July 2019. Compared with non-DIP pilot hospitals, trends in patient age after the DIP reform were similar for DIP and non-DIP hospitals for both conditions, while differences in patient severity grew because severity in DIP hospitals increased more for THA/TKA ($P = 0.036$) or dropped in non-DIP hospitals for AHD ($P = 0.011$) following DIP reform. Treatment intensity (measured via RWs) for AHD patients in DIP hospitals increased 5.5% ($P = 0.015$) more than in non-DIP hospitals after payment reform, but treatment intensity trends were similar for THA/TKA patients in DIP and non-DIP hospitals. When the DIP payment reform in China was introduced just prior to the pandemic, hospitals subject to this reform responded by admitting sicker patients and providing more treatment intensity to their AHD patients. Policymakers need to balance between cost containment and the unintended consequences of prospective payment systems, and the DIP payment could also be a new alternative payment system for other countries.

Keywords: Payment system reform, Diagnosis-Intervention Packet, patient selection, age, severity, relative weight

Introduction

Prospective payment systems (PPSs) have been widely used to contain costs and improve healthcare efficiency worldwide (World Health Organization, Regional Office for the Western Pacific, 2015; Fu *et al.*, 2021). Total health expenditures in China have been growing at a 14.18% rate per year over the period from 2009 to 2020, which has prompted payment reform in recent years (CPC Central Committee and the State Council, 2020). The dominant payment method for inpatient care in China is case-based payment, which has two pillars [also known as the 'dual-track payment' (He, 2023)]: the traditional Diagnosis-Related Groups (DRG) payment and a new 'Diagnosis-Intervention Packet' (DIP) under the global budget (National Healthcare Security Administration, 2021). DIP payments are different from the DRG system because the DIP classification is based on the combination of principal diagnosis and procedures and is more granular than the DRG system with over 10 000 groupings.

The DIP payment system was also designed according to the Theory of Yardstick Competition, which reimbursed each group based on the historical average cost at the market level and incentivized providers to improve efficiency by reducing medical discretion (Hafsteinsdottir and Siciliani, 2010; Bäuml and Kümpel, 2021). DIP payment was developed to address the implementation of DRG payment in China, which was slower than anticipated because of its resource-

intensive requirement of clinical expert judgement to classify patients (Qian *et al.*, 2021). DIP payments have been implemented rapidly in China because it has little dependency on experts. The relative weight (RW) that determines the payment amount for an admission in each DIP group is determined by its historical relative costs (Lai *et al.*, 2022), and the payment varies by city.

Under the global budget, providers do not know the exact payment of each group amount during the treatment process, but have incentives to respond to DIP reform in two ways. First, they may change the services provided for a given patient (treatment policy). Second, providers may respond to DIP implementation by selecting more profitable patients (admission policy or selection effect) (Ellis and McGuire, 1996; Ellis, 1998; Sood *et al.*, 2013), which is the focus of this study. To achieve this target, they may avoid patients with high expected costs (Paddock *et al.*, 2007; Takahara, 2016; Kifmann and Siciliani, 2017; Vallance *et al.*, 2018) as the traditional DRG payment, which has been referred to as 'cream-skimming', 'cherry picking' or 'lemon dropping' (Martinussen and Hagen, 2009; Plate *et al.*, 2019; Ko *et al.*, 2022). This provider response to payment reform will likely limit access to healthcare (Kantarevic and Kralj, 2014) for older patients with more complications or comorbidities (Barnett *et al.*, 2019; Ko *et al.*, 2022). On the other hand, although demographic and clinical characteristics (e.g. age, condition severity and discharge status) are not considered in DIP classification (Lai *et al.*, 2022; He, 2023), providers may admit severe patients to provide them treatments with higher technical sophistication (intensity) that have higher expected payments (Hafsteinsdottir and Siciliani, 2010; Bäuml and Kümpel, 2021). This incentive is expected to be much stronger and easier to implement under the granular classification of DIP than DRG payments. Thus, providers would trade off between the two incentives. A diagram of the theoretical framework of provider responses to DIP payment is presented in Supplementary Figure A1 of Appendix A.

Prior studies of patient selection following payment reforms in several countries provide evidence for patient selection changes. Recent studies evaluating the implementation of bundled payment for joint replacement in America found no evidence of patient selection (Barnett *et al.*, 2019; Plate *et al.*, 2019), but another recent study found systematic avoidance

of patients aged 85 years and older (Ko *et al.*, 2022). Studies of Medicare inpatient rehabilitation facility PPS found no (Paddock *et al.*, 2007) or minimal change in patient composition (Sood *et al.*, 2013). Studies of DRG payment reform in China found that younger and less severely ill patients were deliberately selected for DRG payment (Jian *et al.*, 2015) when providers could decide whether to apply DRG or fee-for-service payment for each patient and that younger patients with lower Charlson Comorbidity Index (CCI) (a measure of patient severity) were selected under a previous case-base payment system when admitted for angina or acute myocardial infarction (Wu *et al.*, 2022). In terms of evaluation of DIP payment, we only found one relevant study, which observed no changes in patient characteristics (Lai *et al.*, 2022). However, this study used the whole sample of hospitalized patients instead of specific diseases and may not be very comparable over time and across different hospitals.

