Journal of Electronics and Information Science (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2025.100220
ISSN 2371-9524 Vol. 10 Num. 2

Causal Optimization Model for Balanced Allocation of Medical Resources and Analysis of Big Data-Driven Robust Decision Support

Sining Chai

Northeastern University, Boston, MA, USA

Keywords: Balanced Allocation of Medical Resources; Causal Optimization Model; Big Data; Robust Decision-Making; Health Management

Abstract: The balanced allocation of medical resources is a core measure to address the issues of "difficulty in accessing medical care and high medical costs" and a key proposition for advancing the Healthy China initiative. Traditional allocation models, which rely on experiential decision-making and correlation analysis, struggle to accurately identify the causal relationship between resource supply and health needs, resulting in insufficient allocation efficiency and fairness. Centering on causal inference and robust optimization theory, combined with the multi-dimensional enabling characteristics of big data technology, this paper systematically reviews the construction logic and core methods of causal optimization models for balanced medical resource allocation, as well as the implementation path of a big data-driven robust decision support system. Following the logical framework of "causal identification - model optimization - decision implementation", the study analyzes the adaptive scenarios of different causal models in resource allocation, explores the application value of big data technology in enhancing decision robustness, and finally points out the current research bottlenecks and future development directions. It aims to provide theoretical references for the scientificization and precision of medical resource allocation.

1. Introduction

As the core of the public service system, the balanced allocation of medical resources is closely related to national health rights and social equity. China faces prominent problems of insufficient total medical resources and structural imbalance: high-quality resources are overly concentrated in large hospitals in major cities, while grassroots and remote areas suffer from resource scarcity. This forms a "siphon effect" and "service gap", which not only reduces medical accessibility but also exacerbates cost increases, hindering the advancement of the hierarchical medical system. Under the Healthy China strategy, resource allocation urgently needs to shift from "scale expansion" to "quality and structural optimization", establishing a scientific model and data-driven decision system. Traditional allocation relies on basic indicators such as population and geography or expert experience, with obvious limitations: it only focuses on variable correlations rather than causality, easily confuses "number of hospital visits" with "actual health needs", and ignores driving factors such as aging; it has low tolerance for uncertainty, leading to passive decision-making in scenarios

such as public health emergencies. In contrast, big data technology integrates multi-source data including electronic health records, and when combined with causal optimization models and robust decision theory, can achieve a leap from "data correlation" to "causal insight", providing support for solving these dilemmas. This paper focuses on core issues such as model construction and data fusion to offer theoretical references for scientific decision-making.

2. Core Contradictions in Balanced Allocation of Medical Resources and the Need for Causal Cognition

2.1 Analysis of Core Contradictions in Medical Resource Allocation

The core contradictions in China's medical resource allocation manifest in three aspects: structural imbalance, low efficiency, and lack of fairness. Structural imbalance presents "three-dimensional characteristics": at the hierarchical level, tertiary hospitals concentrate over 40% of high-quality resources, while grassroots institutions account for less than 20%, resulting in both "minor illnesses treated in large hospitals" and idleness of grassroots resources; at the regional level, the number of hospital beds per thousand people in eastern regions is 2.3 times that in western regions, with high-quality resources aggregating in urban agglomerations; at the category level, specialized medical resources are surplus, while public health resources are insufficient, disconnected from the disease spectrum associated with aging. Efficiency issues stem from mismatches between supply and demand: high-end equipment in large hospitals operates at less than 50% capacity yet outpatient services are overloaded, while grassroots services only account for 35% of the national total. The root cause lies in the failure to accurately identify the causal relationship between "resources and outcomes" and neglect of hidden factors such as the "siphon effect". The lack of fairness is reflected in disparities in medical accessibility: the average distance to medical facilities in rural areas and waiting times for high-quality services in central and western regions are several times those in cities. Traditional allocation models fail to consider differentiated needs, leading to a disconnect between "formal fairness" and "substantive fairness" [1, 2, 3].

2.2 Core Value of Causal Cognition for Resource Allocation

The key to resolving contradictions in medical resource allocation lies in shifting from "correlation analysis" to "causal inference" to clarify the necessary connection between "causes" and "effects". Correlation can only identify statistical associations—for example, "increased resource investment" and "improved health levels" may both be influenced by economic development rather than direct causality. In contrast, causal inference can isolate confounding factors, accurately identifying the net effect of "resource investment" on "health outcomes" to inform decision-making. Its value is reflected in three aspects: first, accurately locating demand targets—for instance, causal models can confirm that unmet medical needs in rural western areas stem from resource scarcity, while insufficient demand in urban communities in eastern regions results from weak health awareness, enabling "allocation based on actual needs"; second, optimizing resource structure—through the causal chain of "resources - diseases - health", determining the optimal intervention effect of general practitioners in chronic disease management to avoid mismatches; third, predicting policy effects—simulating the impact of policies such as "increasing grassroots physician staffing" to proactively mitigate risks.

