INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Management of Fraudulent Use of Medical Insurance Funds Based on Big Data Mining - A Case Study of Huzhou City, Zhejiang, China

WANG Ming

Doctor of Management

Supervisors:
PhD Diana Aldea Mendes, Associate Professor
ISCTE University Institute of Lisbon
PhD LIANG Decui, Professor
University of Electronic Science and Technology of China

December, 2022

BUSINESS
SCHOOL

Marketing, Operations and General Management Department

Management of Fraudulent Use of Medical Insurance Funds Based on Big Data Mining - A Case Study of Huzhou City, Zhejiang, China

WANG Ming

Doctor of Management

Supervisors:
PhD Diana Aldea Mendes, Associate Professor
ISCTE University Institute of Lisbon
PhD LIANG Decui, Professor,
University of Electronic Science and Technology of China

December, 2022

BUSINESS
SCHOOL

Marketing, Operations and General Management Department

Management of Fraudulent Use of Medical Insurance Funds Based on Big Data Mining - A Case Study of Huzhou City, Zhejiang, China

WANG Ming

Doctor of Management

Jury:

PhD Henrique Manuel Caetano Duarte, Associate Professor with Habilitation,
ISCTE University Institute of Lisbon
PhD Andreia Teixeira Dionísio Basílio, Associate Professor with Habilitation,
Universidade de Évora
PhD Anabela Ribeiro Dias da Costa, Assistant Professor,
ISCTE University Institute of Lisbon
PhD Xiao Wen, Professor,
University of Electronic Science and Technology of China
PhD Diana Elisabeta Aldea Mendes, Associate Professor,
ISCTE University Institute of Lisbon

December, 2022

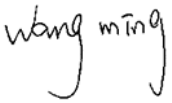
iscte INSTITUTO UNIVERSITÁRIO DE LISBOA

Management Of Fraudulent Use Of Medical Insurance Funds Based On Big Data Mining - A Case Study Of Huzhou City, Zhejiang, China

WANG Ming

Declaration

I declare that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university and that to the best of my knowledge it does not contain any material previously published or written by another person except where due reference is made in the text.

Signed:  Date: 2022.12.30

Name: WANG Ming

作者申明

本人郑重申明：除了论文致谢中明确说明并致以谢意的部分外，所呈交的论文不包含任何他人或作者本人已用于获得任何教育机构的学位和证书而使用过的材料。同时尽我所知，除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。

作者签名：  日期：2022.12.30

姓名(拼音)：WANG Ming

[This page is deliberately left blank.]

Abstract

With the continuous expansion of the basic social security system coverage and the scale of fund income and expenditure, various insurance frauds are frequent, while the social security level continues to improve, posing a serious threat to the security of medical insurance funds. The supervision and prevention of medical insurance funds fraud have been gradually resolved by new technologies, management theories, and regulatory methods. However, Medical insurance fraud is a complex system involving regulators, hospitals, doctors, patients, and even patients' family members. It is necessary to use the scientific judgement method with technical support so that the policy introduction will be more targeted.

This thesis designs a data-algorithm-process-strategy style of medical insurance risk fraud identification model based on the medical data of 2.4 million people in Huzhou City, Zhejiang Province. (1) Various relevant parties were selected for interviews research, which confirmed that the problem of medical insurance fraud did exist and the formation process was relatively complicated. (2) Use big data model with Apriori algorithm to screen medical insurance fraud. (3) Adopt the rule engine method to analyze the use of medical insurance and conclude 11 fraudulent use scenarios of medical insurance fraud models. (4) Use physical portrait technology to analyze medical insurance fraudsters and their motives, and identify seven types of fraudsters' behaviors.

By analyzing the findings of this study, the feasibility of the medical insurance fraud risk identification model in this study is proved, and the combinatory analysis of big data model, rule engine method, and entity portrait technology is effective in fraud identification and can be applied to such kind of problems.

Given the above analysis process and conclusions, we put forward seven solutions and suggestions to facilitate medical insurance fraud management.

Keywords: medical insurance fraud; big data mining; association rule; rule engine; entity portrait

JEL: M10; M15

[This page is deliberately left blank.]

Resumo

Com a contínua expansão da cobertura do sistema de segurança social básico e com o escalamento das receitas e despesas dos fundos, as fraudes no setor de seguros de saúde são cada vez mais frequentes. Enquanto o nível da segurança social continua a melhorar, as fraudes representam uma séria ameaça à segurança dos fundos de seguro saúde. A supervisão e a prevenção de fraudes em fundos de seguros médicos foram gradualmente resolvidas por novas tecnologias, novas teorias de gestão e novos métodos regulatórios. No entanto, a fraude de seguro médico é um sistema complexo que envolve reguladores, hospitais, médicos e pacientes e até mesmo familiares dos pacientes. É preciso utilizar um método de julgamento científico com apoio técnico sofisticado, para que a implementação da política seja mais direcionada.

Esta tese constrói, com base num processo-estratégia de algoritmos de dados, um modelo de identificação de fraude de risco de seguro médico, considerando dados médicos de 2,4 milhões de pessoas na cidade de Huzhou, província de Zhejiang: (1) Várias partes relevantes foram selecionadas para entrevistas, cuja resultado ajudou a confirmar que o problema de fraude de seguro médico existia e o processo de formação era relativamente complicado. (2) Usou-se um modelo de big data com o algoritmo Apriori para rastrear fraudes em seguros médicos. (3) Adotou-se o método do mecanismo de regras para analisar o uso de seguro médico e conclui-se que existem pelo menos 11 cenários de uso fraudulento de modelos de fraude de seguro médico. (4) Usou-se a tecnologia de retrato físico para analisar os fraudadores de seguros médicos e seus motivos e identificaram-se 7 tipos de comportamentos dos fraudadores.

Ao analisar os resultados deste estudo comprova-se a viabilidade do modelo de identificação de risco de fraude de seguro médico, e a análise combinatória do modelo de big data, método de mecanismo de regra e tecnologia de retrato de entidade prova ser eficaz na identificação de fraude, e, portanto, pode ser aplicado a esses tipos de problemas. Dado o processo de análise e as conclusões acima, apresentamos sete soluções e sugestões para facilitar a gestão de fraudes em seguros médicos.

Palavras-chave: fraude, seguros de saúde; big data mining; regras de associação; rule engine; entity portrait

JEL: M10; M15

[This page is deliberately left blank.]

摘 要

随着基本医疗保障制度覆盖面和基金收支规模不断扩大，各类欺诈骗保问题频发，对医疗保障基金的安全造成严重威胁。医疗保险基金的欺诈与监管、防控问题逐步以新技术、新管理理论、新监管手段去解决完善。医保欺诈行为涉及到监管部门、医院、医生和患者甚至患者的家人，是一个复杂的系统。需要在技术手段的支持下采用科技方法判断，从而使政策的出台更有针对性。

本文设计了以“数据-算法-过程-策略”的医保风险欺诈识别模型，以浙江省 HZ 市 240 万人口的就医资料为数据基础：（1）选取各相关方进行访谈调研，证实了医保欺诈问题确实存在并且形成过程比较复杂。（2）采用大数据模型结合 Apriori 算法对医保欺诈行为进行了有效筛查。（3）运用了规则引擎方法对医保费用使用进行分析，得到了 11 种医保欺诈模型的费用使用特征。（4）使用实体画像技术对医保欺诈人员及其动机进行分析，发现了 7 类欺诈人员欺诈行为。通过分析本研究的结论证实了本研究设计的医保风险欺诈识别模型可行性，大数据模型、规则引擎方法和实体画像技术的组合分析在医保欺诈识别中是有效的，可以用于此类问题的解决。

针对以上分析流程和分析结论，本文提出了 7 项对策建议，有助于对医保欺诈行为进行管控。

关键词：医保欺诈；大数据挖掘；关联规则；规则引擎；实体画像

JEL: M10; M15

[This page is deliberately left blank.]

Acknowledgments

It has been almost six years from pursue a doctorate to achieve graduation—which also happens during my 40s to 50s. As the Chinese old saying of the age goes, in the age of knowing destiny, I often look inward and ask why I have chosen this difficult path. However, it is for a belief of trust and a heart that never gives up. No matter how bumpy and windy the path it is, sustained by this belief, I must take an arduous journey to get the final triumph.

Looking back on the whole process, from having no idea to being confident, from collecting information to finalizing the dissertation, I have been frustrated, quitted, restarted, and delighted along the way. But in retrospect, these memories are always full of affection. I have received guidance from teachers, assistance from friends, and consideration from my family in my academic writing. Without them, it would be hardly possible to finish this thesis with high standard. Whenever think about it, I am overwhelmed with gratefulness.

With the completion of my dissertation, my study career at the University of Electronic Science and Technology of China and the Lisbon University of Business Administration are coming to an end. During this period, I express my sincere gratitude to my two supervisors: Professor Diana of the University of Lisbon and Professor Decui Liang from the University of Electronic Science and Technology of China, for the guidance and help they gave me many times on the thesis. Their rigorous academic style and optimistic life attitude gave me deep impressions, which will beacons to me on future work and study. Even though the thesis is now completed, these good habits will shedding a light on the rest of my life and benefit me even more. Also over the selection of the theme their serious advice lies and on the establishment of the structure and on the revision of the content, all of this thesis are enlightened by two supervisors.

In addition, I thank Professor Ma Yongkai, Professor Xiao Wen and all the teachers from the School of Economics and Management of the University of Electronic Science and Technology of China. Also thanks are due to my classmate Pengliang Qiao, my colleagues Dr. Zhu and Dr. Yang, and my old school with providing me with a warm, harmonious and positive learning environment, so that I can learn and think in a good learning atmosphere, and to all the teachers with enriching my knowledge in both In addition, I thank Professor Ma Yongkai, Professor Xiao Wen and all the teachers from the School of Economics and Management of the

University of Electronic Science and Technology of China. Also thanks are due to my classmate Pengliang Qiao, my colleagues Dr. Zhu and Dr. Yang, and my old school with providing me with a warm, harmonious and positive learning environment, so that I can learn and think in a good learning atmosphere, and to all the teachers with enriching my knowledge in both quantity and quality during the school. After six years of study, I have enlarged my horizons, broaden my knowledge, and made many good friends, I also thank all my classmates and friends, will cherish this deep friendship.

Especially, it is because of teachers' inculcation that I can accumulate certain professional knowledge, improve my learning ability, form my own perspective and express my own opinion on current economic phenomenon. Meanwhile, I also thank my excellent classmates working together in the past few years, to create a good shared learning and exchanging environment. It should be mentioned that these six years of study and life have already become one of the most important experiences in my life. As I look back on the study, research, and various processes, these scenes of the past often leap before my eyes, the recalling itself is good learning. Every teacher and classmate is worth learning from, their academic spirit and characters have become my role models. Thanks everyone again for your encouragement and sincere help.

Finally, my cordial thanks also go to my family member and friends, who stand behind me, give me great confidence in face of any difficulties and setbacks. I will study more committed, work harder in the in return for their support.

My acknowledgements also go to experts, scholars, leaders and teachers in the reference. Their selfless dedication has opened another window and let the fresh sunshine come into my life. The completion of the doctoral dissertation is not ending the study, but paving the way for explore a new area.

Thanks to all the people who has helped and encouraged me, and sincerely wish you health, peace and happiness!

Once again, I would like to express my lofty respect to all the leaders and teachers!

Wang Ming in Shanghai, China

December 29, 2022

致 谢

从决定读博士学位到现在毕业，已经快六个年头了，我也从不惑之年步入了知天命的年纪，我常常内心询问自己为何选择这条艰难的路，因为一份信念、一份信任及一份永不放弃的心。道路崎岖坎坷，但信念支持着我一路披荆斩棘，最终凯旋而归。

回想整个过程，从毫无头绪到胸有成竹，从开始的搜集资料到最后的论文定稿，这一路走来，我沮丧过、放弃过、重来过、开心过，但是如今回忆起来，都是满满的情义。在写作的过程中，我得到了老师的指导，朋友的帮助，家人的理解与关怀。如果没有他们，我无法想象我如何高标准完成这篇论文，每每想到这里，我的心里充满了感激。

随着论文的完成，我在电子科技大学、里斯本工商管理大学的这六年学习也即将划上句号。在此期间，我衷心感谢我的两位导师：里斯本大学的戴安娜教授及电子科技大学的梁德翠教授。在完成论文期间多次给与我指导和帮助。两位老师严谨的治学作风、积极乐观的人生态度，给我留下了非常深刻的印象，对我今后的工作和学习将产生深远的影响。即使现在论文已经完成，但这些好习惯将会陪伴我终生，令我受益良多。本文从题材的选定、结构的确立以及内容的修改，都投影着导师认真的身影。

此外，我还要感谢电子科技大学经管学院的马永开教授、肖文教授以及所有任课老师。感谢我的同学乔鹏亮、感谢我的同事朱医生和杨医生，感谢母校为我提供了温馨、和谐、积极向上的学习环境，让我能在良好的学习氛围下有所学、有所思。谢谢所有任课老师，谢谢你们在校期间让我的知识储备有了量和质的双面提升。经过六年的学习，我开阔了眼界，扩大了知识面，结交了许多好朋友，感谢各位同学、朋友，我将珍藏这份深厚的友谊。

正是因为老师们谆谆教诲和认真教学，我才能积累一定的专业知识，提高自己的学习能力，能够对现有的经济现象提出自己的看法、发表自己的观点。同时我也要感谢这几年来一起奋斗的同学们，正是有如此优秀的你们，才会创造出一个良好的学习讨论环境，应该说这六年的学习与生活，都已经成为我人生中最重要的经历之一。回望这六年学习研究时光，各个过程历历在目，这本身就是一种很好的学习！从每一位老师和同学身上我都能学到新的知识，他们的治学精神、为人品格都成为我学习的榜

样，感谢所有人的鼓励和真心相助！

最后，我要感谢我的家人、亲友。不管我经受什么样的困难与挫折，他们都永远在我的身后，给我极大的勇气和信心，这样无私的奉献，我会更加用心的学习，以后越发努力的工作，回报他们。

感谢我参考文献中的各位专家、学者、领导和老师，正是你们无私的奉献，给我的人生打开了又一扇天窗，增加了一片阳光。博士毕业论文的完成，不是学习的结束，而是一个新领域学习与探索的开始。

感谢所有帮助和鼓励过我的人们，真诚祝愿你们健康、平安、幸福！

再次向各位领导和老师表示崇高的敬意！

王明于上海

2022年12月29日

Contents

List of Tables

[Tables in other chapters see Annex B.]

List of Figures

[Figures in other chapters see Annex C.]

List of Abbreviations

AECOPD	Acute Exacerbation of Chronic Obstructive Pulmonary Disease
ACC	Accuracy
AUC	Area under Curve
BP	Back Propagation
CAHD:	Coronary Atherosclerotic Heart Disease
CHD	Cardiovascular Heart Disease
CHNS	China Health and Nutrition Survey
CNY	Chinese Yuan
COPD	Chronic Obstructive Pulmonary Disease
CT:	Computed Tomography
DIP	DIP: Diagnosis-Intervention Packet
DRG	DRG: Diagnosis Related Groups
ENT	ENT: Ears, Nose, and Throat
GMB	GMB: Gradient Boosting Machine
GP:	GP: General Practitioner
HbA1C	HbA1C: Glycosylated Hemoglobin, Type A1C
HDFS	HDFS: Hadoop Distributed File System
LDA	LDA: Linear Discriminant Analysis
MDC	Mapped Diagnostic Contexts
MIMIS	MIMIS: Medical Insurance Management Information System
MRI	Magnetic Resonance Imaging
NER	Named Entity Recognition
NHCAA	National Health Care Anti-Fraud Association
OB	Obstetrics
OECD	Organization for Economic Co-operation and Development
PLA	People's Liberation Army
SAS	Sleep Apnea Syndrome
SVM	Support Vector Machines
TCM	Traditional Chinese Medicine

TSAH	Traumatic subarachnoid hemorrhage
OECD	Organization for Economic Co-operation and Development
PLA	People's Liberation Army
SAS	Sleep Apnea Syndrome
SVM	Support Vector Machines
TCM	Traditional Chinese Medicine
TSAH	Traumatic subarachnoid hemorrhage
XGBoost	Extreme Gradient Boosting

Chapter 1: Introduction

1.1 Background

In China, the social medical insurance fund, also known as Healthcare Security Fund is a special fund for the medical insurance system. It provides the economic foundation designated to serve the primary healthcare needs of all participants. Under the unified management of the administrative agencies, the medical insurance fund consists of enterprise contributions and individual premiums, compensating for insured participants' medical expenses due to sickness.

The insurance fund has played a key role in implementing the national basic medical insurance system and its normal functioning. All the financing, allocation, preservation, appreciation, and management affect the balance between the national rights and obligations, corporate entities, and social members. In other words, all stages of fund management, from raising to payment, in operation and expenditure, are essentially the distribution and redistribution of national income. It embodies the principle of social fairness through balancing and coordinating multiple economic interests (H. H. Xie & Wan, 2014).

The operating status of the social insurance fund is determined by the balance of the fund, and its security depends on the appropriate usage within the safe limit. When the fee is greater than the social insurance fund and yields a slight surplus, there is no mutual influence between the insured objects, and the marginal cost of new users is almost zero. When the expenditure of the social medical insurance fund is greater than the income, the marginal cost of new users increases rapidly, which will inevitably nibble at the reasonable benefit claims of some insurance subjects, which reflects the exclusivity of the insurance fund usage (Ge & Wu, 2013).

The insurance fund, reserve capital in monetary form, is also fundamental to maintaining the sustainable development of primary healthcare and medical insurance.

The information and technology in the 21st century empower insurance fund management. In the last years, digital information and technologies increased exponentially due to the development of methodologies based on data science, with the primary aim of transforming the information into valuable insights. Data science, text mining, data mining, and machine learning are no more buzzwords but necessary tools in any organization possessing big data. Healthcare and insurance domains are following the same path. In this era, all around us are connected,

and everyday life generates vast data (Bao, 2010). According to relevant media statistics, the overall data amount in the recent three years has surpassed 40,000 years before. Under the influence of this data background, it will be another scenario; with the possession of many electric devices in the future, people will be exposed to an ecosystem consisting of these electronic screens, where no media is a silo and the same to devices. We can also imagine the scene as an ecosystem so that the focus of medical insurance cost control should be a single person rather than the system. Big data is a new darling in the internet world, holding data mining as its core when a solution to problems of platforms and hospitals is needed.

In big data, fund management issues emerged from social and medical insurance cause development, such as increasing medical expenses, different degrees of fund waste and fraud, and inefficiency and inefficacy in management. The following three aspects are the background of the current research on the management of medical insurance fund fraudulent use:

1.1.1 China's medical insurance fund plays an important role in the national security

Since deepening the reform of the medical and health system, China has taken steps toward building a national system, including basic medical insurance for urban employees, urban residents, and new rural cooperative medicare insurance as a priority, medical assistance, and commercial insurance as a supplement (W. Deng, 2014). China has established the world's largest primary medical security network covering the entire population, covering 1.372 billion people in urban and rural areas, with the coverage rate staying above 95 percent. The maximum payment limits of the three primary medical insurance systems are more than six times the average annual income of the corresponding population and not less than 50,000 yuan ($ 7,350). And also, the economic burden on medical treatment of urban residents has been eased (H. Zhang, 2014).

The increased demand for medical treatment makes the medical security for significant diseases challenging, based on the general inclusive reimbursement level. In addition, the relatively low reimbursement rate, the starting line and limit line in. policy design, and the residents' poverty reoccurrence due to severe diseases remain prominent. The state has added medical insurance funds to the basic medical security system as a necessary institutional arrangement to share the medical costs and reduce the medical burden of patients with serious diseases.

The urban basic medical insurance fund expenditure of 2015, 2016, and 2017 registered 930.39 billion yuan ($ 136.77 billion), 1.0767 trillion yuan ($ 158.3 billion), and 3.7236 trillion yuan ($ 54.74 billion), respectively (K. S. Huang & Liu, 2017).

Medical insurance fund has become the primary means of payment in the medical system. Related to the rigidity and rapid growth of medical expenditure, the huge pressure faced by medical insurance funds includes the following seven aspects:

1. The medical service demand of insured citizens increases substantially with economic growth. The better the living standard of insured citizens, the higher the need for medical services. Even if economic growth slows down, the rigidity of growth remains unchanged(P. Lu, 2010).

2. The development of science and technology promotes the application of new technologies, drugs, materials, and therapies, which improve clinical diagnosis and treatment while driving up medical fees.

3. The urban demographic structure and an aging population have intensified. Until the end of 2017, the number of Chinese over 60 was more than 240 million, accounting for 17.3% of the total population. Higher life expectancy and evident urban population aging objectively cause an increase in medical expenses.

4. The disease spectrum alteration corresponding to the demographic structure changes, especially in aging, the primary type of disease and major cause of death, has seen priority changing from infectious to a non-communicable chronic disease requiring far more costly treatment. So, the disease spectrum alteration will inevitably raise the healthcare cost.