Prior evaluations have seldom examined whether patient age, severity or treatment intensity change after DIP implementation, so we aimed to extend prior work by exploring the patient selection behaviour of providers after the DIP payment reform in China. Specifically, we hypothesize that, after the DIP reform, providers behaviours of selecting younger patients and patients with lower severity would not be that obvious, and they may even admit more severe patients to increase the treatment intensity (reflected by the average RW of patients). We also explore whether tertiary hospitals have greater responses than secondary hospitals due to their greater ability to adjust clinical and administrative practice to respond to payment incentives. Our study aims to contribute to the hospital payment reform literature by deeply understanding the provider behaviour of inpatient selection following DIP implementation. We investigate these hypotheses from July 2017 to June 2021 when the Coronavirus disease (COVID-19) pandemic complicated provider's ability to react strategically to DIP implementation.

Materials and methods
Institutional background and study population
Two schemes constitute the basic medical insurance system in China, i.e. the employee basic medical insurance scheme (EBMIS) and the resident basic medical insurance scheme (RBMIS). The hospitals under investigation are in a large Tier-1 eastern China city with a population of 24.8 million, of which 16.2 million people are covered by EBMIS and 3.7 million are covered by RBMIS (National Bureau of Statistics of China, 2023). Before the case-based payment reform, it had implemented a global budget scheme at the hospital level.

Hospitalized EBMIS patients in two public tertiary and 11 public secondary hospitals subject to DIP reform starting in July 2019 were the intervention group of interest, since only EBMIS patients are the target population of DIP reform in this city. The control group was EBMIS patients in 8 tertiary hospitals and 19 secondary hospitals in the same city that did not participate in the case-base payment reform until July 2021.

To explore the potential patient selection behaviour, we selected two elective diseases or procedures that could be scheduled on a non-emergent basis, and hence patients have the possibility of being 'dumped' by a hospital and redesignated for care in another one (McNeely *et al.*, 2020): (1) patients receiving total hip/knee arthroplasty (THA/TKA)

and (2) patients with arteriosclerotic heart disease (AHD). We selected these two common reasons for hospital admission (Amanatullah *et al.*, 2019; Aengevaeren *et al.*, 2020; Long *et al.*, 2023) to make our study comparable with previous studies and address inpatient care that is instrumental to policymakers. Joint replacement and heart conditions are commonly studied at work, are prevalent, of high clinical relevance and well defined in hospitalized patients (General Office of the State Council of the People's Republic of China, 2019b), and are both paid close attention by the national health authority in China (General Office of the National Health Commission of China, 2020). THA/TKA is a typical surgery accounting for about 1.6% of operations in the study hospitals, with the third most cases in all the surgeries concerned by the National Health Commission of China (General Office of the National Health Commission of China, 2020). AHD is a representative condition primarily treated by medication, with the second largest inpatient volume in study hospitals and accounting for 3.7% of conditions. We included these two types of diseases to consider their potentially different patient selection effects. THA patients were identified by ICD-9CM-3 code of the principal procedure to be 00.7 and 81.51–81.53; TKA to be 00.80–00.83, 81.54 and 81.55. AHD patients were recognized by the ICD-10 code of principal diagnosis to be I25.1.

Data and variables

We used de-identified discharge records of EBMIS patients in this city from July 2017 to June 2021, which contain information on patient and hospital characteristics, admission and discharge time and status, diagnoses and procedures, and inpatient costs at the discharge level.

We obtained patient age, severity and RW of the DIP group, since they were our outcomes of interest. Patient severity was measured by the CCI, a commonly used measure that is calculated based on secondary diagnoses (Charlson *et al.*, 1987; Sundararajan *et al.*, 2004). Other control variables included gender (dummy variable of male or female) and principal diagnoses (dummy variable of the first three digits of ICD-10 codes).

Statistical analysis

To understand the effect of DIP reform on patient age, patient severity and treatment intensity, we used the difference-in-differences (DID) method at the discharge level. For patient i in hospital h in time t, we employed the equation:

$$Y_{iht} = \alpha Treatment + \beta Treatment \times Post + \gamma X_{iht} + H_h + \tau_t + \varepsilon_{iht} \tag{1}$$

where Y_{iht} is the outcome variable; *Treatment* is a dummy equal to one for patients in 13 DIP pilot hospitals (intervention group) and zero for patients in 27 non-DIP pilot hospitals (control group). *Post* is a dummy equal to one for discharges after 1 July 2019. The coefficient of interaction term between *Treatment* and *Post* is the interested policy effect. X_{iht} is a set of patient characteristics, including gender, age group (when age is not the outcome variable), CCI group (when CCI is not the outcome variable) and principal diagnosis (for THA/TKA patients only). H_h is a vector of hospital fixed effects and τ_t is a vector of year-month fixed effects. ε_{iht} is the error term, and standard errors were clustered at the hospital-year-month

level. That is, when analysing age and RW as the outcome variables, CCI was included as a covariate in the form of CCI group (0,1, 2, ≥3). Similarly, when analysing CCI and RW as the outcome variables, age was included as a covariate in the form of age group (<50, 50 to 60, 60 to 70, 70 to 80 and ≥80 years).

We used the generalized linear model (GLM) with normal distribution and identity link for the outcome of patient age. For severity (CCI) outcome, we used the Poisson model, since it is a non-negative discrete integer. For the RW outcome, we used GLM with normal distribution and identity link for THA/TKA patients and GLM with gamma distribution and log link for AHD patients considering different distributions and model fitness of RW in two groups. We reported coefficients in the GLM model for age and RW and incident rate ratio (IRR) in the Poisson model for CCI.