3. Construction Logic and Core Methods of Causal Optimization Models for Balanced Allocation of Medical Resources

As an integration of causal inference and optimization theory, causal optimization models follow

a core logic: first, identify the causal relationship between variables in medical resource allocation through causal inference to clarify the net effect of "resource investment" on "health outcomes"; then, it constructs an objective function with this causal relationship as a constraint, and solve for the optimal allocation scheme. Based on different causal identification methods, these models are classified into three types: those based on the potential outcomes framework, structural causal models, and double/debiased machine learning.

3.1 Model Based on the Potential Outcomes Framework

The model based on the Potential Outcomes Framework (Rubin Causal Model, RCM) centers on classic causal inference theory. It identifies causal effects by comparing potential outcome differences between the "treatment group" (regions/populations receiving resource investment) and the "control group" (regions/populations not receiving resource investment). In medical resource allocation, "treatment" typically refers to resource investment behaviors such as increasing the number of hospital beds, while "outcomes" correspond to health outputs such as reduced morbidity. By constructing counterfactual scenarios, the model addresses the core challenge of "being unable to observe two treatment outcomes for the same subject simultaneously". Model construction involves three steps: first, defining variables—treatment variables quantify the intensity of resource investment (e.g., increment in hospital beds per thousand people), while outcome variables consider both efficiency and fairness (e.g., medical accessibility index); second, selecting methods such as Propensity Score Matching (PSM) to construct the control group, controlling for confounding variables such as economic level by matching samples with similar scores; third, integrating causal effects and resource constraints to construct an objective function and solve for the optimal allocation ratio. This model is suitable for comparative optimization of inter-regional resource allocation, with advantages of clear causal logic, strong interpretability, low computational cost, and ease of practical application. Its limitations include insufficient matching accuracy for continuous treatment variables, difficulty in characterizing multi-factor interactions, and sensitivity to sample distribution [4, 5, 6].

3.2 Model Based on the Structural Causal Model

The model based on the Structural Causal Model (SCM) intuitively presents variable relationships such as "economic level → resource investment → health outcomes" through causal graphs. It effectively addresses multi-variable causal problems that are challenging for the potential outcomes framework by simulating intervention effects using the Do-operator. Compared with the latter, SCM focuses more on variable causal structures than sample matching, enabling clear identification of direct and indirect causal effects to support precise allocation. The core of model construction lies in causal graph drawing and intervention effect calculation: first, determining variables related to resource supply (e.g., number of hospital beds), demand (e.g., morbidity), confounding factors (e.g., GDP), and outcomes (e.g., healthy life expectancy) based on literature and expert consensus, and drawing causal graphs to clarify variable directions using domain knowledge; second, identifying confounding paths through the backdoor/frontdoor criterion and adjusting for variables such as GDP to eliminate interference; finally, constructing a multi-objective optimization model based on intervention effects calculated by the Do-operator, maximizing health outcomes and fairness under resource constraints. This model is adaptable to complex scenarios, capable of identifying the dual effects of resource investment through multi-variable causal graphs and proposing efficient combined schemes such as "transportation subsidies + grassroots resource investment". Its limitations include reliance on expert experience for causal graph construction, increased difficulty in structure identification with more variables, and high requirements for data quality.

3.3 Model Based on Double/Debiased Machine Learning

The model based on Double/Debiased Machine Learning (DML) is an emerging method in recent years. Its core is decomposing causal effect estimation into two independent machine learning tasks: "outcome prediction" and "treatment assignment prediction". It effectively addresses confounding variable control in high-dimensional data, reducing the risk of model specification bias. Particularly suitable for multi-source high-dimensional data scenarios in medical resource allocation, it breaks through the "curse of dimensionality" bottleneck of traditional models. The model follows a "debiasing + optimization" logic: in the first stage, using algorithms such as random forests and neural networks to separately predict outcome variables (e.g., health outcomes) and treatment variables (e.g., resource investment), obtaining residual terms that eliminate the impact of confounding variables; in the second stage, estimating causal effects free from interference through linear regression based on these residuals; in the third stage, using this effect as a constraint to construct an objective function for "cost minimization" or "benefit maximization", and solving for the optimal scheme using algorithms such as integer programming. Relevant studies, based on multi-source provincial and municipal data containing 87 feature variables, identified that general practitioners have a more significant intervention effect on chronic disease management through DML. Based on this, a scheme of "directing 70% of new physician staffing to general practice" was proposed, which can improve chronic disease control rates and reduce costs. The model's advantages include adaptability to high-dimensional data, high accuracy in causal estimation, strong resistance to model misspecification, and support for multi-resource type analysis. Its limitations are high model complexity, high computational cost, weak result interpretability, and the need for maintenance by professional algorithm personnel [7, 8].

