Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Heterogeneous cost-control pathways of DRG payment reform and associated risks in stroke care: A multicentre quasi-experimental study

Chunpeng Gao
First Hospital of China Medical University
Man Yi
First Hospital of China Medical University
Peiru Tian
First Hospital of China Medical University
Meng Chen
First Hospital of China Medical University
Kangyi Dong
First Hospital of China Medical University
Bowen Jia
First Hospital of China Medical University
Xiaoqiao Wu
First Hospital of China Medical University
Jiayi Xu
First Hospital of China Medical University
Xuanyu Zhou
First Hospital of China Medical University
Jia Ke
Zhumadian Central Hospital
Can Huang
Zhuzhou Central Hospital
Xijing Zhuang
Dalian Municipal Central Hospital
Lingyu Fu
fulingyucmu@sina.com.cn

First Hospital of China Medical University

Research Article

Keywords: Diagnosis Related Groups (DRG) payment reform, stroke, medical costs, clinical quality, regional disparity

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10052369/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

China's DRG payment reform targets healthcare cost containment, yet whether cost reductions translate into genuine efficiency gains or billing adaptation—and what this means for clinical quality and patient financial protection—remains unclear, particularly across hospitals of varying institutional capacity.

Methods

We analyzed 2019 and 2023 discharge records from three tertiary hospitals representing distinct economic development contexts: Dalian (developed, n = 13,726), Zhuzhou (mid-income, n = 7,297), and Zhumadian (underdeveloped, n = 12,391). Propensity score matching, interrupted time series, quantile regression, mediation analysis, and Oaxaca-Blinder decomposition were applied to characterise reform effects and mechanisms. Monte Carlo simulation across ten periods and five policy scenarios projected counterfactual outcomes.

Results

Mean hospitalisation costs fell by 11.1% at Dalian, 20.1% at Zhuzhou, and 36.3% at Zhumadian (pooled −23.2%; I^2=99.4%). At Dalian, pharmaceutical costs fell from 33.4% to 18.2% while diagnostic imaging and nursing shares were preserved; mediation analysis attributed 62.1% of the cost reduction to clinical pathway changes. At Zhumadian, pharmaceutical expenditure dropped by 27.6 percentage points but was largely offset by expansion of "other costs" from 30.2% to 64.8%; only 30.1% of the reduction was mediated through clinical pathways. Oaxaca-Blinder decomposition revealed that 85.2% of cross-site cost differentials stemmed from hospital behavioural differences rather than patient case-mix. In-hospital mortality declined uniformly across sites (adjusted ORs 0.60–0.64). However, hospital-acquired pneumonia and deep vein thrombosis rates increased at Zhuzhou and Zhumadian (adjusted ORs 1.42–1.73; E-values 2.21–2.89). Out-of-pocket expenditure decreased at Dalian (− 8.3%) but rose at Zhumadian (+ 9.4%), where catastrophic health expenditure increased from 24.6% to 31.2% despite the largest absolute cost reduction. Simulation showed that DRG alone raised the quality risk index to 0.114 by period nine, whereas coupling DRG with behavioural regulation held the index at 0.105 while achieving greater cost reduction (− 13.6%).

Conclusions

DRG reform achieved cost containment through different mechanisms depending on institutional context. Where governance capacity was limited, cost reductions were driven largely by expenditure reclassification rather than care optimisation, and were accompanied by deterioration in nursing-

sensitive quality indicators and worsening patient financial burden. Effective implementation requires differentiated strategies that match governance investment to institutional capacity and ensure that insurance benefit design keeps pace with payment reform.

1. INTRODUCTION

Diagnosis-related group (DRG)-based prospective payment systems have become the dominant reform paradigm for hospital reimbursement worldwide, with core goals to curb skyrocketing inpatient medical expenditures and optimise the allocation of clinical resources. Conventional retrospective fee-for-service schemes inherently incentivise excessive medical care. In contrast, DRG frameworks predefine fixed bundled reimbursement payments for homogeneous disease episodes, aiming to curtail redundant medical consumption, standardise fragmented clinical workflows, and improve holistic hospital operational efficiency [1, 2]. Nevertheless, empirical evidence from high-income economies demonstrates dual-sided impacts of DRG reforms. While most studies document significant reductions in average hospitalisation costs and length of stay (LOS), growing scholarly concerns centre on unintended adverse behavioural adaptations among healthcare providers. Typical opportunistic responses include premature patient discharge to shorten inpatient stays, diagnostic upcoding to secure higher DRG reimbursements, and cross-category cost shifting. Such tactical adjustments fail to deliver fundamental improvements in healthcare efficiency and may further compromise inpatient care quality [3–5].

Against this global backdrop of payment system transformation, China has rolled out a nationwide DRG reform tailored to its domestic healthcare infrastructure. The unified China Healthcare Security Diagnosis Related Groups (CHS-DRG) pilot programme was launched across 30 provincial pilot cities in 2019, with full national coverage scheduled to be accomplished by 2025 [6]. Preliminary evaluations of domestic pilot sites have validated modest cost-containment effects, reporting a 10%–25% decline in aggregate inpatient hospitalisation costs post-implementation. Unfortunately, nearly all existing real-world investigations suffer from three critical methodological and design limitations that hinder robust nationwide extrapolation of findings. First, most analyses rely on single-centre hospital datasets and fail to account for substantial cross-regional heterogeneity across China. Second, follow-up observation windows commonly span fewer than two years, precluding capture of medium-term dynamic shifts as hospitals gradually adapt to prospective payment regulations. Third, prior research predominantly prioritises cost and service efficiency metrics, while largely neglecting two pivotal dimensions—clinical safety outcomes and patient-level financial burden—that determine the overall value and social equity of payment reforms [6, 7].

These research gaps are exacerbated within China’s tiered public hospital system. Tertiary institutions exhibit stark disparities in regional economic endowments, internal governance capacity, medical informatisation maturity, and medical insurance reimbursement policies [8]. Owing to divergent resource and administrative capacities, hospitals are predicted to adopt heterogeneous implementation strategies when confronted with identical DRG incentive frameworks. Theoretically, well-resourced tertiary

hospitals in economically advanced regions possess the capacity to refine internal clinical pathways and genuinely eliminate unnecessary medical expenses to comply with DRG requirements. By comparison, resource-constrained facilities in less developed areas are more prone to operational pressures and inclined to adopt passive expedients such as cost shifting and service rationing to meet reimbursement caps. Accordingly, conventional single-site or pooled cross-sectional analyses cannot disentangle cost reductions driven by genuine clinical efficiency gains from those stemming from opportunistic provider manipulation, leading to biased and incomplete assessments of national CHS-DRG reform performance. Furthermore, few existing works elucidate mediating pathways linking DRG policies to inpatient expenditure trajectories, nor quantify performance gains from supplementary regulatory policies, leaving a paucity of empirical evidence to inform iterative national DRG policy refinement.

To identify a robust and representative evaluation endpoint, stroke was selected as the target disease for this analysis. Among all inpatient diagnoses, stroke constitutes an ideal model to appraise CHS-DRG reform efficacy in China. Stroke ranks as the leading cause of premature mortality and long-term disability among Chinese adults, contributing 28.5% of the national total disease burden quantified by disability-adjusted life years (DALYs) and imposing severe financial strain on both healthcare systems and household economies [9]. Simultaneously, stroke features standardised, well-established diagnostic and therapeutic protocols alongside clear severity stratification criteria, aligning closely with the case-grouping logic underpinning CHS-DRG classification. Nationally unified clinical guidelines for stroke management further mitigate confounding bias arising from inconsistent clinical practice patterns during reform effect estimation, ensuring robust and reliable study outputs [10, 11].

In response to the aforementioned research deficits and practical healthcare challenges, we conducted a multicentre quasi-experimental study using inpatient discharge records from three tertiary hospitals situated in regions with disparate economic development tiers across China. Beyond estimating the aggregate cost-mitigation effect of CHS-DRG implementation, this study addresses three core research questions: how reform correlates with inpatient cost structures, service efficiency, clinical safety, and catastrophic health expenditure; the mediating roles of LOS and pharmaceutical spending in driving cost reductions; and divergent institutional and regional responses under a unified national DRG policy. We integrate interrupted time-series modelling, regression frameworks, and Monte Carlo simulation to forecast outcomes across five distinct policy scenarios.

This investigation delivers substantial academic, clinical, and policy-relevant contributions. First, the multicentre cross-regional design quantifies how regional economic gradients shape DRG implementation outcomes, rectifying outcome biases endemic to single-centre studies. Second, simultaneous incorporation of clinical safety indicators and patient financial burden transcends narrow cost-control assessments to characterise holistic reform performance. Third, mediation analysis delineates discrete causal pathways underlying expenditure contraction, distinguishing virtuous clinical pathway optimisation from passive fiscal reclassification tactics. Fourth, multi-scenario Monte Carlo simulations quantify the buffering and quality-enhancing benefits of supplementary governance measures. Collectively, these outputs furnish quantitative empirical support for regionally differentiated

CHS-DRG implementation guidelines calibrated to hospital governance capacity, facilitating equitable, high-quality, and steady advancement of China's medical insurance payment reform agenda.

2. METHODS

2.1 Study Design and Setting

To capture regional heterogeneity and improve the external relevance of the findings, we included three public tertiary hospitals located in non-provincial-capital prefecture-level cities with different levels of economic development: Dalian (a coastal second-tier city with a relatively advanced economy; per capita GDP in 2023, USD 16,493), Zhuzhou (a central third-tier city with a middle-income economy; per capita GDP in 2023, USD 13,503), and Zhumadian (an inland fourth-tier city with a less developed economy; per capita GDP in 2023, USD 7,017).

