Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-8, Issue-3, March 2026
DOI: 10.53469/jrse.2026.08(03).09

Robust Causal Inference and Policy Effect Evaluation and Optimization Based on Large-Scale Medical Insurance Big Data

Sining Chai

Northeastern University, Boston, MA, USA

Abstract: ***With the continuous improvement of the medical security system, large-scale medical insurance big data has become a core resource for insights into the operation of medical services and evaluation of policy implementation effects. However, medical insurance data is plagued by issues such as noise interference and sample selection bias. Traditional causal inference methods struggle to accurately reveal the causal relationship between policies and their effects, hindering the scientific optimization of medical insurance policies. Taking large-scale medical insurance big data as the research object, this paper constructs a causal inference framework of "data preprocessing - model selection - robustness verification", integrating traditional and machine learning methods such as Propensity Score Matching (PSM) and random forest. Combined with a specific case of DRG/DIP payment method reform in a certain city, it conducts effect evaluation from dimensions including medical costs, service utilization, and fund sustainability, and proposes policy optimization strategies such as payment method reform, fund supervision enhancement, and balanced resource allocation based on the evaluation results. The research findings indicate that robust causal inference can effectively improve the accuracy of medical insurance policy effect evaluation, providing reliable data support for the formulation and adjustment of medical insurance policies and contributing to the further improvement of the medical security system.***

Keywords: Medical Insurance Big Data, Causal Inference, Robustness Analysis, Policy Optimization, DRG/DIP Payment Reform.

1. Introduction

By the end of 2024, the participation rate of basic medical insurance in China has remained stably at a high level, and medical insurance data has subsequently experienced explosive growth. Covering multi-dimensional content such as basic information of insured personnel and detailed medical service records, it provides abundant empirical materials for medical insurance policy research. However, medical insurance big data exhibits prominent complexity: disparate data standards across regions lead to high heterogeneity, accompanied by frequent interference from outliers and complex variable correlations. Traditional studies struggle to eliminate the interference of confounding variables, often resulting in significant deviations in evaluation results. For instance, some studies overestimated the cost-control effects of policies due to insufficient consideration of factors such as changes in population structure. Against this backdrop, conducting research from the perspective of medical insurance big data is of great significance. It not only enables the accurate identification of the true effects of policies through more robust causal inference models but also facilitates the formulation of optimization strategies based on scientific evaluation results. This paper conducts research by integrating cutting-edge methods such as data mining, aiming to provide support for the scientific formulation of medical insurance policies.

2. Robust Causal Inference Methods for Large-Scale Medical Insurance Big Data

2.1 Data Collection and Preprocessing

The collection of large-scale medical insurance big data adopts a "multi-source integration" model, with core channels including medical insurance agency databases, hospital information systems (HIS), and third-party platforms. The preprocessing employs a "step-by-step purification + feature enhancement" strategy: first, data is cleaned through "rule-based verification + isolation forest algorithm"; second, data formats are unified in accordance with the HL7 FHIR standard to eliminate coding biases; third, missing values are imputed using multiple imputation methods combined with cross-departmental data; finally, derived features are extracted, and a certain number of core features are selected through L1 regularization to balance computational complexity and information integrity.

2.2 Selection and Application of Causal Inference Models

Targeting the characteristics of medical insurance big data, adaptive improvements are made to three traditional models: Propensity Score Matching (PSM), Difference-in-Differences (DID), and Regression Discontinuity Design (RDD):

1) PSM uses logistic regression to calculate propensity scores, matches samples based on propensity scores, and incorporates stratified balance tests;

2) DID constructs a two-way fixed effects model to control for regional and time fixed effects, with the addition of interaction terms;

3) RDD takes policy thresholds as breakpoints and estimates effects through local linear regression.

Meanwhile, a "multi-model integration framework" is built to fuse machine learning and deep learning: Random Forest-PSM replaces logistic regression, XGBoost-DID screens key variables, and GNN-based causal inference models capture network correlations, effectively enhancing the robustness of causal inference.