To explore the heterogeneous effects of patient selection behaviour between tertiary and secondary hospitals, we used the following equation:

$$\begin{aligned} Y_{iht} = &\alpha Treatment + \beta_1 Treatment \times Post \\ &+ \beta_2 Post \times Tertiary + \beta_3 Treatment \times Tertiary \\ &+ \beta_4 Treatment \times Post \times Tertiary + \gamma X_{iht} + H_h + \tau_t + \varepsilon_{iht} \end{aligned} \tag{2}$$

where *Tertiary* is a dummy equal to one for patients in tertiary hospitals and zero for patients in secondary hospitals, and β_4 would reflect the differential effects between tertiary and secondary hospitals.

We performed additional analyses to further explore the potential mechanism of patient selection under DIP payment guided by the theoretical framework mentioned in the Introduction section. Since the incentive of avoiding high costs is for patients within the same DIP group with the same payment, we chose two DIP groups with the most cases in THA/TKA and AHD patients, respectively, to observe the changes in age and CCI within the same DIP group.

We tested the parallel trend assumption between the intervention and control groups before the DIP reform by replacing post-dummies in the interaction term in (1) with a series of year-month dummies. We also conducted three robustness checks. First, we considered the potential confounding effect of three other policies or events during the study period and re-estimated the DID analysis. Since the inpatient volume in this city dramatically decreased before March 2020 due to the COVID-19 pandemic and was basically recovered afterwards (Chen *et al.*, 2021; Lou *et al.*, 2021), we dropped data from the period when the pandemic most seriously affected China (December 2019 to March 2020) to assess its potential confounding effects on the results. Then, we dropped data for AHD patients from January 2021 (5 months before the end of our study period) when the volume-based procurement (VBP) of coronary stent policy started. Finally, we generated an 'adjusted RW' by assigning the RW in 2021 to the same group for each year to account for a continuous price change of medical services and zero-mark-up policy of medical devices in December 2019 (General Office of the State Council of the People's Republic of China, 2019a) to balance the influence of price change and observe a net change in treatment intensity. Second, considering various principal diagnoses of patients for the procedure of THA/TKA, which

may affect the comparability between the treatment and control group, we used a subsample of single principal diagnosis with the most cases in THA/TKA patients and re-estimated the DID analysis. Third, we conducted a controlled interrupted time series (ITS) analysis as a robustness check for the DID analysis.

Results

Descriptive statistics

Characteristics of hospitalized patients of the THA/TKA and AHD cohorts in DIP pilot and non-DIP pilot hospitals before and after the DIP payment reform are shown in Table 1. Figure 1 shows the unadjusted monthly trends of the three outcomes. The final THA/TKA cohort included 12 845 discharge cases in total, with more than 85% of them in tertiary hospitals. The average age of patients was similar in DIP hospitals and non-DIP hospitals before and after DIP reform (72 years). Patient severity (via CCI) was significantly lower in DIP pilot hospitals than non-DIP hospitals both before (0.21 vs 0.27, $P < 0.001$) and after the reform (0.34 vs 0.38, $P = 0.015$), and unadjusted average CCI increased in both groups over time. The average treatment intensity (via RW) of patients was significantly higher in DIP hospitals than in non-DIP hospitals before the reform while lower after the reform. Unadjusted RW decreased in both groups during the study period. Total cost per case corresponding to THA/TKA discharges is more than 72 000 Renminbi (RMB).

The final AHD cohort included 90 346 discharge records. As shown in Table 1 and Figure 1, the average age in pilot hospitals was always significantly lower (68 vs 75 years, $P < 0.001$ before the reform; 67 vs 73 years, $P < 0.001$ after the reform). For both pilot and non-pilot hospitals, unadjusted age also decreased after the reform. CCI in pilot hospitals was also always significantly lower (0.85 vs 1.87, $P < 0.001$ before the reform; 0.87 vs 1.80, $P < 0.001$ after the reform). After the DIP reform, average CCI increased in the treatment group while decreased in the control group. The average RW, corresponding to more than 20 000 RMB of total costs per case for AHD patients, was significantly higher in pilot hospitals during the study period. After the reform, RW in both groups slightly increased.

Main results of the DID approach

Table 2 presents the DID results of the impacts of DIP reform in two selected diseases. The results of full specifications for three outcomes of the THA/TKA and AHD cohorts are in Supplementary Tables B1–B3 of Appendix B. Compared with non-DIP pilot hospitals, trends in patient age after the DIP reform were similar ($P = 0.730$) for DIP and non-DIP hospitals for THA/TKA patients, while differences in patient severity grew because severity in DIP hospitals increased more [IRR = 1.201, 95% confidence interval (CI), 1.012 to 1.426, $P = 0.036$] following DIP reform. Compared with non-DIP pilot hospitals, trends in treatment intensity (via RW) after the DIP reform were similar ($P = 0.232$) for DIP and non-DIP hospitals.

In comparison with the AHD patients of non-DIP hospitals, trends in patient age after the DIP reform were similar ($P = 0.095$) for patients in DIP hospitals. Differences in patient severity grew because severity in non-DIP hospitals dropped (IRR = 1.063, 95% CI, 1.014 to 1.114, $P = 0.011$) follow-ing DIP reform. Treatment intensity (measured via RWs) for AHD patients in DIP hospitals increased 5.5% ($P = 0.015$) more than that in non-DIP hospitals after payment reform.