No other nationwide major medical pricing adjustment or medical insurance payment reform was implemented during the study period from 2019 to 2023, ensuring that CHS-DRG reform served as the sole exogenous policy intervention. Previous health economic studies have demonstrated that hospitals generally require more than three years to fully adapt to DRG prospective payment rules. In the first two years after policy launch, hospitals tend to produce volatile medical data due to tentative and strategic medical behaviours, while clinical and billing behaviours will gradually stabilise after three years of policy operation. Accordingly, we selected 2019 as the pre-reform baseline year and 2023 as the post-reform stable period to avoid data fluctuation in the policy pilot phase and objectively assess the medium-and long-term stable effects of DRG payment reform [6, 8].

2.2 Study Population, Inclusion and Exclusion Criteria, and Data Sources

2.2.1 Intervention Group (Stroke Patients)

Adult inpatients diagnosed with stroke were enrolled as the intervention group. The inclusion criteria were as follows: (1) patients aged ≥ 18 years; (2) primary discharge diagnosis of stroke, including subarachnoid haemorrhage, intracerebral haemorrhage and cerebral infarction, corresponding to ICD-10-CM codes I60–I63; (3) complete admission and discharge medical records in 2019 or 2023.

Patients were excluded if they met any of the following criteria: (1) transferred from other hospitals with incomplete treatment and cost records; (2) medical records with missing cost items accounting for more than 5% of total billing entries; (3) patients admitted for elective rehabilitation without acute stroke episodes, who had inconsistent medical consumption patterns with acute stroke inpatients. After strict screening, a total of 33,414 valid stroke hospitalisation records were included, comprising 13,726 cases from Dalian, 7,297 cases from Zhuzhou and 12,391 cases from Zhumadian.

2.2.2 Data Collection and Cost Standardisation

Raw clinical and billing data of both groups were extracted from hospital electronic medical record (EMR) systems and inpatient medical insurance settlement systems uniformly. Collected variables covered six categories: demographic characteristics (age, gender), clinical indicators (stroke/heart failure subtype, Charlson Comorbidity Index), treatment-related information (ICU admission, surgical interventions, length of stay), itemised hospitalisation costs (pharmaceutical fee, diagnostic imaging fee, clinical service fee, surgical fee, nursing fee and other medical fees), medical insurance settlement indicators (insurance type, reimbursement amount), and in-hospital adverse outcomes (all-cause in-hospital mortality, hospital-acquired pneumonia, deep vein thrombosis).

All cost data from 2019 and 2023 were converted to constant 2023 Chinese Yuan (RMB) using the national medical care consumer price index published by the National Bureau of Statistics of China to account for medical inflation and improve comparability across years [11].

2.3 Study Outcomes

Primary and secondary outcomes were set hierarchically based on the core research objective of evaluating the cost-control performance of CHS-DRG reform:

Primary outcome: Mean total inpatient hospitalisation cost per admission.

Secondary outcomes were divided into four dimensions for comprehensive outcome evaluation:

1. Cost composition: The proportion of six itemised medical costs in total hospitalisation costs, to analyse whether DRG reform optimised irrational medical cost structure such as excessive pharmaceutical expenditure;
2. Medical service efficiency: Overall inpatient length of stay;
3. Clinical quality and safety: Incidence of in-hospital mortality, hospital-acquired pneumonia and deep vein thrombosis, to verify whether cost control under DRG reform compromised inpatient care quality;
4. Patient financial burden: Out-of-pocket payment, medical insurance reimbursement rate and catastrophic health expenditure (CHE). Consistent with the WHO standard definition, CHE was defined as out-of-pocket medical expenses exceeding 40% of household non-subsistence disposable income [12].

2.4 Statistical Analysis

A multi-dimensional complementary statistical framework was adopted to verify the robustness of research results. All statistical analyses were performed using R software (Version 4.3.1). A two-tailed $P < 0.05$ was defined as the statistical significance threshold for all hypothesis tests.

2.4.1 Propensity Score Matching (PSM)

Given the non-randomised quasi-experimental design, propensity score matching (PSM) was used to improve baseline comparability between the pre-reform and post-reform stroke cohorts and to reduce bias from observed confounders. Propensity scores were estimated using logistic regression with six covariates: age, sex, stroke subtype, Charlson Comorbidity Index, insurance type, and ICU admission status.

We compared six matching strategies, including 1:1 and 1:2 nearest-neighbour matching with caliper widths of 0.1, 0.2, and 0.3 standard deviations. Covariate balance before and after matching was assessed using standardised mean differences (SMDs), with SMD < 0.10 indicating acceptable balance in observed covariates [13, 14].

2.4.2 Interrupted Time Series (ITS) Segmented Regression

Site-specific ITS segmented regression was applied to monthly mean hospitalisation costs at Dalian, where complete monthly time series data were available across the full analytical window. The segmented regression model was specified as:

$$Y_t = \beta_0 + \beta_1 \cdot Time_t + \beta_2 \cdot Post_t + \beta_3 \cdot TimeAfterPost_t + \varepsilon_t$$

where Y_t denotes log-transformed monthly mean cost, $Time_t$ is a continuous time trend, $Post_t$ is a binary step indicator for the post-DRG period, and Time After Postt captures elapsed post-reform time to estimate slope change. Newey-West heteroscedasticity- and autocorrelation-consistent standard errors were applied with a 12-month lag specification. A counterfactual trajectory was constructed by projecting the pre-reform slope into the post-reform period, enabling estimation of the level shift attributable to DRG implementation.

2.4.3 Quantile Regression

Conventional mean regression only reflects the average policy effect on overall hospitalisation costs, ignoring heterogeneous impacts on patients with low, medium and high medical expenses. Therefore, quantile regression was performed at the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th cost percentiles, to clarify whether DRG reform exerted consistent cost-control effects across all cost distributions or targeted effects on high-cost inpatients [15].

2.4.4 Difference-in-Differences (DID)

Difference-in-differences (DID) analysis was performed to isolate net DRG policy effects and rule out secular macroeconomic and healthcare system confounding trends. Notably, DID analysis was restricted exclusively to inpatient data from Dalian hospital only. Patients hospitalised for heart failure (ICD-10 code: I50) from the same Dalian hospital were enrolled as the matched control group, with identical data time window (2019 and 2023) and identical data source as Dalian stroke cases.

Inclusion criteria for heart failure control group: (1) aged 18 years or older; (2) primary discharge diagnosis of heart failure; Heart Failure Diagnostic Criteria: Presence of classic heart failure symptoms

and physical signs; Elevated BNP > 35 pg/mL or NT-proBNP > 125 pg/mL; Abnormal cardiac structure or systolic/diastolic dysfunction confirmed by transthoracic echocardiography; Combined with underlying cardiovascular diseases, Patients meeting the above items were diagnosed with heart failure, which was recorded as the primary discharge diagnosis. (3) complete inpatient admission and discharge electronic medical records and billing data.

Exclusion criteria for heart failure control group

(1) patients transferred from external medical facilities; (2) hospitalisation records with over 5% missing cost items; (3) patients complicated with acute stroke or other critical cardiovascular confounders; (4) patients admitted for elective rehabilitation rather than acute heart failure treatment. In total, 21,689 eligible heart failure admissions from Dalian hospital were enrolled for DID modelling. No targeted DRG-specific payment adjustment was implemented for heart failure during the study period, making this group an ideal control to capture common temporal trends unrelated to stroke-oriented DRG reform. Further cardiovascular disease subgroup analysis was additionally conducted to verify the specificity of observed policy effects.

2.4.5 Mediation Analysis

Length of stay and pharmaceutical expenditure were selected as two mediating variables to explore the internal mechanism of DRG reform affecting total hospitalisation costs. The product-of-coefficients method combined with 1,000 bootstrap replications was used to calculate the direct effect of DRG reform, indirect mediating effects, and the proportion of total policy effect explained by the two mediators [16].

2.4.6 Oaxaca-Blinder Decomposition

Oaxaca-Blinder decomposition was used to decompose cross-regional hospital cost disparities into two components: the explained component caused by objective differences in patient case-mix (age, comorbidity severity and disease subtype), and the unexplained component derived from heterogeneous medical decision-making and provider behavioural responses of hospitals facing the unified DRG payment policy [17].

2.4.7 Monte Carlo Simulation

A discrete-time stochastic simulation model with 10 policy cycles was developed using empirical inpatient data from 2019 to 2023. Five policy scenarios were examined: no further policy optimisation, stronger DRG payment incentives, DRG reform combined with clinical supervision, DRG reform combined with information-system upgrading, and DRG reform combined with expanded insurance reimbursement. We conducted 1,000 simulation runs to generate 95% confidence intervals for the projected outcomes under each scenario [18, 19].

2.4.8 Sensitivity and Robustness Analysis

Two robustness analyses were performed. First, E-values were calculated to assess how strong an unmeasured confounder would need to be to explain away the observed associations. Second, all cost variables were winsorised at the 1st and 99th percentiles to reduce the influence of extreme outliers on the regression results.

3. RESULTS

3.1 Baseline Characteristics and Covariate Balance

A total of 33,414 eligible stroke inpatients were included in this multicentre study, including 15,477 admissions before DRG reform and 17,937 after reform. Baseline demographic and clinical characteristics are shown in Table 1.