2.3 Methods for Robustness Analysis and Verification

Robustness analysis is carried out from two aspects: internal validity and external validity.

For internal verification, three methods are adopted:

1) Placebo test: By randomly assigning policy implementation time or treatment group status 1000 times, if the true estimated value is significantly different from the placebo distribution ($P<0.05$), the effect is reliable. For example, the true cost-control effect of DRG reform falls outside the 99% confidence interval of the placebo distribution;

2) Sensitivity analysis: Adjust key parameters or exclude extreme samples, if the result change range is <10%, the result is robust;

3) Model substitution test: Repeat the evaluation with different models, if the result direction is consistent and the difference is <5%, the result is reliable.

External validity is improved through "stratified analysis + cross-regional verification": stratification by regional economic level, hospital grade, and patient age group. For example, the cost-control effect of DRG reform is higher in high-economic-level regions than in low-level regions; cross-regional verification applies the outpatient mutual aid model of a certain city to another province, with a policy effect difference of <8%; meanwhile, the applicable boundaries are clarified in the context of events such as the pandemic to avoid over-generalization of results, comprehensively ensuring the reliability and applicability of inference results [1, 2, 3].

3. Evaluation of Medical Insurance Policy Effects

3.1 Core Areas and Evaluation Dimensions of Medical Insurance Policies

Current medical insurance policies in China focus on four core areas:

1) Payment method reform (e.g., DRG/DIP, capitation payment), aiming to standardize medical service behaviors and control irrational cost growth;

2) Optimization of benefit guarantees (e.g., outpatient mutual aid, expansion of critical illness insurance), aiming to improve security levels and reduce the burden on patients;

3) Strengthening of fund supervision (e.g., intelligent monitoring, unannounced inspections), aiming to prevent fund loss and ensure sustainability;

4) Cross-regional medical treatment settlement (e.g., direct cross-provincial settlement, national pooling), aiming to improve service convenience and promote fairness.

Based on this, a "three-dimensional and four-level" evaluation index system is constructed:

1) Medical cost dimension: Includes 6 core indicators — average outpatient cost, average inpatient cost, average daily inpatient cost, proportion of drug costs, proportion of consumable costs, and proportion of out-of-catalog costs—used to evaluate the cost-control effect of policies;

2) Service utilization dimension: Includes 5 indicators — outpatient visit rate, hospitalization rate, average length of hospital stay, standardized management rate of chronic diseases, and hierarchical diagnosis and treatment rate—used to evaluate the impact of policies on the accessibility and rationality of medical services;

3) Fund sustainability dimension: Includes 5 indicators — growth rate of fund income, growth rate of fund expenditure, number of months of payable accumulated balance, average fund payment per case, and amount of recovered irregularly used funds—used to evaluate the impact of policies on fund operation;

4) Resident security dimension: Includes 4 indicators — satisfaction of insured personnel, proportion of out-of-pocket expenses, security rate for critically ill patients, and direct settlement rate of cross-regional medical treatment—used to evaluate the improvement effect of policies on residents' security experience.

3.2 Case Evaluation Based on DRG/DIP Payment Reform

3.2.1 Case Background and Data Sources

The DRG/DIP payment reform implemented in a certain city in January 2021 is selected as the case. The policy covers 87 secondary and above hospitals in the city, classifies inpatient cases into 618 DRG groups and 1289 DIP groups, adopts a "total budget management + payment by group" model, and establishes a "retain surpluses and share over expenditures" mechanism. The data sources are the medical insurance big data of the city from 2019 to 2023, including the treatment group (87 hospitals implementing DRG/DIP) and the control group (45 hospitals not implementing it), with a sample size of 10.7 million inpatient records. Core variables include inpatient costs, length of hospital stay, diagnostic codes, and cost structure [4, 5, 6].

