RESEARCH | Article in Press

Payment systems and physician altruism under DRG reform: evidence from a multicenter artefactual field experiment in China

Chao Zheng · Shuyan Lin · Yuting Zhang · Jia Yin · Qiang Sun

Received: 28 June 2026 / Accepted: 20 October 2026
© The Author(s) 2026

Abstract

Background Payment mechanism reform is central to global efforts to balance cost containment and quality of care. We compared service choices under simulated Diagnosis-Related Group (DRG) and Fee-for-Service (FFS) payment across predefined severity levels and estimated the relative preference for experimental patient benefit and physician altruism.

Methods We conducted a multicenter artefactual field experiment with 252 practicing physicians from six tertiary public hospitals in three provinces of China. We summarized service quantity, patient benefit loss ratio (PBLR), and experimental profit, estimated preferences using mixed logit, and integrated 23 key-informant interviews through a joint display.

Results FFS was associated with a significantly higher volume of services overall than DRG (mean 5.88 vs. 4.88; $p<0.001$). Analysis by severity revealed a crossover effect, FFS incentivized overprovision for mild and moderate cases but yielded the model-based definition of optimality provision for severe cases. DRG reduced service volume but resulted in significant underprovision for severe cases ($p<0.001$). Severe-case PBLR was higher under DRG: mean 0.12 versus 0.10 and median 0.05 versus 0.04 ($p=0.040$), reflecting underprovision risk. Physician altruism, quantified using a mixed logit model, yielded a marginal rate of substitution of 5.499 (95% CI 4.532–6.466), in experimental profit tokens per patient-benefit token, indicating physicians forwent 5.499 units of experimental profit to generate one additional unit of patient benefit. Interviews confirmed that DRG-linked performance balances promoted under-provision and risk avoidance.

Conclusions Neither FFS nor DRG is universally superior. The experiment suggests severity-dependent trade-offs between service volume and model-defined patient benefit. Subject to confirmation in real-world implementation studies, these findings support evaluating severity-sensitive blended payment models and safeguards for complex cases.

Keywords Health systems · Provider payment · Physician behaviour · Altruism · Experimental methods

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



https://doi.org/10.1186/s12913-026-15933-w

Published online: 04 November 2026

Background

At the end of 2025, China has achieved near-universal health insurance coverage, with over 1.3 billion enrollees and a coverage rate exceeding 95% [1]. Payment systems constitute a core institutional mechanism through which health systems seek to reconcile cost containment with professional and ethical commitments to patient care. As prospective payment models such as Diagnosis-Related Groups (DRG) continue to expand worldwide, concerns persist regarding how financial incentives reshape physicians' clinical decision-making, particularly clinical uncertainty is high. Understanding these behavioral responses is essential for evaluating not only the efficiency but also the social consequences of payment reform. Many health systems have relied on Fee-for-Service (FFS) as the primary model for provider reimbursement. While straightforward, FFS is consistently associated with strong incentives for service overprovision and supplier-induced demand, posing a significant threat to financial sustainability [2–4].

Payment reform changes the rewards attached to clinical and organizational decisions, but its consequences depend on the payment design and implementation context. Extensive evidence from both high- and middle-income countries corroborates that FFS incentivizes overprovision, whereas DRG payment can reduce unnecessary service use, shorten length of stay, and increase cost-awareness. However, DRG may also induce unintended behaviors such as underprovision, upcoding, risk selection [5–9]. For example, Klein et al. found that the post-reform period was associated with a substantial reduction in both length of hospital stays and medical expenditures, alongside notable improvements in the quality of care [10]. Meanwhile, a study analyzing data from acute myocardial infarction patients demonstrated that after implementing DRG and per-diem reimbursement, both medical costs and hospital stay decreased while mortality rates remained unchanged; however, readmission rates increased [8]. These studies motivate examining both service reduction and potential trade-offs, rather than assuming that reduced volume necessarily improves care. However, empirical studies in China remain limited, especially those capturing real-world behavioural responses of frontline physicians to DRG reform.

The influence of payment systems on physician behaviour is a central theme in health economics, with experimental research providing insights into the incentive mechanisms [11–13]. Studies using FFS and capitation have documented different service-provision responses and have compared participant groups within controlled tasks [14, 15]. Patient benefit and professional norms can influence those choices alongside personal earnings [16, 17]. This interaction between payment structure and professional ethics is an area of complex, unresolved inquiry across health systems [18–21]. However, much of the existing evidence focuses on average treatment intensity or aggregate outcomes, paying limited attention to how payment systems interact with physicians' intrinsic motivation and professional norms across different levels of clinical severity. China provides a setting for examining how prospective payment incentives interact with professional preferences. We conducted an artefactual field experiment with practicing physicians from six tertiary public hospitals and complemented it with interviews with clinicians and administrative staff. Field-experimental evidence must be interpreted in relation to its participant pool and decision environment [22]. Practicing physicians improve the occupational relevance of this task, while the six-hospital sample and hypothetical treatment decisions limit population and clinical generalizability. Throughout, the optimal quantity means the maximum of the experimentally specified benefit schedule.

We address three connected research questions. RQ1 asks how simulated FFS and DRG payment change service volume and model-defined patient benefit across severity levels. RQ2 asks what relative preference for experimental patient benefit versus physician profit is implied by these choices, and how that preference varies across participants. RQ3 asks how clinicians and administrators describe the organizational incentives and governance arrangements that may shape responses to payment reform. RQ1 identifies the behavioral pattern, RQ2 examines a preference-based interpretation, and RQ3 supplies contextual explanations.

This study makes three contributions. First, we provide evidence from a multicenter artefactual field experiment isolating the behavioral effects of DRG versus FFS across clinical severity. Second, we structurally estimate

physician altruism by quantifying the incentive–benefit trade-off. Third, integrating experimental evidence with qualitative interviews, we elucidate the mechanisms through which payment systems shape clinical behavior.

Theoretical framework: physician behaviour under payment reform

We embed our artefactual field experiment within a coherent economic–behavioural framework that explains how provider payment systems interact with clinical severity to shape physicians' clinical choices. This framework synthesizes three core components: principal–agent theory with a dual-utility specification, prospect theory notions of reference points under distinct budget constraints, and motivation crowding by professional norms. Reference dependence and motivation crowding offer additional interpretations of interview accounts; they are not separately estimated preference parameters in this experiment. This integrated model generates comparative predictions for FFS versus DRG payments along a severity gradient and rationalizes our empirical patterns, FFS-induced over-provision for mild to moderate cases with the model-based definition of optimality care for severe cases, and DRG's overall cost-dampening with undertreatment risk at high severity. Qualitative interview evidence further grounds the mechanisms.

