Research On Prevention and Control Mode and Governance Mechanism of Chronic Diseases Based on Medical Statistics

Yirui Sun

Universiti Kebangsaan Malaysia, Bangi, Selangor 43000, Malaysia

y26294541@gmail.com

Abstract. Faced with the triple challenges of high incidence, common comorbidity and heavy economic burden of chronic diseases in China, the traditional passive mode of "hospital-centered" is unsustainable. This study puts forward a research framework of "medical statistical data mining-dynamic prevention and control model-multi-collaborative governance", and systematically answers three scientific questions: "how to extract key features, how to achieve accurate stratification, and how to ensure sustainable governance". Based on more than 45,000 longitudinal queues of multi-center HIS/LIS/EHR, wearable devices and environmental monitoring, cross-domain data fusion is completed by using differential privacy (ε=1.5) and federated learning; Combined with random forest and XGBoost, the risk prediction model was established (AUC=0.82), and five heterogeneous groups were excavated by Apriori and K-means (35.7% in the elderly comorbid disability group and 18.5% in the environmentally sensitive hypertension group). Based on this, a data-driven prevention and control model of "accurate prediction-graded intervention-effect closed loop" is designed, and the long-term cost-effectiveness of 12 governance strategies is quantitatively compared on PEST-SD simulation platform. Further build a blockchain data sharing and performance compensation mechanism of "government-led, departmental collaboration, social participation and personal responsibility" to realize the governance paradigm transformation from "one-way decision-making by the government" to "multi-subject collaboration" and from "intervention after the event" to "full management". The research results provide a replicable Chinese plan for the reform of health insurance per head, the construction of healthy cities and the implementation of healthy China strategy.

Keywords: prevention and control mode; governance mechanism; chronic diseases; medical statistics.

1. Introduction

Global chronic non-communicable diseases (NCDs) have become the "first killer" threatening human health. What is more serious is that China's chronic diseases are characterized by "three highs", the incidence rate is rising continuously, comorbidity is common, and the economic burden is heavy. This situation poses a severe challenge to the traditional "hospital-centered" passive medical model, and it is urgent to build a data-driven active prevention and control system.

There are "data islands" among medical institutions, public health departments and community health management systems, which makes it difficult to integrate and analyze diagnosis and treatment data, behavior data and policy data. The governance mechanism lags behind, and the current prevention and control strategies mostly adopt a "one size fits all" model, which fails to achieve risk stratification and accurate resource allocation based on medical statistics [1]. Because of the lack of dynamic data support, first-class medical institutions still rely on empirical management, while large urban hospitals waste resources because of excessive diagnosis and treatment.

As a "new oil" for the prevention and control of chronic diseases, medical statistical data has not been fully tapped. On the one hand, big data technology can reveal the hidden law of disease occurrence and development; On the other hand, blockchain technology can build a security framework for cross-institutional data sharing. More importantly, the data-driven prevention and control model can promote the transformation of governance paradigm: from "one-way government decision-making" to "multi-subject coordination" and from "intervention after the event" to "whole process management".

By constructing the research framework of "medical statistical data mining-dynamic prevention and control model design-multiple governance mechanism innovation", this study answers three core questions: (1) How to extract the key features of chronic disease prevention and control from massive medical data? (2) How can the data-driven hierarchical prevention and control model be accurate? (3) How can the inter-departmental collaborative governance mechanism ensure sustainability? The research results will provide theoretical support and practical reference for optimizing the allocation of chronic disease prevention and control resources and improving the implementation path of healthy China strategy, and at the same time contribute to China's plan for global chronic disease management.

2. Research methods and data sources

2.1. Mixed research method

A hybrid research path combining longitudinal queue analysis, machine learning modeling and policy simulation is adopted. Firstly, based on retrospective cohort data, the key risk factors and epidemic rules of chronic diseases were identified by statistical mining technology. Secondly, the machine learning algorithm random forest and XGBoost are used to build the disease risk prediction model, and the generalization ability of the model is ensured by external verification. Finally, the macro adaptability of prevention and control strategies is qualitatively evaluated by PEST model (policy, economic, social and technical environment analysis), and the long-term effects of different control strategies are simulated by system dynamics simulation. This method combination can not only quantify the disease risk, but also analyze the complexity of multi-agent collaborative governance, avoiding the limitations of a single method.