5. Medical price increases and hospital profit-driven factors are also particular concerns. In pursuing the maximum interests, wrongdoings like excessive service, over-prescription, and repeated examinations of patients are common, resulting in a continuous increase in medical expenses.

6. Waste, swindling, abuse, fraud, and other fraudulent activities cause severe destruction to the medical insurance fund.

7. The fund's operational efficiency is relatively low, and the control and administration need improvement.

As a basic platform for residents' health, medical insurance funds need more healthy and scientific management, and local medical insurance departments have faced great pressure on medical insurance fund management. Scientific and refined management has become a must of development.

1.1.2 China's medical insurance fund frauds emerge in an endless stream

Due to the significant pressure, adapting the medical insurance system to the current economic situation and social capacity is necessary. However, new threats require more attention. The

increasing population and complex insured personnel - produce a conflict of interest and make the work between hospitals, pharmacies, patients, and medical insurance agencies more challenging. Rising medical insurance expenditure cannot fill the gap in medical services, and the medical expenses are still a major pressure on residents' lives, which means that the cost of medical insurance is not directly proportional to the benefit of medical insurance participants. So there must be non-objective reasons for the improper utilization of medical insurance funds.

The relationship between the medicine industry, patients, and health insurance is close and complicated by various conflicts of interest. Firstly, hospitals provide medical services, offer therapies and save lives. They manage to maximize profit by serving patients. Secondly, the patient hopes to recover as soon as possible through the best treatment option. Also, they hope to reduce their expenses and get the most considerable compensation from the fund. Given the above, medical insurance balances insured participants and hospitals, ensuring both interests and rights (Yu, 2016).

The complex insurance personnel constitution - causes various medical insurance frauds - to emerge endlessly. It shows that 20 to 30 percent of the total fund amount is wasted. Standard waste practices include falsely using others' cards, describing or forging prescription medicare drugs for traffic, reselling and renting the card to fraud gangs, or even joining the intermediary agents to serve as bridges between hospitals and fake patients to defraud the fund. The misbehaviors mentioned above frequently occur, resulting in a massive waste of funds (Cui, 2016 ; Mei, 2016).

Besides medical insurance participants, illegal acts also exist in healthcare institutions. Typically, the patient is subject to treatment in medical practice, and as a service provider, hospitals lead the practice. Whether the hospital practice is standard significantly impacts controlling medical insurance expenses. However, some designated institutions pursue economic profit at the expense of social benefit. Hospitalization period decomposition, false claims, forging hospitalization records, excessive medical treatment, and rejection of insured participants with serious illnesses are common fraud tricks. Some organizations plot with the participant for false hospitalization claims and illegal reimbursement income or declaration of diseases not enrolled in the medical insurance category (Yu, 2016).

Apart from the hospital, some designated medical institutions and employees sell daily goods violating the medical insurance policy. The medicare cardholders cash out by purchasing snacks, daily chemicals, and even rice, flour, oil, and salt, products than drugs.

Unlawful practices for medical insurance fund embezzlement, abuse, and fraud are various, such as false hospitalization, hospitalization decomposition, extraordinary prescription, over-

prescription, arbitrary raising prices, excessive examinations, and other sophisticated means to escape supervisory review.

Table 1.1 Medical insurance fund fraud subjects and types

Subject	Type of medical insurance fraud
Participant（patient）	The insured swipe cards for medicines, falsely prescribes medical insurance drugs, sells and resells medical insurance cards, and rents medical insurance cards to medical insurance fraud rings
Medical institutions	Decompostition of hospitalization period, hospitalization formality, lower admission standard, Falsification of hospitalization record, unnecessary examination, raise prices, refuse critical patients, false claim, overtreatment, big prescription, use high price drugs
Designated pharmacy	Exchange of medicines, sale of daily necessities and drugs not related to diseases

As summarized in Table 1.1, all fraudulent activities lead to a huge loss of medical insurance funds, directly jeopardizing people's interests and lives. To safeguard fundamental insurance rights and interests and reduce waste, rigorous and reasonable fund control should be conducted (Z. R. Huang, 2016).

1.1.3 Big data management of China's medical insurance funds has been gradually carried out

Medical information and technology have significantly improved the implementation of medical insurance and management information systems. During the process, big data on medical insurance was collected since medical institutions at all levels had accumulated a vast amount of data, including the diagnostic record, treatment details, descriptions, and digitalized medical files. Then, a partial artificial intelligence approach for medical insurance came into being. The related business management procedure was regulated via the information system and recruited experts, scholars, and doctors to make artificial judgments and retain conclusions by examining, approving, reviewing, and auditing specific cases in the medical process. This mode plays a key role in the starting phase of the medical insurance system. But when the number of participants and daily average service quantities of authorized medical institutions has increased, the management personnel above can no longer fulfill the massive supervision task.

The big data of medical insurance funds pool many medical pieces of knowledge and regulations and conceal a few records of fraud. Detecting and identifying fraud in big data is of great significance. Thus, it is crucial to mine and learns the hidden medical knowledge and treatment patterns in medical big data, quickly snap the domain knowledge, spot the few hidden fraud records, and formulate a fraud detection model. After that, timely eliminate the illegal

practice by filtering abnormal information, effectively reducing the suspected fraud record aggregate, improving the real-time performance of fraud detection targeted at new medical behavior. Finally, improve inspection accuracy and supervision efficiency to make fraud less frequent and lessen the loss of medical insurance funds.

Data mining has made quick progress in recent 20 years. It includes manipulating huge databases, pattern recognition, statistical learning, artificial intelligence, machine learning, and other methodologies and algorithms widely used in different scientific and business fields. In particular, Data Mining is discovering knowledge hidden within large amounts of data that is hard to be artificially spotted.

A huge amount of data (structured and unstructured) are recorded in datasets and cloud repositories. Extracting valuable information from these datasets will significantly contribute to the diagnosis, medication, treatment, cost control, and other aspects of the healthcare domain. Several researchers have carried out studies on data mining in the medical field, mainly focusing on disease diagnosis, health risk prediction, traditional Chinese medicine prescription analysis, medical cost forecast, and patient visit time analysis.

It is a mega-trend to introduce big data mining methodologies and technology into the medical insurance fund fraud control, which is the best pathway to better management of medical insurance (D. B. Huang et al., 2017 ; Y. Ren et al., 2018 ; W. G. Wang, 2020).

1.2 Research dilemmas

With the increasing population aging and the rapid growth of medical expenses, the medical insurance fund under the new situation is facing unprecedented pressure. It is urgent to detect the leakage of funds based on big data mining and to build an effective control system. At the same time, whether the control system at the macro level and the concrete measures at the spectator level are effective has also become a real challenge for regulators.

In 2021, the Ministry of Public Security, the National Healthcare Security Administration and the National Health Commission jointly deployed and carried out a special crackdown on insurance fraud following the law. By the end of September, the national public security organs had busted 251 criminal gangs, arrested 3,819 suspects, uncovered 1,246 cases of fraud against medical insurance funds, recovered 230 million yuan of medical insurance funds, and closed down 277 medical institutions in conjunction with medical security departments.

According to the contents of the notice, some criminals kept updating their lawbreaker's methods. Some designated hospitals for medical insurance lured people who did not have

hospitalization conditions into hospital treatment with a small favor and defrauded medical insurance funds by making up their condition, forging medical records, and falsely reporting consumables. Some medical insurance fixed point pharmacies with ginseng protect personnel, and illegal personnel colludes fictitious sales records; Some insured persons forged or changed reimbursement vouchers to defraud the medical insurance fund. In the typical case reported, a private medical institution director took the lead in cheating 10 million yuan, shocking and seriously endangering the national medical security fund safety.

The medical security fund has many users and a long chain, which involves many aspects such as medicine, law and audit, and involves many links such as designated medical institutions, designated pharmacies, insured personnel, and medical insurance operators, which objectively brings difficulties to the supervision work. Secondly, some medical security bureaus, other regulatory departments, and their staff still do not seriously perform their duties - such as going through the formalities when reviewing relevant documents - so the criminals who cheat insurance easily succeed. Moreover, some regulatory personnel and lawless inside and outside collusion fraud the insurance system together.

There are many difficulties in solving the fraudulent use of the management medical insurance fund. Although the medical insurance administration institutions have issued many policies, the digitalization of medical insurance patient information has reached version 3.0 of face recognition, and the country has also issued many policies such as DRGS (Diagnosis Related Groups), which divides patients into 500-600 diagnosis-related groups based on factors such as age, gender, length of stay, clinical diagnosis, surgical disease severity, complications and outcomes, and then determines how much compensation should be given to the hospital (Yang, 2016).

However, many medical institutions still use Medicare funds fraudulently every year. In March 2022, the National Healthcare Security Administration found that the Tongji Hospital, affiliated with Tongji Medical College of Huazhong University of Science and Technology, had the problem of swapping and falsely recording high-value medical supplies in orthopedics from January 2017 to September 2020, defrauding the medical insurance fund to pay 233,343,609.64 yuan (Data source: National Healthcare Security Administration).

The illegal use of the medical insurance fund is not easy to be spotted, and patients (medical insurance participants) are the beneficiaries. When filling a prescription or in hospitalization, they can pay through the medical insurance fund, reducing their expenses, and solving their health problems, so patients will not report the problem of illegal and fraudulent use of the medical insurance fund.

The use of medical insurance funds is usually made by doctors, and many medical institutions have special departments for project grouping, packaging to cash out from medical insurance funds. In many hospitals, especially private hospitals, doctors' salaries and bonuses are bonded with their performance, so they will maximize the examination, treatment, and hospitalization for patients to increase their income.

Facing the huge pressure of the medical insurance fund and the severe impact caused by the fraud of the medical insurance fund, it is urgent to verify the feasibility of big data mining technology in the construction of medical insurance cost control models. Can effective medical insurance cost risk control models be created and applied in the macro and micro fields?

First, there are still serious problems in the management system of designated medical institutions of medical insurance in China. For example, designated medical institutions of individual medical insurance funds do not strictly identify the identity of patients entering and leaving hospitals, and use other people's medical insurance cards instead of their names for medical treatment. When there are illegal behaviors such as hanging up the bed and hospitalization, the patients shall go through the admission procedures and make a complete set of false admission medical records and examination reports to defraud the medical insurance expenses. The actual patients were never admitted for testing and treatment. Some medical institutions have lax control over patient entry and exit standards. Patients who do not meet the indications for hospitalization are still admitted to the hospital, which reduces the admission standard. Having the symptomatology of a minor illness helps significantly. For diseases that should be able to be treated in the region, they blindly expand the medical needs of insured personnel and illegally transfer them to tertiary hospitals in Beijing, Shanghai, and other places for treatment, increasing the unnecessary expenditure of overall funds.

Secondly, designated medical institutions abuse expensive antibiotics for some insured patients, and related patients agree with the medical institutions to do so, believing that the more expensive the drug, the better the effect, and do not pay for it. In this way, there will be a certain risk of medication. Careless use of antibiotics can lead to medical errors. And the use of drugs beyond the prescribed scope of the drug list, the use of drugs unrelated to the disease, and some clinicians for personal gain to promote some new specific drugs, increased the burden on patients and medical insurance funds. In order to ensure the large proportion of drugs in designated medical institutions, manual multi-point inspection, no indication or evidence inspection also appeared. Appointed retail pharmacies have the phenomenon of forwarding drugs, such as selling daily necessities, cosmetics, and food, violating the agreement (Wu, 2013).

The research and introduction of big data mining technology started relatively late in China,

and it is still very immature compared with foreign countries. In 1993, under the National Natural Science Foundation of China, some hospitals and research institutions began to study big data mining technology and methodologies (K. Luo & Dai, 2016). With the development of information technology, all fields are initializing and performing the digitalization of information. Following the needs of the times, some companies began to fund some domestic hospitals in 1998 to carry out relevant commercial research on big data mining, and achieved fruitful results. Since then, domestic research on big data mining has boomed. Hospitals and research institutions pay attention to big data mining (L. F. Zhu et al., 2015) and many Chinese scholars have begun to set foot on the road of applying big data mining in the medical field. At present, many Chinese scholars have applied big data mining to the field of risk management in medical institutions (X. Ma et al., 2018).

To sum up, it is necessary to establish a set of feasible medical insurance fund cost risk control models that can be applied in macro and micro fields to verify the feasibility of big data technology.

1.3 Proposal of research questions

Zhejiang, in the south-east coastal area of China, is the fastest-growing province nationwide. However, the city's expenditure on healthcare funds keeps increasing, putting significant strain on the government's fiscal management. Regarding medical insurance, fund overdraft is necessary to create a set of practical cost risk control models on medical insurance funds that can be applied in the macro and micro fields. After proving the feasibility of big data technology, the medical insurance regulatory authorities can analyze the risks of the fund operation in force, investigate and punish fraud in medical insurance funds, and put forward effective management suggestions.

At least, we need to know what flaw exists in medical insurance, how people involved in medical insurance fraud take advantage of the loopholes, and whether there is a more effective way to manage big data of medical insurance fraud.

1.3.1 Preconditions for medical insurance fraud

1.3.1.1 Internal conditions

The defrauded individual is usually a patient or a family member of a patient with severe or multiple illnesses simultaneously. Fraudulent healthcare providers are almost for-profit

pharmacies or private hospitals.

Abuse of power exists in two situations: one is the abuse of position power, such as the use of medical insurance full-time staff, doctors, and pharmacy heads taking advantage of his position to collude inside and outside, and the other is the abuse of technical expertise power. In short, the abuse of power is closely related to a high degree of concentration of power.

1.3.1.2 External conditions

The health insurance card system has problems: the health insurance card is falsely used by others, and alternation problems in the old and new medical insurance cards. For example, the old medicare card can still be used in other places when it is not canceled, or its loss is reported. What is more, the system of medical treatment in different places is not connected, and the regional medical insurance bureau cannot grasp the actual medical treatment situation in another place, such as buying false cases in Beijing to defraud the compensation of the new rural cooperative medical insurance.

There are problems in the healthcare organization audit system for coverage eligibility: the illegal participation, such as paid chronic healthcare or covered by two healthcare at the same time (the urban and rural residents' medical insurance and urban employee’s medical insurance).

Hospital management mistakes offer possibilities for defraud acted alone: in the doctor’s busy time, other’s illegal use of doctor's account for prescriptions.

1.3.2 The formation of medical insurance fraud

1.3.2.1 Mediating factors

In order to recognize and utilize the loopholes in medical insurance to defraud insurance, the insured or healthcare service institutions must be willing to cheat before they can illegally defraud the fund. So, the insured or the personnel of the medical service institutions take a chance to ease the economic burden of the disease or the difficulties of the institution's operation by defrauding the medical insurance fund and even to satisfy personal pleasures. Such mentality also includes the desire and imitation of others’ violation commitment when the fraudulent subject sees that behavior. In addition, some people believe that insurance fraud can be operated from illegal acts to violations, thus evading legal sanctions.

1.3.2.2 Fraudulent methods

Insurance fraud can only be realized through cooperation. The central part of complicity includes cooperation between government departments and insurance participants, collusion

between medical service institutions and insurance participants, and conspiracy to motivate all parties involved in insurance fraud into a community of shared interests. Positive incentives are made by dividing the profits, while negative incentives force individuals to participate in insurance fraud employing punishment. For example, a hospital requires employees to provide at least three medical insurance cards for false hospitalization procedures; otherwise, no salary will be paid (Q. Du et al., 2020).

1.3.2.3 Fraudulent behavior

The specific fraud means are diverse: making false bills for medical services; using other‘s data codes for medical insurance settlement; falsifying medical records to get doctors to overprescribe for inpatients; hospitalization formalities; exchanges for drug service items not on the list, such as using medical insurance card in the pharmacy in exchange for salad oil and other items; checking in for people not covered by medical insurance, such as forging false medical records of others treated in other places; cashing out of the medical insurance card, such as cash the balance of the medical insurance card at a discount price. Even more, some cheaters blatantly offense the medical insurance: hiring people to defraud the insurance, such as taking turns to swipe the card in drug stores, provided that each person pays 30 yuan per day. In order to attract more participants, they use advertising and other means.

Since the varied and diverse medical insurance behavior above is hard to detect and manage, is there a more effective management method for medical insurance fraud based on big data technology?

Targeting the difficulties and the characteristics of medical insurance fraud identification in China, this study digs through and analyzes the real medical insurance big data with mining technology to solve the problems of large-scale, complex, and unbalanced data structure in the identification of medical insurance fraud.

It starts with an analysis of the problem characteristics, and uses text mining to construct the fraud attributes from the perspective of expense, hospital, disease, and behavior. Then, based on the Easy Ensemble sampling method and Light GBM algorithm, the medical insurance fraud identification model was constructed. Comparing the identification effect with SVM, random forest, XGBoost, and other methods, the key characteristics of fraud were further explored and analyzed, and the fraud law was figured out. L. B. Liu (2019) put the desensitized data of 8.36 million real medicare visits from 253 hospitals in China into the experimental test, the results show that the ACC of the model is 0.86, the AUC is 0.81, and the fraud sample

identification rate is 82%. We effectively identify fraudsters when the feature dimension is only 223.

Furthermore, through the analysis of the key characteristics, it is found that cost is the most important indicator of fraud. On the approval amount, from the total amount to every single case, every single time is different. Although the overall fraud amount is high, every single amount is relatively normal, so the fraud can be committed on the normal reimbursement amount, time by time, breaking up the total amount into many reimbursements and many times, to get a higher approval amount of medical insurance on the whole. In terms of the details of various expenses, drug fees and treatment fees accounted for the highest proportion, and the treatment fees of the fraudsters were higher than those of the normal insured people, so their fraud means were to defraud the medical insurance fund by false diagnosis and treatment. The higher the ranking of hospital characteristics, the more serious the medical insurance fraud in China. In general, there are some characteristics of medical insurance fraud: it is difficult to commit insurance fraud alone, so it is usually completed by cooperation between medical service providers and medical insurance providers; the participants neglect the legal and proper use of medical insurance card; there are problems in medical insurance system, hospital management system, financial supervision and so on, which give provisions for medical insurance fraud.

The Social Insurance Law, enacted in 2010, vaguely identifies administrative and criminal liabilities for health insurance fraud. Administrative liability includes ordering returns, penalties, and termination of service agreements, while criminal liability does not have corresponding provisions in the Criminal Law, and there are also certain problems in the interpretation of article 266 of the Criminal Law by the Standing Committee of the National People's Congress in 2014. This leads to lower criminal costs of Medicare fraud. Therefore, it is necessary to strengthen the cohesion between social insurance laws and administrative law, civil law and criminal law, improve the relevant terms and conditions of criminal law, and severely crack down on medical insurance fraud. According to the principle of limitation in criminal law, when a certain behavior that endangers society cannot be effectively stopped and controlled by moral, administrative, and economic means; it is necessary to consider formulating, modifying, and supplementing relevant criminal norms to regulate and increase the criminal cost of insurance fraud (C. B. Luo, 2016).

But it is worth mentioning that the criminal punishment of medical insurance fraud is not the end, but an effective means to raise the awareness of the integrity of all parties and eliminate illegal acts. Medicare fund regulation is a multi-departmental work that requires cross-

departmental collaboration. The medical security department bears the primary responsibility. Departments of development and reform, health, market supervision, drug supervision, auditing, and justice must perform their duties under the division of functions, establishing a coordination mechanism to form a pattern of source prevention and control, process supervision, and strict punishment for insurance fraud.

At the technical level, how to use big data to develop more effective management measures?

1.4 Definition of research objects

According to this research topic, the research subject can be defined as several stakeholders. It mainly analyzes medical insurance data.

On the content, big data refers to an extensive collection of data (volume), hard to manage, both structured and unstructured (variety), generated, collected, and processed fastly (velocity). Big data is so large and complex that no traditional data management method can store and process it efficiently. It is developing into a new generation of information technology and service industries - from which - new knowledge and value are created, and new capabilities are enhanced. In medical insurance data, digital content mainly includes insurance data, payment data, management data, client data, and complaints. The insurance data is primarily the basic personal information of the insured, such as age, employment, income, and family information. The payment data contains information on the insured patient's medical services, which can record his medical behavior more thoroughly, a more critical and unique data process in the medical insurance information system. The management data is mainly based on the existing insurance data and payment data to extract data that can support the decision-making and management of medical insurance, including information on the income and expenditure of the fund at all levels, costs, and diseases. Other data, like socially shared data and some content, provide scientific support. Currently, medical information systems are not unified by the central government but are automatically developed locally, including the information system between hospitals and organizations. Therefore, the progress of China Medical Social Insurance Fund in the overall data upgrade and standardization unification is relatively slow.

With the above problems and the current situation of data decentralization, the research object is clarified mainly as the big data in the medical insurance fund payment system in Huzhou City, Zhejiang Province. Totally 2.4 million insured people in Huzhou are enrolled in this analysis project, covering outpatient clinics, hospitalizations, and pharmacy data from 2017 to April 2018.

1.5 Research contributions

The fraud supervision, prevention, and control of medical insurance funds are not new, but research on a common issue is continuously updated. The importance and originality of this study are that it explores new technologies, new management theories, and new regulations in the social development needed to gradually resolve this issue, especially in the Chinese medical insurance system covering the world's largest population.