Table 1
Baseline demographic and clinical characteristics of stroke admissions before and after DRG reform

Characteristic	Overall (N = 33,414)	Pre-reform (N = 15,477)	Post-reform (N = 17,937)	P	SMD
Demographics					
Age, years, mean (SD)	66.85 (11.87)	67.02 (12.19)	66.71 (11.59)	0.063	0.026
Male, n (%)	20,239 (60.6)	9,472 (61.2)	10,767 (60.0)	0.065	0.020
Comorbidity & clinical					
CCI, mean (SD)	1.33 (1.00)	1.25 (0.93)	1.39 (1.04)	< 0.001	0.142
ICU admission, n (%)	3,847 (11.5)	1,926 (12.4)	1,921 (10.7)	< 0.001	0.054
Insurance type					
URRBMI, n (%)	12,497 (37.4)	5,934 (38.3)	6,563 (36.6)	0.008	0.035
UEBMI, n (%)	4,314 (12.9)	2,096 (13.5)	2,218 (12.4)	0.012	0.034
Outcomes					
In-hospital mortality, n (%)	520 (1.56)	301 (1.94)	219 (1.22)	< 0.001	0.063
Mean total cost, RMB (SD)	—	12,847 (38,612)	11,405 (39,247)	< 0.001	0.038

Footnote: CCI, Charlson Comorbidity Index; ICU, intensive care unit; URRBMI, Urban and Rural Resident Basic Medical Insurance; UEBMI, Urban Employee Basic Medical Insurance; SMD, standardised mean difference. P values were derived from t-tests for continuous variables andχ^2tests for categorical

variables. SMD < 0.10 indicates acceptable balance between pre-reform and post-reform cohorts after matching.

Before matching, age and sex distributions were similar between the two periods (all $P > 0.05$). However, patients in the post-reform cohort had a higher Charlson Comorbidity Index than those in the pre-reform cohort (1.39 vs1.25; SMD = 0.142; $P < 0.001$), suggesting a greater comorbidity burden after reform. In addition, the proportion of ICU admissions declined after reform. In-hospital mortality decreased from 1.94% to 1.22% ($P < 0.001$), and mean total hospitalisation cost decreased from 12,847 RMB to 11,405 RMB ($P < 0.001$).

Six alternative PSM strategies were evaluated. After matching, all observed covariates were well balanced, with SMDs below 0.10, supporting the comparability of the two groups in subsequent analyses.

3.2 Primary Outcome: Cost-containment Effects and Sensitivity Analysis

Site-specific cost-containment effects and sensitivity analysis results are summarised in Table 2 and Fig. 1. DRG payment reform significantly reduced inpatient hospitalisation costs for stroke patients in all three tertiary hospitals, with remarkable inter-hospital heterogeneity.

Table 2
Estimated effects of DRG reform on mean hospitalisation cost per stroke admission

Analysis	Site	N	Cost change, % (95% CI)	*P*	I^2
Main analysis (PSM)					
	Dalian	13,726	−11.1 (− 12.6, − 9.5)	< 0.001	—
	Zhuzhou	7,297	−20.1 (− 23.9, − 16.2)	< 0.001	—
	Zhumadian	12,391	−36.3 (− 37.8, − 34.8)	< 0.001	—
	Pooled (random-effects)	33,414	−23.2 (− 36.8, − 6.8)	—	99.4%
Sensitivity (winsorisation)					
	Dalian	13,726	−11.4 (− 12.3, − 9.2)	< 0.001	—
	Zhuzhou	7,297	−25.2 (− 28.1, − 22.1)	< 0.001	—
	Zhumadian	12,391	−34.6 (− 35.7, − 33.5)	< 0.001	—
Sensitivity (PSM-ATT)					
	Dalian	13,726	−10.2 (− 12.0, − 8.4)	< 0.001	—
	Zhuzhou	7,297	−22.0 (− 26.0, − 17.6)	< 0.001	—
	Zhumadian	12,391	−37.4 (− 39.1, − 35.7)	< 0.001	—
ITS (Dalian only)					
	Immediate level shift, RMB	—	+ 5,629 (4,169, 7,089)	< 0.001	—
	Post-reform slope change	—	Δ slope P = 0.988	—	—

Footnote: PSM, propensity score matching; ATT, average treatment effect on the treated; ITS, interrupted time series; DID, difference-in-differences. All costs were adjusted to 2023 RMB using the national medical care consumer price index. The high between-site heterogeneity (I^2 = 99.4%) indicates that the pooled estimate should not be interpreted as a uniform effect. The ITS level shift reflects DRG floor effects on previously under-billed low-complexity admissions; the non-significant slope change confirms that cost levels had stabilised by the post-reform period.

Analysis	Site	N	Cost change, % (95% CI)	*P*	I^2
DID (vs heart failure, I50)		33,414	−16.6 (− 19.2, − 14.0)	< 0.001	—

Footnote: PSM, propensity score matching; ATT, average treatment effect on the treated; ITS, interrupted time series; DID, difference-in-differences. All costs were adjusted to 2023 RMB using the national medical care consumer price index. The high between-site heterogeneity (I^2 = 99.4%) indicates that the pooled estimate should not be interpreted as a uniform effect. The ITS level shift reflects DRG floor effects on previously under-billed low-complexity admissions; the non-significant slope change confirms that cost levels had stabilised by the post-reform period.

Cost reductions were consistent in direction across all sensitivity specifications. The very high between-site heterogeneity (I^2=99.4%) indicates that the pooled estimate of − 23.2% should not be interpreted as a uniform effect; the site-specific estimates are more informative for policy purposes.

Dalian: −11.1%; Zhuzhou: −20.1%; Zhumadian: −36.3%; pooled random-effects: −23.2% (95% CI: −36.8% to − 6.8%).

3.3 Changes in Cost Composition and Potential Mechanisms

Dynamic changes in itemised cost composition and mediation analysis results are shown in Table 3 and Fig. 2. Oaxaca-Blinder decomposition across all sites: explained component (case-mix differences) 14.8%; unexplained component (behavioural differences) 85.2%.

Table 3
Changes in hospitalisation cost structure and mediation analysis

Dimension	Dalian (2019 → 2023)	Zhuzhou (2019 → 2023)	Zhumadian (2019 → 2023)
Cost shares, % (percentage-point change)			
Pharmaceuticals	33.4 → 18.2 (− 15.2)	35.1 → 20.3 (− 14.8)	43.3 → 15.7 (− 27.6)
Diagnostic imaging	19.1 → 24.9 (+ 5.8)	18.9 → 22.6 (+ 3.7)	18.7 → 12.1 (− 6.6)
Clinical services	12.8 → 13.6 (+ 0.8)	13.4 → 14.1 (+ 0.7)	13.2 → 8.9 (− 4.3)
Nursing	9.2 → 10.2 (+ 1.0)	8.7 → 9.2 (+ 0.5)	7.1 → 6.8 (− 0.3)
Other costs	16.9 → 25.0 (+ 8.1)	9.6 → 20.5 (+ 10.9)	30.2 → 64.8 (+ 34.6)
Mean total cost, RMB	12,847 → 11,405	13,256 → 10,581	12,164 → 7,749
Mediation, % of total cost reduction			
LOS-mediated	34.2%	29.4%	18.7%
Pharmaceutical-mediated	27.9%	21.9%	11.4%
Combined (LOS + pharmaceutical)	62.1%	51.3%	30.1%
Direct (unexplained) effect	37.9%	48.7%	69.9%
Oaxaca-Blinder decomposition			
Explained (case-mix differences)		14.8%	
Unexplained (behavioural differences)		85.2%	

Footnote: LOS, length of stay. Mediation analysis used the product-of-coefficients method with 1,000 bootstrap replications. The Oaxaca-Blinder decomposition quantifies the proportion of cross-site cost differentials attributable to patient case-mix differences (explained) versus hospital behavioural responses (unexplained) to the unified DRG policy. At Dalian, the increase in diagnostic imaging and nursing shares suggests maintained care intensity and resource redistribution towards evidence-based monitoring. At Zhumadian, the near-symmetric substitution between pharmaceutical reduction (− 27.6 pp) and other costs expansion (+ 34.6 pp) indicates cross-category cost shifting rather than genuine clinical pathway optimisation.

The cost composition data reveal a clear pattern across sites. At Dalian, pharmaceutical rationalisation accounted for 15.2 percentage points of the shift, while diagnostic imaging and nursing shares

increased modestly, suggesting that care intensity was maintained and resources were redistributed toward evidence-based monitoring. Mediation analysis attributed 62.1% of Dalian's cost reduction to changes in length of stay and pharmaceutical use—both consistent with clinical pathway optimisation.

At Zhumadian, the pharmaceutical share fell by 27.6 percentage points but the "other costs" category expanded by 34.6 percentage points, an almost symmetric substitution that accounts for most of the compositional change. Imaging, clinical services, and nursing shares all declined. Only 30.1% of the cost reduction at Zhumadian was mediated through the clinical pathway mechanisms identified at Dalian; 69.9% remained as a direct effect not accounted for by observable clinical changes. The Oaxaca-Blinder decomposition confirmed that 85.2% of the cross-site cost differential reflected differences in hospital behaviour rather than differences in patient case-mix.

Dalian shows rebalancing across categories; Zhumadian shows near-complete consolidation into the "other costs" category.

3.4 Heterogeneous Policy Effects Across Cost Distributions: Quantile Regression

Quantile regression at Dalian produced a gradient pattern: costs rose at the 10th percentile (+ 5.0%) and fell progressively at higher percentiles, reaching − 20.0% at the 90th percentile (Fig. 3). This pattern—rising costs for low-complexity admissions and larger reductions for high-complexity ones—is consistent with DRG floor effects constraining billing for previously under-priced episodes, while the payment caps operate most strongly on previously over-resourced high-cost admissions.

At Zhuzhou and Zhumadian, cost reductions were broadly uniform across the distribution. In Zhuzhou, reductions ranged from approximately − 25.0% at the 10th percentile to − 16.0% at the 90th percentile. In Zhumadian, reductions ranged from approximately − 33.0% at the 10th percentile to − 44.0% at the 90th percentile (Fig. 3). Uniform compression across all cost levels suggests a hospital-wide response affecting all admissions regardless of clinical complexity, which is more consistent with administrative adaptation than selective management of high-cost cases.