3.2.2 Evaluation Process and Result Analysis

The "XGBoost-DID" integrated model is used for evaluation, with control variables including hospital grade, number of beds, doctor-nurse ratio, regional population density, and per capita GDP. The evaluation results are as follows:

1) Medical cost effect: The DRG/DIP reform reduced the average inpatient cost from 12,860 yuan to 11,920 yuan, a decrease of 7.3%. Among them, the proportion of drug costs dropped from 38.2% to 32.5%, and the proportion of consumable costs decreased from 22.1% to 18.3%, indicating a significant cost-control effect. However, the decrease in average costs of tertiary hospitals (9.5%) was higher than that of secondary hospitals (4.2%), reflecting the limited cost-control space for secondary hospitals.

2) Service utilization effect: The average length of hospital

stay shortened from 8.6 days to 7.8 days, a decrease of 9.3%, and the hospitalization rate fell from 15.2% to 14.8%, a decrease of 2.6%. This indicates that the reform has improved the efficiency of medical services and reduced unnecessary hospitalization demand. However, the decrease in the hospitalization rate of chronic disease patients (1.2%) was lower than that of acute disease patients (3.8%), reflecting the strong rigid demand of chronic disease patients.

3) Fund sustainability effect: The average medical insurance fund payment per case decreased from 8,920 yuan to 8,350 yuan, a decrease of 6.4%, the growth rate of fund expenditure dropped from 12.5% to 8.2%, and the number of months of payable accumulated balance increased from 9.8 months to 10.5 months, alleviating the pressure on fund operation. However, the amount of recovered irregularly used funds increased from 12 million yuan before the reform to 15 million yuan, reflecting that hospital irregular behaviors have become more concealed after the reform, requiring strengthened intelligent monitoring.

4) Resident security effect: The satisfaction of insured personnel increased from 82.3 points to 87.5 points, the proportion of out-of-pocket expenses decreased from 29.1% to 27.8%, and the security rate for critically ill patients rose from 89.5% to 92.3%, improving residents' security experience. However, the increase in satisfaction of cross-regional medical treatment patients (3.5 points) was lower than that of local patients (5.2 points), reflecting that the connection between cross-regional settlement and DRG/DIP still needs optimization [7, 8, 9].

3.3 Policy Implications of Evaluation Results

Three key policy implications can be derived from the evaluation results:

1) The cost-control effect of DRG/DIP reform exhibits "hierarchical differences". Differentiated payment standards need to be formulated for secondary hospitals, such as appropriately increasing the payment weight for some common diseases;

2) The improvement of service efficiency driven by the reform depends on the level of hospital informatization. Hospitals that have not realized the automatic matching of electronic medical records and DRG codes only achieved a 3.1% reduction in the average length of hospital stay, necessitating the acceleration of hospital information system upgrades;

3) Fund supervision should focus on "hidden irregularities" such as overcoding/upcoding DRG groups and split admissions. A dual supervision mechanism of "intelligent verification of DRG codes + monitoring of abnormal cost fluctuations" needs to be established.

4. Suggestions for Medical Insurance Policy Optimization

4.1 Optimization Directions Based on Evaluation Results

Combined with the DRG/DIP reform case and other policy evaluation results, the optimization of medical insurance policies should focus on three core directions:

1) Precise payment: Targeting the differences in policy effects among different levels of medical institutions and disease types, break the "one-size-fits-all" payment model, and improve the adaptability of payment methods to the actual needs of medical services;

2) Intelligent supervision: Address the concealed trend of hospital irregular behaviors after the reform, build a full-process monitoring system using medical insurance big data, and realize the transformation from "post-event investigation" to "pre-event early warning and in-event intervention";

3) Balanced security: Solve the security shortcomings of special groups such as cross-regional medical treatment patients and chronic disease patients, narrow the gap in medical insurance benefits between regions and groups, and improve the fairness and accessibility of policies.