Heterogeneous effects by hospital level

Supplementary Tables B1–B3 present the heterogenous effects of different hospital levels in Panel B. Age trends for THA/TKA patients were similar for tertiary and secondary hospitals ($P = 0.898$). The incident rate of CCI was significantly higher in tertiary hospitals than secondary hospitals (IRR = 2.367; 95% CI, 1.624 to 3.448; $P < 0.001$). Finally, treatment intensity (via RW) decreased significantly in tertiary hospitals by 168.851 (95% CI, –314.612 to –23.089; $P = 0.023$) compared with secondary hospitals.

In AHD patients, age significantly increased by 3.73 years (95% CI, 1.25 to 6.20; $P = 0.003$) in tertiary hospitals compared with secondary hospitals, as did severity (IRR = 1.150; 95% CI, 1.046 to 1.264; $P = 0.004$) and treatment intensity (14.1%, 95% CI, 5.0% to 24.0%; $P = 0.002$).

Additional analyses

The distribution of DIP groups in two diseases is presented in Supplementary Table C1 of Appendix C. We also plotted bar charts for total costs per case by age group and CCI group for selected DIP groups as shown in Supplementary Figure C1. Total cost per case was indeed generally higher in groups with higher age and CCI, but age and CCI within the same group did not decrease in pilot hospitals after DIP reform (Supplementary Tables C2 and C3).

Validity test and robustness checks

A key premise of the DID approach is that the DIP and non-DIP outcomes exhibit parallel trends before the reform. Otherwise, the validity of DID analysis would be doubtful. We observed similar pre-reform trends for age and CCI in AHD patients between DIP and non-DIP hospitals (Supplementary Figure D1, Appendix D). The parallel trend assumption for RW in AHD patients and all three outcomes in THA/TKA patients were not satisfied. Results for the sensitivity analyses that considered COVID-19 were consistent with the main results (Supplementary Table E1). In the RW analysis for AHD patients when accounting for the VBP policy, a consistent positive coefficient was found, but the result became insignificant (Supplementary Table E2). In the analysis of adjusted RW, we observed an opposite change in direction in THA/TKA patients and consistent results for AHD patients (Supplementary Table F1, Appendix F). In the subsample analysis of the single principal diagnosis of THA/TKA patients with the most cases (fracture of femur, accounting for more than 40% of total THA/TKA cases), the direction and significance of coefficients for all outcome variables were consistent with the whole sample of THA/TKA group, except that the increase in CCI changed from significant to insignificant (Supplementary Table G1, Appendix G). The parallel assumption became satisfied for age, RW and adjusted RW (Supplementary Figure G1). The results of the sensitivity analysis of the controlled ITS were generally consistent with the main results (Supplementary Tables H1, Appendix H).

Table 1. Sample characteristics of hospitalized patients with THA/TKA or AHD before and after the DIP reform

	Pre-DIP reform (July 2017–June 2019)				Post-DIP reform (July 2019–June 2021)			
	DIP pilot hospital ($n=13$)	Non-DIP pilot hospital ($n=27$)	t/χ^2	P value	DIP pilot hospitals ($n=13$)	Non-DIP pilot hospitals ($n=27$)	t/χ^2	P value
Panel A: THA/TKA								
Patient characteristics								
Number of discharges	2 782	3 357			2 901	3 805		
Age/years, mean (SD)	72.02 (11.72)	72.16 (11.19)	0.460	0.646	71.92 (10.86)	71.76 (12.64)	−0.526	0.599
Male sex, No. (%)	746 (26.82)	958 (28.54)	2.250	0.134	772 (26.61)	1052 (27.65)	0.893	0.345
CCI, mean (SD)	0.21 (0.59)	0.27 (0.58)	3.728	<0.001	0.34 (0.71)	0.38 (0.72)	2.436	0.015
Hospital level, No. (%)			0.685	0.408			0.045	0.831
Tertiary ($N=10$)	2393 (86.02)	2912 (86.74)			2474 (85.28)	3252 (85.47)		
Secondary ($N=30$)	389 (13.98)	445 (13.26)			427 (14.72)	553 (14.53)		
Intensity of care								
Total costs per case/RMB, mean (SD)	72 393.54 (19 651.39)	72 044.57 (19 663.29)	−0.692	0.489	77 647.76 (20 071.83)	76 667.22 (22 605.61)	−1.846	0.065
RW per case, mean (SD)	4150.87 (651.36)	4057.98 (708.53)	−4.659	<0.001	3837.92 (618.36)	3866.27 (782.75)	1.604	0.109
Panel B: AHD								
Patient characteristics								
Number of discharges	21 678	29 780			16 000	22 888		
Age/years, mean (SD)	68.15 (11.13)	74.84 (11.90)	64.603	<0.001	66.58 (13.95)	72.52 (13.69)	41.760	<0.001
Male sex, No. (%)	12 211 (56.33)	15 037 (50.49)	171.477	<0.001	9330 (58.31)	12 395 (54.16)	66.016	<0.001
CCI, mean (SD)	0.85 (1.14)	1.87 (1.48)	83.988	<0.001	0.87 (1.13)	1.80 (1.54)	64.747	<0.001
Hospital level, No. (%)			3.8×10^3	<0.001			3.6×10^3	<0.001
Tertiary ($N=10$)	16 172 (74.60)	14 144 (47.49)			12 347 (77.17)	10 725 (46.86)		
Secondary ($N=30$)	5506 (25.40)	15 636 (52.51)			3653 (22.83)	12 163 (53.14)		
Intensity of care								
Total costs per case/RMB, mean (SD)	20 833.53 (25 973.66)	20 820.92 (20 871.86)	−0.061	0.951	21 922.52 (25 237.37)	22 799.74 (23 271.99)	3.532	<0.001
RW per case, mean (SD)	1130.94 (1053.87)	1011.56 (905.86)	−11.672	<0.001	1166.88 (1066.86)	1025.43 (922.91)	−13.918	<0.001