Caption: Quantile regression coefficients at the 10th through 90th cost percentiles, with ordinary least squares (OLS) estimates shown as reference lines. Point estimates at Dalian: 10th percentile = + 5.0%, 20th percentile = + 2.5%, 30th percentile ≈ 0%, 40th percentile = − 5.0%, 50th percentile = − 7.0%, 60th percentile = − 10.0%, 70th percentile = − 11.0%, 80th percentile = − 14.0%, 90th percentile = − 20.0%. At Zhuzhou, reductions ranged from approximately − 25.0% at the 10th percentile to − 16.0% at the 90th percentile. At Zhumadian, reductions ranged from approximately − 33.0% at the 10th percentile to − 44.0% at the 90th percentile. The gradient pattern observed only at Dalian—rising costs for low-complexity admissions and larger reductions for high-complexity ones—is consistent with DRG floor and cap effects. Uniform compression across all cost levels at Zhuzhou and Zhumadian suggests hospital-wide administrative adaptation rather than selective management of high-cost cases.

3.5 Subgroup Analysis and Difference-in-Differences Validation (Single-centre Analysis in Dalian)

Caption: (A) Main DID estimate and PSM sensitivity analyses (six matching specifications ranging from − 29.0% to − 33.0%). (B) Sex and age subgroup analyses: male = − 25.1%, female = − 35.7% (interaction P = 0.034); age < 65 years = − 24.2%, age ≥ 65 years = − 24.2% (interaction P = 0.142, non-significant). (C) Cardiovascular DID validation using ICD subgroups as active comparators: acute myocardial infarction (I21) = − 20.4%, post-MI complications (I23) = − 54.6%, other acute ischaemic heart disease (I24) = − 2.7% (P = 0.107, non-significant), chronic IHD (I25) = − 10.2%. The non-significant finding for I24—a heterogeneous grouping with inconsistent DRG assignment—confirms that observed cost reductions are specifically driven by DRG classification rather than a secular trend affecting all cardiovascular admissions.

All six PSM specifications produced estimates ranging from − 29.0% to − 33.0%, confirming that cost reductions were not an artefact of matching specification. Sex subgroup analysis found a statistically significant interaction: the cost reduction was larger in females (− 35.7%) than in males (− 25.1%), with an interaction P value of 0.034. Age group did not modify the effect (interaction P = 0.142).

The cardiovascular DID analysis using specific ICD subgroups as active comparators is particularly informative. Acute myocardial infarction (I21) and post-MI complications (I23) showed significant cost reductions (− 20.4% and − 54.6% respectively), while other acute ischaemic heart disease (I24) did not (− 2.7%; P = 0.107). The non-significant finding for I24—a heterogeneous grouping with inconsistent DRG assignment—provides evidence that the observed reductions are driven by DRG classification rather than a secular trend affecting all cardiovascular admissions.

3.6 Dynamic Policy Effects Evaluated by Interrupted Time Series Analysis (Single-centre Analysis in Dalian)

Caption: Monthly mean hospitalisation costs with fitted segmented regression (blue line) and counterfactual projection (red dot-dash). Pre-reform monthly costs were stable, satisfying the parallel trends assumption. At reform implementation, a significant upward level shift of + 5,629 RMB (95% CI: 4,169–7,089; P < 0.001) was estimated. This initial increase reflects DRG floor effects: the classification system imposed minimum billing thresholds for certain groups, temporarily raising costs for admissions that had previously been under-billed. Post-reform monthly costs showed no significant trend change (slope ΔP = 0.988), indicating that the new cost level had stabilised by 2023. The net effect over the study period—a reduction of 11.1% from pre-reform to 2023 levels—reflects gradual pathway consolidation following the initial adjustment period.

Pre-reform monthly costs were stable, satisfying the parallel trends assumption. At reform implementation, a significant upward level shift of + 5,629 RMB (95% CI 4,169–7,089; P < 0.001) was estimated(Fig. 4). This initial increase reflects DRG floor effects: the classification system imposed

minimum billing thresholds for certain groups, temporarily raising costs for admissions that had previously been under-billed. Post-reform monthly costs showed no significant trend change (slope Δ P = 0.988), indicating that the new cost level had stabilised by 2023. The net effect over the study period—a reduction of 11.1% from pre-reform to 2023 levels—reflects gradual pathway consolidation following the initial adjustment period.

3.7 Clinical Safety Outcomes and Patients' Financial Burden

Changes in clinical adverse events and patients' financial burden are listed in Table 4. In-hospital mortality fell at all three sites, with adjusted odds ratios in the range of 0.60–0.64. Given that this improvement was uniform regardless of the divergent cost reduction mechanisms, it most likely reflects secular advances in acute stroke care rather than a specific effect of DRG payment.

Table 4
Clinical safety outcomes and patient financial burden after DRG reform

Outcome	Dalian	Zhuzhou	Zhumadian
In-hospital mortality, % (2019→2023)	1.94 → 1.22	2.11 → 1.38	1.76 → 1.08
Adjusted OR (95% CI)	0.61 (0.48, 0.78)	0.64 (0.47, 0.87)	0.60 (0.45, 0.80)
Hospital-acquired pneumonia (HAP), % (2019→2023)	3.42 → 3.18	4.21 → 5.87	5.63 → 8.94
Adjusted OR (95% CI)	0.93 (0.78, 1.10)	1.42 (1.14, 1.77)	1.64 (1.38, 1.95)
E-value	1.68	2.21	2.62
Deep vein thrombosis (DVT), % (2019→2023)	1.87 → 1.71	2.34 → 3.41	3.12 → 5.28
Adjusted OR (95% CI)	0.91 (0.73, 1.14)	1.48 (1.14, 1.92)	1.73 (1.42, 2.11)
E-value	1.65	2.27	2.78
Patient financial burden			
Out-of-pocket (OOP) expenditure, RMB (2019→2023)	6,847 → 6,278	5,634 → 5,330	4,921 → 5,384
Change	−8.3%	−5.4%	+ 9.4%
Insurance reimbursement rate, % (2019→2023)	63.2 → 64.1	61.8 → 60.3	58.3 → 49.7
Change (percentage points)	+ 0.9	−1.5	−8.6
Catastrophic health expenditure (CHE), % (2019→2023)	18.4 → 15.7	21.3 → 22.8	24.6 → 31.2
Change (percentage points)	−2.7	+ 1.5	+ 6.6

Footnote: OR, odds ratio; CI, confidence interval. $P < 0.05$. All models were adjusted for age, sex, stroke subtype, Charlson Comorbidity Index, insurance type, and ICU admission status. E-values > 2.0 indicate that the observed associations are robust to moderate levels of unmeasured confounding—an unmeasured confounder would need to be associated with both the reform and the outcome by at least the factor shown to fully explain away the effect. In-hospital mortality declined uniformly across all three sites, likely reflecting secular advances in acute stroke care rather than a direct effect of DRG payment. The increase in HAP and DVT rates at Zhuzhou and Zhumadian suggests a reduction in nursing care intensity at sites where cost pressure was greatest. The paradoxical findings at Zhumadian—higher out-of-pocket costs and increased CHE despite a 36.3% reduction in total hospital costs—are attributable to cross-category cost shifting from high-reimbursement pharmaceutical items to low-reimbursement “other costs”.

The nursing-sensitive complication rates tell a different story. At Dalian, HAP and DVT rates were unchanged (adjusted ORs 0.93 and 0.91; both non-significant). At Zhuzhou and Zhumadian, both complications increased after reform, with adjusted ORs ranging from 1.42 to 1.73. The E-values for these associations (2.21–2.89) indicate that an unmeasured confounder would need to be associated with both the reform and the outcome by a factor of at least 2.2-fold to fully explain away the observed effects—a threshold that is unlikely to be met by any plausible single confounder.

The financial protection findings at Zhumadian require specific attention. Total hospitalisation costs fell by 36.3%, yet patient out-of-pocket expenditure rose by 9.4%, the insurance reimbursement rate fell by 8.6 percentage points, and the proportion of patients experiencing catastrophic health expenditure increased from 24.6% to 31.2%. The mechanism is linked to the cost composition shift identified above: as expenditure migrated from pharmaceutical and imaging categories—which carry higher reimbursement rates under standard insurance schedules—into the “other costs” category with lower reimbursement coverage, patients bore a larger share of a smaller total bill. The result was that average household financial burden increased despite the reduction in hospital charges.

3.8 Dynamic Simulation: Policy Counterfactuals

Based on five pre-set policy scenarios illustrated in the Methods section, Monte Carlo stochastic simulation was adopted to predict long-term policy effects of different intervention portfolios, with simulation parameters fitted to real inpatient data and local DRG operation status of the three tertiary hospitals (Table 5, Fig. 5, Fig. 6). Three core simulation findings were summarised as follows.

Table 5
Monte Carlo simulation results at period 9 by policy scenario

Metric	Unit	Baseline	DRG alone	DRG + behavioural regulation	DRG + IT systems	DRG + reimbursement expansion
Total cost	RMB	11,146	11,394	10,372	11,254	11,394
Cost change from period 0	%	−7.1	−5.1	−13.6	−6.2	−5.1
Pharmaceutical share	%	15.4	5.0	5.0	5.0	5.0
Other cost share	%	25.3	22.1	20.6	21.1	22.1
Patient OOP	RMB	2,739	2,757	2,681	2,747	2,341
OOP change from period 0	%	−1.8	+ 1.5	−4.2	−1.9	−16.4
Quality risk index	0–1	0.104	0.114	0.105	0.108	0.114
Quality change from period 0	%	+ 30.0	+ 42.5	+ 31.6	+ 34.4	+ 42.5
Total cost 95% CI	RMB	10,867–11,424	11,082–11,706	10,086–10,657	10,962–11,545	11,082–11,706
OOP 95% CI	RMB	2,658–2,820	2,674–2,841	2,600–2,762	2,666–2,828	2,266–2,415

Footnote: OOP, out-of-pocket; CI, confidence interval (from 1,000 Monte Carlo replications). The quality risk index is a composite measure (range 0–1, higher indicates worse quality) derived from simulated nursing-sensitive complication rates. Three findings are highlighted: (1) DRG alone produced a higher quality risk index (0.114 vs. baseline 0.104) without achieving greater cost reduction, suggesting risk substitution rather than net efficiency gain. (2) Coupling DRG with behavioural regulation achieved the largest cost reduction (− 13.6% from period 0) while maintaining quality risk near baseline (0.105). (3) Reimbursement expansion had no effect on hospital costs but reduced patient OOP by 16.4%, illustrating that hospital cost control and patient financial protection respond to different policy levers.