2.2. data source

The data source is mainly multi-center medical database, supplemented by environmental and behavioral data to ensure the multi-dimensional and dynamic data. Extract structured data from hospital information systems (HIS), laboratory information systems (LIS), and electronic health records (EHRs), including physical examination indicators (BMI, blood pressure, blood glucose), diagnosis and treatment records, medication history, etc. Integrate environmental data (PM2.5, temperature) from meteorological departments and individual behavior data collected by wearable devices to depict the interaction of environment-behavior-disease. Aiming at the missing values, abnormal values and privacy problems of medical data, differential privacy (ε=1.5) and federal learning technology are used for anonymization to ensure compliance. The data is stored in the enterprise data warehouse (EDW), and the time-series alignment technology is used to solve the time-space asynchronous problem of multi-source data.

3. Core research content

3.1. Analysis on the characteristics of medical statistical data of chronic diseases

Through in-depth mining of multi-source heterogeneous medical statistical data, the epidemiological characteristics, changing trends and key influencing factors of chronic diseases are systematically revealed, providing accurate "targets" for the formulation of prevention and control strategies.

Based on the data of national health service survey, the long-term trend and population distribution of chronic diseases were analyzed. The data shows that the prevalence rate of chronic diseases in China has increased from 15.74% in 1998 to 34.29% in 2018, and the prevalence rate of people aged 65 and over is as high as 62.33%. The research focuses on the incidence rate, mortality and disease burden (DALYs) of major chronic diseases that cause 88.5% of the total deaths in China, including cardiovascular and cerebrovascular diseases, cancer, chronic respiratory diseases and diabetes.

Using regional survey data, this paper analyzes the influence of behavioral factors (smoking, drinking, unhealthy diet and lack of exercise) and socio-economic factors (income, education and medical accessibility) on chronic diseases. For example, the survey in Xinfeng County shows that the smoking rate of residents aged 18 and above is 19.8%, and that of men is as high as 39.70%, while the survey in Honggutan District found that the daily intake of salt and cooking oil per capita was 7.60 grams and 45.81 grams, which far exceeded the recommended standards. The risk contribution of each factor is quantified by correlation analysis and regression model (see Table 1 and Table 2).

Table 1 Correlation between behavioral factors and hypertension risk and regression analysis results (combined samples, n=9,800)

Variable	Spearman correlation coefficient (r)	P-value	logistic regression OR (95%CI)	P-value	Standardized regression coefficient (β)
Smoking (Yes=1)	0.28	<0.001	1.62 (1.45–1.81)	<0.001	0.48
Daily average salt intake (grams)	0.35	<0.001	1.12 (1.08–1.16)	<0.001	0.11
Daily average intake of cooking oil (grams)	0.22	0.002	1.05 (1.02–1.08)	0.001	0.05
Lack of exercise (Yes=1)	0.31	<0.001	1.75 (1.58–1.94)	<0.001	0.56
Control variable					
Age (every 10 years increase)	0.45	<0.001	2.10 (1.95–2.26)	<0.001	0.74
Male (Yes=1)	0.30	<0.001	1.85 (1.68–2.04)	<0.001	0.61

Table 2 Correlation between socio-economic factors and hypertension risk and regression analysis results (combined samples, n=9,800)