We model practicing physicians as dual-motive agents choosing a service intensity $q \in \{0,\ldots,10\}$ for a case with severity s ∈{mild, moderate, severe} under a payment mode m ∈{FFS, DRG}. For physician i, severity s, and payment mode m, deterministic utility is $U_i(q; s,m) = \alpha_i B_s(q) + \beta_i \pi_{s,m}(q)$. Its maximizing quantity is obtained by comparing all 11 alternatives. For $\beta_i > 0$, α_i / β_i is the positive magnitude of the utility trade-off in profit tokens per benefit token; along an indifference contour, $d\pi/dB = -\alpha_i / \beta_i$. The benefit-maximizing quantities q*=3, 5, and 7 are task benchmarks for mild, moderate, and severe vignettes. The schedules are single-peaked but not globally concave, so a continuous first-order condition is not the exact decision rule. Detailed model derivations, identification assumptions and Testable implications are provided in Supplementary File 1, methods S1.

The five predictions organize the analysis: P1 concerns FFS mean quantity above q^* for mild and moderate vignettes; P2 concerns proximity of the severe-case FFS mean to $q^*=7$; P3 concerns a severe-case DRG mean below $q^*=7$; P4 concerns preference heterogeneity in payment responses; and P5 concerns organizational transmission of payment pressure. They are conditional empirical propositions, not universal consequences of the utility model. Figure 1 identifies the evidence and remaining tests for each proposition.

The Fig. 2 shows how payment mode determines price/budget constraints, which shape marginal incentives and hence the deviation of provided care q from the clinical optimum $q^*(s)$. Outcomes feed back into organizational incentives, which then update behaviour over time. Moderators—severity, reference dependence, and motivation crowding—alter the transmission at specific links.

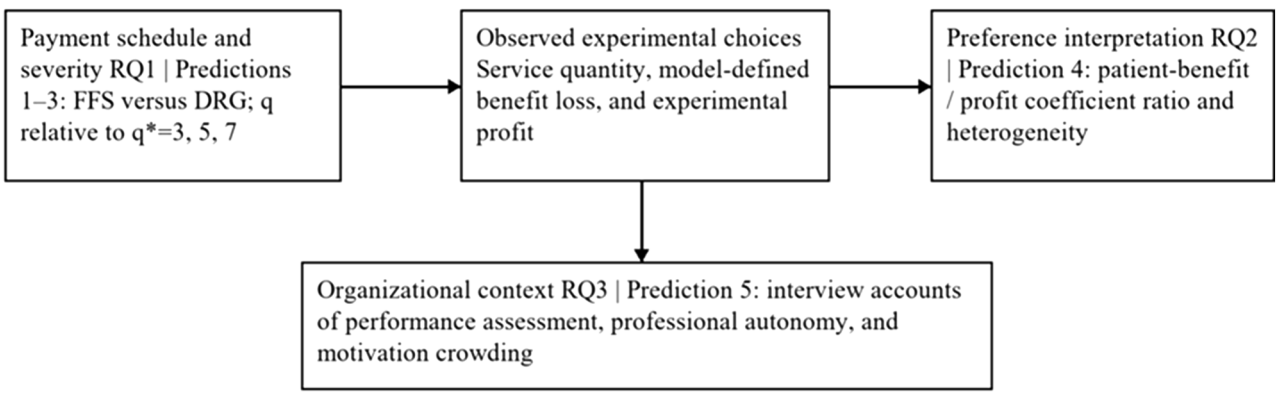


Fig. 1 Joint interpretation

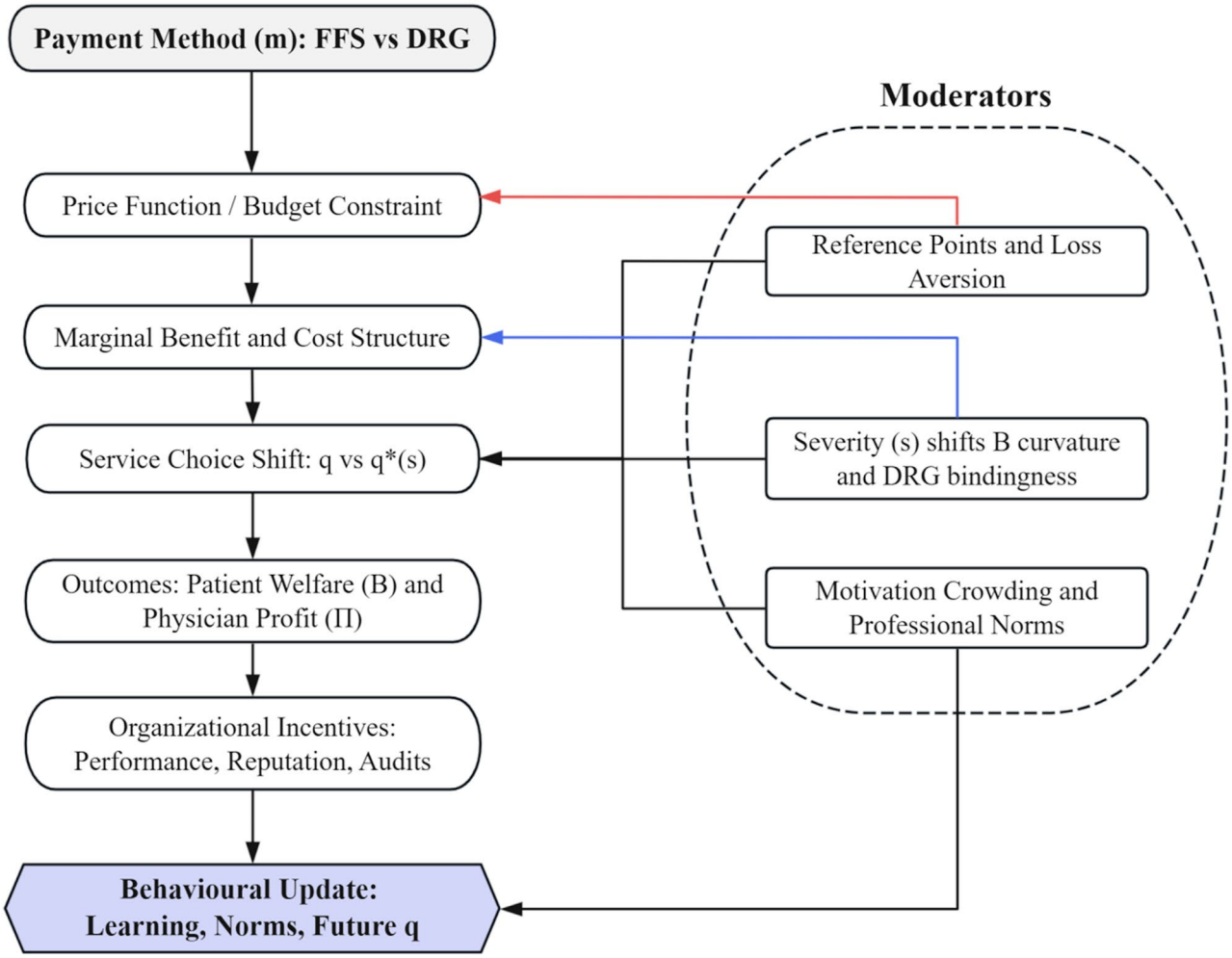


Fig. 2 Theoretical pathway from payment method to behavioural updates

Methods

Study design and procedures

Experimental design

We used an artefactual field experiment in which practicing physicians selected service quantities for hypothetical patients under alternative payment schedules. The task abstracts the treatment decision to a single decision-maker with a finite set of quantities.