3.4 Comparison and Adaptive Scenarios of the Three Models

Table 1: Comparison of Medical Resource Allocation Causal Optimization Models and Their Applicable Scenarios

Model Type	Theoretical Basis	Causal Identification Method	Data Requirements	Core Advantages	Main Limitations	Adaptive Scenarios
Model based on the Potential Outcomes Framework	Rubin Causal Model, counterfactual inference theory	Sample matching methods such as Propensity Score Matching, entropy balancing, and Coarsened Exact Matching (CEM) to control confounding variables	Low dimensionality (≤20 dimensions), relatively balanced sample distribution, observable confounding variables	Clear causal identification logic, strong result interpretability, low computational cost, easy practical application	Insufficient matching accuracy for continuous treatment variables, difficulty in characterizing multi-factor interactions, sensitivity to sample distribution	Inter-regional resource allocation comparison, single resource investment effect evaluation, decision scenarios with simple data dimensions
Model based on the Structural Causal Model	Pearl's causal graph theory, Do operator intervention theory	Constructin-g causal graphs, identifying confounding paths through backdoor/ frontdoor criteria, calculating direct/ indirect causal effects	Causal relationships between variables can be defined through domain knowledge, high requirements for data integrity, clear variable association directions required	Clear display of multi-variable causal structures, distinction between direct/indirect effects, support for precise intervention strategy design	Reliance on expert experience for causal graph construction, increased difficulty in structure identification with more variables, sensitivity to data quality	Resource allocation under multi-factor interaction, scenarios requiring clear intervention paths (e.g., analysis of the "resource - accessibility - health" chain)
Model based on Double/ Debiased Machine Learning	Machine learning theory, causal inference debiasing ideology	Dual models predicting outcome and treatment variables separately, estimating causal effects through residual regression to isolate high-dimensional confounding impacts	Supports high-dimensional data (≥50 dimensions), handles multi-type feature variables, requires large sample size (≥1000 samples)	Adaptability to high-dimensional big data scenarios, high accuracy in causal estimation, strong resistance to model misspecification, support for multi-resource type analysis	High model complexity, high computational cost, weak result interpretability, need for maintenance by professional algorithm personnel	Resource allocation with multi-source big data fusion, multi-feature scenarios such as chronic disease management, complex resource combination optimization

There are significant differences among these three types of causal optimization models in terms of their theoretical foundations, core advantages, and application scenarios. A comparative analysis of these models is presented in Table 1.

4. Empowerment of Big Data for Causal Optimization Models and Construction of Robust Decision Support Systems

The accuracy and practicality of causal optimization models depend on data support. The development of big data technology provides "full-dimensional, real-time, and fine-grained" data sources for the models. Meanwhile, through technologies such as data cleaning, feature engineering, and uncertainty analysis, it enhances model robustness, constructing a closed-loop system of "data - model - decision".

4.1 Big Data Sources and Feature Extraction for Medical Resource Allocation

Big data for medical resource allocation exhibits "multi-source and heterogeneous" characteristics, mainly covering four categories: medical service data, medical insurance settlement data, public health data, and socio-economic data. The realization of big data value relies on effective feature extraction. Core features in medical resource allocation scenarios are divided into three types: demand features, supply features, and correlation features. The key to feature extraction lies in "denoising" and "dimensionality reduction"—processing noise through methods such as missing value imputation and reducing dimensionality via techniques like principal component analysis to provide high-quality inputs for causal optimization models.

4.2 Paths to Enhancing Robustness of Big Data-Driven Causal Optimization Models

Robustness refers to a model's ability to maintain stable outputs under uncertainties such as data noise, and is a core requirement for medical resource allocation decisions that need to respond to unexpected situations like public health emergencies. Big data technology enhances model robustness through "data augmentation", "uncertainty quantification", and "dynamic updating". Data augmentation expands the sample space—processing imbalanced samples from remote areas using the SMOTE algorithm, or generating simulated data for scenarios such as sudden epidemics with Generative Adversarial Networks (GANs), and training the model with fused real data to reduce overfitting risks. Uncertainty quantification identifies data uncertainties such as demand fluctuations and model uncertainties such as causal effect estimation errors through methods like Monte Carlo simulation and Bayesian inference, clarifying the confidence intervals of decision results to optimize scheme stability. Dynamic updating relies on stream computing to process real-time data such as medical services, monitoring allocation effects such as consultation rates, and updating model parameters through online learning to realize a closed loop of "decision - implementation - feedback - optimization". For example, a central hospital in a city shortened the response time for medical resource allocation from 24 hours to 8 hours using this approach [9, 10].