This study achieves good analysis and supervision results via big data mining technology and in-depth physical population persona labeling.

1.5.1 Academic contribution

First, through qualitative investigation and systematic bibliographic research - this dissertation sums up the connotation and performance of medical insurance fund fraud and sorts out its characteristics - unveiling the root causes of medical insurance fraud problems that are difficult to eradicate and construct a conceptual framework based on Chinese social security insurance fund fraud.

Second, this study is exploratory and interpretative. Embracing the related theory of quasi-public goods and medical insurance-related theories, it analyzes the medical insurance cost control management model and the risk based on big data mining resources. It then puts forward the technique application in medical insurance cost control. Thereby it carries out the practical application and analysis of big data mining in the medical insurance cost control model, for the first time clarifying the goals and principles of the system design and the practical application of big data mining in the medical insurance cost control model. During the process, it provides the application environment and tools for big data mining, making full use of the data, and finally analyzes the practical application efficacy of big data mining in medical insurance cost control.

Third, this study presents available methods of audit and payment works for current medical insurance cost control and expenditure in China. Thereby audit of medical insurance costs and detection of illegal use of medical funds can be performed quickly, proving the role of medical big data mining in medical insurance cost control.

1.5.2 Management contribution

First, this study explores safeguarding the social benefits. It focuses on medical insurance fund management to prevent unlawful and fraudulent use, reduce expenditure, enhance the reasonable and legal use of medical insurance funds, and ease the economic burden of patients.

The results of this dissertation can be adopted to ensure the rational use of medical insurance funds and control the payment of medical insurance expenditures. Also, it provides methods for medical insurance fund expenditures cost control and institution management and presents references for organizations that formulate policies for the use of medical insurance funds.

1.6 Research route and contents

This study examines questions and conducts research according to the methodology. The research route is as follows.(See figure 1.1)

1.6.1 Proposing questions

First, the background of China's medical insurance fund is explained. With the acceleration of the construction of medical insurance and healthcare information system, the number of the insured population and management service objects surged, which puts forward higher requirements for the operation and supervision of the medical insurance fund. Thereby it puts forward the analysis of the risks and fraud of the current fund operation, comes up with practical management suggestions for it, and defines the relevant concepts, research objects, and methods.

1.6.2 Literature overview

This chapter describes the operation mode of China's medical insurance fund, the fee control model, and foreign medical insurance fee control. It discusses the existing problems of social medical insurance, the current situation of social medical insurance anti-fraud, the causes of social medical insurance fraud, the research on social medical insurance fraud assessment, and the methods of evaluating fraud.It also analyzes the research status of social medical insurance in China, the basic medical insurance system for urban workers, the fairness theory of social medical insurance, the welfare economics theory of social medical insurance, the theory of public goods of social medical insurance, the theory of externalities of social medical insurance, the efficiency theory of social medical insurance, the theory of producer surplus and consumer

surplus of social medical insurance, the information asymmetry theory of social medical insurance ere.

At the same time, the study establishes and analyzes the management model of medical insurance big data, then focuses on technical analysis methods based on big data technology, especially the important technical means of fraud research. Finally, the research analysis model was drafted, and the literature review of the analysis technology field in the model was reviewed (J. J. Zong, 2018).

1.6.3 Research method

Through literature review, investigation and research, data mining, and other research methods, the study uses the technical programs of data mining on research objects. Following the application of big data analysis methods in the medical insurance policy and the medical insurance cost risk control model, it gets the data analysis results and obtains valuable research findings. The data analysis results are re-mined by the entity portrait algorithm, and the conclusion of entity portrait population labeling under different data mining result rules is formed.

1.6.4 Research data collection and interviews

This chapter collects the medical insurance data of the insured person in Huzhou City, Zhejiang Province, from 2018 to 2019 (50365776 outpatients, 598204 inpatients). The data covers all the records of social insurance participants in Huzhou City, including outpatient clinic visits, hospitalization, and drug store purchase (Data Source: Huzhou Municipal Medical Security Bureau, Zhejiang Province).

It surveys 16 medical institutions in Huzhou through interviews. Hospital presidents, doctors, patients and family members are all interviewed by giving personal and institutional interpretation opinions on different channels of health insurance violations, the frequency and scope, and the severity of the violations.

1.6.5 Fraud detection based on big data model

In this chapter, a complete model was established for medical records in Huzhou City, from January 2018 to April 2019 to analyze the fraudulent use of medical records. For example, falsified inpatient records, the medical insurance card of hospitalized patients appeared in other

pharmacies to fill medicine, and men's medical insurance cards used to prescribe gynecological drugs for women.

According to the literature review and technical data mining programs, data analysis results and valuable conclusions are obtained. Moreover, based on the data analysis of the entity persona algorithm, the data analysis results are re-mined through the algorithm, and the entity portrait population labeling conclusions under the conditions of different data mining result rules are formed.

1.6.6 Solutions and advice

Given the status of medical insurance research at home and abroad, this chapter uses research methods such as literature review, investigation, and data mining. It summarizes and analyzes the fraud scenarios and characteristics of illegal participators based on the healthcare insurance data of Huzhou City from 2018 to 2019. Then it explains the reasons for fraud occurrence, qualitatively discovers the data characteristics of various types of medical insurance fraud, clarifies the characteristics of the fraud behaviors, constructs a data model of medical insurance fraud, proposes a medical insurance fraud prediction identification factor. Z. W. Zhu (2017) discusses the results. Through the efforts above, the study combed the research thoroughly for management advice. It concluded the academic and management significance of the research, and the research limitations, shedding light on the direction of future research.

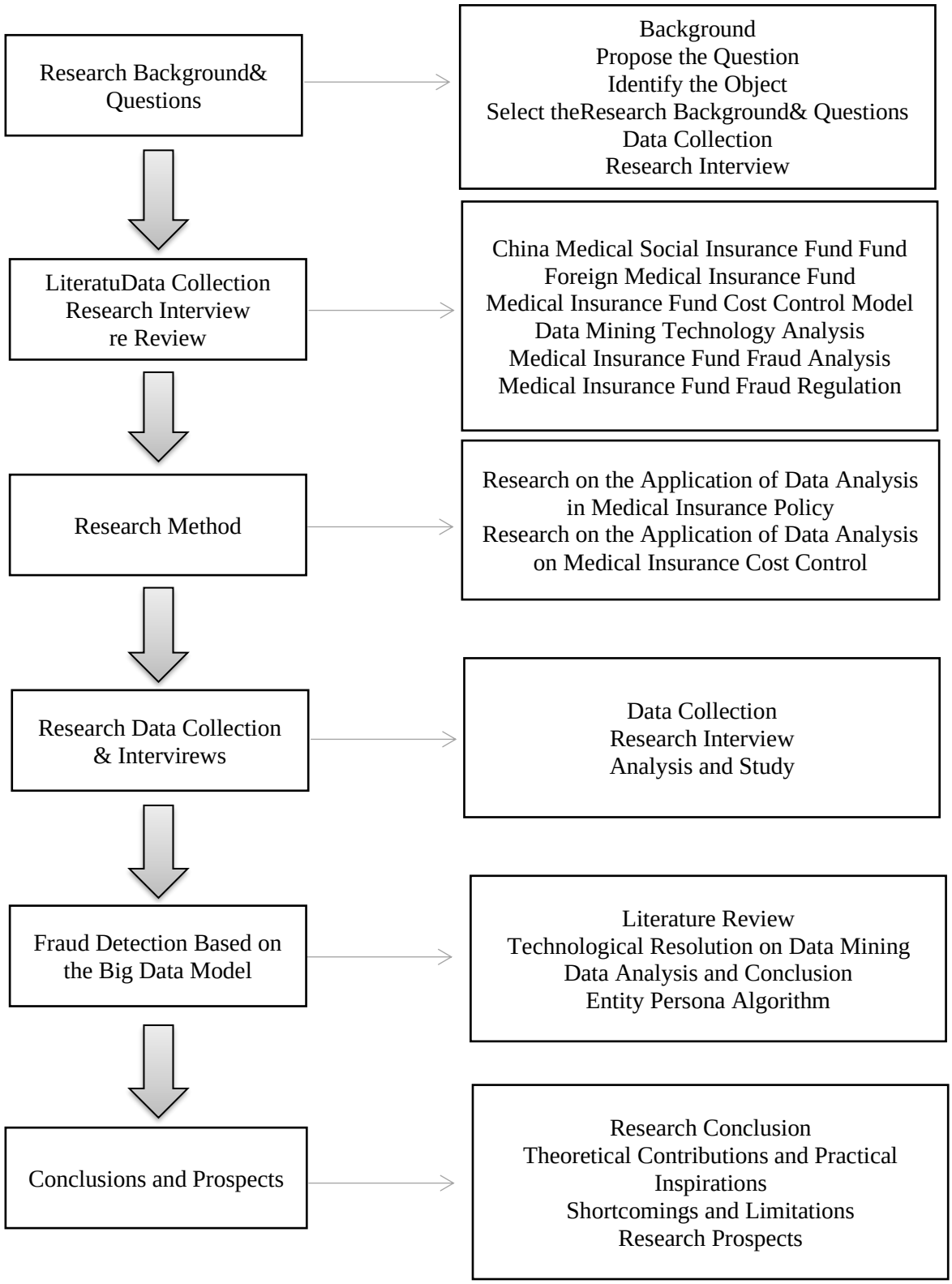


Figure 1.1 Research route

Chapter 2: Literature Review

2.1 A brief history of social security insurance

Social medical insurance refers to a social security system enforced through State legislation. It is jointly financed by employees, enterprises (employers), communities and the state. Establish a social insurance fund, which provides basic allowance for workers or their immediate family when they are disabled to work or temporarily lose their job (for oldness, sickness, unemployment, work-related injury, maternity). Thus it guarantees the primary livelihood of workers, maintains social stability, and promotes economic development (Blank, 2020 ; Z. B. Chen et al., 2022 ; Jiang, 2018).

Dawson (1912) studied social insurance in Germany from 1883 to 1911 and compared its history, operation, and results with the contents of the Insurance Law of 1911. In 1883, German enacted the world's first law on social security called Labor Sickness Insurance Act, starting the social insurance legislation. The Act has followed by France, England, the United States, and other countries, especially major industrial countries, which had developed a complete system of legal social security. The Act made it compulsory for workers in certain industries whose wages are less than the limit to joining the health insurance fund, including mandatory levies on workers and employers. That marked health insurance as a compulsory social insurance system comes into being. After that, German promulgated a series of relevant regulations and kept updating and refining those laws. The Work-related Injury Insurance Act of 1884 and the Old Age and Disability Insurance Act of 1889 marked the institutionalization of social security (S. F. Liu, 1995). So far German's medical insurance system is underpinned by three pillars, social meical insurance, long-term care insurance and commercial medical insurance (including commercial long-term care insurance), featuring legal social medical insurance as the main and its compulsion. For this sake, after 2009, German required all German citizens and long-term foreign residents to participate in health insurance and long-term care insurance program (Greenger et al., 2013). The statutory social medical insurance is compulsory for low-income people to participate, covering many people and a wide range of responsibilities. The funds are mainly pooled from wages and salaries and are administrated by the national health fund. The statutory social medical insurance allows most residents -

company employees, college students, farmers, retirees and the unemployed - to participate, and low-income people are also required to participate. Due to its mandatory requirements, it covers more than 90% of the population in Germany (T. Yu, 2005). S. F. Liu (1995) points out that since the late seventy's of the last century, thanks to the high-speed development of the economy, the country was gradually moving from a closed backward agricultural towards an open modern country. Especially since the ninety's, the transition from a planned economic system to a socialist market economic system, society has dramatically changed, and the traditional social norms, social values, and cultural systems can no longer adapt to the changing modern society. Moreover, the new social order, norms, values, and cultural systems have not yet been formed. So much undesirable disorder phenomenon make the public suffers, resulting in a lot of social problems such as public nuisance, crime commitment, ecological destruction, food safety issue, social climate change, weathering, public security, unemployment, traffic jam, aging population, labor disputes and so on" In order to solve the social problems and to prevent such social problems from happening again and developing furtherly, it requests emphasis on balance the personal life safeguard with the whole social stability, carry forward social security legislation, or social security system, it can relieve people's worries, keep the society in good order, to promote the development of production.

The theory of the state welfare put forward by Schmuller and Brentano in 1970s believes that in addition to maintaining social order and national security, the state also has a "cultural and welfare purpose", that is, the implementation of social insurance, the development of public education, the improvement of health. As the European industrial revolution began to move toward capitalism, the peasants lost land and transformed into the proletariat; the pension problem also came into being, changing the pension model from a traditional intergenerational model for thousands of years to a socialized model. Schmuller challenged the mindset of raising children to support old age left by feudal society, which was the progress of human civilization and a historic change. The pension system originated in Germany, establishing the world's first official pension system. In 1872, in his speech at the Eisenach Conference of the School, Schmuller pointed out the need to reform the economic proposition and the existing forms of production and the upbringing and psychological state at all stages of society. They also advocate strengthening state functions, emphasizing the legal system's importance, the economy's relevance and ethics (Q. Liu, 2013). Germany's five major insurance plans for pension, medical, unemployment, work-related injury, and long-term care all stipulate a host of family-friendly policies. These policies facilitate the achievement of the objectives of the German family policy and are in line with the needs of the development of the German social

insurance system. Studying and learning those policies will significantly benefit China's construction of a family-friendly social insurance system and the overall two-child policy.

China's Social Medical Insurance has undergone three historical stages: from 1978 to 1997, 1998 to 2008, and 2009 to 2018. The 1978-1997 period was characterized by parallel transformation for reform and development, the 1998-2008 period was the game between medical insurance and providers, and the 2009-2018 period was healthcare reform and overall management beyond the game (J. T. Liu, 2019). Xiang (2013) analyzes the development-oriented social policy, introduces the correspondent ideas of famous foreign scholars, and puts forward the principles and paths for constructing China's development-oriented social policy. (Y. P. Tian, 2010) discussing the relationship between social policy and social theory, proposes the development path of building a social policy system in China in the transition period. Y. Du and J. Wen (2010) introduced the development path of social policy from welfare-oriented to asset-oriented, and came up with a social policy construction path that integrates social welfare policy and economic development. There is also a review of the trajectory of China's social policy development since the reform and opening up, summarizing the characteristics of China's social policy development and discussing the trend of China's social policy (Xiang, 2010). P. He (2017) conducted a study on the development and reform of the German social insurance law, and proposed that the original path is to adjust and amend the derailed social legislation, to re-guide the principle of social solidarity and insurance, readjusting and strengthening the decentralization system of insurance organizations with risk structure sharing mechanism and open insurance organization competition, subsidizing the external burden of social insurance with federal subsidies, and sorting out the national tasks achieved by social insurance. Germany Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (2013) points out that since the 20th century, Germany has emphasized the autonomy and priority of private non-profit welfare groups and the autonomy of social security personnel.

F. C. Zhao (2005) studied the Social Security Act deliberated by the Legislative Council in the United States in 1935, which was internationally the first time to eye social security in the strategic horizon of national social policy. The social security system became a policy tool for the state to intervene in the redistribution of national income, promote the effective needs of society, and drive economic growth. The study discusses the necessity of existence and the design of social insurance, such as emphasis on insurance or welfare, pay-as-you-go and cost control, payroll tax, coverage, mandatory property. Roosevelt in 1933 indicates that a government should not exist if cannot take good care of the elderly and the sick, cannot provide jobs for the strong, cannot involve young people into the industrial system, and allows the

shadow of insecurity to hang over every family. The purpose of the bill is that the federal and states should share the responsibilities for social security. The policy not only acts as the means to cope with the unpredictable disasters and disruptions for the protected social members, but also alleviates the serious social problem facing the United States at that time – poverty. Thus, it eases social contradictions and advances the recovery of social productivity. Therefore, the act played an important role in recovering the US economy as a safety valve and stabilizer for social development. At the same time, regarding the development of the US social security system, the Social Security Act of 1935 is a milestone in the history of welfare in the United States (Manchester, 1974).

Samuelson in 1945 proposed that medical insurance products belong to the category of common quasi-public goods, which are typically effective and congestible, so if the government provides free medical services or charges nominal fees, people will over-consume the product, which aggravates the congestion situation. In conclusion, the theory of social welfare is Samuelson's classic theory as a welfare economist, and its ideas are mainly embodied in the theory of fairness and efficiency.

W. Liu (2004) by analyzing the concepts of fairness and efficiency of the neoclassical economic school, and Keynes's view of fairness and efficiency, it pointed out that Samuelson's concept of fairness and efficiency is not a simple compromise between the above concept and views (Miao, 2015). In fact, Samuelson advocates efficiency first, when fairness and efficiency conflict, excessive pursuit of the fair result will hinder the development of the economy, so it is better to take economic development as the ultimate goal (Xiao & Shen, 2010). Miao (2015) indicate the source of Samuelson's thought on economic ethics, and analyzes Samuelson's economy ethics from equal stress on science and conscience, efficiency and fairness, economic development and citizen well-being, policy propositions in economic development and improvement of people's livelihood, and finally they make a brief evaluation of Samuelson's economy ethics.

With the advancement of industrialization, countries continue to introduce medical insurance systems to meet social needs, reduce social risks, and maintain social stability. In addition, with the development of social politics and the economy, the social security system work has gradually become the focus of the governments. Under the rapid economic development, the widening gap between the rich and the poor is impacted by their respective historical traditions, cultural backgrounds, economic levels, social development, and many other factors. And economic globalization and digital technology will further aggravate the income gap between developed and developing countries. Eventually, the widening gap

between the rich and the poor will deeply affect the evolution of the international political and economic pattern.

China's social medical insurance fund can also be explained by its general and special definition. Generally, the medical insurance fund should include the medical insurance fund under all medical security systems, that is, the basic medical insurance fund for urban employees managed by the medical insurance administration, the basic medical insurance fund for urban residents, and the new agricultural cooperative fund managed by the ministry of health administration. Specifically, the medical insurance fund usually refers to the primary medical insurance fund for urban employees and urban residents managed by the medical insurance administration, not including the new rural cooperative medical insurance (G. Qu, 2014).

2.2 Problems of the social security insurance

2.2.1 Advantages in terms of policy

The social security insurance is a system enforced by national laws and government decrees, to ensure that employees can obtain substantial assistance from the state and society for temporary or permanent work incapacity due to birth, old age, illness, death, injury, disability and unemployment. It is an administrative social insurance system established by the state under the provisions of the Constitution and is backed by finance with the surplus of insurance funds used as a reserve. The insurance participants have permanent protection rights and no suspension of service since the losses incurred are made up by the state allocation. Therefore the social insurance can play the role of social stabilizer. When considering some objective phenomena such as: oldness, weakness, sickness, disability, pregnancy, and incapacity to work among social members that cannot be avoided in any epoch and under any social system; the purpose of social insurance is to give appropriate compensation to ensure the basic living standard by accident of the above situations, thereby preventing the emergence of insecurities. Social insurance ensures the smooth reproduction of social labor. Since the laborer is vulnerable to various accidents in the workplace, resulting in the stagnation of reproduction, they should be given the necessary economic compensation and allowance to restore the workforce in case of accidents, as mentioned earlier, happen.

There are many ways to think about the positive effects. We will explore three of the most relevant, each providing different insights into the function of social security insurance.

Economic perspective: Participation in security insurance brings about many benefits. First, individual employees only need to contribute 8% of their wages every month, adding 5% of their wages paid by the enterprise (the proportion increasing by 1% in each payment year later and remaining unchanged after reaching 11%). Then employees can see the state establishes a personal social insurance account for them in proportion to 11% of wages (the account is owned by an individual and can be inherited and transferred). As the amount of personal account is 3% more than the individual contribution, they can pocket ten to dozens more yuan per month. Second, social security premiums do not charge interest tax and are more economical than bank deposits. Third, if the accumulated payment years are more than 15 years, and the official retirement age is reached, it can be applied to enjoy monthly pensions. Fourth, pensions only increase while the social living standards rise, so that life after retirement is guaranteed, and the more contributions are paid by the employees, the more they enjoy.

Social environment perspective: On one hand, today, the whole country has formed a general climate of strengthening the social insurance system, and from the State Council to the provincial government have successively promulgated the "four rules and one regulation", which clearly stipulates the mandatory participation, making participation in social insurance a necessity. In the past, some thought social insurance was not guaranteed and did not know whether the money could be claimed when it was payable. This concern can now be eliminated, as the State has comprehensively improved its social insurance mechanism, backed by strong financial muscle, to ensure that the insured can enjoy it as scheduled. On the other hand, social insurance is a compulsory social system, which is an important part but also the guarantee of the market economy system, so all countries in the world, especially developed countries, have generally established a perfect social insurance system. Some countries also include social insurance in the scope of tax collection, strengthening the social insurance system as a major measure in line with international standards. Moreover, the law has further intensified penalties for those who refuse to participate in insurance or those who owe insurance premiums, to ensure that laws are observed and strictly enforced.