Experimental framework

To create the decision task, we generated 15 unique clinical scenarios by crossing five abstract disease archetypes (k = A-E) with three levels of patient severity (s=*mild*, *moderate*, *severe*), guided by literatures in experimental economics [14, 15, 23, 24]. For each scenario, they were asked to choose a level of service intensity (q) on a discrete scale from 0 to 10. This task was performed under two distinct, counter-balanced payment simulations: FFS and DRG, resulting in 30 unique decision points per physician.

Model parameterisation, analytical approach, data quality control

The experiment was calibrated to reflect real-world financial incentives and clinical trade-offs. The experiment assigns financial consequences directly to the physician decision-maker; actual DRG payment typically reaches the hospital and is transmitted through organizational arrangements. Under FFS, physician revenue increased with service intensity (q); under DRG, revenue was fixed by disease and severity. In both models, cost was specified as a convex function. Parameters for Cost $C_{sk}(q) = 0.1*q^2$, $q \in [0,10]$, reflecting increasing marginal effort and resource use. Cost parameters were independent of payment model, disease type, and severity. FFS parameters were adapted from validated international models [14] and aligned with Chinese fee schedules, while DRG parameters were based on Zhang et al. [21]. DRG payments were set at 110% of the FFS reimbursement at the clinically optimal service level, $q^*(s)$, creating incentives to manage service volume. Parameters were specified for all 30 scenarios. Parameters for Physician Net experimental earning $\pi_{sk}(q)$, net earnings were defined as reimbursement minus cost. Under FFS, reimbursement varied by disease but not severity; under DRG, it varied by both disease and severity. Parameters for Patient Benefit $B_{jk}(q)$, Patient benefit was modeled independently of payment as a concave, non-monetary function of service intensity. It peaked at $q^*(s) = 3,5$, and 7 for mild, moderate, and severe cases, respectively, enabling measurement of over- and under-provision. Analytical Approach, Data Quality control was presented in Supplementary File 1, Methods S2 and Table S1.

Study setting

The six study hospitals were comprehensive tertiary public hospitals selected from Shandong in eastern China, Heilongjiang in northeastern China, and Ningxia Hui Autonomous Region in western China to include different geographic and economic settings. This was purposive site selection. Participants were required to have at least three years of clinical experience and direct professional experience under both FFS and DRG payment. Such eligibility indicates some prior exposure, but does not quantify its duration, intensity, or local payment version. We recruited participants for two complementary study components: a quantitative artefactual field experiment and a qualitative key informant interview series (Quantitative & Qualitative questionnaires were presented in Supplementary File 2).

To ensure verisimilitude and incentivize realistic decision-making, we implemented a dual-channel, real-stakes reward system. First, physicians received a performance-based monetary payment corresponding to the total net profit they generated across all scenarios (conversion rate: 10 experimental tokens = 1 *CNY*). Second, to explicitly activate and measure professional altruism, a donation was made in each physician's name to a real-world patient charity fund, with the amount determined by the total patient benefit they generated. Physicians received documented proof of their respective donations to ensure transparency and trust. The arrangement may therefore capture concern for the beneficiary alongside image or warm-glow motives; their separate contributions are not identified.

The target was 270 physicians, informed by prior exploratory experimental studies. The source description reports 270 participants before screening and 252 retained submissions. Eighteen were excluded for incompleteness or failure of a severity-monotonicity criterion. We therefore describe 93.3% as the fraction retained for analysis, rather than a recruitment-response rate. The pre-screening count is shown in Fig. 3. The questionnaire instructed physicians to decide individually, avoid discussion, and request assistance if needed. Each payment block included five comprehension questions about the physician role, diseases, severity, earnings, and the beneficiary of patient-benefit donations. The stated protocol required correct answers before proceeding, but the exact passing rule, correction/retry procedure, and associated counts are not documented. Payments from both blocks were scheduled to be settled together at the end. This timing does not rule out learning, fatigue, anticipated-wealth effects, or carryover.

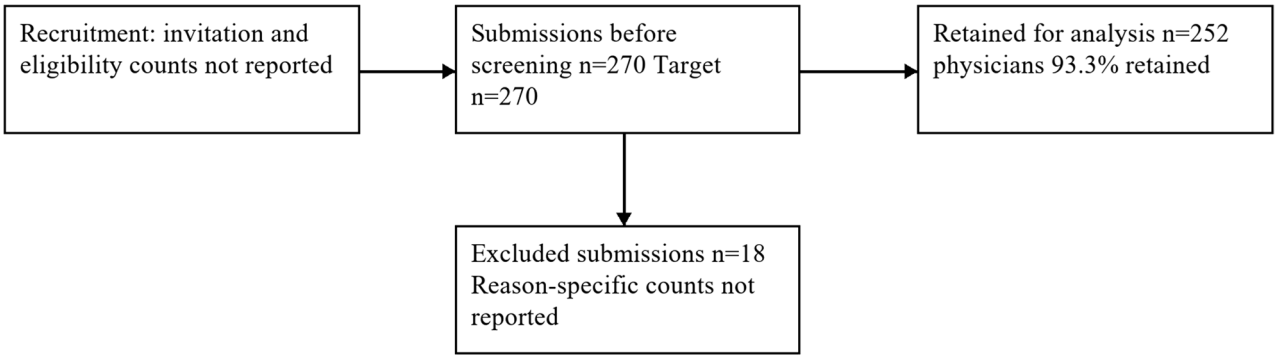


Fig. 3 Screening count

Statistical analysis

The primary outcome for this study was the volume of medical services (an integer, q, from 0 to 10) selected by the physician for each clinical vignette. This served as our direct measure of physician clinical behaviour. These decisions are repeated observations from the same physicians, not independent participants. Overprovision and underprovision refer to q above or below the corresponding experimental benchmark q^*.