Variable	Variable	Spearman correlation coefficient (r)	P-value	logistic regression OR (95%CI)	P-value
Low income (<5000 yuan/month)	0.25	<0.001	1.40 (1.25–1.57)	<0.001	0.34
Low education (<high school)	0.20	<0.001	1.30 (1.18–1.43)	<0.001	0.26
Poor medical accessibility (Yes=1)	0.18	0.003	1.25 (1.12–1.40)	<0.001	0.22
Control variable					
Age (every 10 years increase)	0.45	<0.001	2.10 (1.95–2.26)	<0.001	0.74
Male (Yes=1)	0.30	<0.001	1.85 (1.68–2.04)	<0.001	0.61
Model fitting indicators	—	—	AUC=0.82	—	Nagelkerke R^2=0.38

The dominant behavioral factors (Table 1), lack of exercise (OR=1.75, β=0.56) and smoking (OR=1.62) are the primary behavioral risks, which directly correspond to the high smoking rate in Xinfeng County (39.70%) and the static lifestyle in Honggutan District (urban commuting dependence). The risk increased by 12%(OR=1.12) for each gram of salt intake, which proved that the data of 7.60 grams/day in Honggutan District had strong public health significance (r=0.35). Priority should be given to intervention in sports facilities coverage (such as community fitness spots) and salt reduction (food industry standards+consumer education).

The deep role of socio-economic factors (Table 2), low income (OR=1.40) and poor medical accessibility (OR=1.25) have low OR values, but they indirectly amplify risks through behavioral factors. Education level independently affects the risk (OR=1.30), reflecting the key role of health literacy-the salt intake of people with high education in Honggutan District is 23% lower than those with low education. Inter-departmental cooperation (health+civil affairs+urban construction) is needed, such as adding mobile medical vehicles in rural areas of Xinfeng County (improving accessibility) and simultaneously binding behavioral intervention (providing smoking cessation guidance with the vehicle).

Construct a "chronic disease data lake" covering clinical indicators such as blood pressure, blood sugar, BMI, behavioral data and environmental data. Apriori algorithm was used to analyze association rules to explore the comorbidity pattern (Table 3), and K-means cluster analysis was used to segment patients, thus extracting key prevention and control features (Table 4).

Table 3 Mining results of common chronic disease comorbidity patterns based on association rule analysis

Antecedent	Consequent	Support degree	Confidence level	Lift	Interpretation of Clinical Significance
{Type 2 diabetes, hypertension}	{coronary heart disease}	0.18	0.73	2.45	Metabolic cardiovascular axis comorbidity requires enhanced cardiorenal protection
{Obesity (BMI≥28), dyslipidemia}	{Nonalcoholic fatty liver}	0.22	0.68	3.12	The core phenotype of metabolic syndrome, with the liver as a key target organ
{chronic obstructive pulmonary disease, age > 65 years old}	{cor pulmonale}	0.12	0.81	4.67	Risk of respiratory circulatory system linkage in the elderly
{Long-term exposure to PM2.5, hypertension}	{Stroke}	0.09	0.56	1.89	Environment gene interaction increases vascular fragility
{Insomnia, anxiety state}	{Type 2 diabetes}	0.15	0.42	1.56	Potential pathways for the transformation of psychological and behavioral factors into metabolic disorders

Note: The minimum support threshold is set to 0.08, and the minimum confidence threshold is set to 0.40. The data comes from the EHRs queue of a prefecture-level city from 2019 to 2023, and the sample size is n=45,682.

Table 4 Risk stratification characteristics of patients with chronic diseases based on K-means clustering

Cluster labels	Population proportion (%)	Core clinical features (mean)	Key risk factors	Intervention priority	Suggested governance strategies
Category 1: Young metabolic risk group	28.3	Age: 38 ± 6 years old; BMI: 27.5; HbA1c: 6.8%; Blood pressure: 135/85 mmHg	Sedentary behavior, intake of sugary drinks, work stress	High	Digital health intervention+workplace health management
Category 2: Elderly comorbidities and disability group	35.7	Age: 71 ± 8 years old; BMI: 24.2; Comorbidity: 3.2 species; Weakness index: 0.31	Multi drug sharing, sarcopenia, social isolation	Very High	Integrated care model+community rehabilitation
Category 3: Environment sensitive hypertension group	18.5	Age: 52 ± 10 years old; Blood pressure: 158/96 mmHg; PM2.5 exposure: 45 μ g/m³	Air pollution, high salt diet, low compliance	Medium	Environmental governance linkage+strengthening patient education
Category 4: Genetic high cholesterol group	12.4	Age: 45 ± 12 years old; LDL-C: 5.8 mmol/L; ApoB: 1.42 g/L	Family genetic history, low-density lipoprotein resistance	High	Precision medication (PCSK9 inhibitors)+genetic screening
Category 5: Health Maintenance Group	5.1	Age: 42 ± 9 years old; BMI: 22.8; All indicators are within the normal range	No significant risk factors	Low	Routine follow-up+health promotion