Social fairness perspective: Social insurance serves social equity. The literacy divide and labor ability differences cause the income gap. Social insurance funds can be channeled from compulsorily levied insurance premiums to subsidize workers with low incomes or no sources of income, improve their living standards, achieving fair public allocation to a certain extent. Social insurance helps promote social progress. Also, social insurance mobilizes mutual assistance, a better expression of mutual assistance and cooperation.

2.2.2 The fraudulent behaviors in the implementation of social security insurance

There are many fraudulent acts in implementing medical insurance (K. An, 2020). Medical insurance fraud refers to the illegal reimbursement of expenses and the deception of medical insurance funds by Chinese citizens, legal persons, or other organizations in the process of participating in medical insurance within the framework of the system stipulated by laws and regulations (Z. H. Huang, 2020), and the specific means include but are not limited to: false medical treatment, false drug purchase, falsification of materials, concealment of the truth, impersonation (Y. Zhao, 2020). With the population and deepened implementation of China's basic medical insurance, medical insurance fund fraud incidents also occur at the same time. Some institutions or individuals do the wheeling and dealing, cash out from the medical insurance funds by forging cases, false hospitalizations, fabricating information, bills, impersonation, making up stories. Currently, there are growing problems in managing and using China's medical insurance funds. Medical insurance fraud management has become increasingly important in implementing China's basic medical insurance.

The fraudulent behaviors of medical insurance involve several parties and related misbehaving defined on them: First, the fraud of medical insurance-insured patients (demand side); Second, fraudulent acts in collusion between providers of health insurance services (medical institutions and pharmacies) and insured patients; Third, fraud by the provider of medical insurance services (X. Liu, 2008). So we analyze the subjects that may be involved in Medicare fraud and the manner of fraud and violation as follows:

(1) The insured obligor or legal beneficiary medical insurance applicant—the insured person, including units, organizations, urban workers, urban residents, and new rural cooperative insurance participants and other individuals, the insured obligor needs to pay for the received diagnosis and treatment services and purchased drugs and equipment in the middle of medical behavior. The legal beneficiary of medical insurance, the insured person and his relatives, refers to the individual who can obtain benefit compensation by participating in medical insurance. Medical insurance obligators and legal beneficiaries are the most important and direct stakeholders in the medical insurance system. To maximize their interests, these groups of people have illegally embezzled the medical insurance fund and benefited from it by forging Medicare cards, using imposters, forging invoices, falsely making out issuing fee bills, and over-prescribing drugs (F. Y. Yu, 2018).

(2) The medical insurance administration department and its staff. A medical insurance agency refers to the institution handling the whole process of the social security fund in the

medical insurance system. The fraudulent means mainly include forging false certification materials to make unqualified enrollment or participating in the process of fraudulent personal treatment (F. Y. Yu, 2018).

(3) Medical insurance service organizations and their staff. Medical insurance service organizations are organizations authorized to provide medical services. They are the service provider and economic organizations that master the medical services expertise, such as hospitals and pharmacies. The hospital's fraudulent means mainly include: falsifying hospitalization records; falsely making out medical expenses. Increasing the bill of treatment for patients through overtreatment, privately using the insured's information to falsify medical records to cash out, and deliberately changing non-reimbursed items to make a claim. The fraudulent methods of pharmacies mainly include: deliberately not verifying the identity of the insured; no control of prescription medicine; neglecting the compliance of medical insurance card use; including non-Medicare drugs and even daily necessities in the list of medical insurance; stealing the medical insurance cards to cash out of personal accounts (F. Y. Yu, 2018).

(4) Fraud gangs. Medical insurance fund fraud gangs refer to criminal rings illegally occupying medical insurance funds through organized and fraudulent practices. Their fraud methods mainly include: the illegal reselling of medical insurance drugs; assisting the insured to lend the medical insurance card to others to defraud the fund; using the identity of others to seek medical advice and prescribe drugs; colluding with prescribing physicians to forge medical records and hospitalization certificates; complicit with pharmacies in stealing a large number of medical insurance cards (F. Y. Yu, 2018).

2.3 The Status quo of the social security insurance anti-fraud

2.3.1 The anti-fraud practice in China

Depending on the overall anti-fraud practice of medical insurance funds in different parts of China, at this stage, the current practice is devoted to the following aspects of medical insurance fund fraud.

(1) The construction of laws and regulations: the laws or regulations implemented earlier in China regarding the anti-fraud of medical insurance funds mainly include the Social Insurance Law of the People's Republic of China, Provisional Regulations Social Insurance Premiums, and the Labor Security Supervision Regulations. In 2016, in order to strengthen the management of social insurance fraud cases, the Ministry of Human Resources and Social

Security formulated the Measures for the Management of Social Insurance Fraud Cases, which requires social insurance administrative departments to establish a standardized and effective management system for social insurance fraud cases, and strengthen the scientific, standardized, and whole-process, information-based management, vigorously crack down on medical insurance fraud and insurance fraud. Each region has established corresponding medical insurance anti-fraud management measures according to the characteristics of regional management, led by human resources and social security departments, and jointly implemented by public security departments, financial departments, and health departments. However, at present, China still lacks perfect laws and regulations for anti-fraud medical insurance, and there is no clear and complete definition of medical insurance fraud. There are no relevant clauses, charges, or basis for conviction and sentencing in Criminal Law.

(2) Medical insurance management system construction: the standardization construction of various medical insurance systems in China has formed a relatively uniform fund payment standard and cost control standard. By standardizing medical service behavior, strengthening the management of the use of the Diagnosis and Treatment Catalog, Drug Catalog, and built-in materials, formulating scientific settlement methods and reasonable fixed settlement standards; in solving the problem of settlement of medical expenses in different places, most of them have completed the provincial level. The system is unified, and the system of provincial-level regional settlement has been unified. Unify the medical treatment and drug catalogs of the three different medical insurance systems for employees, residents, and new rural cooperative medical insurance, providing a basis for the settlement of medical treatment in different places nationwide. A credit evaluation system for medical institutions has been preliminarily established, and designated medical institutions are subject to supervision, inspection and assessment. The amount of medical insurance budget and the number of supervision times vary according to different levels. However, there is still a lack of an integrity assessment system for insured individuals and measures to manage the untrustworthiness of insured individuals.

(3) The effectiveness of anti-fraud management work at the technical level also depends on the progress of medical insurance anti-fraud technology. At present, some regions in China have begun using anti-fraud software systems to identify, summarize and analyze medical insurance fraud. However, there is still a lack of a real and effective platform for sharing electronic diagnosis and treatment information, and the anti-fraud software system needs to improve the identification accuracy of suspected cases, making the practicability of the invested system difficult to meet expectations (C. Tian, 2014).

Medical insurance fraud is a complex problem involving many factors, such as groups, people, and behaviors. Its identification and prevention are faced with a lot of problems and challenges in regulations, systems, and technologies. To sum up, there are mainly the following aspects: First, the composition of the people and groups involved in medical insurance fraud is complex. The persons and groups involved include insured persons, medical institutions, employers and professional fraud gangs. Often the insured person and the personnel of the medical institution jointly defraud under the condition of convergence of interests. There are even collective behaviors organized and led by medical institutions, such as falsification of etiology, falsification of prescriptions, falsification of medical details, falsification of hospital beds, making identification very difficult. Second, the methods of medical insurance fraud are complex, and there are many types. According to statistics, there are as many as 30 types of medical insurance fraud methods, such as false medical qualifications, false medical treatment qualifications for outpatient thresholds, and false annual medical expense balances of insured persons, false hospitalizations using patient identities, false causes, and non-medical insurance payments. With the advancement of technology and the temptation of illegal income, it is believed that new medical insurance fraud detection methods will appear. Third, medical insurance fraud is hidden and difficult to identify. With the increase in administrative penalties for medical insurance violations and frauds in recent years, fraud has been suppressed, but the fraudulent methods are constantly changing and renovating, and it is challenging to detect serious fraud problems only by relying on independent verification by medical insurance institutions. Due to the lack of unified diagnosis and treatment standards, admission and discharge standards, hospitals and doctors may deliberately reduce hospitalization conditions and deceive medical insurance institutions. When medical insurance management institutions identify whether medical institutions, doctors, and insured personnel are fraudulent, it is sometimes difficult to determine. Fourth, the basic data format of medical insurance is not uniform, and the current utilization rate is not high. Currently, the level of medical insurance informatization across the country is inconsistent, and the establishment of medical insurance systems in different regions is inconsistent. Some regions are centralized at the city level, and some are centralized at the county level. The standards and formats of basic data are not uniform, and the current accumulated data utilization rate is not high (X. C. Gu, 2017a).

All of the above have created obstacles to the large-scale use of medical insurance data for big data research and analysis. To sum up, although researchers and contents of similar fraud topics have achieved many research results, the research mainly stays at the system design level and lacks empirical research. The research on anti-fraud related issues in foreign developed

countries is already in the stage of application of technology developments, and the institutionalization of fraud is also more complete. Therefore, this thesis focuses on the technical research concerning medical insurance anti-fraud, including medical insurance anti-fraud models based on big data, medical insurance anti-fraud technical system research, and so on. At the technical level, we will discuss effective medical insurance anti-fraud methods and models, and provide technical support for establishing and improving the medical insurance system (X. C. Gu, 2017 ; M. Q. Sun, 2020).

With the frequent occurrence of medical insurance fund fraud incidents, various medical insurance anti-fraud actions have been studied and carried out in all parts of China. Medical insurance measures have been formulated and introduced in terms of policy formulation, supervision, analysis and monitoring, punishment, to identify and prevent medical insurance fraudulent behaviors.

2.3.1.1 The practice of Tianjin City

In response to medical insurance fraud, the Tianjin Municipal Labor and Social Security Department has formulated new measures to prevent medical insurance fraud and form a long-term mechanism for medical insurance "anti-fraud". Implement a creditworthiness evaluation system for designated medical establishments, and announce the evaluation results to the public. For designated medical institutions with high integrity levels, as excellent cases, provides special channels to deal with their newly launched diagnosis and treatment projects. For designated medical institutions with poor creditworthiness levels and all kinds of doctors who violate medical insurance regulations, a gray list system will be established, and the fixed medical institutions and doctors included in the gray list will be punished by stopping their service agreements and refusing to reimburse their prescriptions. Strengthen the management of outpatient medical treatment in designated medical institutions by implementing the medical insurance outpatient medical manual system. Insured persons must present a manual for outpatient medical treatment; otherwise, they will not be eligible for insurance treatment. The number of drugs prescribed at a time is strictly limited by the prescription drug dosage specifications. For violations of the above essential regulations, the handling agency will gradually upgrade the means of handling, from inquiries to the system-wide notification to refusal to pay. Further improve the list of prescription drugs. Guide patients to use drugs rationally, guide doctors to prescribe drugs reasonably, give priority to the use of domestic drugs, and prevent the excessive use of expensive imported drugs. Set the proportion of personal self-responsibility for class B drugs in the catalog, and increase the restrictions on prescribing drugs

according to the hospital level, indication type and other conditions for drugs with higher unit prices. By improving the settlement method of medical insurance costs, implementing different payment settlement methods, and adopting double-entry settlement combined with different methods such as total prepayment, single-disease payment, and per-project payment, the efficiency of the use of medical insurance funds is improved. Through the computer network settlement of outpatient medical expenses, real-time monitoring of outpatient medical treatment is realized, and the insured person's outpatient medical treatment realizes card consumption. Increase the intensity of supervision and inspection of designated medical institutions. Through daily supervision, special governance, and around the phenomenon of unreasonable expenses and malicious insurance fraud, strengthen the rectification of community medical service institutions.

2.3.1.2 The practice of Yunnan City

The Department of Human Resources and Social Security of Yunnan Province, in communication and cooperation with the Provincial Development and Reform Commission, the Provincial Public Security Department, the Provincial Department of Finance, the Provincial Health Department, the Provincial Food and Drug Administration and other departments, jointly formulated the Yunnan Provincial Medical Insurance Anti-Fraud Management Measures, promulgated and implemented it. The Measures clearly define the scope of medical insurance fraud, and the targets of the restrictions cover citizens, legal persons, or other types of organizations, and all kinds of arbitrage of medical insurance funds are applicable. The Measures for the Administration of Medical Insurance Anti-Fraud in Yunnan Province apply to various medical insurance anti-fraud work, such as basic medical insurance for urban workers within the administrative region of Yunnan Province. Seek medical treatment to achieve card consumption. Increase the intensity of supervision and inspection of designated medical institutions. Through daily supervision, special governance, and around the phenomenon of unreasonable expenses and malicious insurance fraud, strengthen the rectification of community medical service institutions.

2.3.1.3 The practice of Fuzhou City

Fuzhou Bureau of Human Resources and Social Security strengthen cooperation with Fuzhou Public Security bureau, both signed the memorandum of the anti-fraud cooperation on medical insurance fund, joining law enforcement to crack down on the medical insurance fund fraud. The cooperation includes intelligence interaction, consultation on major issues and establishing joint prevention and control mechanism, to strengthen the social insurance anti-fraud. The

Fuzhou Bureau of Human Resources and Social Security established reporting rules of to-do list work, cooperating with the economic investigation team of the Public Security Bureau, to further intensify the anti-fraud work within the system. The Economic Investigation Detachment of the Fuzhou Municipal Public Security Bureau paid more attention to the related work involving medical insurance fund fraud and adopted zero tolerance for economic crimes in social security. Both parties shall specify the working institutions responsible for the linkage, establish a list of persons responsible for contact, and regularly inform the information resources sharing. The Fuzhou Municipal Bureau of Human Resources and Social Security shall improve the pre-prevention and in-process control of fraud by establishing and improving the information monitoring system. Public security departments shall organize and carry out special control activities from time to time. For the cases that have a large social impact, the two sides jointly carried out case analysis and early warning training. The cooperative law enforcement actions of the two departments effectively maintained the orderly progress of the medical insurance work.

2.3.1.4 The practice of Zhenjiang City

The Zhenjiang Municipal Human Resources and Social Security Bureau, the Public Security Bureau, and the Health and Family Planning Commission jointly launched the "show sword" operation of medical insurance anti-fraud. Through the leadership of the municipal government, jointly crack down on illegal acts of defrauding medical insurance funds, focusing on checking the medical reimbursement behavior of insured persons and the medical insurance service behavior of designated medical institutions, designated retail pharmacies, and medical insurance practitioners. During the inspection, the focus was on the inspection of large-scale reimbursement of the insured personnel and high-frequency reimbursement of single persons; and, through on-site verification, supervised whether there were various acts such as using the medical insurance card to make profits or defrauding the medical insurance fund, and deal with it according to law.

To sum up, despite the continuous improvement of the system, the Internet is full of fraud cases. Just searching for the field of medical insurance fraud in scientific article repositories, we found: 57 articles in CNKI, 284 articles in Wanfang, and 509,000 articles in Baidu Academic. Several experts have made great efforts to solve medical insurance fraud and provided some solutions that Government departments implemented. Although various anti-fraud work practices have been carried out, the effect is still not ideal. The theoretical and practical research on anti-fraud is still a huge social problem.

2.4 Research on the causes of social security insurance fraud

2.4.1 Cause of medical insurance fraud

L. Einav and A. Finkelstein (2018) put forward a clear concept of moral hazard in medical insurance, he believes that medical services is highly specialized, thus patients can hardly get as much knowledge as the healthcare providers, which can easily lead to excessive medical services and unnecessary service supply. Therefore, in order to fundamentally eliminate and avoid the problem of insurance fraud, the insurance contract should be improved and amended. M. Feldstein (1970) and R. G. Evans (1974) argue that providing too many health services may be due to the provider's advantages in having the Internet and information technology, or it may be that the patient is asking for additional demand. Fuchs (1978) and Dionne (2000)believe that information asymmetry is one of the causes of insurance fraud, and finds that socio-economic development is closely related to information asymmetry, and uneven resource allocation is an important reason for promoting the rapid growth of insurance fraud. Cooter and Ulen (2008) from a legal perspective on the causes of health insurance fraud, the study found that the perpetrator will regard the fraud as a venture capital behavior, and when the perpetrator's own cost is lower than the legal cost, the perpetrator's enthusiasm for fraud will increase. The study developed by Menon (2018)concluded that incomplete partnerships between health care providers and hospitals have led to an increase in health care fraud. L. Y. Li (2011) argues that the main reason for insurance fraud is that the benefits of fraud are more significant than the costs paid by the insured. From the theoretical and empirical perspectives, the relevant literature and empirical results of moral hazard in the insurance market are reviewed, the problems caused by moral hazard in the insurance market are summarized, and corresponding countermeasures are proposed (G. Q. Zhao & Wu, 2011). In recent years, some scholars. C.C Deng (2012) and G. Pei (2011) have researched relevant regulatory and legal regulations to analyze the causes of insurance fraud under the problem of lack of problems. J. W. Xie (2013) argues that the leading cause of insurance fraud is the asymmetric nature of insurance products, resulting in different situations in which insurance holders and insurers have different understandings of insurance products. The combination of qualitative and empirical analysis reveals the inherent law of health insurance fraud to a certain extent, indicating the ability to improve medical insurance anti-fraud (C. Li, 2015). It is conducive to the stability of the entire insurance market and the implementation of insurance policies. Fu et al. (2019) believes that the current medical insurance fraud damages the credit system of China's medical insurance system and undermines

the relevant management system of medical insurance. Therefore, it is imperative to establish a sound medical insurance credit system. Relevant countermeasures and suggestions have been put forward for the creditworthiness system.

2.4.2 Detection of social security insurance fraud

Medicare fraud is a very complex social phenomenon. There are many types and revelations; from the cause of fraud - fraud is both primitive and speculative; from the perspective of the main body of fraud - fraud can be single or conspiratorial, and the indicators of fraud can be related to the individual characteristics of the perpetrator or unrelated to the individual characteristics of the actor. Currently, there is no uniform understanding of the definition of Medicare fraud.

In commercial insurance practice，it refers to policyholders may not truthfully report the size of their losses, or report an accident that never occurred; There may also be fraud when the policyholder enters into the policy without truthfully informing the relevant information; or knowingly expand the loss in order to expand the claim. In addition, medical insurance fraud can be divided into two categories: medical insurance fraud in the narrow sense and medical insurance fraud in the broad sense. The narrow sense refers to the behavior of the insured (or related beneficiaries) by fabricating false materials, exaggerating the degree of loss and a series of fraudulent acts to defraud insurance money. Medical insurance fraud in the broad sense refers to the conduct of the underwriter, middleman, or third party of the insurance contract to carry out fraudulent means to obtain illegal benefits and defraud medical insurance funds.

Insurance fraud has different severity levels, from fraud bud to planned crime of fraud and opportunistic. Y. Yao et al. (2020) put forward that medical insurance fraud is an extreme form of moral hazard. A lot of domestic and foreign experience shows that medical insurance fraud undermined the fairness and efficiency in the process of the medical treatment cost of financing and hindered the development of the insurance industry. This dissertation constructs the fraud factor model in the theory of health insurance moral hazard, and on this basis, uses the data of China's commercial health insurance and uses various forms of empirical models to dig out the fraud influencing factors in the reality of health insurance in line with China's national conditions, such as the total cost of medical treatment with hospital qualification, the extension of the age of the insured and so on.

The National Health Care Anti-Fraud Association (NHCAA) defines Medicare fraud as the primary purpose of insurance fraud by individuals or organizations that use deliberate deceptive

or false means to obtain illegal benefits. The fraudulent subject of medical insurance can be either the provider of the medical act, that is, the medical institution, or the demander of the medical act, that is, the insured object.