Note: *N* is the number of hospitals.
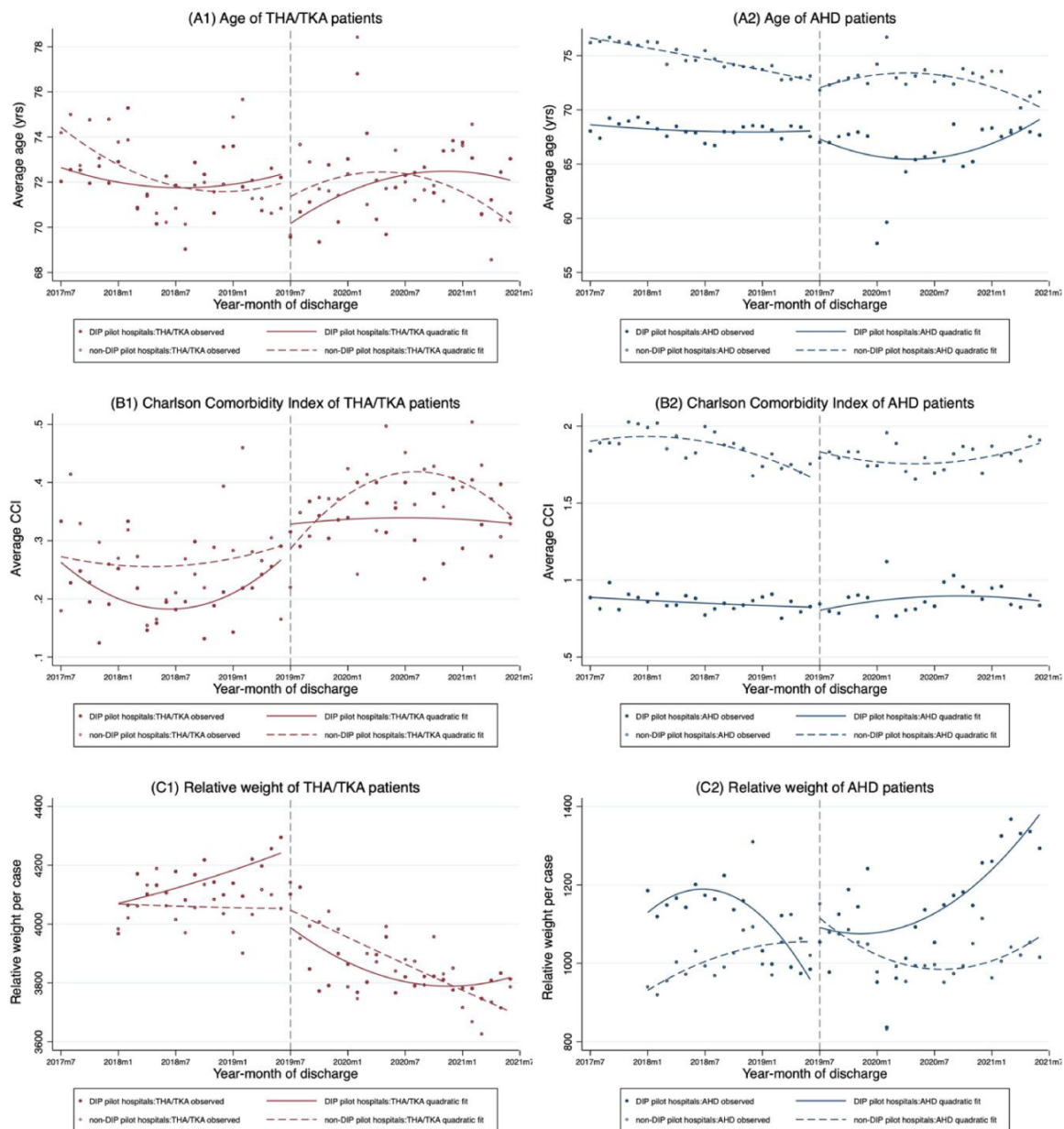
Figure 1. Trends of average age, severity and RW of hospitalized patients with THA/TKA or AHD from July 2017 to June 2021

Notes: Trends of the outcomes (A for age, B for CCI and C for RW) for THA/TKA and AHD in DIP pilot and non-DIP pilot hospitals are plotted. The vertical dashed line denotes the implementation of the DIP payment reform on 1 July 2019.

Discussion

Following the implementation of the DIP payment in China, changes in patient age and treatment intensity in DIP hospitals and non-DIP hospitals were similar for THA/TKA patients, while patient severity increased in DIP hospitals, which did support our hypotheses. The treatment intensity (as measured by RWs) in AHD patients significantly increased, which makes sense given that the patient severity increased (consistent with our hypothesis).