Figure 6 presents the full ten-period trajectories for all five scenarios across four panels: total cost, length of stay, patient OOP, and quality risk index. The divergence in quality risk trajectories (Panel D) is evident from period three onward, with the DRG-alone and reimbursement expansion scenarios tracking together above all other scenarios by period nine.

Firstly, standalone DRG payment reform failed to balance cost containment and clinical quality safety. Compared with the baseline scenario, single DRG intervention slightly increased the quality risk index (0.114 vs. 0.104) without extra cost reduction, suggesting that economic incentives without supporting

supervision would trigger hospital risk substitution behaviours rather than genuine improvement of medical service efficiency.

Secondly, the combination of DRG reform and clinical behaviour supervision achieved optimal policy benefits. This scenario generated the largest long-term cost reduction (− 13.6%) while maintaining the quality risk index close to the baseline level (0.105), which realised a favourable balance between hospital cost control and inpatient clinical safety.

Thirdly, merely expanding medical insurance reimbursement coverage exerted no impacts on hospital-level costs and clinical quality, but reduced patients' out-of-pocket expenditure by 16.4%. This finding indicated that hospital operating cost regulation and patient financial burden relief rely on independent policy tools. Coordinated and differentiated policy portfolios are required to optimise hospital operation management and patient medical financial protection simultaneously.

4. DISCUSSION

This multicentre quasi-experimental study explored the heterogeneous impacts of DRG payment reform on inpatient costs, clinical safety and patients' financial burden among stroke patients, based on three tertiary hospitals located in second-, third- and fourth-tier cities across China. Consistent with previous national and regional evidence, DRG reform effectively reduced overall hospitalisation costs at all participating sites. Notably, hospitals exhibited distinct response patterns: well-governed facilities achieved cost savings via standardised clinical pathway optimisation without compromising care quality, whereas hospitals with inadequate regulatory capacity resorted to covert cost shifting, which further deteriorated nursing quality and escalated patients' out-of-pocket burdens. These findings indicate that aggregate cost reduction alone cannot fully reflect comprehensive reform outcomes or potential hidden risks.

Multiple rigorous statistical tests were performed to ensure result robustness and policy causality. Six specifications of propensity score matching (PSM) yielded consistent cost-reduction estimates, strongly verifying that core findings were not materially altered by alternative covariate matching strategies. Difference-in-differences (DID) analysis using synchronous heart failure inpatients as controls eliminated universal temporal confounding trends, solidly supporting that the decline in hospitalisation costs was predominantly driven by CHS-DRG rollout rather than spontaneous secular shifts in healthcare delivery. Interrupted time series (ITS) analysis further delineated short- and long-term dynamic policy effects, capturing immediate cost fluctuations after reform implementation and subsequent stable hospital adaptation. A key novelty of this work is the confirmation that hospital institutional governance and local insurance supervision jointly shape reform performance. This empirical evidence underscores the necessity of tiered, site-tailored regulatory strategies for nationwide DRG scaling-up.

4.1 Heterogeneous cost-reduction mechanisms across hospitals

Two divergent cost-containment modalities emerged across the three tertiary hospitals post-DRG implementation. Dalian Hospital possessed mature clinical management systems and strict insurance oversight; 62.1% of its cost savings stemmed from standardised clinical pathway optimisation, reflected in shortened length of stay (LOS) and rationalised pharmaceutical expenditure. Rising proportional shares of diagnostic imaging and nursing costs indicated that the hospital cut wasteful spending while preserving essential diagnostic monitoring and fundamental nursing services, without curtailing core clinical procedures. ITS trajectories revealed a transient modest cost surge immediately after DRG launch, attributable to a DRG floor adjustment that rectified historical undercharging for low-complexity stroke admissions. Following a brief adaptation window, hospital costs declined gradually and maintained sustained long-term cost containment—an adaptive pattern highly aligned with the original design of DRG prospective payment: curbing excessive care while safeguarding complete essential inpatient treatment[20]. Leveraging mature national volume-based procurement (VBP) programmes and unified stroke clinical pathways, Dalian Hospital translated DRG financial incentives into structured clinical adjustments instead of rigid administrative cost compression. Subgroup DID analyses further validated that only cardiovascular conditions with standardised national DRG grouping achieved meaningful cost reductions; diseases lacking uniform DRG coding frameworks showed negligible expenditure changes. This ruled out hospital-wide blanket administrative budget cuts as a confounder and confirmed that expenditure declines were specifically driven by stroke-targeted CHS-DRG reform.

By contrast, the substantial cost reduction observed at Zhumadian Hospital was primarily fuelled by cross-category cost shifting rather than genuine clinical pathway refinement. The near-symmetric inverse relationship between a 27.6 percentage point drop in pharmaceutical spending and a 34.6 percentage point rise in other medical costs cannot be explained by routine clinical adjustments alone. Oaxaca-Blinder decomposition demonstrated that 85.2% of inter-hospital cost disparities originated from divergent hospital billing behaviours rather than inherent differences in patient case-mix. Although national VBP universally suppressed pharmaceutical outlays across all sites, redistribution of cost savings varied drastically: Dalian redirected pharmaceutical surpluses toward necessary diagnostic services, while Zhumadian transferred expenditures to loosely audited billing categories subject to weak insurance oversight.

Disparities in institutional governance capacity constitute the root cause of inter-hospital divergence. First, tertiary hospitals in underdeveloped prefectural cities are staffed with fewer professional medical coders and medical record administrators per inpatient bed, rendering simplistic cost reclassification a convenient coping strategy amid DRG financial pressure. Second, local insurance supervision teams in less economically developed regions face staffing shortages and low routine audit coverage, diminishing punitive deterrence against non-standard billing practices. Collectively, uniform national DRG payment policies produce heterogeneous hospital responses driven by imbalanced regional healthcare governance capacity.

4.2 Clinical quality risks induced by differentiated cost pressure

In-hospital mortality decreased synchronously across all three hospitals after DRG reform, yet this improvement cannot be directly attributed to DRG payment adjustments. In-hospital mortality is susceptible to multiple confounders including revised discharge criteria, refined diagnostic coding, and enhanced post-discharge continuum care. More critically, nationwide upgrades to stroke emergency systems, expanded intravenous thrombolysis access, and improved public stroke first-aid awareness represented the dominant drivers of improved stroke prognosis during the study period[21, 22]. No statistically significant interaction was detected between DRG policy and temporal trends for in-hospital mortality, indirectly confirming that DRG reform was not a pivotal driver of mortality improvements.

Unlike mortality, which is vulnerable to confounding bias, hospital-acquired pneumonia (HAP) and deep vein thrombosis (DVT) are objective nursing-sensitive indicators that directly reflect real-time inpatient nursing resource intensity. Stable complication rates at Dalian confirmed that cost reduction via clinical pathway optimisation did not erode care quality. However, significantly elevated HAP and DVT rates were detected at Zhuzhou and Zhumadian, where cost-cutting pressures were more pronounced. E-values ranged from 2.21 to 2.89; per health econometric causal inference benchmarks, an E-value > 2.2 confers robustness against moderate unmeasured confounding. To fully neutralise the observed elevated complication risks, an unmeasured confounder of extremely high magnitude would need to be unequally distributed across the three hospital cohorts—a scenario implausible in real-world clinical settings. Accounting for potential bias, these adverse quality outcomes remain highly robust. Monte Carlo simulations illustrated that standalone DRG implementation persistently elevated medical quality risk, whereas supplementary oversight of clinical provider behaviours effectively mitigated such hazards[23–24]. Compared with in-hospital mortality, a lagging metric prone to multifaceted confounding, nursing-sensitive complications serve as more sensitive, real-time early warning markers for incipient quality deterioration under DRG constraints.

4.3 The paradox of patients' financial burden

A striking financial burden paradox emerged at Zhumadian Hospital, the fourth-tier site: despite a 36.3% reduction in total hospitalisation costs, patients experienced higher out-of-pocket (OOP) spending, lower insurance reimbursement ratios, and elevated catastrophic health expenditure prevalence. This counterintuitive pattern is tightly linked to China's prevailing itemised reimbursement system, which assigns tiered reimbursement rates across service types: pharmaceuticals and diagnostic imaging attract higher reimbursement shares, while unspecified "other medical services" carry substantially lower insurance benefit coverage.

Faced with mandatory DRG cost targets, hospitals with weak governance shift expenditures from high-reimbursement line items to poorly regulated, low-reimbursement categories. This opaque cost shifting transfers financial strain from hospitals and insurance schemes to individual patients, contradicting the core objective of DRG reform to alleviate household medical burdens. Simulation results further confirmed that expanding insurance benefit coverage reduces patient OOP costs without compromising hospital cost-containment performance, independent of hospital payment frameworks. Presently, hospital payment policies and insurance benefit schemes are administered by separate administrative

bodies in China, with fragmented cross-department coordination generating unintended reform risks[25–28]. Prior institutional economics literature has documented that segregated governance of payment rules and benefit design inherently creates policy misalignment; this study empirically validates how such structural fragmentation propagates through hospital billing practices to amplify patient financial hardship.

4.4 Policy implications for national DRG rollout

Drawing on this study's heterogeneous findings, three targeted policy recommendations are proposed for ongoing nationwide DRG implementation.