The Health Insurance Commission of Australia uses Back-Propagation (BP) neural networks for fraud detection. Subsequently, Hall et al. (1996) added an artificial neural network model to the fraud identification study of medical service providers. Ortega et al. (2006) applied artificial neural network recognition technology to the study of medical resource abuse through the Chilean health insurance claim sample and further promoted the techique. Sokol et al. (2001) used data mining technology to establish a fraud identification model, the principle of which is to identify and filter suspected fraudulent behaviors according to the characteristics of medical services (such as preventive medicine, reflection process). L. Wan et al. (2006) used the clinical pathway principle to build a data mining-based identification model to identify supplier fraud, and applied it to an empirical study of citizen health insurance. Liou et al. (2008) used logistic regression, artificial neural network, and classification tree to identify the fraudulent behavior of medical service providers in Taiwan's medical insurance system, compared and analyzed the identification results, and found that the logistic regression method has strong identification ability. In addition, electronic fraud identification technology has been widely used in health insurance research. The technique relies on heuristics and machine learning behavioral rules to complete fraud identification in medical insurance claims, such as genetic algorithms Z. Hong et al. (2000) and Bayesian networks (Viaene et al., 2004). Another example is the research developed by (Vera-Hernandez, 2003), where moral hazard can be measured by differences in health status and treatment costs. A travel insurance company has developed an electronic fraud detection system based on behavioral heuristics and information theory. Hawkins et al. established a three-layer BP neural network to identify medical insurance fraud. Gholamreza et al. (2011) used the Probit model to try to find the relationship between high insurance amount and high insurance loss rate through the model. Viaene et al. (2004) used various recognition models such as fuzzy set clustering and logistic regression model. Kou et al. (2013) applied RIDIT model to insurance fraud detection study for the first time. Brockett et al. (2000) advocated the use of principal component analysis to identify medical insurance fraud. Ortega et al. (2006) built a fraud identification system based on neural networks, which was applied to a Chilean medical insurance company to help insurance companies identify medical claims, affiliates, medical personnel, and the employer's fraud risk. Banderier (2000) used three data processing methods to analyze and model fraudulent data: neural network, logistic regression, and classification decision tree. On this basis, scholars have comparatively studied the

fraudulent behavior of Taiwan's medical insurance system, and the results show that logistic regression is the best method. Pauly (1986) by considering the information asymmetry between insurance parties, he believes that the focus of health insurance anti-fraud should be to control the moral hazard of policyholders. Caudill et al. (2005) believe that information asymmetry leads to medical insurance fraud. Taking the particular insurance subjects into account, we can be effectively avoid the insurance fraud. Blesch (2008) believes that cooperation fraud between the policyholder and the agent is also one of the reasons for the increasingly severe insurance fraud. Rudman et al. (2009) argues that medical fraud is also related to physician fraud. The survey shows that more than 15% of doctors are involved in insurance fraud, largely related to the imperfection of China's medical insurance system, which provides new ideas for controlling medical insurance fraud. Herb and Tom in 1995 selected the indicators of medical insurance fraud based on experience and relevant professional knowledge and combined the expert consultant method to identify fraud characteristics and study the subsequent fraud formation mechanism. On this basis, an expert system is established that is conducive to identifying fraudulent behaviors by insurers. Depending on Stefan et al. (2008), Joudaki et al. (2015) emphasized the important role of government regulation and related policies in the insurance anti-fraud process and in maintaining the insurance market's stability.

Compared with some developed countries, China's insurance identification technology is not yet mature. Due to the difficulty of obtaining insurance fraud data; and the complexity and concealment of medical insurance fraud, there are difficulties in identifying and measuring medical insurance fraud (Q. Y. Wan, 2020). Nevertheless, many scholars in China have constructed a theoretical, conceptual model of insurance fraud, and even some scholars have conducted empirical research on insurance fraud. Yin and Liu (2020) by using the insurance data of a large domestic insurance company, a probit model was established for empirical analysis and research. The results show a moral hazard in China's health insurance market (Y. Lin, 2015). Based on the artificial BP neural network model, an empirical study of the abusive medical behavior of designated medical service institutions was conducted by constructing a new fraud identification index system for rural cooperative organizations, and the Logistic regression model recognition ability was compared. The results show that the BP neural network model has a higher degree of recognition than the Logistic recognition regression model. Therefore, applying the BP neural network model to the fraud identification of the new agricultural cooperative can effectively improve the accuracy of the medical insurance payment review. Y. Liu (2017) believes that the insurer's information on medical insurance behavior is challenging to grasp, and it is easy to produce false applications. Such false applications are

also difficult to identify because of incomplete information; on this basis, a Logit model of insurance fraud is established. C. Liu and Zhu (2018) based on the BP neural network model, by embedding logistic regression analysis, further reduced the interference ability of weak factors on the recognition of artificial neural networks,.

In 2010, China promulgated the Social Insurance Law, marking the basic coverage of universal medical insurance. In the process of gradually establishing a base in China, due to the imperfection of the system itself and the imperfection of regulatory measures, medical insurance fraud cases are not uncommon. System defects are gradually emerging, but the existence of fraud in social medical insurance significantly impacts the entire social security system and even the fairness and justice of the entire society. Since 2010, more than 20 studies on the identification of social security fraud have been published every year, which is also consistent with the development of medical insurance, especially in recent years, social security fraud has become a major obstacle to the development trend of social security, and has a close relationship with the great attention of the whole society. Judging from the research content, China's research on social security fraud is still in its infancy. Currently, it is mainly focused on qualitative analysis, and most identification methods used are information asymmetry, game theory, imperfect system, imperfect law, and lack of supervision system (P. Fei, 2021).

2.5 Research on the assessment of social security insurance fraud detection: Methods of assessment

In recent years, various types of medical insurance fraud and insurance fraud have considerably impacted the healthy development of China's medical insurance funds, seriously affecting the sustainable development of China's medical system. Given the endless occurrence of medical insurance fraud, improving the monitoring of medical insurance fraud is the focus of developing the current medical insurance system informatization construction (Y. Liu, 2020).

In China, universal medical insurance has been basically realized, followed by a more severe challenge to the monitoring situation of medical insurance services, that is, the cases of fraud and insurance fraud in basic medical insurance (referred to as medical insurance) have increased sharply, which has brought huge losses to state property and also undermined the fairness of society and the protection of medical care. In order to strengthen the efficient and intelligent monitoring of various medical service links such as outpatient clinics, hospitalization, and drug purchase, the national human resources and social security department has adopted many means, including manual diagnosis and treatment rule screening, expert intervention and

data comparison. These methods have achieved certain results, but there are still significant limitations due to the relatively backward technical means.

The risk identification of medical insurance fraud mainly focuses on the following types: First, identity identification, to prevent others from illegally enjoying medical insurance treatment by fraudulently using the identity of the insured; The second is to identify the matching degree of disease with medical items and drug types, to avoid the parties illegally applying for medical insurance reimbursement for medical expenses caused by third-party liability injuries, or "upgrading" and "exceeding the scope" to apply for reimbursement items that are inconsistent with the actual disease type; The third is the quantitative identification of drugs and medical services to avoid the parties from obtaining medical insurance support above the prescribed quantity and frequency, and then obtaining illegal benefits. The settlement of basic medical insurance includes outpatient (emergency) settlement and hospitalization settlement, and the subject and process of risk identification of the two are not the same (C. Shen, 2021).

Risk point of outpatient (emergency) diagnosis settlement. The first is that pharmacies purchase drugs from the medical insurance list. The main fraudulent methods of insured persons include fraudulently using other people's social security cards to buy drugs, transferring drugs to others or even selling them after purchasing drugs in excess, and transferring drugs to others or even selling them after illegally purchasing drugs that they do not need. The second is the general outpatient clinic of the hospital. The risk points of the insured who need to be identified include fraudulently using other people's social security cards to obtain medical services, over-prescribing diseases or over-prescribing drugs, transferring them to others, or even selling them. The third is the special chronic disease clinic. Due to the fixed monthly reimbursement amount and the limitation of disease types, patients cannot easily over-purchase and exceed the range of drugs. However, the special chronic disease retains the study manually as a qualification certificate because the patient's physical condition allows others to purchase drugs instead, resulting in an increased risk of fraudulent use of other people's information. The fourth is the cost of off-site emergency treatment. When the insured person seeks medical treatment in an emergency department in a different place, some areas require the insured person to bring medical records and other materials back to the insured place for reimbursement afterward. Currently, information-sharing platforms have not yet been established between different medical insurance co-ordination areas, so the risk of falsifying reimbursement materials is greater (C. Shen, 2021).

Risk points for hospitalization settlement. The first is the cost of hospitalization in the

insured place. In this process, the links prone to fraud risk include fraudulent registration with other people's social security cards when they are in the hospital. For health injuries caused by a third-party liability accident, the insured applies for medical insurance payment after the responsible party pays for medical expenses. In some links that require the provision of materials for reimbursement afterward, materials such as invoices, discharge summaries, and hospitalization details are forged. The second is the cost of hospitalization in different places. Similar to off-site emergency departments, the risk points are also reflected in the falsification of materials (C. Shen, 2021).

Because of the above risk points, the countermeasures to be solved are recommended: anchor the information of the insured to avoid identity fraud. First, artificial intelligence should be popularized to upgrade the security level of social security cards, improve the identity recognition system, increase the face and fingerprint recognition functions of medical insurance cards for drug purchase and registration, strengthen the identification of the first gate, and avoid manual recognition errors. Second, some areas still use paper vouchers for particular chronic diseases, manual invoices for medical services, that should be upgraded to electronic information cards or electronic vouchers as soon as possible, and used by designate relatives and friends to prescribe drugs on patient’s behalf. Achieve legal sharing of information and improve symmetry. The key to improving the accuracy of medical insurance fraud risk identification is to enhance the information symmetry between relevant entities. First of all, we should improve the big data monitoring system, limit the type, quantity, and frequency of drugs and medical services by disease, and carry out advance control and background supervision; Establish an information sharing platform across medical institutions, and timely early warning of insured persons in the event of abnormal multi-frequency payments. Secondly, the information docking of medical insurance and work injury insurance and accidental injury compensation should be promoted, and after the reimbursement of medical expenses by work injury insurance, the information should be timely fed back to the medical insurance department and relevant medical institutions. Encourage the party responsible for accidental injury to promptly inform the medical institution of the relevant information while bearing the liability for compensation, and the medical institution shall carry out system identification and bill identification to avoid the wrong payment of the medical insurance fund. Third, establish an information-sharing platform across the overall planning area to achieve instant settlement and information exchange for medical treatment in different places. Attach importance to the construction of "information sources" for insured persons. By incentivizing the insured and insiders to disclose information on time, the source of fraud risk can be remediated. First of all,

the special rectification activities of medical insurance inspections should be normalized, the cost of fraud should be increased, continuous pressure should be formed, the fraudulent entities should be urged to turn themselves in, and the majority of insured persons should be promoted to comply with the rules and discipline.

Second, improve the reward mechanism, smooth the channels for providing information, encourage those in the know to take the initiative to provide risk information, and ensure the safety of information providers. Third, break the chain of conspiracy between fraudulent entities and medical service personnel and other personnel, increase the punishment of accomplices in accordance with the law, and encourage relevant staff to perform their duties following regulations (C. Shen, 2021).

The main tasks of the design and implementation of the medical insurance fraud identification system are as follows: (1) because of problems such as a large amount of time and manpower required for labeling massive medical insurance data and the difficulty of finding new types of fraud, an unsupervised medical insurance fraud identification model based on weighted isolated forest algorithm is proposed. The path length standard deviation of the isolated binary tree is used to weight the calculation formula of the abnormal score in the isolated forest algorithm, and the parallelization of the weighted isolated forest algorithm is realized based on the Spark platform. Experiments show that the parallelized weighted isolated forest algorithm has higher anomaly detection accuracy and efficiency than the original isolated forest algorithm; (2) Since traditional classification algorithm has low accuracy in processing high-dimensional unbalanced medical insurance data, a supervised medical insurance fraud identification model based on the weighted random forest algorithm (hereinafter referred to as HSFS_WRF) based on mixed sampling and feature selection is proposed. This thesis proposes to improve the HSFS_WRF algorithm combined with hybrid sampling, Relief F feature selection, and weighted random forest, and realize the parallelization of HSFS_WRF algorithm based on the Spark platform. Experiments show that HSFS_WRF has higher classification accuracy than other hybrid classification algorithms, and the parallelized algorithm has higher computational efficiency; (3) To assist relevant regulatory agencies to carry out medical insurance review work and alleviate the manual review workload, design and implement a medical insurance fraud identification and review system. The above two models are applied to quickly detect new samples and discover new types of fraud through user actions to assist auditors in online audits. In addition, the use of visual chart libraries to display the use of medical insurance funds from many aspects makes it convenient for auditing institutions to supervise the use of funds (X. Di, 2021).

2.6 The research status of social security insurance in China

The development of social medical insurance in China has gone through three stages. From free medical treatment in the era of the planned economy to the establishment of the workers' medical insurance system in line with the socialist market economy in 1998 to accelerate the reform of the medical insurance system and guarantee the essential medical treatment of employees.

2.6.1 The period of China's publicly health service system and labor insurance system

The Labor insurance medical system has been the main medical insurance system covering state functionaries and urban workers in China for nearly half a century. It has made a great contribution to improving people's health. According to the sixth population census in 2010, China's infant mortality rate dropped from 34.7‰ years in 1981 to 13.1‰ in 2010, and the average life expectancy rose from 67.9 years in 1981 to 73.5 years in 2010.

With the advancement of economic system reform, China's economy, society, environment and other aspects of profound changes, the past public funds, labor insurance and medical system showed a variety of drawbacks, including unscientific financing mechanisms, low degree of socialization, and unreasonable payment methods.

2.6.2 Basic medical insurance system for urban workers

China's basic medical insurance system for urban workers points out that basic medical insurance premiums are jointly paid by employers and individuals, reflecting the mandatory characteristics of national social insurance and the unity of rights and responsibilities. The basic medical insurance fund consists of social pooling and personal account funds. China's urban workers' basic medical insurance system was piloted in Zhenjiang City, Jiangsu Province, and Jiujiang City, Jiangxi Province, in 1994 and has been implemented for over 20 years. Basic medical insurance for urban workers has played a specific role in cultivating the concept of personal health responsibility, establishing a moral hazard constraint mechanism, and promoting the optimal allocation of health resources.

The direct purpose of establishing a basic medical insurance system for urban workers is to control the rapid growth and waste of medical expenses by reasonably sharing medical expenses, reducing the burden on government and enterprise institutions, and achieving relative fairness in basic medical insurance. The other purpose is to reduce the social functions of enterprises

(affairs) and organizations through this system, create conditions for the market-oriented reform of the employment system and the innovation of the enterprise system, and establish a basic medical system that is compatible with the level of China's economic development.

The framework of the basic medical insurance system for urban workers includes six parts. First is establishing a common payment mechanism for reasonable burdens; the basic medical insurance premiums are jointly paid by employers and individuals, reflecting the mandatory characteristics of national social insurance and the unification of powers and obligations. The employer's wage rate controls about 6% of the total wages of the employees, and the wage rate of the employees is generally 2% of their wage income. The second is to establish a pooled fund and an individual account system. The basic medical insurance fund comprises a social pooling fund and an individual account fund. All individual payments are transferred to individual accounts, unit payments are transferred to individual accounts at about 30%, and the rest are established as pooled funds. The third is to establish an independent unified account and a clear scope of a payment mechanism. The pooled fund and the individual account determine their respective payment limits, the pooled fund mainly pays for inpatient (large) medical expenses, and the personal account mainly pays for outpatient (small) medical expenses. The fourth is to establish an effective medical service management mechanism. The scope of basic medical insurance payment is limited to the medical expenses in the drug catalog, diagnosis and treatment items, and medical service facility standards stipulated in basic medical insurance. The fifth is to establish a unified socialized management and service system, and the grass-roots medical insurance implements a certain level of social management. Sixth, establish a sound and effective supervision mechanism and implement special financial account management for the basic medical insurance fund.

X. B. Gao (2006) thinks that the establishment of the basic medical insurance system for urban workers has achieved some results, including the gradual expansion of the coverage of social medical insurance, the gradual improvement of the social medical insurance policy system, and the gradual acceptance of the operating mechanism of the social medical fund by the society. However, at the same time, some deficiencies are embodied in the trend of China's aging population, which will greatly challenge the implementation of medical insurance. The marketization of medical institutions, the high prices of medical care and the prices of drugs, the excessively rapid growth rate of medical expenses, the difficulty of regulating and controlling the social medical insurance pooling fund, the imperfection of China's medical security system, the single system structure, the narrow coverage, and the low-security function have become increasingly prominent. It is recommended to speed up the legislative pace of

social insurance, implement the "law of large numbers" of social insurance, strive to expand the coverage of social medical insurance, enhance the anti-risk ability of social medical insurance funds, and include social medical insurance funds in the scope of financial budgets, and the financial department will bear the final bottom.

H. X. Ren (2012) regards the issue of moral hazard as an important factor in the excessive growth of medical expenses. Whether the moral hazard can be effectively and reasonably controlled is the key to determining whether medical insurance can develop smoothly. X. L. Wang (2005) and X. Z. Yao (2013) both believe that the coverage of basic medical insurance for former urban workers is not wide enough to alleviate the economic pressure brought about by the aging of the population, resulting in the transfer of economic risks to medical service providers. It is recommended to establish a reasonable medical cost increase mechanism as soon as possible to help the government expand the source of funds for the medical insurance fund, increase its total income, and alleviate the economic pressure it may face (N. H. Gao, 2010) Starting from the establishment and development of China's current medical insurance system and the many existing problems, it is believed that the establishment and improvement of the medical insurance system are of great significance to the deepening of the market economy.

C. Y. Kuang and Chai (2000) believe that a social medical insurance system suits the socialist market economic system. Finances, enterprises, and individuals can afford it and guarantee the basic medical needs of employees and workers in all urban organs, enterprises, and public institutions and their staff and workers. Y. J. Zhu et al. (2002) believes that from the perspective of the operation of the basic medical insurance system for employees, and ensuring the balance of the basic medical insurance fund, the system setting and operation must consider the opportunistic behavior of relevant entities, summarize the existing experience at home and abroad, use the law to clarify the rights and responsibilities of the relevant subjects of the basic medical insurance system for employees, and establish an effective restraint mechanism through in-depth reform to ensure the rational use of basic medical insurance funds. Establish a disease control mechanism based on prevention, a systematic cost control-payment mechanism for basic medical insurance for employees (Glied & Stabile, 2001), and a supporting medical management system. F. Li (2013) taking the urban workers' medical insurance in Hohhot as a case, the city's current situation and reform plan for the urban workers' medical insurance are analyzed. This has theoretical significance for the reform of the basic medical insurance system for urban employees and has an excellent reference for the reform of the basic medical insurance system for urban employees in China through the discussion of typical cases.

2.6.3 Basic medical insurance system for urban residents

The basic medical insurance system for urban residents (hereinafter referred to as urban medical insurance) is the basic system for safeguarding residents' right to health. In 2018, the National Medical Security Administration proposed that a unified medical insurance system for urban and rural residents will be implemented nationwide in 2019. How to promote the improvement of urban residential medical insurance in China and effectively realize the equalization of public services has become a hot issue in academic circles. Shanghai has excellent medical insurance service conditions for urban residents, while the unsatisfactory service conditions in Hangzhou, and Hefei confirm the above hypothesis to a certain extent. The unbalanced level of medical insurance for urban residents in various cities is contrary to the ultimate goal of equalizing basic medical services. It is not conducive to implementing the strategy of comprehensive, coordinated, and sustainable social development. M. Ran (2009) through a comparative analysis of the policies of seven pilot cities, suggested expanding the system's coverage and including general outpatient clinics in the overall planning scope. It is not necessary to establish a "personal account" like Changzhou; moderately increase the level of security and improve the payment of medical expenses mechanism to effectively solve the problem of residents' "expensive medical treatment". The practice of Jiuquan City can be used as a reference to establish moderate and large risk adjustment funds across the country to cope with the shortage of general funds and ensure the stability and continuity of the system. J. J. Zhou et al. (2015) used data from the China Health and Nutrition Survey (CHNS) to empirically examine the impact of urban residents' medical insurance system on residents' demand for medical treatment and preventive health care services. Z. P. Wang et al. (2007) analyzed the effectiveness of my country's urban residents' basic medical insurance system in pilot areas and found some problems, such as the unclear definition of urban residents, low overall planning level, lack of personal accounts, single-level and poor fund supervision ability. Cao and Gu (2012) studied the effectiveness and deficiencies of China's urban residents' basic medical insurance system in pilot areas and put forward a clear definition of the protection objects, a connection of the three systems, and increased publicity efforts to improve the urban residents' basic medical insurance system. H. Zhang and Xu (2007) introduced the basic medical insurance system for urban residents in some cities and analyzed the adverse selection problem caused by the high risk of insured objects and the form of voluntary participation. The basic medical insurance system for employees is gradually integrated, and medical insurance adopts mandatory measures to diversify risks. The following countermeasures and suggestions are put forward: first, to issue

unified national implementation guidance to improve the overall level; second, the government strengthens public health and community investment in health services (X. Wang, 2008). Through the research on the characteristics of the resident medical insurance system and the foreign medical insurance payment system, combined with the practice of the resident medical insurance payment system in Zhenjiang City, and based on the premise and reference that the resident medical insurance is different from the employee medical insurance payment system design, it is proposed that residents medical insurance should implement a compound payment system of "total budget management" centered on "insured fixed-point heads", and provide residents with high-quality, continuous basic medical services by building a new mechanism of "funds follow patients" and government purchases of services. J. Li and L. Xu (2007) believed that the basic medical insurance system for urban residents was not perfect, which led to problems such as "difficulty in seeing a doctor" and "expensiveness in seeing a doctor", which affected the construction of a harmonious society in China. They proposed that the government give full play to the leading role and establish a scientific public Hospital compensation and incentive mechanism to make China's health service sustainable, stable, and coordinated development. Some scholars also discussed the aspects of college students' medical care and medical expenses. In short, so far, the academic research on the basic medical insurance system for urban residents is generally in the macro-analysis stage, and there are few quantitative studies. The comprehensive and scientific analysis of the basic medical insurance system for urban residents needs to be further improved (J. Ma, 2011). The basic medical insurance system model for urban residents is characterized by the gradient financing model, and put forward corresponding countermeasures and suggestions: First, increase the government's financial investment in the medical insurance of urban non-employed residents, according to the growth of fiscal revenue. Increase the proportion of financial investment year by year. Second, adjust the residents' financing model to encourage young residents to participate in insurance. Third, improve community medical service institutions, encourage residents to seek medical treatment in community-level medical institutions, improve the utilization rate of medical resources and alleviate the fund payment pressure. Fourth, based on promoting the connection of the three basic medical insurance systems in China's urban and rural areas, institutional arrangements should help coordinate the relationship between the basic medical insurance system for urban residents and other systems.