4.2 Specific Optimization Strategies and Implementation Paths

4.2.1 Constructing a Hierarchical and Classified Medical Insurance Payment System

To address the "hierarchical differences" in the DRG/DIP reform, a "three-dimensional hierarchical" payment optimization strategy is proposed:

1) Hospital stratification: For tertiary hospitals, refine DRG group classification, separate complex cases to add special DRG groups, and increase the payment weight by 15%-20% to avoid losses; for secondary hospitals, simplify DIP grouping, classify common diseases into basic groups, and increase the payment standard by 5%-8% to ensure reasonable returns; for primary medical institutions, implement "capitation payment + performance incentives", incorporate indicators such as family doctor signing rate, and provide 5% fund subsidies to qualified institutions to guide resource sinking.

2) Disease classification: Establish a special payment mechanism for chronic disease patients, calculate costs according to "basic expenses (fixed payment) + personalized treatment expenses (80% reimbursement)", and coordinate outpatient and inpatient expenses to avoid duplicate payments; for acute disease patients, optimize the time sensitivity of DRG/DIP grouping. For example, adjust the upper limit of the length of hospital stay for "acute myocardial infarction" to 7-10 days, and pay 50% of the average daily cost for exceeding the limit to force efficiency improvement.

3) Regional adaptation: Based on the conclusions of robustness analysis, set the basic payment standard at the provincial level, and dynamically adjust the coefficient according to per capita GDP and medical cost index—1.1-1.2 for central and western rural areas and 0.9-1.0 for eastern coastal areas—to narrow the regional cost-benefit gap and promote the balanced implementation of policies [10, 11, 12].

4.2.2 Building a Full-Process Intelligent Fund Supervision System

To address the "hidden irregularities" after the DRG/DIP reform, a "four-in-one" full-process intelligent fund supervision system is built relying on medical insurance big data:

1) Intelligent coding verification: Connect to the audit system of hospital HIS to conduct real-time verification of the matching between diagnosis and surgical codes. Immediately issue warnings for abnormalities such as "cholecystectomy" corresponding to "lung cancer", and suspend DRG/DIP settlement for institutions with a coding error rate exceeding 3%.

2) Cost fluctuation monitoring: Use models such as ARIMA to predict cost ranges, generate abnormal lists for institutions with a single inpatient cost exceeding ±20% or a monthly average disease cost growth rate exceeding 10%, and construct a three-level indicator system of "institution - department - doctor". Suspend the medical insurance prescription rights of doctors who exceed the standards.

3) Medical behavior tracing: Identify abnormalities based on graph neural networks. For example, patients hospitalized for "coronary heart disease" in 5 hospitals within 3 months will be verified; recover funds and include violators in the blacklist for irregularities.

4) Post-event evaluation: Measure effectiveness by comparing indicators. After the implementation of the system in a certain province, the recovered irregular funds increased from 12 million yuan per year to 35 million yuan, and the irregular rate dropped from 4.2% to 1.8%. Meanwhile, optimize supervision thresholds to improve accuracy.

4.2.3 Promoting the Balanced Allocation of Medical Resources and Security Services

To address issues such as low satisfaction of cross-regional medical treatment patients and insufficient chronic disease management, a "two-dimensional balance" optimization strategy is proposed:

1) Balanced allocation of medical resources: Combined with the conclusion that the improvement of hierarchical diagnosis and treatment rate is not obvious, establish a linkage mechanism between medical insurance fund allocation and medical resource allocation. Link the proportion of medical insurance fund allocation for tertiary hospitals to the primary referral rate—increase by 5% for those with a referral rate of over 30% and decrease by 3% for those failing to meet the standard; provide 50-100 million yuan of medical insurance fund subsidies to support the construction of regional medical centers in central and western regions, attracting high-quality resources to flow to underdeveloped areas.

2) Balanced coverage of security services: Optimize direct settlement of cross-regional medical treatment, increase the number of special outpatient diseases for direct cross-provincial settlement from 33 to 50, and implement a dual-channel model of "online independent registration + offline assisted registration" to improve the registration success rate. For chronic disease patients, build an "Internet + medical insurance" service model, supporting online follow-up visits, prescription circulation, and drug delivery with the same reimbursement ratio as offline services. Family doctors can receive 10-20 yuan of medical insurance subsidies for each effective online follow-up visit [13, 14, 15].