Secondary outcomes included the algorithmically determined patient benefit and physician net profit associated with each decision. We also derived a measure of physician altruism, operationalized as the marginal rate of substitution (MRS) between patient benefit and physician profit. This MRS value quantifies the amount of personal profit a physician is willing to forgo to generate one additional unit of patient benefit. All outcome variables were treated as continuous data. To evaluate the impact of physicians' clinical decisions on patient welfare, we introduced the Patient Benefit Loss Ratio (PBLR). This metric is calculated using the formula: PBLR= *(Optimal Patient Benefit−Actual Patient Benefit)/Optimal Patient Benefit.* As the PBLR data exhibited a skewed distribution, non-parametric tests were utilized for the analysis.

Qualitative component

Semi-structured interviews were used to contextualize the experimental choice patterns. Twenty-three informants were purposively selected from clinicians and administrative staff involved in DRG implementation; interviews lasted approximately 30 min. Guides addressed implementation arrangements, clinical behavior, financial incentives, documentation, and operational challenges (Supplementary File 2). The analysis is described as thematic and organized into two broad themes with 13 sub-themes (Supplementary File 3). Reporting is mapped against COREQ [25] (Supplementary File 2).

Results

Demographic characteristics and doctors' personality traits

Demographic characteristics

A total of 252 clinical physicians from comprehensive tertiary Grade A hospitals were included in this study. Participants were recruited from three provinces: Shandong (n = 119, 47.3%), Ningxia Hui Autonomous Region (n = 68, 27.0%), and Heilongjiang (n = 65, 25.8%) (Supplementary File 1, Table S2). Slightly more than half of the participants were male (51.2%). The majority of physicians were in the 31–40 age group (49.6%), followed by

the ≤30 age group (32.9%); only 17.5% were aged 41 years or older. A high level of educational attainment was observed, with 70.2% of physicians holding a master's degree or higher. The participants were predominantly from the departments of internal medicine (40.5%) and surgery (42.1%). Regarding professional experience, the largest group of physicians (41.3%) had been practicing for five years or less. In terms of professional titles, most participants held junior (33.7%) or intermediate (38.1%) positions. A smaller proportion held associate senior (11.9%) or senior (6.7%) titles, with the remainder (9.5%) classified as other.

Doctors' personality traits

The TIPI (Ten-Item Personality Inventory) comprises five personality dimensions, with two items per dimension. For each dimension, the score is calculated as the average of its two items. Statistical analysis revealed the following order of mean scores (from highest to lowest) among physicians: Extraversion, Openness, Emotional Stability, Agreeableness, and Conscientiousness. The specific scores are presented in Fig. 4. These descriptive scores do not establish a causal influence of personality on payment responses.

Volume of medical services by payment model

FFS model

Under FFS, mean service quantity was 5.88 (median 5.70), compared with the equally weighted task benchmark of 5. Mild and moderate means were 4.75 and 5.89, versus $q^*_{1k}=3$ and $q^*_{2k}=5$ ($p<0.001$ for both benchmark comparisons). The severe-case mean was 6.99 (median 7.00), versus $q^*_{3k}=7$ ($p=0.238$). These are aggregate summaries. These findings indicate that the FFS model is associated with a general tendency toward service over-provision, primarily driven by excessive services for patients with mild to moderate conditions.

DRG model

Under DRG, mean service quantity was 4.88 (median 4.83), or 0.12 units below the equally weighted task benchmark of 5 ($p<0.001$). Mild and moderate means were 3.38 and 4.95, versus $q^*_{1k}=3$ and $q^*_{2k}=5$; the reported benchmark-comparison p-values were <0.001 and 0.144, respectively. For severe vignettes, the mean was 6.29, or 0.71 units below $q^*_{3k}=7$ ($p<0.001$). These comparisons refer to experimental quantities and do not measure the proportion of clinically necessary treatment withheld. Therefore, the DRG model was associated with an overall under-provision of care, characterized by over-servicing mild cases, providing appropriate service levels for moderate cases, and significantly under-servicing severe cases.

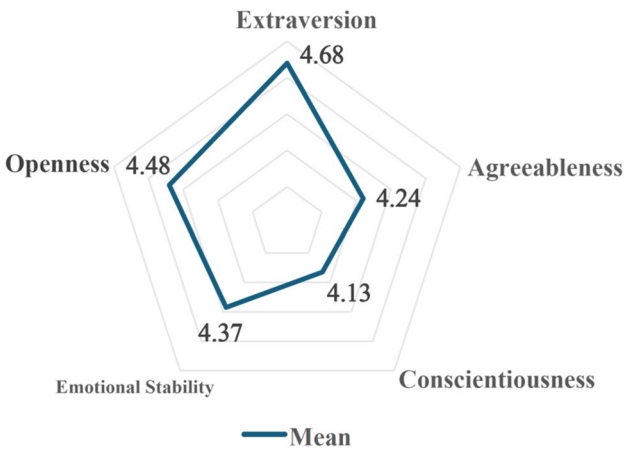


Fig. 4 Mean scores of physicians' personality traits

Table 1 Volume of medical services provided by physicians under FFS & DRG medical insurance payment models

Variable	Observations	Mean	Median (P25,P75)	Comparison with the optimal service level q^*_{sk} (*P*)
FFS	3780	5.88	5.70(5.07,6.60)	<0.001
mild	1260	4.75	4.60(3.20,5.80)	<0.001
moderate	1260	5.89	5.80(5.00,6.60)	<0.001
severe	1260	6.99	7.00(6.60,7.20)	0.238
DRG	3780	4.88	4.83(4.53,5.07)	<0.001
mild	1260	3.38	3.00(3.00,3.40)	<0.001
moderate	1260	4.95	5.00(4.60,5.00)	0.144
severe	1260	6.29	6.20(5.80,7.00)	<0.001
FFS versus DRG (*P*)	**<0.001**			

Table 2 Model-defined patient benefit loss ratio and physicians' net profit under two payments

Variable		FFS		DRG		*P*
		Mean	Median(P25,P75)	Mean	Median(P25,P75)	
Patient Benefit Loss Ratio	Overall	0.11	0.07(0.04,0.12)	0.11	0.05(0.01,0.15)	**0.010**
	mild	0.11	0.09(0.02,0.16)	0.09	0.01(0.00,0.10)	**<0.001**
	moderate	0.09	0.06(0.02,0.12)	0.12	0.03(0.00,0.17)	**0.019**
	severe	0.10	0.04(0.01,0.10)	0.12	0.05(0.02,0.15)	**0.040**
Physicians' Net Profit	Overall	7.85	7.62(6.70,8.94)	7.56	7.69(7.43,7.98)	**0.028**
	mild	6.61	6.24(4.86,8.08)	4.44	4.82(4.50,4.82)	**<0.001**
	moderate	7.93	7.84(6.24,9.08)	7.03	7.11(7.11,7.11)	**<0.001**
	severe	9.00	9.08(8.79,9.98)	11.21	11.52(10.48,12.00)	**<0.001**
Comparison of three severity levels *(P)*		**<0.001**		**<0.001**		