The similar age trend and relative increase in severity for THA/TKA and AHD patients suggest that there was no evidence of cream-skimming after the DIP reform. These findings are consistent with a previous study of DIP payment reform in Guangzhou, China (Lai *et al.*, 2022) and several studies on

bundle payments in America for patients with cardiac procedures (McNeely *et al.*, 2020) or joint replacement (Barnett *et al.*, 2019; Plate *et al.*, 2019; Ryan *et al.*, 2019), which did not observe differential changes in patient characteristics. Moreover, results of additional analyses demonstrate that although total cost per case was higher for patients with higher age and severity within the same DIP group, providers did not select younger or less severe patients. Since the increase in CCI within the same group was not significant in the whole sample, it could be inferred that the selection of sicker patients mainly occurred among DIP groups for higher payments, which is consistent with our theoretical framework. Along with the significant increase in (adjusted) RW, our study suggests that the granular classification system in DIP payment has indeed generated strong provider incentives to

Table 2. Impacts of the DIP payment reform on patient age, CCI and RW: DID analysis

	THA/TKA patients (*n* = 12 845)			AHD patients (*n* = 90 346)	
Panel A: outcome = age	**Coefficient (95% CI)**	**P value**		**Coefficient (95% CI)**	**P value**
Treatment (DIP pilot hospitals = 1)	0.901 (−1.379, 3.181)	0.439		−80.213 (−82.149, −78.277)	<0.001
Treatment × Post (pilot hospitals after DIP reform = 1)	0.178 (−0.833, 1.190)	0.730		0.759 (−0.132, 1.650)	0.095
Patients controls	Y			Y	
Hospital fixed effect	Y			Y	
Year-month fixed effect	Y			Y	
Panel B: outcome = CCI	**IRR (95% CI)**	**P value**		**IRR (95% CI)**	**P value**
Treatment (DIP pilot hospitals = 1)	0.362 (0.234, 0.56)	<0.001		0.347 (0.309, 0.391)	<0.001
Treatment × Post (pilot hospitals after DIP reform = 1)	1.201 (1.012, 1.426)	0.036		1.063 (1.014, 1.114)	0.011
Patients controls	Y			Y	
Hospital fixed effect	Y			Y	
Year-month fixed effect	Y			Y	
Panel C: outcome = RW	**Coefficient (95% CI)**	**P value**		**Coefficient (95% CI)**	**P value**
Treatment (DIP pilot hospitals = 1)	292.755 (−6.206, 591.716)	0.055		0.075 (−0.011, 0.161)	0.087
Treatment × Post (pilot hospitals after DIP reform = 1)	−34.181 (−90.220, 21.858)	0.232		0.054 (0.011, 0.097)	0.015
Patient controls	Y			Y	
Hospital fixed effect	Y			Y	
Year-month fixed effect	Y			Y	

Notes: Impacts on age were estimated via the general linear model with normal distribution and identity link, adjusting for gender, CCI group and principal diagnosis (for THA/TKA patients only) at the discharge level. Impacts on CCI were estimated via the Poisson model, adjusting for gender, age group and principal diagnosis (for THA/TKA patients only) at the discharge level. Impacts on RW were estimated via the general linear model with normal distribution and identity link for THA/TKA patients and general linear model with gamma and log link for AHD patients, adjusting for gender, age group, CCI group and principal diagnosis (for THA/TKA patients only) at the discharge level. All specifications also controlled for hospital fixed effect and year-month fixed effect, with standard error clustered at the hospital-year-month level (see Supplementary Appendix Tables B1–B3 for full specification). Y, which is 'yes' for short, indicated that the listed covariates were included in the model. The sample size of RW was 11 490 for THA/TKA patients, and 75 654 for AHD patients, due to the lack of data in 2017.

admit sicker patients and provide higher treatment intensity, instead of working hard on cost control by selecting profitable patients within the same group, which has been typically observed in evaluation studies of DRG payment (Jian *et al.*, 2015; Wu *et al.*, 2022). Our result is consistent with other DIP payment reform evaluations, including providing more intensive medical investments for sicker patients (Qian *et al.*, 2021), an increase in the overall treatment intensity (Lai *et al.*, 2022) and an increase in total costs per case in older patients (Chen *et al.*, 2023). It should be noted that the increase in RW may also be partially contributed by an increase in treatment intensity for a given patient ('upcoding' behaviour as a treatment policy). It is also worth noting that the increase in intensity of care would not necessarily increase inpatient expenditures, since the global budget provides a fixed cap either at the regional level or at the hospital level. However, whether the increase in treatment intensity is a beneficial practice clinically (even for severe patients) and whether this behaviour would affect the quality of care are still noteworthy.

In the heterogeneous effect analysis, we found a greater increase in tertiary hospitals for age and RW in AHD patients and CCI in both diseases. This was in accordance with our expectation, where tertiary hospitals may have more capability to provide higher treatment intensity for severe patients. This result was inconsistent with another study that showed that average RW increased more in non-tertiary hospitals after DIP reform in another city (Lai *et al.*, 2022). That study included both public and private hospitals, and non-tertiary

hospitals particularly contained many private hospitals, which may confound the results. This confounding is potentially significant since a recent study suggested that public and private hospitals had different responses after the DIP reform (Tang *et al.*, 2023). In our study, the larger response in tertiary hospitals was more obvious for AHD patients than for THA/TKA patients, possibly because THA/TKA was already a procedure with high intensity of care, and most procedures were concentrated in tertiary hospitals (as shown in our descriptive analysis). Thus, there were fewer opportunities for tertiary hospitals to further increase the treatment intensity for THA/TKA.