5. Conclusion and Outlook

Taking medical insurance big data as the core, this study constructs a causal inference framework of "data preprocessing - model selection - robustness verification", integrating traditional models such as PSM and DID with machine learning models such as random forest and GNN. This solves problems such as heterogeneity in medical insurance data and improves the accuracy of policy effect evaluation. The evaluation of the DRG/DIP reform case in a certain city shows that this framework can accurately identify the true effects of policies in aspects such as cost control and reveal the heterogeneous characteristics of policy effects. The proposed optimization strategies provide a feasible path for the adjustment of medical insurance policies.

The research has limitations: the data does not fully cover social factors, and the model has insufficient adaptability to public health emergencies such as the pandemic. In the future, data sources can be expanded to build multi-dimensional models, dynamic algorithms can be developed to improve adaptability to special scenarios, international comparative studies can be conducted to learn from experience, and the medical insurance policy evaluation and optimization system can be improved to promote the high-quality development of medical security.

References

[1] Raita Y, Camargo Jr C A, Liang L, et al. Big data, data science, and causal inference: a primer for clinicians[J]. Frontiers in Medicine, 2021, 8: 678047.
[2] Xu W. Recent Theoretical Advances and Emerging Trends in Empirical Research on the Difference-in-Differences Method [J]. Guangdong Social Sciences, 2023(5):51-62.
[3] Kim D W. Medicine General & Health Policy[J]. J Korean Med Sci, 2025, 40(8): e110.
[4] Deng Q, Wang C, Jin J, et al. Empirical study on the operational effects of DIP pilot cities based on the difference-in-differences method [J]. China Hospital Management, 2023,43(06):6-9.
[5] Sanchez P, Voisey J P, Ta T, et al. Causal machine learning for healthcare and precision medicine[J]. Royal Society Open Science, 2022, 9(8): 220638.
[6] Ye Z, Zhang Z, Zhang D J, et al. Deep Learning-Based Causal Inference for Large-Scale Combinatorial Experiments: Theory and Empirical Evidence[J]. Management Science, 2025.
[7] Wu Q. Exploration of the Integration Path and Model of Medical Consumables Management System under DRG/DIP Payment Reform [J]. China Hospital, 2025,29(12):30-32.
[8] Yang W, Yi D, Zhou X H, et al. Translational analysis of data science and causal learning in real-world clinical

evaluation of traditional Chinese medicine[J]. Science of Traditional Chinese Medicine, 2024, 2(1): 57-65.

[9] Yao L, Chu Z, Li S, et al. A survey on causal inference[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2021, 15(5): 1-46.

[10] Tang L, Li C. Research on the Effect of Inpatient Medical Cost Control under the DRGs Medical Insurance Payment Reform [J]. Journal of Fujian Medical University (Social Sciences Edition), 2024,25(05):15-20.

[11] Li T. An Empirical Study on the Impact of Reimbursement Policy Adjustment for Out-of-Town Medical Treatment on Patients' Medical Insurance Payment Costs: Based on Data from a Tertiary General Hospital in Sichuan Province [J/OL]. Health Soft Science, 1-7[2025-12-05].

[12] Gao S, Wang X, Lu Y, et al. Current scenario and challenges of clinical pharmacists to implement pharmaceutical care in DRG/DIP payment hospitals in China: a qualitative interview study[J]. Frontiers in Public Health, 2024, 12: 1339504.

[13] Wang X, Tao Y, Gao S, et al. Effects of DRG/DIP payment reform on hospital pharmacy administration and pharmaceutical services in China: a multicenter cross-sectional study[J]. Frontiers in Public Health, 2025, 13: 1585279.

[14] Zhou T, Zhang F, Wang Y, et al. Visualized analyses of the research on DRG and DIP in China[C]//Proceedings of the 2023 8th international conference on intelligent information processing. 2023: 106-113.

[15] Cui S, Dai R, Huang W, et al. Comparing DRG and DIP: Analysis of Bundled-Payment Healthcare Schemes in China[J]. Available at SSRN, 2024.