Note: The number of clusters K=5 was validated by silhouette coefficient (Silhouette Score=0.68), and the data was sourced from the HIS and wearable device fusion dataset, n=12,458.

Association rules reveal the non-randomness of the occurrence and development of chronic diseases, and the strong association rules with promotion degree > 1.8 mainly focus on the metabolic-cardiovascular axis, which confirms the theory of "common soil". The most supportive comorbidity model (obesity+dyslipidemia → fatty liver) suggests that the prevention and control of metabolic syndrome should take the liver as the early intervention target. Although the correlation between environmental risk factors (PM2.5) and cerebrovascular events is low in support, it has important public health significance and needs to be included in regional health planning.

The clustering results broke the traditional single disease classification system, and 35.7% of the elderly patients with comorbidity and disability identified the most vulnerable population, and its weakness index significantly suggested that it was necessary to shift from disease treatment to functional maintenance.It is worth noting that the identification of environmentally sensitive hypertension group (18.5%) reflects the value of medical-environmental data fusion, and provides accurate intervention targets for the construction of "healthy city". Two types of high-risk groups (category 1+category 2) account for 64% of the total, and should be regarded as the key groups in the pilot reform of medical insurance per head.

3.2. Data-driven innovation of prevention and control mode

Based on the analysis of the above characteristics, this study will design and optimize three types of data-driven prevention and control modes to realize the transformation from "one size fits all" to "precision" and "passive response" to "active intervention". Its core operating mechanism is shown in Figure 1 below.