4.3 Construction and Application of Big Data-Driven Robust Decision Support Systems

Based on causal optimization models and big data technology, a robust decision support system for the balanced allocation of medical resources with four layers—"data layer, model layer, decision layer, and feedback layer"—can be constructed to achieve full-process empowerment. The data layer serves as the foundation: integrating multi-source data to build a data warehouse through data middle platform technology, and completing processing such as cleaning and desensitization with data

governance tools to break "data silos" and provide high-quality data services. The model layer is the core, including causal identification and optimization solution modules: the former selects adaptive models to estimate the causal effects of resource investment, while the latter constructs multi-objective functions with these effects as constraints and solves for optimal schemes using intelligent algorithms, with built-in modules to evaluate scheme stability. The decision layer is responsible for result transformation: displaying allocation ratios through visualization platforms to provide personalized support for different stakeholders, with integrated risk early warning functions. The feedback layer ensures the closed loop: collecting real-time post-implementation data to evaluate effects, and iterating the model by incorporating expert and public opinions. This system has been applied in multiple regions: identifying causal effects through multi-source data and models, proposing resource allocation schemes, which have increased grassroots consultation rates and reduced cross-regional medical visits after implementation, demonstrating practical value.

5. Conclusion

The balanced allocation of medical resources is a core path to achieving health equity, and the integration of causal optimization models with big data technology provides a scientific tool for addressing resource imbalance dilemmas. This review shows that the potential outcomes framework is suitable for causal identification in simple scenarios, the structural causal model excels in analyzing complex causal structures, and double/debiased machine learning performs prominently in high-dimensional data. Big data ensures model accuracy and practicality through empowerment, efficiency improvement, and dynamic updating, constructing a "data - model - decision - feedback" closed loop that shifts allocation from "experience-driven" to "data and model dual-driven". Current research still faces bottlenecks such as complex causal identification, insufficient data quality, and disconnection between models and practice. Future efforts need to break these limitations through technological innovation. With the development of related technologies, decision-making for medical resource allocation will become more accurate and efficient, providing guarantees for the Healthy China initiative and helping to achieve the goal of "universal access to equitable and high-quality medical services".

References

[1] Dai T, Yuan J, Dai K, et al. Unequilibrium evolution and driving mechanism of medical resource allocation: An empirical study based on the SBM-Dagum model[J]. Health Economics Research, 2025, 42(09): 48-52+57.

[2] Atento R G, Quinto L, Espelita C A M, et al. Integrating Business and Health Analytics: A Conceptual Framework for Dual Outcomes in Healthcare[J]. International Journal of Health & Business Analytics, 2025, 1(1).

[3] Keya K N, Islam R, Pan S, et al. Equitable allocation of healthcare resources with fair survival models [C]//Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). Society for Industrial and Applied Mathematics, 2021: 190-198.

[4] Zhang C. Equitable resource allocation in health emergencies: addressing racial disparities and ethical dilemmas[J]. Journal of Medical Ethics, 2024.

[5] Liu Y, Zhao Y, Chen S, et al. Research on the allocation of medical resources and service utilization in TCM hospitals in China based on the coupling coordination model[J]. Modern Preventive Medicine, 2024, 51(22): 4147-4152, 4158.

[6] Martinez S, Al-Mansoori N. Optimizing Resource Distribution in Healthcare: A Framework for Equitable Allocation [J]. Nvpubhouse Library for Journal of Social Sciences and Humanities Research Fundamentals, 2025, 5(08): 1-13.

[7] Sun Y, Wu S, Cao Z. Research on regional differences and spatiotemporal evolution of the fairness of high-quality medical resource allocation in China[J]. Chinese Hospitals, 2024, 28(12): 29-35.

[8] Li J, Wu Y, Lu Y. Analysis of medical resources for allocation equity using traditional Chinese medicine resource as a model[J]. The International Journal of Health Planning and Management, 2022, 37(6): 3205-3217.

[9] Zhang Q, Ouyang Y. Research on the coupling coordination relationship between China's multi-level medical security and medical resource allocation[J]. Chinese Journal of Health Policy, 2025, 18(09): 48-56.

[10] Li G, Feng C, Zhang T, et al. Spatially Equitable Allocation of Medical Resources for Pandemic Containment: A Service Level-Based Approach[J]. Transportation Research Record, 2025: 03611981251359295.