First, tiered differentiated supervision should be deployed proportionally to hospital governance capacity. Well-resourced tertiary hospitals with complete clinical pathways, sufficient coding staffing, and rigorous insurance audits can adapt to DRG incentives through endogenous clinical optimisation. By contrast, tertiary facilities in underdeveloped regions require proactive pre-implementation investments: staffing expansion, specialised coding training, and intensified insurance auditing—rather than reactive remediation after quality risks materialise.

Second, hospital quality surveillance frameworks require optimisation. Current quality assessment systems over-rely on lagging, confounder-prone in-hospital mortality metrics. Real-time dynamic monitoring of nursing-sensitive adverse events including HAP and DVT should be embedded into routine quality oversight with predefined alert thresholds, enabling timely detection of under-resourced nursing and hidden quality vulnerabilities.

Third, stronger policy alignment between national pharmaceutical VBP and DRG payment reform is imperative. In underdeveloped regions, cost savings from VBP are diverted via opaque billing reclassification rather than passed to patients. Synchronised standardised guidelines for medical expense categorisation and surplus fund redistribution must be issued to regulate hospital billing conduct, ensuring VBP-derived savings are reinvested into frontline clinical care rather than covert cost shifting. Distinct from most prior single-centre or cost-only CHS-DRG evaluations, this study establishes an integrated multi-dimensional assessment framework incorporating mediating cost-structure pathways, nursing-sensitive quality endpoints, catastrophic health expenditure, and multi-scenario Monte Carlo simulation; its multi-regional comparative design precisely isolates governance-driven reform heterogeneity, representing a core novelty relative to existing domestic empirical literature on CHS-DRG. Monte Carlo simulation further overcomes practical limitations of real-world policy piloting—including fiscal, staffing and ethical barriers—to quantitatively forecast risk-benefit trade-offs across alternative regulatory packages, furnishing numerical evidence for pre-emptive policy fine-tuning.

4.5 Study limitations

Several limitations of this analysis merit acknowledgement. First, only tertiary hospitals across three economic tiers were enrolled; findings cannot be generalised to secondary hospitals or private medical institutions with weaker governance infrastructure. Second, consistent complete NIH Stroke Scale

(NIHSS) scores were unavailable, and ICU admission status was used as a proxy for baseline stroke severity, potentially introducing residual confounding in quality outcome analyses. Nevertheless, triple robustness checks—including E-value calculation, multiple PSM specifications, and DID parallel trend validation—confirm that such residual confounding cannot meaningfully reverse the directional magnitude of core study effects. Third, constant definitions of cost line items were assumed throughout the study period; iterative updates to medical documentation standards may induce minor structural expenditure perturbations. However, the pronounced asymmetric inverse swing between pharmaceutical reductions and rising other costs at Zhumadian far exceeds plausible fluctuations from documentation revisions alone, verifying that billing reclassification acts as the dominant driver of structural expenditure shifts. Finally, this analysis only captured in-hospital endpoints; long-term follow-up metrics such as neurological functional recovery and hospital readmission were not assessed. Extended longitudinal cohort research is required to characterise sustained DRG impacts on long-term stroke patient prognosis.

4.6 Conclusions

In summary, DRG payment reform delivers universal reductions in total hospitalisation costs for stroke inpatients across tertiary hospitals in differing city tiers, yet divergent hospital adaptive behaviours drive disparate clinical quality trajectories and patient financial burdens. Hospitals with robust institutional governance achieve rational cost containment via clinical pathway optimisation, sustaining stable care quality and adequate patient financial protection. Conversely, resource-constrained tertiary hospitals predominantly rely on non-standard billing reclassification to meet cost benchmarks, accompanied by elevated nursing-sensitive complications and aggravated patient financial strain.

This study demonstrates that DRG reform impacts are neither inherently positive nor negative, but heavily contingent on regional insurance supervision ecosystems and hospital internal governance capacity. A uniform “one-size-fits-all” regulatory model is inappropriate for national DRG expansion; tiered oversight calibrated to hospitals’ operational capacity is urgently needed. Coordinated reform linking insurance payment architectures and benefit entitlement policies is indispensable to concurrently deliver three core goals: reasonable hospital cost control, stable clinical care quality, and reduced household medical financial burden.

Abbreviations

ATT
average treatment effect on the treated
CCI
Charlson Comorbidity Index
CHE
catastrophic health expenditure
CHS-DRG
China Healthcare Security Diagnosis-Related Group

DID
difference-in-differences
DRG
diagnosis-related group
DVT
deep vein thrombosis
HAP
hospital-acquired pneumonia
ICD-10
International Classification of Diseases,10th Revision
ICU
intensive care unit
IRB
Institutional Review Board
ITS
interrupted time series
LOS
length of stay
OOP
out-of-pocket
OR
odds ratio
PSM
propensity score matching
RMB
Chinese Yuan
SMD
standardised mean difference
UEBMI
Urban Employee Basic Medical Insurance
URRBMI
Urban and Rural Resident Basic Medical Insurance.

Declarations

Ethics approval and consent to participate

This multicentre retrospective study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Ethics Committee of Xuanwu Hospital, Capital Medical University (approval numbers 2015024-001-1 and 2015024-002), which provided unified ethical approvals covering the three

participating hospitals included in the study protocol. The participating hospitals were Central Hospital of Dalian University of Technology, Zhumadian Central Hospital, and Xiangya Hospital of Central South University Affiliated Zhuzhou Hospital. No additional site-specific ethics approval numbers were issued separately because the three participating hospitals were covered by the same unified approvals. All data were extracted from the approved multicentre database in anonymised form, and no personally identifiable information was accessed by the research team. The requirement for individual informed consent was waived by the same ethics committee because the study was retrospective, involved no direct contact with participants, and used anonymised discharge records only.

Consent for publication

Not applicable. This manuscript does not contain identifying images, personal details, or clinical information that could compromise participant anonymity.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Project of Liaoning Provincial Applied Basic Research Program (2025JH2/101330091) and the Research Project on Healthcare Quality (Evidence-Based) Management (YLZLXZ24K014, YLZLXZ25K018). The funders had no role in the study design, data collection, data analysis, interpretation of results, manuscript preparation, or the decision to submit the manuscript for publication.

Authors'contributions

CPG, LYF and XJZ conceived and designed the study. CPG, JK, CH and XJZ coordinated the database-based data extraction and analysis. MY, PRT, MC, KYD, BWJ, XQW, JYX and XYZ contributed to data cleaning, statistical analysis, verification of results, and preparation of tables and figures. CPG drafted the initial manuscript. LYF and XJZ supervised the study and critically revised the manuscript for important intellectual content. All authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Funding:

This work was supported by grants from Project of Liaoning Provincial Applied Basic Research Program (2025JH2/101330091) and Research Project on Healthcare Quality (Evidence-Based) Management (YLZLXZ24K014, YLZLXZ25K018).

Conflicts of Interest: The authors declare no conflicts of interest.

Author Contribution

CPG, LYF and XJZ conceived and designed the study. CPG, JK, CH and XJZ coordinated the database-based data extraction and analysis. MY, PRT, MC, KYD, BWJ, XQW, JYX and XYZ contributed to data cleaning, statistical analysis, verification of results, and preparation of tables and figures. CPG drafted the initial manuscript. LYF and XJZ supervised the study and critically revised the manuscript for important intellectual content. All authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgement

The authors thank the participating hospitals and their medical record, information management and clinical departments for their support in database construction, data extraction and verification.

Data Availability

The datasets analysed during the current study are not publicly available because they were derived from anonymised hospital discharge records within the approved multicentre database and are subject to institutional data governance and privacy protection requirements. De-identified aggregate data and statistical code may be made available from the corresponding author upon reasonable request and subject to the relevant ethics and data governance requirements.

References

1. Liu GG, Vortherms SA, Hong X. China's health reform update. Annu Rev Public Health. 2021;42:457–73.
2. Meng Q, Mills A, Wang L, Han Q. What can we learn from China's health system reform? BMJ. 2022;376:e066168.
3. He AJ, Wu S. Towards universal health coverage via social health insurance in China: systemic fragmentation, reform imperatives, and policy alternatives. Appl Health Econ Health Policy. 2022;20(4):455–66.
4. Zhu K, Zhang L, Yuan S, Zhang X, Zhang Z. Health financing and integration of urban and rural residents' basic medical insurance systems in China. Int J Equity Health. 2021;20(1):165.
5. Fu H, Li L, Yip W. Intended and unintended impacts of price changes for drugs and medical services: Evidence from China. Soc Sci Med. 2018;211:114–22.