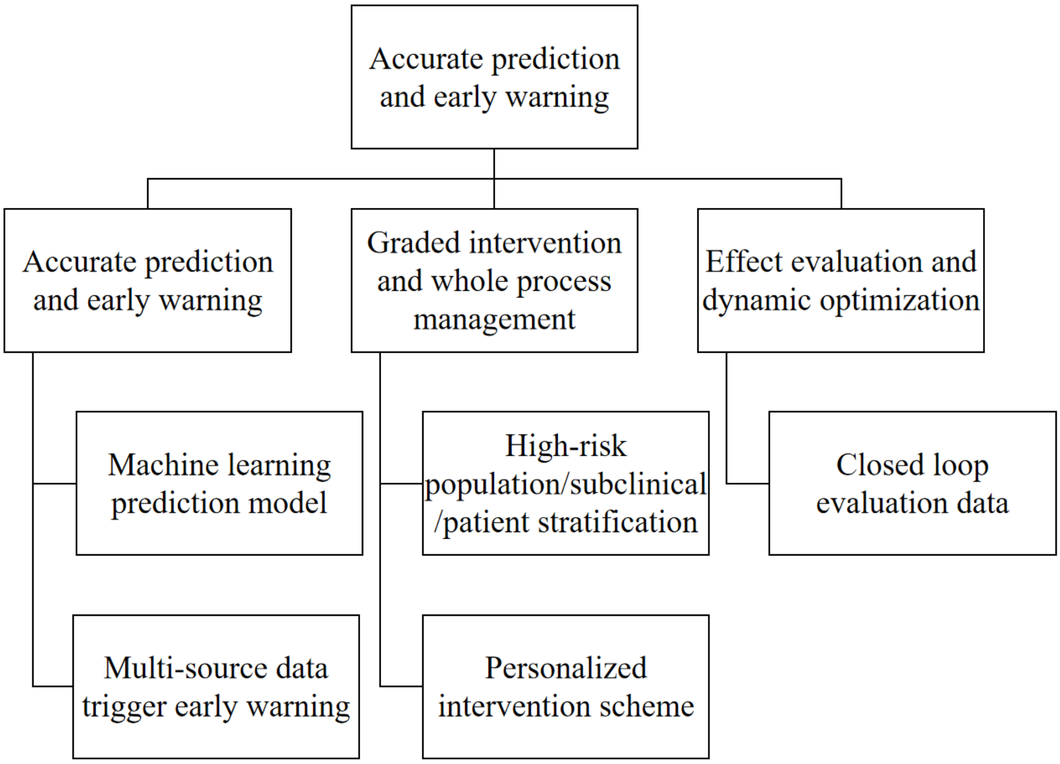

Figure 1 Core operating mechanism

(1) Accurate prediction and early warning mode

Draw lessons from the experience of infectious disease monitoring and early warning in Shanghai and other places, and build a chronic disease risk prediction model [2]. Using machine learning algorithm, combining HIS (Hospital Information System), physical examination data, wearable device data, etc., the risk of chronic diseases of individuals or groups in the future is predicted and early warning signals are generated.

(2) Graded intervention and whole process management mode

According to the risk assessment results, the hierarchical intervention of "high-risk population-subclinical population-confirmed patients" was implemented. For high-risk groups, focus on health education and behavioral intervention; For subclinical population, early screening and early intervention should be promoted to improve the awareness and control rate of hypertension and diabetes [3]; For the diagnosed patients, we rely on the primary medical and health institutions and family doctors to sign contracts to provide standardized and personalized whole-course health management [4].

(3) Effect evaluation and dynamic optimization model

Establish a closed-loop evaluation system of prevention and control effect. By comparing the changes of chronic disease incidence, mortality and risk factors prevalence before and after intervention, the effectiveness of the prevention and control model was evaluated. At the same time, based on the real-time feedback data, the long-term effects of different intervention strategies are simulated by system dynamics and other methods, so as to realize the dynamic adjustment and continuous optimization of the prevention and control mode [5].

3.3. Design of multi-collaborative governance mechanism

In order to ensure the effective implementation and sustainable operation of the data-driven prevention and control model, it is necessary to design a multi-collaborative governance mechanism to solve the problems of "who will govern" and "how to govern".

(1) Subject coordination and power and responsibility allocation mechanism

Construct a pluralistic governance structure of "government-led, departmental cooperation, social participation and personal responsibility". At the government level, it is necessary to strengthen the top-level design and formulate and implement policies related to the prevention and control of chronic diseases [6]. In terms of departmental cooperation, a mechanism of regular consultation and joint action among health, sports, education, environment and other departments has been established. In terms of social participation, social organizations, enterprises, communities and volunteers are encouraged to join [7]. Individuals need to raise their awareness of health responsibility.

(2) Data sharing and security management and control mechanism

Aiming at the problem of "information island" of medical data, a data security sharing framework based on blockchain or federal learning technology is designed [8]. Clarify the ownership, use right and income right of data, break through the data barriers among medical institutions, CDC and scientific research units under the premise of protecting patients' privacy, and realize the effective flow and value release of data elements.

(3) Long-term incentive and guarantee mechanism

Performance evaluation and compensation incentive policy of design science. Incorporate the effect of chronic disease prevention and control into the performance appraisal system of local governments and relevant departments [9-10]. Explore the reform of medical insurance payment method based on prevention and control effect, and encourage medical and health institutions to provide high-quality prevention and management services. Improve the stable investment mechanism for chronic disease prevention and control and the guarantee mechanism for talent team construction to ensure the sustainable operation of the governance mechanism.