Comparative analysis and interpretation

Direct comparison revealed that the mean service volume under FFS was significantly higher than under DRG, both overall (5.88 vs. 4.88, a difference of 20.49%; $P<0.001$) and across all patient severity strata: mild (40.00% higher), moderate (18.98% higher), and severe (11.13% higher). In summary, while the implementation of a DRG payment model appears to mitigate the general over-provision observed under FFS, its effect is not uniform across patient populations. For patients with mild and moderate conditions, service volumes under DRG align more closely with optimal levels than under FFS. Conversely, for patients with severe conditions, the FFS model resulted in service volumes more consistent with the clinical optimum. Full details are presented in Table 1.

Impact of physician decision-making on patient benefit

Patient benefit under different payment models

Overall median PBLR was 0.05 under DRG and 0.07 under FFS ($p=0.010$), whereas both overall means rounded to 0.11. For mild vignettes, both mean and median PBLR were lower under DRG: means 0.09 versus 0.11 and medians 0.01 versus 0.09 ($p<0.001$). For moderate vignettes, the median was lower under DRG (0.03 versus 0.06), but the mean was higher (0.12 versus 0.09; $p=0.019$). For severe vignettes, PBLR was higher under DRG than FFS: mean 0.12 versus 0.10 and median 0.05 versus 0.04 ($p=0.040$). Detailed statistical results are presented in Table 2.

Impact of physician decision-making on net profit

Physician' net profit under different payment models

To ensure a direct and standardized comparison between the two payments, we analyzed physicians' experimental net profit as the primary financial outcome. Given the skewed distribution of profit data, non-parametric tests were employed for all statistical analyses.

Our analysis of overall physician net profit revealed a statistically significant difference between the two payment models ($P=0.028$). The mean experimental profit under the fee-for-service (FFS) model was 7.85 (median: 7.62), while under the DRG model, it was 7.56 (median: 7.69). This disparity in profitability between FFS and DRG was statistically significant across all patient severity strata—mild, moderate, and severe ($P<0.001$ for all comparisons). Furthermore, within each payment system, experimental net profit varied significantly based on patient severity. Under FFS, profit scaled with patient severity. Under DRG, a similar, but more pronounced, financial gradient was observed. Full details are available in Table 2.

Characterizing physician decision-making and altruism

Physician decision-making profiles

To characterize physicians' decision-making, we first established a classification framework. Given the 11 discrete service volumes (0–10) a physician could provide, each choice corresponds to a specific outcome for both patient benefit and physician profit. We developed a six-profile typology by first ranking the 11 potential outcomes for patient benefit and physician profit, respectively. These ranks were categorized into three tiers: Tier 1 (rank 1: "highest"), Tier 2 (ranks 2–6: "high"), and Tier 3 (ranks 7–11: "low"). Physician decisions were classified into one of six mutually exclusive profiles based on the combination of these tiers: *(a)* Highest Patient Benefit, *(b)* High Patient Benefit, High Physician Profit, *(c)* High Patient Benefit, Low Physician Profit, *(d)* Highest Physician Profit, *(e)* Low Patient Benefit, High Physician Profit, *(f)* Low Patient Benefit, Low Physician Profit.

Analysis of these profiles revealed distinct patterns based on payment model and patient severity. Under the FFS model, the proportion of decisions yielding the *Highest Patient Benefit* increased with patient severity, from 30.56% for mild cases to 40.00% for moderate and 44.60% for severe cases. Conversely, under the DRG model, a decreasing trend was observed: the proportion of Highest Patient Benefit decisions fell from 68.33% for mild cases to 60.32% for moderate and 31.83% for severe cases.

Notably, decisions prioritizing the Highest Physician Profit were infrequent across both models. The respective proportions for mild, moderate, and severe cases were 10.24%, 2.38%, and 6.75% under FFS, and 0.63%, 0.16%, and 0.00% under DRG. A comprehensive breakdown of all six decision profiles is provided in Table 3.

Table 3 Physician decision-making profiles

Physician Decisions	FFS, *n*(%)			DRG, *n*(%)		
	mild	moderate	severe	mild	moderate	severe
a	385(30.56)	504(40.00)	562(44.60)	861(68.33)	760(60.32)	401(31.83)
b	589(46.75)	596(47.30)	521(41.35)	204(16.19)	34(2.70)	182(14.44)
c	105(8.33)	14(1.11)	70(5.56)	91(7.22)	171(13.57)	575(45.63)
d	129(10.24)	30(2.38)	85(6.75)	8(0.63)	2(0.16)	0(0.00)
e	18(1.43)	52(4.13)	14(1.11)	91(7.22)	288(22.86)	21(1.67)
f	34(2.70)	64(5.08)	8(0.63)	5(0.40)	5(0.40)	81(6.43)

Table 4 Effects of sociodemographic characteristics on physician altruism

Variable		Coefficient		MRS(95%CI)
		Patient Benefit	Physician Net Profit	
Physicians' Overall Altruism		1.918****	0.349**	5.499(4.532,6.466)
Gender	Male	2.079***	0.233***	8.909(6.735,11.080)
	Female	1.702***	0.404***	4.213(3.321,5.106)
Age (years)	≤30	2.457***	0.353***	6.965(5.478,8.452)
	31–40	1.854***	0.344***	5.394(4.242,6.545)
	≥41	1.664***	0.258***	6.454(3.208,9.700)
Education	Bachelor's degree or below	1.873***	0.230***	8.143(4.985,11.301)
	Master's degree or above	1.595***	0.348***	4.579(3.726,5.432)
Department	Internal Medicine	2.457***	0.353***	6.965(5.478,8.452)
	Surgery	1.854***	0.344***	5.394(4.242,6.545)
	Other	1.664***	0.258***	6.454(3.208,9.700)

Note: ***$P<0.001$

Overall physician altruism

Results from a mixed logit model analysis provided quantitative insight into physician motivations. Both physician net profit and patient benefit were positive and highly significant predictors of decision-making ($P<0.001$), confirming that physicians are influenced by both financial incentives and patient welfare. Crucially, the regression coefficient for patient benefit was substantially larger than that for physician profit, demonstrating a stronger preference for maximizing clinical outcomes. Furthermore, the statistical significance of the coefficients' standard deviations ($P<0.001$) confirms the existence of significant heterogeneity in these preferences across the physician population.