To our knowledge, this is the first empirical study to explore the impacts of DIP payments under the global budget on patient selection behaviour using a comparable sample of specific diseases. When implementing PPSs like DIP, policymakers may need to balance between cost containment and unexpected outcomes, such as the damage of health benefits of disadvantaged patients or the deterioration of quality.

Our study has several limitations. First, the results of the parallel assumption test for several outcomes for the THA/TKA cohort were not that satisfying, so these results must be interpreted carefully. This assumption was satisfied for the AHD cohort. However, after using a subsample of single principal diagnosis with the most cases of THA/TKA patients in the robustness check, the model results were consistent and the parallel assumption was satisfied except for

CCI. Thus, although the variety of principal diagnoses within the THA/TKA group may mainly affect the comparability between pilot and non-pilot hospitals, the results of the main models were basically robust. Second, we selected two diseases in our study, and responses observed here may not generalize to other diseases or procedures. Third, the additional analyses of the potential patient selection mechanism were exploratory rather than conclusive, which could be a direction for future research. Fourth, although inpatient services basically recovered after March 2020 in the study city (Chen *et al.*, 2021; Lou *et al.*, 2021) and the robustness check considering COVID-19 was consistent with the main results, the longer-term effects of DIP beyond June 2021 may be different. With hospitals completely having recovered from COVID-19 and having more ability to adjust hospital management strategies in response to DIP payments, and providers having learned more about the new payment system and its incentives, they may have stronger responses to the reform. Further research should conduct in-depth interviews with hospital administrators and physicians to understand their intended long-term strategy to address DIP. Long-term monitoring of unexpected results is also necessary.

Supplementary data

Supplementary data is available at *Health Policy and Planning* Journal online.

Data availability

The data that support the findings of this study are available from Healthcare Security Administration of the study city. However, restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. Data are, however, available from the corresponding author upon reasonable request and with permission of Healthcare Security Administration of this city.

Funding

Author contributions

X.Z., S.T., M.Q., X.Y. and M.L.M contributed to conception or design of the work. X.Z. and X.Y. contributed to data collection. X.Z., R.W., M.Q. and M.L.M contributed to data analysis and interpretation. X.Z. contributed to drafting the article. S.T., X.Y. and M.L.M contributed to critical revision of the article. X.Z., S.T., R.W., M.Q., X.Y. and M.L.M contributed to final approval of the version to be submitted.

Reflexivity statement

The authors represent a range of gender identities, locations and seniority. The authors are equally represented in terms of gender (three males and three females). Authors are based in China and the USA, representing upper-middle-income countries and high-income countries, respectively. A majority, four out of the six authors, are senior academics, but the first author is a postdoctoral associate, and the remaining co-author is a PhD student who is in their early career.

Ethical approval. This study received ethical approval from the institutional review board of the School of Public Health, Fudan University (IRB#2020-TYSQ-03-20).

Conflict of interest. None received from all the authors.

References

Aengevaeren VL, Mosterd A, Sharma S *et al*. 2020. Exercise and coronary atherosclerosis: observations, explanations, relevance, and clinical management. *Circulation* **141**: 1338–50.

Amanatullah DF, McQuillan T, Kamal RN. 2019. Quality measures in total hip and total knee arthroplasty. *Journal of the American Academy of Orthopaedic Surgeons* **27**: 219–26.

Barnett ML, Wilcock A, McWilliams JM *et al*. 2019. Two-year evaluation of mandatory bundled payments for joint replacement. *New England Journal of Medicine* **380**: 252–62.

Bäuml M, Kümpel C. 2021. Hospital responses to the refinement of reimbursements by treatment intensity in DRG systems. *Health Economics* **30**: 585–602.

Charlson ME, Pompei P, Ales KL, MacKenzie CR. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* **40**: 373–83.

Chen Y, Cai M, Li Z, Lin X, Wang L. 2021. Impacts of the COVID-19 pandemic on public hospitals of different levels: six-month evidence from Shanghai, China. *Risk Management and Healthcare Policy* **14**: 3635–51.

Chen YJ, Zhang XY, Tang X *et al*. 2023. How do inpatients' costs, length of stay, and quality of care vary across age groups after a new case-based payment reform in China? An interrupted time series analysis. *BMC Health Services Research* **23**: 160.

CPC Central Committee and the State Council. 2020. *Opinions on Deepening the Reform of Health Security System*. http://www.nhsa.gov.cn/art/2020/3/5/art_37_2808.html, accessed 19 November 2020.

Ellis RP. 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics* **17**: 537–55.

Ellis RP, McGuire TG. 1996. Hospital response to prospective payment: moral hazard, selection, and practice-style effects. *Journal of Health Economics* **15**: 257–77.

Fu R, Shen Y, Noguchi H. 2021. The best of both worlds? The economic effects of a hybrid fee-for-service and prospective payment reimbursement system. *Health Economics* **30**: 505–24.

General Office of the National Health Commission of China. 2020. *Notification on Further Strengthening the Quality Management and Control of Single Disease*. http://www.nhc.gov.cn/yzygj/s7657/202007/b31755433b8e4a50b23906b9f1a07393.shtml, accessed 24 October 2020.