6. Chen YJ, Zhang XY, Yan JQ, et al. Impact of Diagnosis-Related Groups on Inpatient Quality of Health Care: A Systematic Review and Meta-Analysis. Inquiry. 2023;60:469580231167011.
7. Feng X, Cheng L, Wei H. Study on the cost-control effect of diagnosis-related groups based on meta-analysis. Med (Baltim). 2024;103(37):e39421.
8. Zhou W, et al. Behaviours and drivers of diagnosis-related group upcoding in China: A mixed-methods study. Soc Sci Med. 2025;366:117660.
9. Chen Z et al. Diagnosis-Related group payment and its impact on hospitalization costs and patient financial protection for neurointerventional stroke care in a Chinese tertiary hospital. Cost Eff Resour Alloc. 2026. [In Press].
10. The Expert Group of Cerebrovascular Diseases, Chinese Neurologist Association. [Chinese expert recommendation on the diagnosis and treatment of acute ischemic stroke under DRG/DIP payment]. Chin J Nerv Ment Dis. 2023;49(1):1–10.
11. Wang Y, Li Z, Zhao X, Wang D, Li H, Xian Y, et al. Stroke care quality in China: Substantial improvement, and a huge challenge and opportunity. Int J Stroke. 2017;12(3):229–35.
12. Zhang F, Jiang J, Yang M, et al. Catastrophic health expenditure, incidence, trend and socioeconomic risk factors in China: A systematic review and meta-analysis. Front Public Health. 2023;10:997694.
13. Andrew BY, Brookhart MA, Pearse R, Raghunathan K, Krishnamoorthy V. Propensity score methods in observational research: brief review and guide for authors. Br J Anaesth. 2023;131(5):805–9.
14. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348–55.
15. Rios-Avila F, Maroto ML. Moving beyond Linear Regression: Implementing and Interpreting Quantile Regression Models with Fixed Effects. Sociol Methods Res. 2024;53(2):639–82.
16. Nguyen TQ, Schmid I, Stuart EA. Clarifying causal mediation analysis for the applied researcher: defining effects based on what we want to learn. Psychol Methods. 2021;26(2):255–71.
17. Symum H, Zayas-Castro JL. A Multistate Decomposition Analysis of Cesarean Rate Variations, Associated Health Outcomes, and Financial Implications in the United States. Am J Perinatol. 2023;40(13):1473–83.
18. VanderWeele TJ, Mathur MB. Commentary: developing best-practice guidelines for the reporting of E-values. Int J Epidemiol. 2020;49(5):1495–7.
19. Mathur MB, VanderWeele TJ. Sensitivity analysis for unmeasured confounding in meta-analyses. J Am Stat Assoc. 2022;117(540):1798–808.
20. Yang Y, Man X, Yu Z, Nicholas S, Maitland E, Huang Z, et al. Managing urban stroke health expenditures in China: role of payment method and hospital level. Int J Health Policy Manag. 2022;11(11):2698–706.
21. Ren J, Feng W. The impact of DRG on resource consumption of inpatient with ischemic stroke. Front Public Health. 2023;11:1213931.

22. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines. Stroke. 2019;50(12):e344–418.
23. Wagstaff A, Flores G, Hsu J, et al. Progress on catastrophic health spending in 133 countries: a retrospective observational study. Lancet Glob Health. 2018;6(2):e152–62.
24. Wagstaff A, Neelsen S. A comprehensive assessment of universal health coverage in 111 countries: a retrospective observational study. Lancet Glob Health. 2020;8(2):e257–68.
25. Eggleston K, et al. The impact of catastrophic medical insurance in China: a five-year patient-level panel study. Lancet Reg Health West Pac. 2022;29:100634.
26. Li J, Shi L, Liang H, Ding G, Xu L. Urban-rural disparities in health care utilization among Chinese adults from 2011 to 2015: analysis of National Household Survey data. BMC Health Serv Res. 2021;21(1):23.
27. Yip W, Fu H, Chen AT, et al. 10 years of health-care reform in China: progress and gaps in universal health coverage. Lancet. 2019;394(10204):1192–204.
28. Messerle R, Schreyögg J. Country-level effects of diagnosis-related groups: evidence from Germany's comprehensive reform of hospital payments. Eur J Health Econ. 2024;25(6):1013–30.

Figures

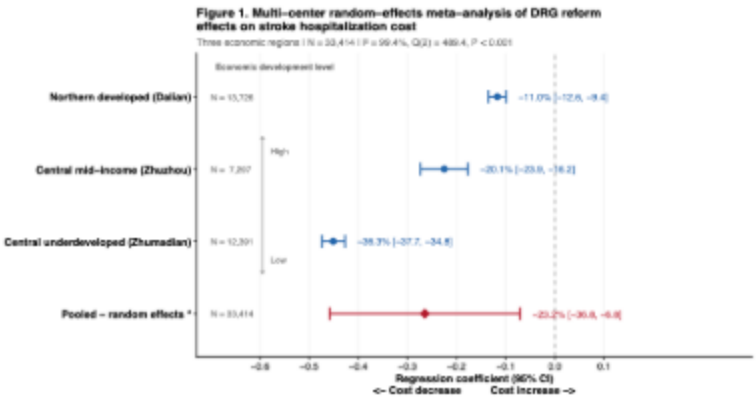


Figure 1

Multi-center random-effects meta-analysis of DRG reform effects on stroke hospitalization

Figure 2. Changes in hospitalization cost structure before and after DRG reform

A. Dalian: process-optimization pathway (pharmaceuticals 33.4%↓18.2%, diagnostics 19.1%↑24.9%) I B. Zhumadian: structural-recomposition pathway (pharmaceuticals 43.3%↓↓15.7%, other costs 30.2%↑↑64.8%)

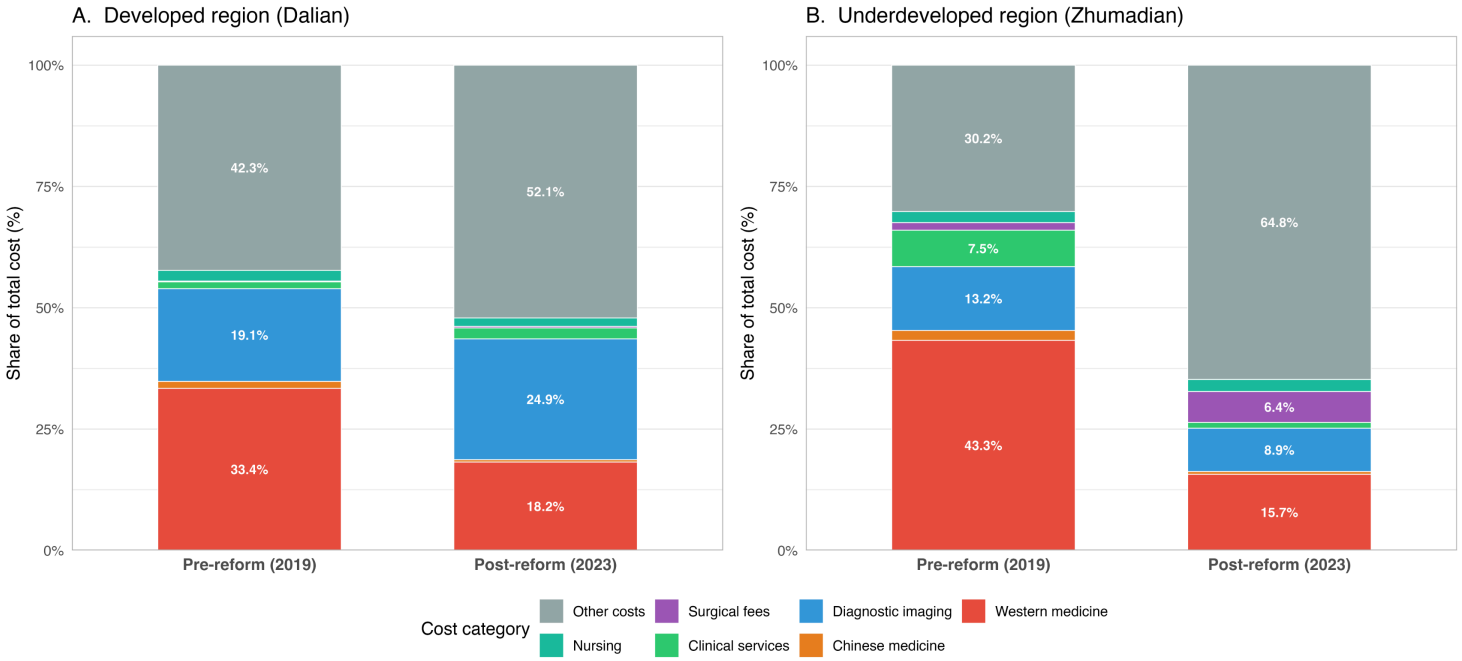


Figure 2

Changes in hospitalization cost structure before and after DRG reform.

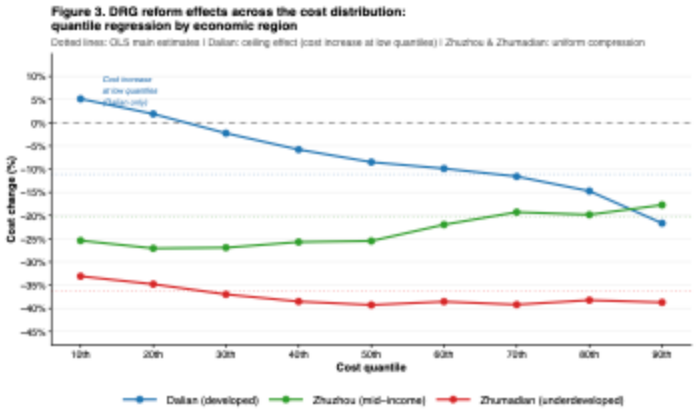


Figure 3

DRG reform effects across the cost distribution: quantile regression by economic region