This preference was quantified by the MRS. We found an MRS of 5.499, which signifies that, on average, physicians placed a 5.5-fold greater weight on patient benefit relative to personal profit. This provides strong evidence for a high degree of overall physician altruism in their clinical decision-making. In subgroup regression analyses stratified by gender, age, education and department, the regression coefficients for patient benefit exceeded those for physician net profit, indicating that physicians across all subgroups placed greater weight on patient benefit. Detailed estimates and full model results are presented in Table 4.

The influence of DRG on physician clinical behaviour: a qualitative analysis of crowding effects and unintended consequences

To achieve a comprehensive, real-world understanding of the impact of the DRG payment system, we conducted semi-structured interviews with 23 key informants across the six participating hospitals. Each interview lasted approximately 30 min. Participants were purposively selected to represent a range of critical perspectives, including clinicians from both DRG-surplus and DRG-deficit departments, as well as administrators from the offices of medical insurance, medical records, and finance. Thematic analysis of the interview transcripts yielded two overarching themes and 13 sub-themes.

Positive consequences: standardization of care and efficiency gains

As summarized in Supplementary File 3, a primary positive consequence of the DRG system reported by informants was the standardization of clinical practice. Physicians noted that the reform promoted adherence to clinical pathways, curtailed unnecessary diagnostic tests and procedures, and mitigated overtreatment. This, in turn, led to demonstrable reductions in both inpatient costs and length of stay for patients. Furthermore, informants highlighted a significant ancillary benefit: the improved accuracy of medical record-keeping. They emphasized that because the "front page" of the medical record-which contains the principal diagnosis and other key

information-is the primary determinant of DRG assignment and, consequently, reimbursement, the reform created a powerful incentive for precision in clinical documentation.

Adverse effects and unintended consequences

Despite these benefits, interviews revealed substantial adverse effects. Hospital policies linking departmental performance or physician remuneration to DRG financial outcomes intensified cost-control pressure, creating incentives to omit necessary tests, use cheaper drugs, or discharge patients prematurely. Informants also reported ethically concerning practices, including "cherry-picking" complex, high-cost patients, "upcoding" severity to obtain more profitable DRG, and "service unbundling," such as splitting one episode of care across multiple admissions. Conversely, overly generous DRG payments could encourage excessive service provision to match reimbursement levels. The system also increased physicians' administrative burden, as monitoring costs and navigating complex regulations contributed to dissatisfaction, reduced morale, and concerns about slower adoption of new technologies. Supplementary File 3 provides a full summary of these themes.

Discussion

Contrasting impacts of payment models on physician clinical behaviour

Our findings reveal that FFS and DRG payment models exert opposing influences on physician clinical behaviour. Under FFS, we observed a systematic over-provision of medical services relative to the model-defined optimal volume. This result is consistent with laboratory experiments conducted with medical students [14, 23] and has been corroborated by numerous empirical and theoretical studies [3, 5]. The underlying mechanism is the retrospective nature of FFS, where financial risk is borne primarily by the payer. This creates powerful incentives for volume expansion—leading to larger prescriptions and more extensive diagnostic and therapeutic interventions—which drives rapid healthcare cost inflation.

Conversely, the DRG prompted clinical decisions that more closely aligned with optimal care, effectively mitigating the over-provision endemic to FFS. This aligns with our qualitative findings that DRG help standardize clinical practice. The shift is attributable to the prospective nature of DRG payment, where a predetermined fee is set before service delivery. This creates a shared financial risk between the payer and the provider, fundamentally transforming the hospital's operational model from revenue maximization through volume to cost containment through efficiency. This incentivizes hospitals to control costs and eliminate unnecessary services, theoretically leading to higher-quality, more efficient care. Our findings validate previous work highlighting the role of DRG in standardizing physician behaviour and reducing unnecessary procedures [26, 27].

Our qualitative findings corroborate and extend the quantitative results. Informants reported that DRG implementation improved the accuracy of medical records, as the record's "front page" largely determines reimbursement. DRG were also associated with lower inpatient costs and shorter hospital stays, thereby encouraging the adoption of clinical pathways. These outcomes are consistent with studies demonstrating DRG' effectiveness in curbing expenses and hospital days [28] and underscore how cost-containment pressure can accelerate the integration of efficiency tools like clinical pathways.

However, these benefits are not without significant trade-offs. Our quantitative data showed that, overall, the DRG model was associated with a net under-provision of services. Our qualitative analysis revealed the mechanisms: to control costs and secure financial surpluses, physicians reported under-provision of care, patient selection ("cherry-picking" severe cases), "upcoding" to more profitable DRG, and service unbundling ("decomposing" hospital stays). These ethically fraught behaviours align with warnings from other researchers [29] who note that cost-containment pressures can lead to insufficient treatment, such as reduced service intensity or premature discharge. The experiment did not allow admission refusal, coding changes, or fragmented hospitalizations,

so these behaviors cannot be attributed to the experimental payment contrast. Thus, while metrics like average length of stay and cost per admission may improve, they must be interpreted with caution, as they may mask a decline in quality or access for the most vulnerable patients.

Intrinsic factors influencing physician clinical behaviour

Our study provides evidence that physician altruism is a dominant mediator of responses to financial incentives. Using the TIPI, we found that higher emotional stability was associated with greater service volume, plausibly because emotionally stable individuals are less swayed by situational pressures or personal biases and can focus more objectively on patients' clinical needs. This aligns with broader evidence on personality traits and professional behaviour [30].

Our analysis provides strong evidence of physician altruism. The mixed logit model yielded a MRS of 5.499 between patient benefit and physician profit, indicating that physicians were, on average, willing to forgo 5.499 units of personal profit for one additional unit of patient benefit. This was triangulated by two findings: (1) the coefficient for patient benefit exceeded that for physician profit in regression models, and (2) decision profiles showed choices maximizing patient benefit far outnumbered those maximizing physician profit.

Prior experiments also found patient benefit influences medical service choices [18], but our effect (MRS 5.499) is substantially larger than the 1.078 reported in laboratory experiments with medical students [31]. This disparity likely reflects the difference between practicing physicians and trainees: real-world practice exposes physicians to the profound impact of illness on patients and families, fostering a sense of mission and professional identity that likely amplifies altruistic motivation. This supports the altruism hierarchy proposed by Brosig-Koch et al., in which practicing physicians are more altruistic than medical students, who are in turn more altruistic than non-medical students [12].