The current academic analysis of the urban resident’s basic medical insurance system and the system in the operation of the existing problems of the study is precious and has a good reference for the study. There are currently more studies on the operation of the urban residents'

basic medical insurance system. First of all, the scholars (can list a few representatives) - from the residents' healthcare coverage, fund-raising, and compensation payment aspects - have carried out thorough research, basically covering the basic medical insurance system for urban residents in the operation of the main problems. Secondly, in terms of improving residents' health care system, it gives the corresponding countermeasure and suggestion. Thirdly, the analysis of the satisfaction of the basic medical insurance system for urban residents and the study of the national medical insurance system for national medical insurance for all abroad have broadened the thinking and perspective for the study of the basic medical insurance system for urban residents in China. At the same time, some scholars have put forward the resident medical system from the whole medical insurance system. The connection between the insurance system and the basic medical insurance system for urban workers and the new rural cooperative medical system is analyzed. The ideas and methods of the connection between the three systems are analyzed to sort out the literature and provide very valuable materials for the further research of this study.

Based on the literature collected, there are still some deficiencies in the current academic research on the basic medical insurance system for urban residents. First of all, most scholars mainly support theoretical research. Although the ideas of many scholars are worthy of reference, in actual operation, it seems impossible to start, and the operability is not strong. Secondly, most scholars start from the problems existing in the operation of the basic medical insurance system for urban residents, and few scholars conduct in-depth research on the root causes of the problems. Finally, from the perspective of research methods, most scholars only conduct qualitative research on existing problems and less quantitative research.

For the current deficiencies, first of all, starting from the actual implementation of the basic medical insurance system for urban residents, analyze the problems that arise in the operation of the system, analyze it in depth, and study the operation effect of the system. Secondly, by analyzing the operational effect of the basic medical insurance system for urban residents in Shanghai, drawing on the domestic and foreign experiences and beneficial attempts, the target model for improving the resident medical insurance system is proposed. The feasibility of the target model is analyzed, and finally, the supporting reform measures for improving the basic medical insurance system for urban residents in Shanghai are proposed.

2.6.4 New rural cooperative medical system

Wagstaff et al. (2009) argues that among all forms of equality in health care systems that reflect equality, equality in financing is more important than in other forms. Bloom (2002) studied

various collaborative healthcare funding models and found that governments at all levels should play an important role in the fundraising of cooperative healthcare. Wagstaff et al. (1999) using the Cavani index and the concentration index, a comparative study of some systems in Asian and countries of Organization for Economic Co-operation and Development (OECD) is regressive or progressive in financing for public health services. The article measures the equality of health financing in progressiveness, given the low proportion of cash disbursements in OECD countries. Yip (2010) Compared with the new agricultural cooperative system and other medical security systems, it is found that the new system does not fundamentally solve the problems that arise in participating farmers' medical treatment. The new agricultural cooperative reimburses relatively few types of severe diseases and does not have a sound guarantee for complicated diseases. Many peasant families will still return to poverty due to illness, resulting in a significant decline in living standards.

In view of the increasingly prominent and objective reality of the problem of Chinese peasants falling poor due to illness and slip back into poverty due to illness, in order to fundamentally solve the problem of medical security for rural residents in China - at the national rural health work conference held in 2002 - the Party Central Committee and the State Council made the Decision on Further Strengthening Rural Health Work, which proposed to gradually establish a new rural cooperative medical system in rural areas across the country, and required all localities to pilot projects, sum up experience, and then gradually promote it. By 2010, the goal of covering rural residents with a new rural cooperative medical system will be achieved. In January 2003, the General Office of the State Council forwarded the Opinions of the Ministries of Health, Finance, Agriculture, and others, regarding the Establishment of a New Rural Cooperative Medical System and making a comprehensive deployment for the pilot work of the new rural cooperative medical system, and requiring that the pilot project be carried out in some counties (cities, districts) across the country starting from 2003. Thus, the new rural cooperative medical system, which concerns the interests of peasants who account for the overwhelming majority of China's population, has opened the prelude to the implementation process. The new rural cooperative medical system is organized, guided, and supported by the government, with the voluntary participation of peasants, funded by individuals, collectives, and the government, and with the main focus on the overall planning of serious diseases. Implementing the new rural cooperative medical system is another "moral and political project" that benefits many peasants after reforming land contracting and rural taxes and fees. Since the pilot work of the new rural cooperative medical system was launched in early 2003, it has developed rapidly with the support of governments at all levels, and the pilot work has

progressed smoothly. In 2003, there were 304 pilot counties (cities, districts) of new rural cooperative medical treatment in China, with 0.43 billion farmers participating, and the participation rate of farmers in pilot counties (cities, districts) was 74% (Y. J. Liu, 2009).

After 13 years of development, the pilot and implementation of the new rural cooperative medical system has achieved many results, such as basically covering the rural population in an all-round way, the amount of financing base has increased significantly, the proportion of outpatient hospitalization compensation and the maximum reimbursement amount has been continuously improved, and the threshold has been continuously reduced, effectively ensuring the right of farmers to enjoy medical services (Y. Liu, 2016).

At present, the research on the new rural cooperative medical system in academic circles at home and abroad mainly focuses on the following aspects: research on the financing mechanism of the new rural cooperative medical system, research on the compensation mechanism of the new rural cooperative medical system; study on the responsibilities of the government in the establishment and improvement of the new rural cooperative medical system; research on the practice model of the pilot of the new rural cooperative medical system; research on evaluation indicators for the implementation effect of the new rural cooperative medical system.

Mao & Jiang (2005) , Mao (2003) , G. L. Wang (2004) , R. J. Yu et al. (2003) and C. Y. Zhang (2004) have all conducted detailed research on the new rural cooperative medical system. Judging from the literature available on such research, the academic community generally summarizes the difference between the new rural cooperative medical system and the traditional rural cooperative medical system as follows: the new rural cooperative medical system has increased government support and improved the financing mechanism that combines individual payment, collective support and government funding; Require provincial and prefectural-level governments to establish rural cooperative medical coordination groups composed of health, finance, agriculture, civil affairs, auditing, poverty alleviation and other departments, set up special rural cooperative medical management institutions within health administrative departments, form rural cooperative medical management committees at the county level, set up handling agencies under them, and set up dispatch agencies (personnel) in townships (towns) or entrust relevant institutions to manage them as needed; Clarify the institutional requirements to strengthen the supervision of rural cooperative medical funds; Emphasize the "overall planning of major diseases", focusing on solving the problem of "poverty caused by illness and returning to poverty due to illness" that arises from major diseases such as infectious diseases and endemic diseases among farmers; The system of overall planning and management with counties (cities) as units has been adopted, and the scope of overall planning has been expanded,

and the ability to resist risks has been strengthened. In summary, the new rural cooperative medical system is an innovation of systems and mechanisms, which is more extensive in financing channels, more scientific in management, and more rigorous in the organization, overcomes the shortcomings of the traditional rural cooperative medical system to a certain extent and has more robust vitality.

2.7 Basic theories of the research

Modern medical security involves the allocation of social and economic resources, social justice, government responsibility and other issues, and its success or failure seems to depend on the actual institutional arrangements and policy practices on the surface, but in fact it is profoundly affected by the basic theory and value preferences of medical security.

2.7.1 Equity theory

The fairness theory of social medical insurance is characterized by public services, mainly reflected in the fairness of the rights, responsibilities, and rights of all parties involved.

In terms of equity in access to insurance, Lagomarsino et al. (2012) reported progress towards universal coverage in nine countries in Africa and Asia, enriching research on equity in access to insurance and equity in health care delivery. Grobler et al. (2015) conducted an in-depth study of inequities in the regional distribution of health workers in several countries, and pointed to the prevalence of low proportions of health workers in rural and other underserved areas. This study examines health service utilization and related factors in Brazil, and assesses differences between socioeconomic groups, in terms of financing benefit equity. Ataguba and McIntyre (2012) uses representative data from South Africa, using health care financing progress rates and benefits. Incidence rates assess the fairness of healthcare financing benefits in South Africa, Bommier and Stecklov (2002) and Jehu-Appiah et al. (2011) suggest that financing equity is not based on the patient's cost of illness but on the patient's actual ability to pay. In terms of the application of general indicator evaluation methods, Fernandes et al. (2018) analyzed and evaluated the inequality of medical insurance between urban and rural areas by using the catastrophic health incidence indicator. Mills et al. (2012) used progressive rate of financing system to fully analyze the health system financing and equity in service access in Ghana, South Africa, and Tanzania, found that low-income groups had a greater burden of disease, but the distribution of health services in these three countries is more convenient for the wealthier. Coleman (2014) uses the population cancer survival rate as a key indicator to

measure the effectiveness of the health system in cancer management and then measures inequality in different regions. In terms of the application of comprehensive index evaluation methods, Kim et al. (2017) used the Gini coefficient to explore the unequal trend of household medical expenditure distribution and the factors related to household catastrophic medical expenditure. Hajizadeh and Nghiem (2011) used the concentration index to study Iran Inequality and determinants of out-of-pocket expenditures for hospital services and related catastrophic expenditures, finding that related catastrophic expenditures are concentrated in households with poorer financial circumstances and are catastrophic with longer hospital stays, lower household wealth indices, and use of private hospitals, reasons for the increased probability of spending. Munge and Briggs (2014) analyzed the progressiveness of the deviation between healthcare financing sources and affordability using the Cavani index. In addition to the above two evaluation methods, Albanese et al. (2011) used Poisson regression and fixed-effects meta-analysis to deal with the research indicators and surveyed the elderly over 65 years old in China, India, Nigeria, and Latin America. Roll et al. (2013) used a multivariate logistic regression model to analyze differences in access to mental health services among different groups in the United States.

The existing research mainly focuses on the content and measurement of medical insurance equity. The study of medical insurance equity is mostly about the access, healthcare resource allocation, health services utilization and financing proceeds. In the measurement, many scholars use general indicators, such as gini coefficient and lorenz curve, concentration index and other comprehensive index for medical insurance equity,

Above all, the shortage of the research on current urban and rural residents basic medical insurance fairness mainly has the following three aspects: (1) in the research methods, theoretical research articles is much more than the empirical research articles, few scholars use quantitative method to evaluate the fairness of medical treatment insurance; (2) in the research indicators, concentration index and gini coefficient are often applied in medical insurance equity research, but the value of these indicators is limited, and the micro theory support is insufficient; (3) in research perspective, it often take advantage of the regional available data, seldom consult to the audience by public questionnaire survey to analyze the health equity evaluation.

2.7.2 Public goods theory

The publicity of social medical insurance products is mainly manifested in the following aspects:

2.7.2.1 Non-exclusivity and exclusivity coexist in consumption

Non-excludability. That is, there is technically no way to exclude individuals or vendors who refuse to pay for them from benefiting from public goods. In other words, public goods or services cannot be prevented by individuals or manufacturers who refuse to pay, and no one can exclude public goods or services that they do not like from their enjoyment by refusing to pay. Exclusivity means that after a consumer has purchased and acquired the right to consume a good, he can exclude other consumers from the benefit of the goods. The non-exclusivity of public goods makes the mechanism of consumption power obtained through market exchange dysfunctional. Two reasons determine non-exclusivity: First, it is not technically easy to exclude many beneficiaries, such as defense products; Second, although technically exclusive, exclusivity is so expensive that it is not economically feasible (J. Gao, 2007). Although the various types of insurance in social medical insurance have their non-exclusive side, such as not excluding anyone from participating in social medical insurance, their exclusivity is more obvious and prominent. In the field of social health insurance, if people do not pay social health insurance premiums in advance, they will be excluded from consuming social medical insurance products - medical services;

2.7.2.2 Non-competitiveness and competitiveness coexist in consumption

Non-competitive means that at a given level of the public good output, increasing the consumption of the product by one person does not cause any increase in the cost of the product, such as health education in community health services, and the consumption of these products by one more or one less person does not cause a change in the cost of the product, and is therefore non-competitive. However, other products of social medical insurance are competitive, and competitive refers to the increase in the production cost of goods caused by the increase in consumers or the number of consumers. However, public goods are not competitive for consumption, such as radio, television, and navigation lights; their common feature is that the increase in the number of consumers does not affect the cost of production. Non-competitiveness stems from three points: First, public goods generally have an indivisible nature; Second, due to the indivisibility of public goods, the marginal cost of each additional consumer is equal to zero before it becomes crowded; Third, competition mechanisms such as national defense cannot be introduced, and economic theory holds that pure public goods can only be invested, produced, and provided by the state and the government. Because private investment in and public goods production may be difficult to get a return.

2.7.2.3 Non-rejection and rejection coexist in consumption

Social medical insurance is a legal and compulsory public undertaking organized by the government following the law, and all social citizens within the scope of the law, whether they like it or not, unconditionally participate in and pay the insurance premiums under regulations, showing a typical non-refusal. However, for the choice of supplementary medical insurance in social medical insurance, individuals have complete freedom to participate or refuse to participate.

Therefore, the provision of social medical insurance products cannot be taken over by the government, and we should actively promote a variety of guarantee models of state insurance basic, enterprise insurance supplement, and individual investment and commercial insurance, mobilize the enthusiasm of the whole society to participate in insurance, reduce the financial burden of national social medical insurance, reduce the waste of social medical insurance resources, and make social medical insurance play a greater positive effect.

2.7.3 External benefit theory

Externalities refer to the fact that the consumption of a certain product by a certain number of people can indirectly affect those who do not consume such a product (Shi, 2011). External effects are divided into two kinds: external positive and negative effects, the former refers to the private benefits that individuals obtain from their economic activities that are less than the social benefits brought by the activities, and the opposite is the external negative effects.

Social medical insurance is a system in which the state or society provides necessary medical services or financial compensation to people who need treatment for illness, injury, or childbirth. It is an important part of the social security system and is related to constructing the entire national welfare system. However, in the operation of this system, due to the defects of the congenital design of the entrusted agent model adopted by the system, the reform of drug circulation and the reform of hospital management in the market economy reform are not coordinated. The professionalism of medical and health knowledge leads to serious information asymmetry between the doctor and the patient and the medical management side, which gives the designated hospitals of important subjects of medical insurance the space to defraud the medical insurance fund. This problem has become increasingly prominent, harming the interests of most insured persons. Follow and carry forward Pigou's research ideas, especially in the field of private products and public goods to distinguish between externalities. The prominent representatives were: Meade (1952) held that in dealing with externalities, it is

necessary to distinguish between social benefits and the production factors of each enterprise, and to adopt adjustment methods such as taxes, subsidies, or joints for different situations. Baumol and Ordover (1988) building an analytical model, explains the effectiveness of the "Pigou tax" in addressing externalities. Komisar and Buchanan (1969) combined the Pareto principle of optimality to analyze the welfare consequences of the party affected by externalities, starting from the behavior of the passive party subject, and finally proposed the argument of reverse configuration. Cleeton and Knight (1924) pointed out that by ignoring the binding effect of private property rights, the market could not maximize the interests of society. Coase (1992) proposed, in *The Problem of Social Costs*, that externalities arising from economic activity are most effectively resolved through the channels of the market without the intervention of government means, which is summarized by its elaboration as the famous Cease's theorem. Arrow (2001) explains in The Organization of Economic Activity the idea of internalizing externalities by creating additional markets. Y. G. Huang and Zhang (2016) proposed the conscience effect, which holds that conscience influences the subject that produces externalities. Stiglitz (1988) believes that people's behavior is mutually influenced, that social norms can be used to demand people to carry out productive activities, and that the propagation and education of social norms and social values is another way to solve external uneconomical.

Pigou (1920) argues that according to welfare economics, there is no reason to think that the private costs of economic activity are equal to social costs, which the externalities of economic activity occur frequently, and that public goods and services can have positive or negative externalities. Welfare economics argues that health care has positive externalities, especially in preventive care, and its spillover effects are more prominent. The most typical example is planned immunization of infectious diseases, which is given to a subset of people within a community, and the likelihood of vaccinators developing infectious diseases is significantly reduced. As the incidence of infectious diseases in the community decreases, the sources of infection decrease, and the chances of unvaccinated people being infected are reduced, benefiting both those who are vaccinated and those who are not vaccinated. Direct consumers of quasi-public goods with external effects have much smaller estimates of consumption benefits than social benefits. Under the market mechanism, consumer demand for quasi-public goods is always less than the best demand in society, so it needs to be solved through government intervention and non-market methods.

2.7.4 Equity and efficiency theory

Fairness and efficiency are the dual value concepts followed by the development of human

society. The lack of any of them is not conducive to the harmonious development of human society, and how to deal with the relationship between fairness and efficiency has become the eternal theme of human society. As an important socio-economic system, social security cannot avoid the game between fairness and efficiency. The emergence of the social security system stems from people's demand for fairness, and maintaining social fairness is the core institutional concept of the social security system. At the same time, the social security system is also an objective requirement for developing socially productive forces to reach a certain level. In practice, it inevitably impacts economic growth, and promoting efficiency has become a natural attribute of the modern social security system. Therefore, fairness and efficiency are the two fundamental principles that run through the development process of the social security system (Y. Zheng, 2011).

The social medical insurance system alleviates social contradictions, resolves various social risks, and directly promotes social fairness, social unity and social harmony, which is the social effect produced by the system. The social health insurance system protects human rights, promotes political civilization and democratic progress, increases people's satisfaction with the government and the ruling party, and achieves political efficiency. The spirit of cooperation in the social medical insurance system of mutual assistance and risk sharing is directly conducive to enhancing people's sense of security and is of great benefit to improving the social morality of fraternity, mutual assistance and sharing.

From the perspective of fairness and efficiency, we can improve China's social security system from eight aspects, such as values and system implementation, to better play its fairness and efficiency functions. Further, the development and improvement of the social security system in China is also the institutional embodiment of the basic requirements of socialism to achieve shared prosperity for the whole country, benefit the people, and let all the people share the fruits of reform and development. Compared with capitalist society, the socialist social security system truly reflects the people's fundamental interests (Y. Zheng, 2011).

2.7.5 Information asymmetry theory

There are two kinds of uncertainties in the health insurance market, one is about treatment, and the other is morbidity. These two uncertainties in the health insurance market are characterized by a high degree of information asymmetry.

Caudill et al. (2005) argues that policyholders are likely to exaggerate insured accidents due to information asymmetry caused by the insurer's inability to obtain accurate information about the insured accident (Watt et al., 2006). After summarizing the research results of

previous studies, it is believed that in the insurance market, the insured always pursues the maximization of the profits of the insurance contract and hopes that the expected profit brought by the insurance contract is greater than zero, which will lead to the moral hazard of the insured. Davies (2017) argues that the odds of patients being utilized become very high due to an imbalance in information. He proposed an information system that uses network technology to reduce information asymmetry in the insurance market. First, there is an information asymmetry between the insured and the insurance institution.

The endowment of different people to contract diseases is different. Still, insurance institutions are at an information disadvantage because they cannot know each insured person's health status. Because policyholders can hide private information, it is difficult for insurance institutions to judge the risk profile, so actively insured people are likely to be inherently weak, making the risk underwritten by the insurance company higher than the average risk. Second, there is an information asymmetry between providers and demanders of health care services. Demanders of health care services are almost ignorant of the knowledge of the medical professional field, do not understand the content, quality, effect, price, and other information of health care services, and can only turn to doctors to obtain this information. Doctors are the people who sell health care services, so the provider of health care services as a proxy chooses the type and method of health care services. Healthcare providers may be driven by their economic interests and use their information advantages to provide excessive and irrational services to those who demand healthcare services, waste health resources, and harm the interests of patients. Finally, there is an information asymmetry between insurance institutions and medical institutions. Due to the uncertainty of disease occurrence and treatment, it is difficult for insurance institutions to quantify and contract the quality and quantity of treatment of medical institutions, and hospitals have information advantages. Healthcare providers may be driven by interests to induce patients to over-consume, thereby increasing the cost of care and the burden on insurance institutions.