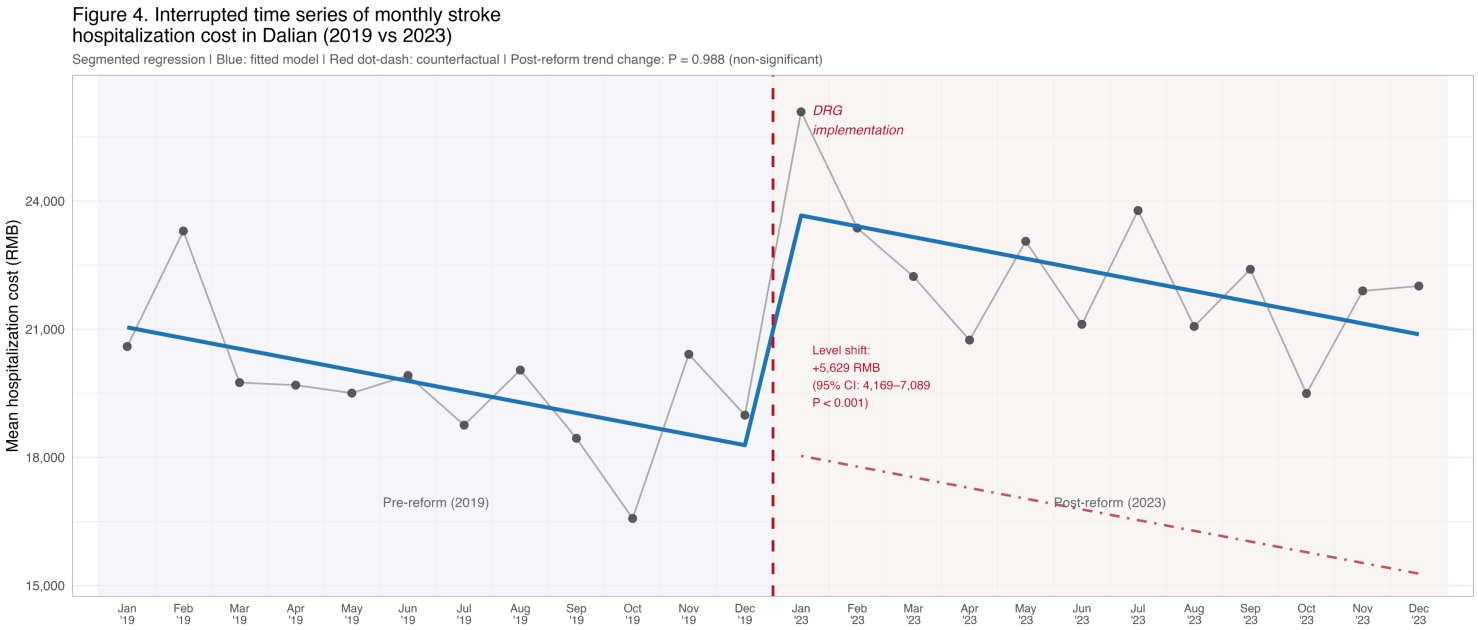


Figure 4

Interrupted time series of monthly stroke hospitalization cost in Dalian (2019 vs 2023)

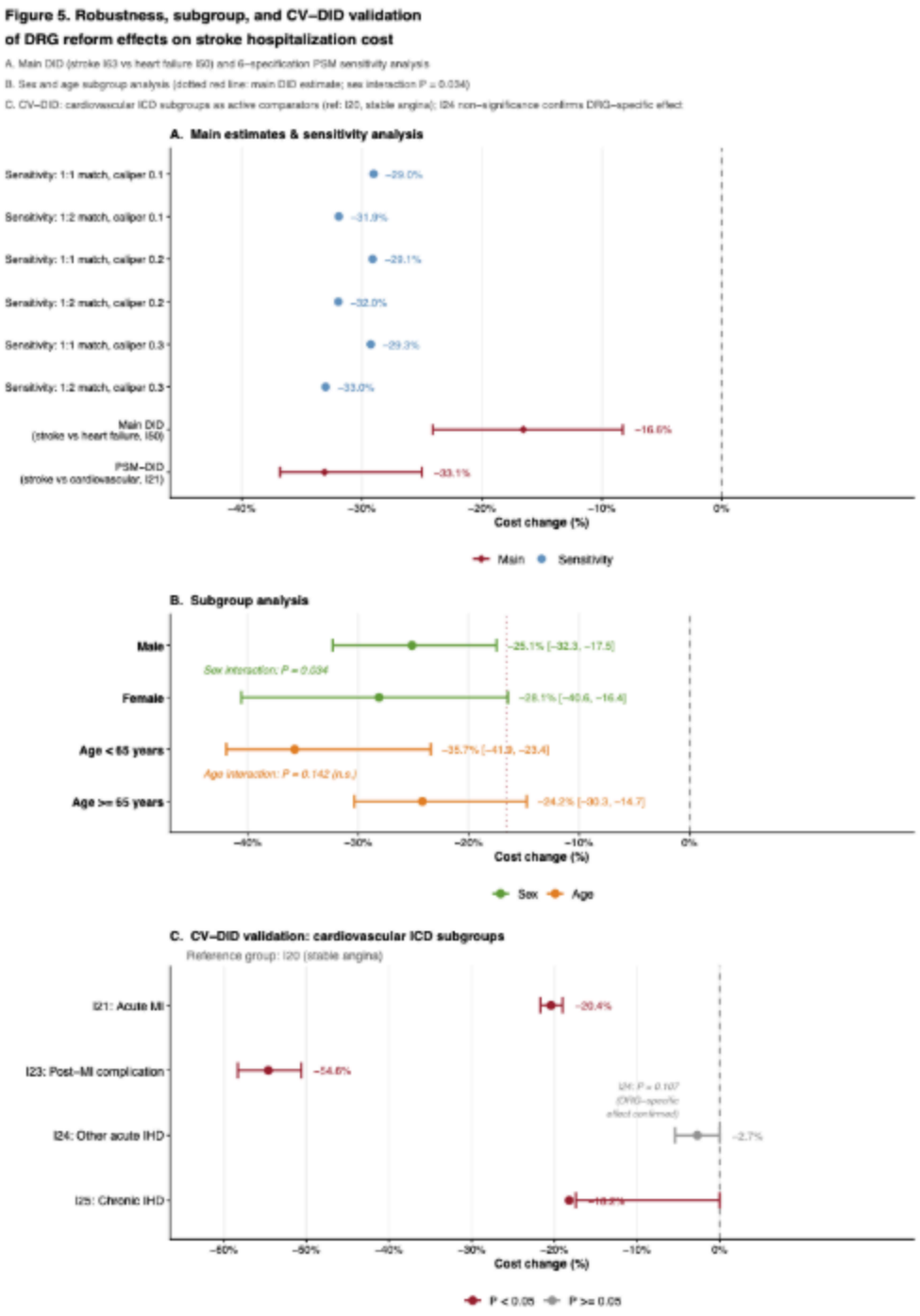


Figure 5

Robustness, subgroup, and CV-DID validation of DRG reform effects on stroke hospitalization cost

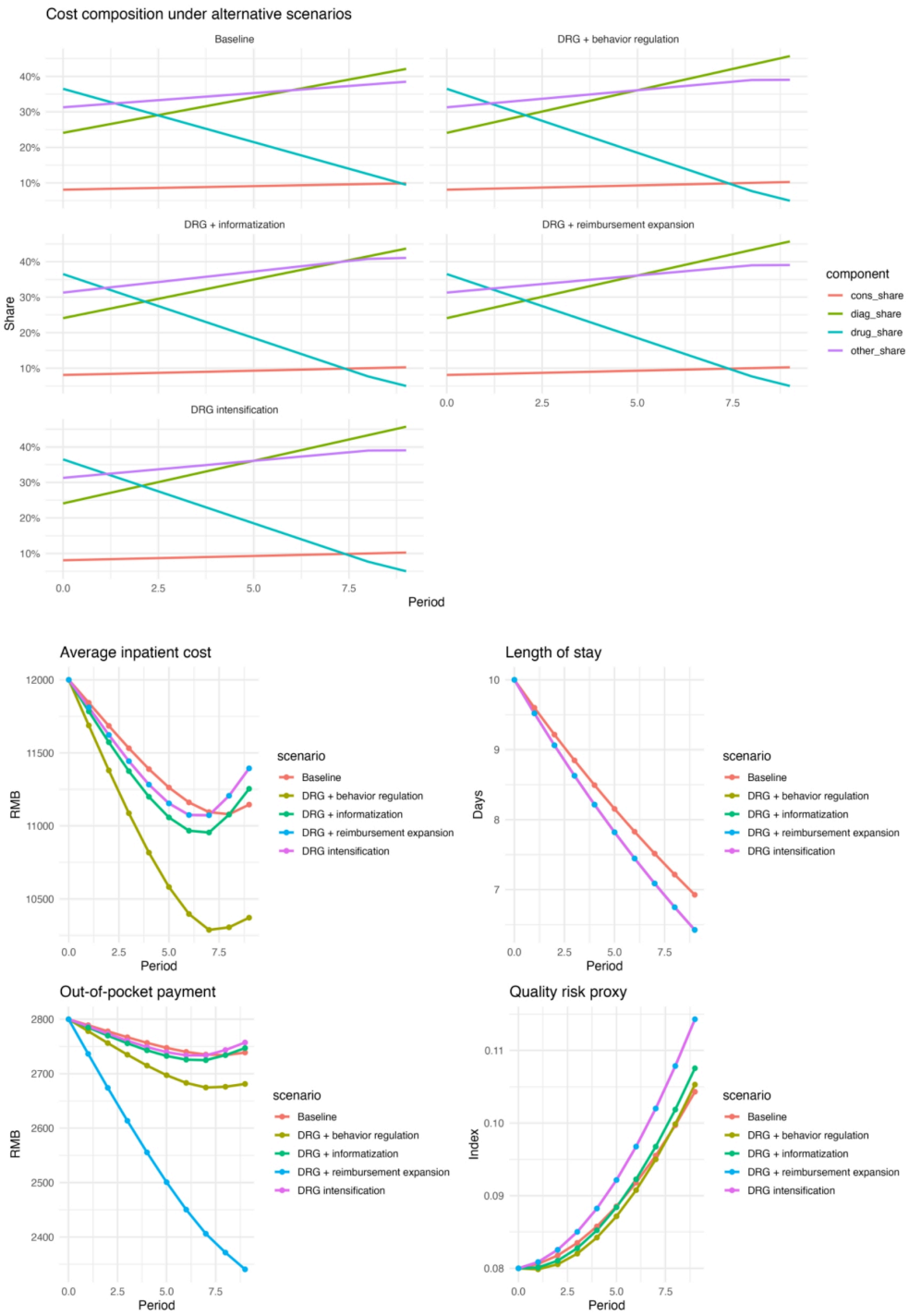


Figure 6

Cost composition under alternative scenarios