General Office of the State Council of the People's Republic of China. 2019a. *Notification on Reform Plan for the Governance of High-cost Medical Devices*. https://www.gov.cn/zhengce/content/2019-07/31/content_5417518.htm, accessed 22 April 2021.

General Office of the State Council of the People's Republic of China. 2019b. *Opinions on Strengthening the Performance Evaluation of*

Tertiary Public Hospitals. http://www.gov.cn/zhengce/content/2019-01/30/content_5362266.htm, accessed 30 November 2020.

Hafsteinsdottir EJ, Siciliani L. 2010. DRG prospective payment systems: refine or not refine? *Health Economics* **19**: 1226–39.

He AJ. 2023. Scaling-up through piloting: dual-track provider payment reforms in China's health system. *Health Policy and Plan* **38**: 218–27.

Jian W, Lu M, Chan KY *et al.* 2015. Payment reform pilot in Beijing Hospitals reduced expenditures and out-of-pocket payments per admission. *Health Affairs* **34**: 1745–52.

Kantarevic J, Kralj B. 2014. Risk selection and cost shifting in a prospective physician payment system: evidence from Ontario. *Health Policy* **115**: 249–57.

Kifmann M, Siciliani L. 2017. Average-cost pricing and dynamic selection incentives in the hospital sector. *Health Economics* **26**: 1566–82.

Ko H, Martin BI, Nelson RE, Pelt CE. 2022. Patient selection in the comprehensive care for joint replacement model. *Health Services Research* **57**: 72–90.

Lai Y, Fu H, Li L, Yip W. 2022. Hospital response to a case-based payment scheme under regional global budget: the case of Guangzhou in China. *Social Science & Medicine* **292**: 114601.

Long H, Zeng C, Shi Y *et al.* 2023. Length of stay and inpatient charges of total knee arthroplasty in China: analysis of a national database. *Chinese Medical Journal (Engl)* **136**: 2050–7.

Lou TF, Ren Z, Sun ZH, Wang W, Fan CY. 2021. Full recovery of elective orthopedic surgery in the age of COVID-19: an 8-month retrospective cohort study. *Journal of Orthopaedic Surgery and Research* **16**: 154.

Martinussen PE, Hagen TP. 2009. Reimbursement systems, organisational forms and patient selection: evidence from day surgery in Norway. *Health Economics, Policy and Law* **4**: 139–58.

McNeely C, Orav EJ, Zheng J, Joynt Maddox KE. 2020. Impact of medicare's bundled payments initiative on patient selection, payments, and outcomes for percutaneous coronary intervention and coronary artery bypass grafting. *Circulation Cardiovascular Quality and Outcomes* **13**: e006171.

National Bureau of Statistics of China. 2023. *China Statistical Yearbook (2023)*. Beijing: China Statistics Press.

National Healthcare Security Administration. 2021. *Three-Year Action Plan of DRG/DIP Payment System Reform.* http://www.nhsa.gov.cn/art/2021/11/26/art_37_7406.html, accessed 30 November 2021.

Paddock SM, Escarce JJ, Hayden O, Buntin MB. 2007. Did the Medicare inpatient rehabilitation facility prospective payment system result in changes in relative patient severity and relative resource use? *Medical Care* **45**: 123–30.

Plate JF, Ryan SP, Black CS *et al.* 2019. No changes in patient selection and value-based metrics for total hip arthroplasty after comprehensive care for joint replacement bundle implementation at a single center. *The Journal of Arthroplasty* **34**: 1581–4.

Qian M, Zhang X, Chen Y, Xu S, Ying X. 2021. The pilot of a new patient classification-based payment system in China: the impact on costs, length of stay and quality. *Social Science & Medicine* **289**: 114415.

Ryan SP, Plate JF, Black CS *et al.* 2019. Value-based care has not resulted in biased patient selection: analysis of a single center's experience in the care for joint replacement bundle. *The Journal of Arthroplasty* **34**: 1872–5.

Sood N, Huckfeldt PJ, Grabowski DC, Newhouse JP, Escarce JJ. 2013. The effect of prospective payment on admission and treatment policy: evidence from inpatient rehabilitation facilities. *Journal of Health Economics* **32**: 965–79.

Sundararajan V, Henderson T, Perry C *et al.* 2004. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *Journal of Clinical Epidemiology* **57**: 1288–94.

Takahara T. 2016. Patient dumping, outlier payments, and optimal healthcare payment policy under asymmetric information. *Health Economics Review* **6**: 57.

Tang X, Zhang X, Chen Y *et al.* 2023. Variations in the impact of the new case-based payment reform on medical costs, length of stay, and quality across different hospitals in China: an interrupted time series analysis. *BMC Health Services Research* **23**: 568.

Vallance AE, Fearnhead NS, Kuryba A *et al.* 2018. Effect of public reporting of surgeons' outcomes on patient selection, "gaming," and mortality in colorectal cancer surgery in England: population based cohort study. *BMJ* **361**: k1581.

World Health Organization, Regional Office for the Western Pacific. 2015. *Case-Based Payment Systems for Hospital Funding in Asia: An Investigation of Current Status and Future Directions.* Manila: WHO Regional Office for the Western Pacific.

Wu J, He X, Feng XL. 2022. Can case-based payment contain healthcare costs? A curious case from China. *Social Science & Medicine* **312**: 115384.