2.8 Big data management model

Big Data (big data technology) is a recent technology hotspot. In the history of computer technology, databases, data warehouses, data marts, and other technology are all designed to solve the access and utilization of large-scale data. With the establishment of information systems and the rapid growth of data in recent years, Big Data has become a hotspot of business studies in various fields. At present, distributed processing methods such as Hadoop,

MapReduce, and Spark are the general technical framework and foundation. Every year, many new technologies emerge from the big data field platform, which completes the process of mining hidden information and knowledge in large-scale data. It is generally believed that traditional machine learning is a shallow-level learning architecture, while deep learning refers to machine learning techniques that automatically learn deep-level architectures under supervised or unsupervised conditions and are used for classification or data mining. Inspired by the human brain's natural signal processing patterns, the concept of deep learning has been proposed and has received more and more attention due to its superior processing performance in many fields (H. Liu et al., 2017). Generally, machine learning algorithms can be divided into three categories. The first one is known as supervised learning. The algorithm generates a function that maps input values to output values from a series of variables. Through the repeated training process, the model acquires the expected Accuracy. The second is unsupervised learning, in which cluster analysis is performed in different groups, and there is no target variable to be predicted in the algorithm. This type of analysis is widely used to segment customers. The third is reinforcement learning. The algorithm trains the machine to make decisions and trains itself through repeated iterative mistakes and retry. Machines try to make precise judgments by learning from experience. It can be used in almost all data-based problems. Different data analysis and mining targets select corresponding algorithms, including: classification algorithms, clustering algorithms, prediction and estimation algorithms, decision-making algorithms, association rule analysis algorithms and recommendation algorithms, support vectors machine, Naive Bayes, K-nearest Neighbor , K-means, and Random Forest Algorithm. In addition, the most frequently used big data algorithms include the resource allocation algorithm, path analysis algorithm, similarity analysis algorithm, sorting and search algorithms. X. L. Li (2016) and J. B. Xu (2007).studied how to apply the classification and integration algorithm of data mining to medical insurance anomaly detection to improve the detection ability. Due to the imbalanced medical insurance samples, the data set should be balanced first before analyzing the medical insurance data and then pairing the relatively balanced data together (H. S. Huang, 2020). Through a detailed analysis of traditional clustering algorithms, an interval image fuzzy clustering algorithm based on data volume is proposed and applied to medical insurance anomaly detection. The main research contents are as follows: (1) Investigates the current medical insurance abnormal audit methods and analyzes the deficiencies in the existing related research; (2) Based on the unbalanced characteristics of medical insurance reimbursement data, analyzes its impact on clustering and make improvements; (3) Based on the improved fuzzy clustering algorithm, it puts forward a

complete set of medical insurance anomaly detection models and proves its accuracy and global optimality.

Chapter 3: Research Methods

Through the investigation and interview analysis in Chapter four, we find that the risk of medical insurance funds does exist. Because medical insurance fraud involves various stakeholders, in the specific implementation, doctors deliberately neglect their duties, and patients take the chance to defraud, making the fraud behavior systematic, hidden and complex. This chapter will examine the source of existing health insurance risk problems, that is, how to identify health insurance risks. At first, this study builds a process model for healthcare fraud identification and then conducts healthcare fraud identification based on scenarios and data.

3.1 Interview

3.1.1 Overview of interview

Interviews are an important method of gathering information. According to the degree of formatting of the interview content and procedure, it can be divided into semi-structured and structured interviews. Semi-structured interviews, also known as in-depth interviews or free interviews, are opposed to structured interviews, which do not follow a fixed procedure designed in advance but only have a subject and scope of the interview. The interviewee and the interviewer have a relatively free conversation around the topic and scope (Feng, 2022). The role is to obtain rich qualitative information through interviews and summarize conclusions from them. The primary purpose of this dissertation by using semi-structured interviews is to focus on the research theme by collecting the response of hospital presidents, patients, and family members about the medical insurance card problems faced in the process of medical treatment, and to technically process the answers of the interview subjects by setting up the main questions face-to-face. To explore the logical relationship between the qualitative answers of the respondents, further focus on the research topics from the perspective of the research objects, further enrich the content, and clarify the dimensions and their relationships.

The Interview methods are divided into structured and semi-structured interview methods, semi-structured interviews are informal interviews conducted according to a rough outline of the interview prepared in advance rather than a fine, accurate quantitative questionnaire.

Compared with structured interviews, semi-structured interviews have higher degrees of

freedom and more flexible processes. Through semi-structured interviews, we can deeply understand whether medical insurance violations are common in specific places (medical service providers) and different groups. In the interview, the words and questions are flexibly adjusted according to the interviewee's conditions, answers, and situations. There are no specific and mandatory requirements for the way and order of the questions, the way the interviewees answer, and the time and place of the interview. The interview follows the principle of voluntariness and confidentiality, and after the consent of the interviewee, it is conducted on the spot without rehearsal and reservation, and the recording is carried out for the collation, summary, analysis, and writing of subsequent materials

3.1.2 Implementation process

Based on the methodological points of the unstructured interview, the implementation process is divided into the following steps:

1. Draw up an interview outline. Around the research theme, literature review, according to the dimension of the variable design problem, the question revolves around the interviewer's understanding and definition of the variable, including three contents: variables should contain dimensions, the relationship between variables, and the relationship between dimensions.

2. Identify the interviewee. The interviewees were selected from medical institutions, medical patients, and family members, and to obtain valuable information, they were paid special attention when choosing the representative interviewees. The hospital level is limited to Grade III and Grade II hospitals, and the preliminary selection of subjects for the interview was based on the conditions required for the research theme, according to the size and region of the hospital.

3. Make an appointment for the interviewee's time in the form of a formal interview.

4. Process interview materials. The entire interview is recorded and adjusted at the end of the interview, then analyzed the text and extracted the frequent words.

3.2 Big Data management and application

3.2.1 Overview of big data management

3.2.1.1 Big data mining model

Big data in a broad sense covers multi-disciplinary fields, multi-source and mixed data, involving text and language, video and image, time and space, network and graphics; in short,

massive data plus complex types of data. Big data mining aims to transform data into knowledge, explore the mechanism of data generation, identify patterns, make predictions, and support decision-making. There are several algorithms beyond big data mining, depending on the purpose of the study. With the rapid development of information technology, accumulated data has increased dramatically, and digging out valuable and essential content from massive data information has gradually become an important task nowadays.

3.2.1.2 Research on the application problems of medical insurance big data

The general status quo of healthcare big data mining and management applications is that hospitals and medical insurance management departments have collected a large amount of data but lack of the knowledge, means, and tools necessary to deal with the huge amount of information. Big data mining technology is still in its infancy, mainly manifested in incomplete historical data and imperfect management information systems. The medical insurance management departments can only achieve lower-level functions, such as data entry, query, and statistics.

With the continuous population aging, the demand for medical treatment is also increasing, and the risk situation that medical insurance supervision is emerging, the medical insurance control fees urgently need to be continuously improved.

3.2.1.3 Research on the application of medical insurance big data technology to medical insurance cost control

Most medical institutions in China have realized the importance to risk control and system management. The structure of the risk analysis and management system consists of three parts: the data analysis layer, the intermediate data processor, and the database.

An efficient and comprehensive information system platform is necessary for implementing medical insurance cost control. Although China's medical insurance management department has invested consistent capital in information systems - compared with the business development needs of China's medical institutions or the medical industry in developed countries - the informatization of medical insurance cost control risks in China is still relatively backward. In the final analysis, this happens because the quality of medical management is poor, and the investment in medical data digitalization is insufficient, which impacts the informatization level of medical insurance control costs in China.

In addition to the above problems, the degree of vigilance against risks is not high, so it is imperative to carry out the application of big data technology in the field of medical insurance cost control.

3.2.2 Analysis method based on association rules model

3.2.2.1 Introduction to the association rules mining model

In data mining, the association rule pattern is one of the most important. The concept of association rules, proposed by Agrawal, Imielinski, and Swami, is a simple but practical rule in data mining (X Q. Mei & Lu, 2016). The association rule pattern is a descriptive model, and the algorithm that discovers the association rule belongs to unsupervised learning (Chen et al., 2016; L. Liu et al., 2012).

Examine some transactions that involve many items: item A appears in Transaction 1, item B appears in Transaction 2, and both items A and B appear in Transaction 3. So, is there a law between the appearance of items A and B in the transaction? In knowledge discovery in a database, the association rule is a knowledge pattern that describes the laws that co-occur between items in a transaction. More precisely, the association rule describes how much effect item A's appearance has on item B's appearance through quantitative measures.

By using life insurance as an example, a policy is a business. Insurance companies often need to record detailed information about the insured before accepting insurance and sometimes go to the hospital for a physical examination. The policy records the age, gender, health status, work unit, work address, and salary level of the insured person. The personal information of these insured persons can be regarded as items in the transaction. By analyzing this data, we can get a correlation rule such as that 45% of insured persons over 40 who work in Zone A have ever filed a claim with an insurance company. In this rule, "over the age of 40" is item A, "working in Area A" is item B, and "has claimed a claim from the insurance company" is item C. It can be seen that Area A may be more polluted, and the environment is relatively poor, resulting in poor health and relatively high claims rates for people working in the area.

Generally, the following four characteristics are used to describe the properties of an association rule: credibility, support, expected confidence and effect.

In the mining of association rules, the following points should be noted:

(1) Fully understand the data.

(2) Clear objectives.

(3) Data preparation should be done well. Data readiness depends on the first two points. Data preparation will directly impact the complexity of the problem and the achievement of the goal.

Discovering association rules goes through the following three steps:

(1) Connect data and prepare data;

(2) Given the minimum support and least trustworthiness, use the algorithm provided by the data mining tool to discover the association rules;

(3) Visualize, understand, and evaluate association rules.

3.2.2.2 Mining process of association rules

The association rule mining process consists of two main stages:

Stage 1: All Frequent Itemsets must first be identified from the data set.

Stage 2: Association Rules are then generated from these high-frequency project groups. The frequency of a project group is called support; taking a 2-itemset containing A and B as an example, we can find the support degree of the project group containing {A, B}. If the support degree is greater than or equal to the minimum support threshold value, then {A, B} is called a high-frequency project group.

The second stage of association rule mining is to generate Association Rules. If the metric values in the original database are taken from continuous data, then appropriate data discretization should be carried out before the association rule mining (in fact, the value of a certain interval corresponds to a certain value). The discretization of the data is an important step before data mining, and whether the discretization process is reasonable will directly affect the mining results of the association rule.

3.2.2.3 Classifications of association rules

Depending on the situation, association rules can be classified as follows:

Based on the categories of variables processed in the rule, association rules can be divided into Boolean and Quantitative association rules. Boolean association rules deal with discrete and categorized values, showing the relationship between these variables. In a database with multi-dimensional and multi-level association attributes, the quantitive association rule processes numeric fields, dynamically segments them, or directly processes the original data. For example: gender = "female" = > occupation = "secretary", is a Boolean association rule; Gender = "female" = >avg (income) = 2300, the income involved is a numeric type, so it is a quantitive association rule.

3.2.2.4 Related algorithms on association rule mining

This research mainly applies the Apriori algorithm, which uses candidate sets to find frequent item sets. It is one of the most influential algorithms for frequent itemsets mining for Boolean association rules, and its core is a recursive algorithm based on a two-stage frequency set. The association rule is a categorically single-dimensional, single-layer, and Boolean association rule.

3.2.3 Anti-fraud model based on rule engine

3.2.3.1 Definition of the rule engine

The rule engine is an inference engine. It maps rules with facts already existing in the rule knowledge base through a matching algorithm. It uses certain conflict resolution and mechanisms to deal with conflict rules in a certain order and execute them (Li.& Cai, 2003). The technical architecture of the rule engine is derived from artificial intelligence, with the characteristics of selective judgment, and rich knowledge, and its origin is the expert system, which is a system that imitates the human way of thinking, and can explain and prove the results and conclusions obtained in a description that humans can understand and are accustomed to. For further analysis of the rules engine, an expert system is introduced. The rules-based expert system consists of three parts: the Inference Engine, the Rule Base and Working Memory. Specifically, the inference engine contains the pattern matcher, the agenda), and the execution engine. In the interference engine, the pattern matcher matches rules with fact data, and the results of the matching determine which rules in the rule base need to be executed; the agenda is responsible for arranging the implementation order of the selected rules. The necessity is that for a system with many rules and factual data and rules, in the matching process, when successfully matching fact data, multiple rules will inevitably appear, and the actions taken between multiple rules may differ. There is a conflict, and the agenda adopts a specific conflict resolution strategy. The rules engine generally performs the following steps:

Read into the fact data; Conflicting rules into the collection; Conflict resolution; Repeat steps above until all the rules in the Conflict collection in the agenda have been executed.

3.2.3.2 Training methods of Anti-fraud rules

The common fraud and illegal abuse behavior patterns found by experts in actual operation are collectively referred to as violation scenarios. From the violation scenario, combining rules with models is used to precipitate previous expert experience. Currently, the training method of the rule engine is to define and classify common fraud patterns through medical experts, form rule logic, and then customize relevant rules for relevant violation scenarios through the database language and rule engine. The idea of medical insurance rule-making is as follows:

First, experts summarize and define risk control scenarios through rich experiences, such as insured people max out Meidcare cards for assigned drugs.

The experts then define the expected behaviors in this risk scenario, such as if the insured person prescribes more than ten different Medicare drugs in a week, it is suspected that the drug

is over-prescribed.

Next, the experts verify the results produced by the rules, continuously adjust the rule thresholds, accumulate positive cases, and provide annotation data for subsequent modeling. In this process, some validated rules are summarized into the rule set.

The modeling process will begin when the labels (positive cases) accumulate to a certain extent. In general, the risk model is a classification problem in supervised learning. In this process, experts will provide model engineers with guidance on feature engineering, such as the insured who generally swipes cards for drugs frequently, and these features can be obtained through physical portraits.

Before the model is officially launched, experts judge the results of modeling risk control to ensure the accuracy and recall of the model. The risk control results can be fed back to the physical portrait for the graph link analysis of the subsequent risk control model (see below for details).

We start from the fraudulent motives of insured persons, pharmacies, doctors, hospitals and other entities, and gradually build a library of rules and models for risk scenarios. Every day, the risk scenario database is used to comprehensively sort out and explore the entities with updated data, and aggregate the risk results for review by audit experts. In chapters five, six, and seven, we collate some typical risk scenarios for insured persons and healthcare organizations and illustrate their modeling ideas.

3.2.3.3 Introduction to the related rules engine

The relevant rule engine of this study mainly applies the two technical architectures of Hadoop and Spark, which are introduced in what follows:

(1) Hadoop

Hadoop implements Google's Map, reduces programming model and framework.

(2) Spark

Spark can be seen as a supplement to Hadoop, it is good at is handling streaming workloads, interactive queries, and machine learning.

3.2.4 Fraud detection based on entity portrait technology

The medical insurance fund is a special fund dedicated by the state to protect the basic medical behavior of insured personnel, share risks, and benefit the country and the people. The medical insurance fund has been running smoothly in recent years. Still, individual insured persons or designated medical institutions regard the pooled fund as a cake of interests. In order to better

establish the supervision mechanism of the medical insurance fund, the National Medical Insurance Bureau in February 2019 released the document *Notice of the National Medical Security Bureau on Doing a Good Job in the Supervision of the Medical Security Fund in 2019*, which clarified the focus of future work: local regulatory departments will work together from multiple angles such as daily supervision, special governance, and flight inspection to establish a high-pressure situation against fraud and insurance fraud.

Big data is the technology emerging in the new era, the current medical insurance stocked data has reached an astonishing magnitude - it is necessary to introduce new technical means to dig deep into the medical insurance data, from which to find out the physical portrait of the hospitals, the insured person, better understand the changes in the expenditure of the local medical insurance fund, find out the question points, take the initiative to find and solve the problem - these are the new requirements of the current new situation.

3.2.4.1 Application of entity portrait on medical insurance anti-fraud

Physical portrait (or user portrait) is one of the core data technologies of internet companies and is the data cornerstone that drives company field sales, product optimization, operation optimization, and risk prevention. The basic principle of the physical portrait is to collect the user's all-time contact points on different pages (such as one user visiting the women's sneaker page at 23:12), and by following their behavior trajectory in each rule-based business scenario while using machine learning or deep learning model to label the user, it can finally build a panorama of the user’s behaviors at all businesses.

Physical profiling technology can also greatly assist the supervision of medical insurance funds through the data profiling of insured persons, hospitals, pharmacies, doctors, and other entities of the medical insurance system to achieve all-around supervision of medical insurance funds (see Chapter Four for details). Considering the insured people, it is possible to record their historical medical trajectory, medication habits, and other behaviors. For the insured people detected to have abused medical insurance in the system, we can extract the characteristics of medical treatment frequency, reimbursement cost, disease type, and other characteristics from the user portrait and find similar insured groups for identification.

3.2.4.2 Application of graph-link analysis on medical insurance anti-fraud

Graph link analysis is a natural extension and enhancement of the medical insurance entity profile. We can take the insured person, doctor, hospital, pharmacy, and other entities as the node, the interaction between the entities as the edge, build the medical insurance entity association map and use the graph link analysis to carry out anti-fraud modeling. Graph link

analysis is a technology widely used in tax, telecommunications, and social media to discover fraud patterns through business links in business correlation graphs.

3.2.4.3 Proactive defense self-learning medical insurance monitoring network based on entity portrait

This program is divided into three main steps: establishing physical profiles of different health insurance entities, such as insured persons, doctors, and pharmacies. Next, based on physical profiling, we use different modeling techniques to detect and warn of health insurance risks, and finally, we use data visualization to display and process risk results. Below we describe the functional modules and technical selection for each step.

(1) Entity Persona Portraits of Medical Insurance

Entity profiling technology takes the entity of interest, such as the insured, doctors, and hospitals, as the starting point, deeply integrates omnichannel data, and deeply analyzes entity attributes and the correlation between different entities from various dimensions. Entity profiling allows us to quickly obtain various entity attributes and entity associations while constantly precipitating new analysis results (such as labeling an insured person as high-risk). The physical portrait provides strong support for designing various anti-fraud functions and risk control models downstream and is the data basis of the medical insurance intelligent monitoring network. Constructing a perfect physical portrait requires a deep understanding of the risk control business and a long-term iterative process. In the following, we take the portrait of the insured as an example to briefly explain how to gradually create a physical portrait of medical insurance risk control.

(2) Integrate participant data

Medicare data is rich in information: each reimbursement document contains information such as specific time, the insured person, medical service provider, doctors, diseases, medical program drug costs, and diagnosis and treatment (outpatient or inpatient). For a given insured person, all the documents generated can be grouped according to the disease, arranged in chronological order, and preliminary statistical analysis can be made.

(3) Starting from the needs of risk control, build a label and indicator system

The construction of the tag system needs to start from the business scenario, which is a certain correlation description of the entity within the scope of the scene. It is necessary to grasp the main contradictions through in-depth communication with medical insurance audit experts and build an indicator system for insured persons according to actual needs. For example, if we want to build a risk control model for chronic disease people to over-arbitrage drugs, then we

should consider the insured's past diseases, age groups, frequency of visits, commonly used drugs, and other information; these needs are our guidelines for building a labeling system. After preliminary analysis, we can establish the insured label.

By perfecting the labeling system of the physical portrait, we can easily analyze the entity from different angles. Moreover, these indicators or labels can be used for feature engineering, which greatly facilitates the construction of downstream risk control models.

(4) Entity association map

An important attribute of an entity is its association with other entities (such as pharmacies and doctors that the insured often visits), and digging deeper into the entity association relationship is the key to Medicare's anti-fraud methodology.

We use the graph to characterize and describe the association relationship between entities from the data level: take the entity as the node, and build a graph model when the correlation relationship between the entities is the edge. By communicating with healthcare experts and carefully observing the health insurance data structure, we propose virtual entities of "medical events". A medical event represents an outpatient or inpatient behavior of an insured person coming to a medical institution on a specific date. We can integrate the relationships between insured patients, doctors, hospitals, diseases, and medical projects through medical events.

(5) Proactive Warning Based on the Anomaly Detection

Because Medicare fraud is highly evolutionary, the existing expert experience cannot fully protect against unknown risks, and anomaly detection is a common means of defending against unknown risks in anti-fraud methods. The rationale is that a small number of fraud cases differ between the data level and a large number of normal cases, so we can model normal behavior, and any sample that deviates from normal behavior can be considered at higher risk.

There are three types of exceptions in total:

1. Point-wise anomaly: A sample that differs significantly from other samples in properties. For example, if the general outpatient frequency of diabetes type 2 patients is 1 to 4 times a month, spending 300 to 500 yuan, then a patient's monthly frequency of 10 times and spending 2,000 yuan will be a single-point anomaly.

2. Contextual anomaly: Refers to an abnormality that behaves in a specific situation. For example, if the daily sales curve of a drug in a pharmacy is normal, seasonal, and cyclical but suddenly surges on a certain day, the date should be considered abnormal.

3. Group anomaly: refers to the group anomaly that appears when a set of normal samples appear as a cluster. For example, dozens of hospitalization receipts handled by a doctor are exactly the same in terms of cost, which should be considered abnormal.

4. Starting from the physical portrait, we model the normal behavior of many insured persons, doctors, and medical institutions, build an anomaly detection model for the above three abnormal situations, and present the detected samples to experts for review. In general, the Isoforest, DBscan algorithm in classical machine learning, or the AutoEncoder algorithm in deep learning can be used for anomaly detection.

[This page is deliberately left blank.]