We also found this altruism to be heterogeneous and associated with specific sociodemographic factors. Physicians aged 31–40 exhibited lower altruism, potentially due to greater financial pressures at this life stage. Male physicians appeared more altruistic than female physicians, though the literature on gender and altruism is mixed [32]. Interestingly, physicians with lower education (bachelor's or below) showed higher altruism, perhaps reflecting different career expectations regarding the trade-off between personal sacrifice and financial reward. Professional norms also play a crucial role. As others have found [17, 19, 33], a strong professional ethos can mitigate the conflict between financial incentives and patient interests. This suggests that strengthening medical professionalism through ethics education and robust evaluation systems can be a powerful tool to counteract negative behaviours associated with DRG payment systems.

Extrinsic and patient-related factors in physician decision-making

Crucially, the effects of payment models varied substantially by patient severity. FFS was more favorable for severe cases: although it led to overprovision in mild and moderate cases, service volume remained clinically appropriate for severe cases. Accordingly, patient-benefit loss under FFS declined as severity increased. Among severe patients, FFS produced significantly less benefit loss—and lower physician profit—than DRG, suggesting that it may be better suited to complex illnesses requiring comprehensive care. By contrast, DRG was more favorable for mild and moderate cases. Although it led to underprovision in severe cases, service volume was appropriate for moderate cases and only moderately excessive for mild cases. Patient-benefit loss under DRG increased with severity. For mild and moderate cases, DRG generated significantly less benefit loss—and lower physician profit—than FFS, indicating its suitability for less complex conditions with more standardized treatment protocols [21].

This quantitative finding was corroborated by our qualitative interviews. Informants explained that when a DRG payment is substantially higher than the expected cost, physicians are incentivized to over-service the patient to justify the payment level. Conversely, to avoid financial losses on costly cases, they may resort to

under-provision. This mirrors game-theory models predicting that DRG can lead to both over- and under-treatment, depending on the profitability of the case.

Organizational governance and transferability

Linking departmental DRG balances to performance assessment may transmit financial pressure, while separating assessment from those balances may protect clinical autonomy. These accounts do not distinguish changes in intrinsic motivation from strategic responses to changed rewards or constraints. Evidence of pressure or risk avoidance is therefore insufficient by itself to establish motivational crowding-out. Upcoding, fragmented admissions, and demoralization were also described, but their prevalence and causal determinants were not measured. Policy evaluation should examine both the payer-to-hospital rule and the way hospitals translate it into incentives for departments and clinicians. Transferability to China's parallel DRG and Diagnosis-Intervention Packet (DIP) arrangements is at the level of incentive mechanisms. DRG and DIP have different grouping and payment-calculation structures and should not be treated as interchangeable labels. The DRG/DIP policy package included special-case review to accommodate complex cases and valuable new technologies. Our task does not simulate the full grouping rules, regional point-value settlement, hospital budget negotiation, or special-case review process. These policy safeguards provide context for future evaluation, not evidence that the present experimental effects apply unchanged to any local version or implementation period.

Limitations

First, physicians made 30 abstract vignette choices on a discrete 0–10 scale. Actual care involves continuous or multidimensional treatment options, diagnostic uncertainty, time pressure, team decisions, accountability, and malpractice exposure. There were no real clinical treatment consequences, although physician payments and patient-benefit donations were real. The benefit maxima and severity categories were imposed by the task rather than clinically validated. Welfare comparisons and their severity pattern may change under alternative benefit schedules; robustness to such alternatives has not been established.

Second, the 252 physicians came from six tertiary public hospitals in three Chinese provincial-level regions. Purposive site selection may overrepresent those interested in research or patient-oriented decisions. The sample therefore cannot support a representative population estimate of altruism. Generalizability to other regions, hospital tiers, private providers, and systems with different compensation, liability, or organizational arrangements requires testing.

Third, incentives were paid at 10 tokens per CNY and may have been modest relative to clinical income. The actual earnings-to-income ratio was not measured. Named donations could activate image motivation or warm-glow giving in addition to concern for patients. Thus, the parameter is task-specific and should not be interpreted as pure clinical altruism.

Finally, the qualitative findings reflect selected informants' accounts, not independently verified patient outcomes. The joint display is interpretive and cannot establish organizational mediation. The brief TIPI dimensions and imprecise subgroup preference estimates further limit claims about individual traits and an altruism hierarchy. Real-world validation should evaluate payment effects on care quality, access, patient selection, coding, and expenditure together.

Conclusions

No single payment method can simultaneously maximise efficiency, quality for complex illness, and equitable access. A pluralistic, severity-sensitive system is therefore needed. FFS supports comprehensive care for severe or uncertain cases but encourages over-provision, whereas DRG standardise care and control costs for routine cases

but risk under-provision, risk avoidance, unbundling, and strategic coding in complex cases. Although clinical decisions reflect payment design, patient severity, experience, and intrinsic traits, physician altruism consistently prioritises patient benefit over personal gain. We therefore recommend retaining DRG for routine, lower-severity care while introducing carve-outs, risk-adjusted add-ons, and outlier payments for critical illness, multimorbidity, and novel technologies. Rigorous audits, system-wide coordination, and clinician-led governance should deter gaming, decouple clinician incentives from DRG outcomes, and safeguard quality and safety.

Abbreviations

CI	Confidence interval
DRG	Diagnosis-related groups
FFS	Fee-for-service
MRS	Marginal rate of substitution
PBLR	Patient benefit loss ratio

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s12913-026-15933-w.

Acknowledgements We sincerely thank the participating physicians and hospital administrators for their time and invaluable contributions to this study.

Author contributions Chao Zheng: Conceptualization, Funding acquisition, Investigation, Methodology, Writing-original draft. Shuyan Lin: Data curation, Formal analysis, Writing – original draft. Yuting Zhang: Conceptualization, Data curation, Formal analysis, Writing – review & editing. Jia Yin: Conceptualization, Data curation, Formal analysis, Supervision, Validation, Writing – review & editing. Qiang Sun: Conceptualization, Data curation, Formal analysis, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

Funding This work was supported by the grants from the Taishan Scholars Program of Shandong Province (Grant No. tsqn202507382).

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate According to the Measures for Ethical Review of Life Science and Medical Research Involving Humans issued by the National Health Commission of the People's Republic of China, research using anonymized data or biological samples may qualify for exemption from ethical review, provided that it does not cause harm to participants and does not involve sensitive personal information or commercial interests. In this study, all respondents provided their written informed consent prior to participation, and all data were de-identified before analysis to preserve participant anonymity. The study used only de-identified data and was therefore deemed exempt from the requirement of additional ethical review and informed consent by the Institutional Review Board (Academic Research Ethics Committee) of Shandong University School of Public Health. All procedures were conducted in accordance with relevant local legislation and institutional requirements.