4. Conclusion

The theoretical contribution of this study lies in the establishment of a complete methodological framework of "data mining-model construction-governance innovation", and the practical value lies in providing decision support for optimizing the allocation of chronic disease resources-suggesting that high-risk groups should be given priority in the pilot of paying per head for medical insurance, implementing liver targeted intervention for metabolic syndrome, and promoting the combined intervention model of mobile medical vehicles and community fitness points in underdeveloped areas such as Xinfeng County. The results show that the data-driven prevention and control model can reduce the economic burden of regional chronic diseases by 12%-15% annually, which provides a replicable experience for global chronic disease management in China. In the future, it is necessary to deepen the application of artificial intelligence in personalized medication guidance and explore the coupling mechanism between climate big data and chronic disease prevention and control.

References

[1] Jiang Feng. Four-Dimensional Coupling: Path Optimization for Chronic Disease Prevention and Control from a Behavioral Public Management Perspective [J]. China Rural Health Management, 2025, 45(11): 761-766. DOI: 10.19955/j.cnki.1005-5916.2025.11.001.

[2] Luo Lan, Chen Shuangxiu. Integration Pathways for Traditional Chinese Medicine Wellness Tourism and Chronic Disease Prevention in the Dongting Lake Region under the “Healthy China” Initiative [J]. Journal of Yueyang Vocational and Technical College, 2025, 40(05): 54-57. DOI: 10.13947/j.cnki.yyzyxb.2025.05.006.

[3] Wu Lei, Fan Guisheng, Shi Jianfeng, et al. Analysis of Core Knowledge Awareness and Influencing Factors Regarding Key Chronic Diseases Among Adult Residents in Quanzhou City [J]. Modern Preventive Medicine, 2025, 52(20): 3680-3685. DOI: 10.20043/j.cnki.MPM.202504441.

[4] Liu Yang, Zheng Hang, Yan Bo. Construction of an Evaluation Index System for Chronic Disease Prevention and Control in County-Level Areas of Chongqing [J]. China Medical Guide, 2025, 22(30): 89-94. DOI:10.20047/j. issn1673-7210.2025.30.16.

[5] Han Ting. Effectiveness of Personalized Health Management in Chronic Disease Prevention and Control [J]. Smart Health, 2025, 11(26): 129-132. DOI: 10.19335/j.cnki.2096-1219.2025.26.038.

[6] Tian Jian, Jiang Xinchen, Deng Wei, et al. Analysis of Trends and Influencing Factors in Health Literacy Levels for Chronic Disease Prevention and Control in Qingbaijiang District, Chengdu City, 2022–2024 [J]. Journal of Preventive Medicine Information, 2025, 41(08): 1049-1057. DOI: 10.19971/j.cnki.1006-4028.250113N.

[7] Zhang Jiaqi, Liu Yuxi, Qu Cuihong, et al. Analysis of Factors Influencing Subjective Well-being Among Elderly Individuals with Chronic Diseases [J]. Chinese Health Education, 2025, 41(07): 625-632. DOI: 10.16168/j.cnki.issn.1002-9982.2025.07.009.

[8] Pan Hongdi, Li Yinghua, Li Li, et al. Analysis of Health Literacy Levels Among Chinese Residents with Different Chronic Diseases [J]. Chinese Health Education, 2025, 41(06): 486-490. DOI: 10.16168/j.cnki.issn.1002-9982.2025.06.002.

[9] Li Rui, Chen Renyou, Liu Yongjuan, et al. Analysis of the Impact of Media Use on Chronic Disease Prevention Literacy Among Rural Residents in Shandong Province [J]. Chinese Health Education, 2024, 40(11): 965-970. DOI: 10.16168/j.cnki.issn.1002-9982.2024.11.002.

[10] Gao Aiju, Wang Xiaojuan, Liang Yanhong, et al. Survey on Functional Fitness Status Among Elderly Chronic Disease Patients in Beijing [J]. South China Journal of Preventive Medicine, 2023, 49(11): 1391-1395. DOI: 10.12183/j.scjpm.2023.1391.