Consent for publication Not Applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's

Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. 2025 national statistical bulletin on the development of medical security in China. https://www.nhsa.gov.cn/art/2026/7/16/art_7_21451.html.
2. Zhang CH, He J. The effect of provider payment methods in medical insurance on physicians' behaviour [in Chinese]. J Chongqing Med. 2012;41(6):618–20.
3. Liu LH. Motivation and constraints: the objectives of medical payment system reform [in Chinese]. China J Mod Med. 2010;20(22):3509–11.
4. Lu Y, Meng QY. Review on the impact of provider payment reforms on physicians' behaviour [in Chinese]. Chin Health Econ. 2014;33(2):36–8.
5. McClellan M. Reforming payments to healthcare providers: the key to slowing healthcare cost growth while improving quality?[J]. J Economic Perspect. 2011;25(02):69–92.
6. Grytten J, Sorensen RJ. Patient choice and access to primary physician services in Norway[J]. Health Econ Policy Law. 2009;4(01):11–27.
7. Dumont E, Fortin B, Jacquemet N, et al. Physicians, multitasking and incentives: Empirical evidence from a natural experiment[J]. J Health Econ. 2008;27(06):1436–50.
8. Hamada H, Sekimoto M, Imanaka Y. Effects of the per diem prospectivepayment system with DRG-like grouping system(DPC/PDPS)on resource usage and healthcare quality in Japan[J]. Health Policy. 2012;107(2–3):194–201.
9. Zeynep Or. Implementation of DRG payment in France: issues and recent developments[J]. Health Policy. 2014;117(02):146–50.
10. Klein-HitpaßU, Scheller-Kreinsen D. Policy trends and reforms in the German DRG-based hospital payment system[J]. Health Policy. 2015;119(03):252–7.
11. Jun L, et al. Toward Experimental Economics in the Real World: A Review of Field Experiments [J]. China Economic Q. 2015;14(03):853–84.
12. Brosig-Koch J, Hennig-Schmidt H, Kairies-Schwarz N, et al. The effects of introducing mixed payment systems for physicians: Experimental evidence[J]. Health Econ. 2017;26(02):243–62.
13. Lagarde M, Blaauw D. Physicians' responses to financial and social incentives: A medically framed real effort experiment[J]. Soc Sci Med. 2017;179:147–59.
14. Hennig-Schmidt H, Selten R, Wiesen D. How payment systems affect physicians' provision behaviour–an experimental investigation[J]. J Health Econ. 2011;30(04):637–46.
15. Brosig-Koch J, et al. Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision[J]. J Economic Behav Organ. 2016;131:17–23.
16. Hennig-Schmidt H, Wiesen D. Other-regarding behaviour and motivation in health care provision: an experiment with medical and non-medical students[J]. Soc Sci Med. 2014;108(05):156–65.
17. Kesternich I, Schumacher H, Winter J. Professional norms and physician behaviour: Homo oeconomicus or homo hippocraticus?[J]. J Public Econ. 2015;131(11):1–11.
18. Godager G, Wiesen D. Profit or patients' health benefit? Exploring the heterogeneity in physician altruism[J]. J Health Ecomomics. 2013;32(06):1105–16.
19. Chen YF, Ding YL, Pan YW, et al. Salary incentives and medical service supply: a real-effort experiment [in Chinese]. Economic Res. 2020;55(1):132–48.
20. Li JJ. Experimental Economics Research on the Influence of Payment Methods on Physicians' Behaviour [dissertation] [in Chinese]. Jinan: Shandong University; 2016.
21. Zhang XY, Han YL, Bo YQ, et al. An Experimental Study on the Impact of Switching from Postpayment to Prepayment Systems on Physicians' Behaviour [in Chinese]. Chin Health Econ. 2020;39(4):28–32.
22. Levitt SD, List JA. Field experiments in economics: the past, the present, and the future. Eur Econ Rev. 2009;53:1–18.
23. Tan Q, Liu S, Liu D, Liu X, Liao Y. A study on the impacts of diagnosis-intervention packet on physicians' behaviours: base on experimental economics. Chin J Health Policy. 2021;14:14–9.
24. Lin S, Sun Q, Zhou H, Yin J, Zheng C. How medical insurance payment systems affect the physicians' provision behaviour in China—based on experimental economics. Front Public Health. 2024;12:1323090.
25. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349–57.
26. Kim SJ, Han KT, Kim SJ, et al. Impact of a diagnosis-related group payment system on cesarean section in Korea[J]. Health Policy. 2016;120(06):596–603.

27. Bie FS, Wang S, Rao KQ, et al. Reform and practical application of diagnosis related groups (DRG) payment system at home and abroad [in Chinese]. Chin J Med Libr Inform Sci. 2018;27(8):43–7.
28. Wang FY, Wang JQ, Wang QH. Analysis of the effect of clinical pathway management on type 2 diabetic patients under DRG payment system [in Chinese]. Jiangsu Healthc Adm. 2022;33(11):1511–3.
29. Rogers WH, Draper D, Kahn KL, et al. Quality of care before and after implementation of the DRG-based prospective payment system. A summary of effects[J]. JAMA. 1990;264(15):1989–94.
30. Eagly AH, Steffen VJ. Gender and aggressive behaviour: a meta-analytic review of the social psychological literature[J]. Psychol Bull. 1986;100(03):309–30.
31. Zhang Y, Li X, Lin X, et al. An experimental economics study on measuring physicians' altruism in medical decision-making scenarios [in Chinese]. South China J Econ. 2022;(2):88–105.
32. Andreoni J, Vesterlund L. Which is the fair sex? Gender differences in Altruism[J]. Quart J Econ. 2001;116(01):293–312.
33. Lin JY, Joo JH, Zhou L, Goswami R, Liao JM. The Impact of Payment Reform on Medicaid Access and Quality: A National Survey of Physicians. Perm J. 2025;29(2):41–7.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Chao Zheng[1] · Shuyan Lin[2] · Yuting Zhang[3] · Jia Yin[4,5,6] · Qiang Sun[4,5,6,7]

✉ Jia Yin
yinjia@sdu.edu.cn

✉ Qiang Sun
qiangs@sdu.edu.cn

[1] Shandong Provincial Third Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China

[2] Army 73rd Group Military Hospital, Xiamen, Fujian, China

[3] Melbourne Institute of Applied Economic and Social Research, Health Analytics, Leadership, and Economics (HALE) Hub, Faculty of Business and Economics, The University of Melbourne, 111 Barry St, Carlton, VIC 3010, Australia

[4] Department of Social Medicine and Health Management, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China

[5] NHC Key Lab of Health Economics and Policy Research, Shandong University, Jinan, China

[6] Center for Health Management and Policy Research, Shandong University (Shandong Provincial Key New Think Tank), Jinan, China

[7] China National Health Development Research Center, Beijing 100044, China