Chapter 4: Interview Research on Medical Insurance Defraud Behavior

4.1 Interview research

4.1.1 Concept of interview research concept of interview research

The Interview methods are divided into structured and semi-structured interview methods. Semi-structured interviews are informal interviews conducted according to a rough outline of the interview prepared in advance rather than a fine, accurate quantitative questionnaire. Compared with structured interviews, semi-structured interviews have higher degrees of freedom and more flexible processes. This study focuses on the research content and outlines, through understanding the actual phenomenon of the interviewee's understanding of medical insurance violations and the relationship between medical insurance violations and supervision. Through semi-structured interviews, we can deeply understand whether medical insurance violations are common in specific places (medical service providers) and different groups. In the interview, the words and questions are flexibly adjusted according to the interviewee's conditions, answers, and situations. There are no specific and mandatory requirements for the way and order of the questions, the way the interviewees answer, and the time and place of the interview. The interview follows the principle of voluntariness and confidentiality. After the interviewee's consent, it is conducted on the spot without rehearsal and reservation. The recording is carried out for the collection, summary, analysis, and writing of subsequent materials.

4.1.2 Content of interview research

As shown in the interview outline in Appendix 1, this study is divided into two parts. One is an interview with patients and families, and the other is with doctors and deans. In each, through five to seven questions, to understand the identity of the interviewer, their understanding of medical insurance violations, and medical insurance violations in the daily operations. The researchers mainly explore the following two questions:

1. The specific operation and behavior of medical insurance violations in daily life.

2. The internal relationship between regulatory laws and medical insurance violations.

4.1.3 The implementation process of interview research

According to the requirements of semi-structured interviews, the implementation process is divided into the following steps:

1. Draw up the outline of the interview

Around the research content and theme, relevant literature and documents were studied. Depending on the literature and actual needs, the outline of the interview was formulated, including whether medical insurance violations were common in the medical practice and the relationship between regulatory agencies and medical insurance violations.

2. Conduct interviews

Interviews were conducted at Hospital H for patients, family members, and doctors on behalf of the hospital. The interviewees were randomly selected in the hospital and interviewed separately to ensure the quality and independence of the interview.

3. Process the materials

After the interview, the recording is timely dictated, the original materials are archived, and the documents are sorted out to provide proof in the following thesis.

4.2 Qualitative results

4.2.1 A basic description of the interviewee

In order to make the research content accurate, the selection of research subjects will be based on hospital H, which is also the researcher’s training hospital. And a randomized, unlimited number of interviews are conducted with the doctors, patients, and their families in the hospital until the survey results show no new situations have emerged. In this study, a total of 32 doctor-patient groups were interviewed, as shown in Table 4.1 in Annex B.

Table 4.1 Basic information of interviewees

No.	Interview Location	Interviewee	Record Method
1	Random locations in the hospital\ reception room in the department and inpatient area	Dr.1	Voice recorder transcription
2		Dr.2	
3		Dr.3	
4		Dr.4	
5		Dr.5	
6		Dr.6	
7		Dr.7	
8		Dr.8	
9		Dr.9	
10		Dr.10	
11		Patient 1	
12		Patient 2	
13		Patient 3	
14		Patient 4	
15		Patient 5	
16		Patient 6	
17		Patient 7	
18		Patient 8	
19		Patient 9	
20		Patient 10	
21		Patient 11	
22		Patient 12	
23		Patient family member 1	
24		Patient family member 2	
25		Patient family member 3	
26		Patient family member 4	
27		Patient family member 5	
28		Patient family member 6	
29		Patient family member 7	
30		Patient family member 8	
31		Patient family member 9	
32		Patient family member 10	

After communicating with the interviewees and describing their interview needs and relevant confidentiality measures, we entered the hospital's internal medicine department to interview the patient. Before the interview, the researchers presented the background, purpose and confidentiality principle of the interview to the patient. And confirming that the interviewee fully understood the purpose and confidentiality principle of the interview, the interviewer interviewed each for three to five minutes and recorded the dialogue with the interviewee's consent. Among the interviewees, patients and their families understand the role and function of medical insurance and get some knowledge about medical insurance violations. The doctor group represents the hospital, has a wealth of experience in dealing with patients, knows something about the illegal operation of medical insurance in the field and patient groups, and has fully understood the regulatory requirements of the regulatory authorities.

4.2.2 Description of patient and family member

Table 4.2 Specific behavior of medical insurance violations

Subjects	Violation Method	Specific Behaviors	Description
Analysis Result on the Insurance participants	Inpatient or outpatient abnormal behavior	Super short hospitalization	The length of stay was calculated according to the discharge date and admission date to determine whether the length of stay was less than the threshold
		Lower standard for hospitalization admission	Identify documents that may not meet the criteria for admission: The number of days of hospitalization is less than or equal to the threshold and the proportion of drug costs, laboratory and examination costs, and treatment costs, in addition to bed costs, exceeds the allocated threshold
		Frequent outpatient	Statistics within a time period, the number of insured outpatient visits exceeds the threshold
		Frequent inpatient	Statistics within a time period, the number of insured hospitalization exceeds the threshold
		Hospitalization decomposition	The insured person is admitted to the hospital many times in a row in the same department of the same hospital due to the same disease, and the interval is less than the threshold
		Suspected imposter for medical service	Swipe others' cards, mostly borrowed from family relatives
	Abnormal behaviors on medication	Drug abuse	The type and number of the same drugs in one prescription exceeds the threshold
		Abnormal descriptions in Chinese herbal tablets	The type of Chinese herbal tablets in one prescription exceeds the threshold, with suspicion of taking the tablets as a "placebo"
		Abnormal cost of single drug purchase	For the same insured person, the amount of drugs purchased at a single pharmacy exceeds the threshold
		Prescribing medicine without a diagnosis	The drug is prescribed without diagnostic information, and the drug is prescribed based on experience or imagination
		Take medicine in advance(chronic disease)	For the insured person's chronic disease, when the same drug is prescribed multiple times at the same medical institution, the duration of the remaining drug prescribed last time is longer than the threshold
		Take medicine in advance(inter-agency for chronic disease)	For the insured person with chronic illness, when the same drug is prescribed multiple times in different medical institutions, the duration of the remaining drug prescribed last time is longer than the threshold

		Frequently take medicine	The number of times the same insured person in the same medical institution or all medical institutions in the selected statistical period (same, week, month) repeatedly prescribes the same drug (generic name and the labeled dosage form are the same) is greater than the limit value
		Restrictions on medical materials	The gender of the insured person does not match the gender applicable to the medical materials
		Medication does not much the gender	The gender of the insured person does not match the gender applicable to the medical materials
		Overdose	The prescribed dosage exceeds the maximum daily dose
	Abnormal behaviors on examinations and consumables	Double examinations	Statistics over a time period, the same insured person in t Statistics over some time show the same insured person in the same medical institution repeatedly carries out the same examination more than the limit
		Cash out from the medical insurance card	Use medical insurance card to buy daily necessities at the drug store or medical institutions; use health insurance card to buy large quantities of certain drugs and resell them for cash
		Examinations out of scope	Whether the examination items in the list of diagnosis and treatment items which are restricted on the level and type of medical institutions violate the restrictions

According to the content of the interview, the illegal content is divided into two parts for a detailed description. Table 4.2 is based on interviews with patients and their families. From the standpoint and perspective of patients and their families, their daily violations are unconscious, and they act with other accomplices in some planned insurance violation behaviors. These behaviors fall into three categories: outpatient or inpatient abnormalities, medication abnormalities, and abnormal tests or consumables.

1. Outpatient or inpatient abnormal behavior: According to the interview content, the patient's outpatient and inpatient abnormal behavior mainly includes frequent hospitalization, frequent outpatient clinics, suspected impersonation, and other behaviors in the short term. Specifically, it is manifested in frequent outpatient visits in a calendar year cycle that does not match the severity of the disease and is very keen on admission to hospitals for patients with their medical insurance benefits or commercial insurance. There are cases of mixed-use and misappropriation of others' cards; on the one hand, the medical insurance card quota is commonly used by the family members, and the other is to brush the medical insurance card

for others. For the above circumstances, it is suspected of unreasonable admission, treatment, and insurance fraud for reimbursement according to regulations of Opinions on Further Promoting the Reform of Medical Insurance Payment Methods and Prescription Management Measures.

2. Abnormal medication: Medication abnormality is a more common practice among patients because it is difficult and inconvenient to check. It is often impossible to confirm the patient's words, whether they have real needs for the required medication and other factors. Abnormal medication includes: agreeing to an excessive number of drug types in the prescription, abnormal types of Chinese medicine tablets in the same prescription, no diagnostic prescription, patients with chronic diseases requiring the drugs beyond the description, descriptions with apparent gender inconsistencies, and over-dose medication without following the doctor's instructions. Medical insurance experience constitutes the above as insurance fraud, medical treatment, and medical insurance fraud according to the Prescription Management Measures, Clinical Drug Instructions, Characteristics of Clinical Routine Use, Notice of the State Council Planning on Deepening the Reform of the Medical and Health System during the Twelfth Five-Year Plan Period and Implementation Plan and other documents.

3. Examination or consumables abnormalities: according to patient interviews, examination or consumables abnormalities cannot be identified by non-professionals. The vast majority of patients and their families trust the checklist issued by the doctor. Still, some patients use insurance cards to doublecheck and consumables out for their own interest. For example, cardholders go to designated pharmacies or medical institutions to swipe for drugs, use medical insurance cards to buy daily necessities, or use medical insurance cards to buy many drugs for resale.

4.2.3 Description of the medical service provider

Table 4.3 Specific violations of medical institutions

Subjects	Violations	Specific Behaviors	Descriptions
Medical Institution	Abnormal inpatient or outpatient behaviors	Lower the standard of hospital admission	Identify documents that may not meet the criteria for admission: The number of days of hospitalization is less than or equal to the threshold and the proportion of drug costs, laboratory and examination costs, and treatment costs, in addition to bed costs, exceeds the allocated threshold
		Suspected impersonate for medical service	The card information does not match the person
	Abnormal Medication	single-prescription Western medicines and proprietary Chinese medicines	The number of drug types within the same prescription exceeds the threshold
		Abnormal type of Chinese Herb tablets	In the same prescription, the types of Chinese medicine tablets exceed the threshold, and there is a suspicion that Chinese medicine tablets are regarded as placebo
		Gender restrictions for medical materials	The gender of the insured person does not match the gender applicable to the medical materials
		Gender discrepancy	The gender of the insured person does not match the gender applicable to the drugs
	Abnormal examinations or consumables	Double examinations	Statistics over a time period, the same insured person in the same medical institution repeatedly carries out the same examination more than the limit
		Swipe for drugs and resale	Buy daily necessities with a medical insurance card at designated points or medical facilities; Use the card to buy lots of a certain drug and sell them to cash out.
		Tests beyond the scope	Whether the examination items in the list of diagnosis and treatment items that have restrictions on the level and type of medical institutions violate the restrictions

Table 4.3 describes the medical insurance violation concluded from the interview content from the perspective of the doctor and the dean. According to the above, medical insurance violations in medical institutions mainly fall into three categories: outpatient or inpatient abnormal behavior, abnormal medication, and abnormal examination or consumables.

1. Outpatient or inpatient abnormalities: Medical institutions may take advice or actions for some patients on low standard admission to reduce the burden on patients or gain profit. If a patient uses other's card, the doctor will sometimes neglect it and not take the drug as a sensitive issue. In public hospitals, such long-term, and frequent problems will be marked or tracked in the medical insurance accounts. In private hospitals or pharmacies, there are frequent violations.

2. Abnormal medication: On the sensitive issue of drug prescription - whether it is to prescribe more drugs, the same class of drugs, Chinese herbs proprietary medicines, drugs and equipment that do not match the patient's gender and condition - public hospitals have strict supervision and excellent doctor training on medical insurance, so, the incidence is low. More attention should be paid to the misconduct of private hospitals pursuing profit, especially the abuse of traditional Chinese medicine and some healthcare drugs suspected of being used as placebos.

3. Examination or consumables abnormalities: In medical practice, the unreasonable examination and use of medical consumables are not easy to detect. Given the impact of medical institutions as authorities, it is difficult for patients and non-stakeholders to question them validated. However, in interviews, some patients, alongside their trust in doctors, also questioned medical equipment that could not be explained and tracked. It shows the lack of openness and transparency, and supervision in this link needs to be improved.

In summary, irregular medical insurance use is common in the insured and the institutions providing medical services. Whether for good intentions or interest, their actions hurt the shared interests of national health insurance. For insured patients and their families, chronic patients do not want to move back and forth primarily due to simple mindsets such as disliking trouble themselves. Consequently - family members will use the patient card, resulting in wrong behavior, and some patients will intentionally make illegal profits through private channels. Violations by medical institutions are generally private institutions with inadequate supervision and unclear training of physicians. Hospitals will take the initiative to provide patients with a seemingly win-win solution, taking advantage of the asymmetry of patient information to induce patients to cooperate. There are also individual doctors' private actions, using the authoritative platform and identity of the hospital to perform some irregularities.

4.2.4 Interview materials collection and analysis

The original materials of the interview were saved through audio recording and transcribed the audio into written materials, collecting a total of 16,958 words. Then proud of each interview screen, delete meaningless "ah", "um", "well", "na" and other tone words, repeated words, long

pauses and other invalid information. After the necessary logical supplements, word and sentence filling, refining and sorting out, the effective document contained about 14,000 words. To summarize the differences and connections between each interviewer, we do not just record the interviewer's accidental structure, but describe them in an invisible, more logical structure (Kvale & Brinkmann, 2009). Medical insurance is the link between medical institutions and medical services demand. It is entrusted by the government to monitor the hospital, control the excessive growth of medical costs, and ensure the essential medical services of patients. It is mandatory but a kind of incomplete commodity (Z. M. Li, 2002). In this study, the research content is collated and analyzed through interviews from the medical demand side and the medical service side, that is, patients and families, hospitals, and doctors. First, learn from the interviews between the two parties about the specific ways and behaviors of widespread and actionable health insurance violations that patients and doctors are exposed to in their daily lives. Second, learn about the intrinsic link between health insurance violations and local regulatory and regulatory behavior through interviews.

4.2.5 Interview record and analysis

For the specific operations and behaviors of medical insurance violations in daily life, this interview mainly sorts out violation phenomenons, the doctors' and patients' understanding, medical insurance payments, the implementation of doctors in the use of medical insurance cards, and the doctors' diagnosis and treatment behaviors;

(1) Some of the doctors interviewed described the doctors' and patients' cognition of medical insurance payments, and from the doctor's point of view, the patient's understanding of the use and protection of the medical insurance card is relatively unclear. The main descriptions are as follows:

“Patients are not very clear about how the insurance works; they know that health insurance could fill a part of their medical expenses, and they do not care whether the insurance card can be used alone. They care more about using it to get more drugs.” (Dr. 1)

“Most patients are aware of using health insurance cards, but some patients may use the health insurance account as something similar to a wallet. Just seeing a doctor and then using up that money, they may not take it as health coverage for him and not realize it's insurance – overall protection.“ (Dr. 2)

(2) Some research subjects have described the implementation of medical insurance payments, and some have shown that patients will make certain groups cheat to take reimbursement or receive gifts for medical insurance use. For many ordinary individuals, these

behaviors may be accomplices carelessly, or they may not be able to identify that they gain a little advantage. The main descriptions are:

"There are such patients. Especially some customers in their 60s to 70s – for problems that can be solved in the outpatient clinic, they want to be hospitalized for a more thorough examination so that the reimbursement ratio in medical insurance will be more economical for them." (Dr. 3)

"Seeing from an annual cycle, sometimes the patient's condition does not meet the criteria for hospitalization, but the patient requests to be hospitalized. The drug will be prescribed for several months when they are discharged, also because this patient generally has better insurance benefits, may want to be hospitalized, and generally, we will also help." (Dr. 4)

"The private medical examination institutions will send rice, oil, or eggs, to attract the elderly." (Family member No. 1)

"This has never happened. In no top three but private pharmacies, it seems to be more common. Especially lure the old man near the pharmacy to measure blood pressure with free toothpaste, toothbrush, or something like this." (Patient 1)

(3) Some research subjects described the impact of medical insurance payment on doctors' diagnoses and treatment behavior. The interview found that some doctors were vague about implementing the medical insurance payment system. They were accustomed to being strict with some diagnosis and treatment services, and the reimbursement drug list and reimbursement process were relatively straightforward, but implementing the dosage regulations for prescribing drugs was more relaxed. The main descriptions are:

"Mostly, we will bend the rules. But there are some very important tests or treatments that we do not allow it. If it is simply some commonly used drugs that do not affect it, we can handle the case as one sees fit, but it could happen to anybody, right?" (Dr. 5)

"Speaking of dispensing drugs, as long as the patients take his card and medical record, the drugs are basically within the medical insurance list; we will help. But when it comes to the anesthetic drugs such as painless gastrointestinal surgery, we will strictly investigate and not allow mistakes." (Dr. 6)

"Some patients ask for some loquat syrup or compound herb mixture. Some medicinal ingredients will have some sedative effects, and it may be that some people need to prescribe this drug for some reason (addiction). But we won't do that. For example, an overdose of some diabetes injections and oral antidiabetic agents will cause serious complications or consequences, so we will not prescribe much." (Dr. 7)

"In my personal experience, some doctor suggested to let me put my card in him for some time, so that he can prescribe many drugs for me in the following months. Because normally for chronic disease, he cannot prescribe many drugs at once." (Family member No. 2)

(4) Some service providers have violated the norms of diagnosis and treatment for excessive diagnosis and treatment and excessive examination. The main descriptions are as follows:

"In single hospitalization, repeat blood type, blood lipids, tumor markers, and other test items are not necessary." (Doctor 8)

"Charges for Chinese herbal fumigation, traditional Chinese steam bath, and traditional Chinese herbal steaming for the same body part at the same time." (Family member 3)

(5) Repeated charges, over-standard charges, and breakdown items charges in violation of price regulations are specifically described as follows:

"In intravenous drip, extra items such as disposable infusion apparatus, disposable drug dispensing syringes and other material fees are charged. When the vein detained needle is used for intravenous drip, charges for venous catheterization and arteriovenous catheterization." (Doctor 9)

"The nursing fee includes the vital monitoring signs such as body temperature, pulse, blood pressure, and respiration, and repeated charges for measurement of them." (Patient 2)

"In Color doppler echocardiography examination, charges for extra items such as ordinary cardiac M-mode ultrasound examination and ordinary two-dimensional echocardiography."

"Exchange of items cannot be charged for special materials, such as: disposable laminar flow hood, disposable labour bag, disposable dressing bag, nursing bag, tape printing wristband, skin preparation bag, specimen bag, sterile protective cover, nursing pad, calcium lime, disposable sterile gloves, microscope covers, calming strips, gynecological materials, surgical irrigators, disposable orthopedic surgery kits, disposable surgical kits, disposable nail boxes)." (Doctor 10)

(6) Inducing or assisting others in seeking medical advice or purchasing medicines under false names, providing false certification, and fabricating medical service items, the specific description is as follows:

"Even knowing the actual patients are children, the doctor still prescribe for adults and use adult's medical insurance cards to seek medical advice and purchase medicines for children." (Patient 3)

(7) The medical insurance policy of basic medical insurance is cumbersome, and the reimbursement procedures are complicated, which is much affected by the age and physical condition, and the below is perceptions of the patients and their families:

“The patient's understanding, I often use the medical insurance card to buy medicines, and I will also register in the hospital, but I am not very clear about the other reimbursement procedures you mentioned. I always ask my children to do it, and I cannot make it.” (Patient 4)

“Maybe because I rarely go to the hospital. I hardly use my medical insurance card, and don’t even think of swiping my card to buy medicine.” (Patient 5)

“The scope of medical insurance reimbursement varies in different regions. In some regions, you can reimburse chronic drugs, while in others cannot.” (Family member 4)

“The medical insurance reimbursement procedure is so tedious that it needs the assistance of young people, especially offsite reimbursement. Throughout the reimbursement process, you need the accompany of children.” (Family member 5)

“If you suffer a chronic disease and take the same medicine many times in the same medical institution, while there is still much remaining drugs from the last time.” (Family member 6)

“Count the number of times one insured person repeatedly prescribed the same drug (in the same generic name and dosage form) in the same medical institution and all the other medical institutions within the selected statistical period (certain day or week or month), exceeding the limit.” (Family member 8)

“Count the number of repeated examinations exceeding the limit within a period of time, in the same medical institution.” (Family member 9)

“Use my medical insurance card to buy daily necessities at pharmacies or medical places.” (Patient 6)

“Use the medical insurance card to buy a large amount of certain medicines and resell them for cash.“ (Family member 10)

“Both point-of-care B-ultrasound examination and B-ultrasound routine examination were charged on the hospitalization form.” (Patient 7)

“I usually go to the hospital every year for maintenance. Even if there is nothing wrong with my body, the doctor will find a way to help me with the hospitalization and use the medical insurance for reimbursement.” (Patient 7)

“I don’t know exactly how to use medical insurance. I have participated in the cooperative medical system; every time I take it for hospitalization is enough. Then the doctor can prescribe any physical check for me at a price lower than self-pay.” (Patient 8)

“I don't have medical insurance. Instead, I use my old friend Mr. Zhang’s card to get medicine in the hospital, only need to pay a little each time.” (Patient 9)

"I had a broken bone, and no one can take care of the home. I stayed in the hospital for three months without discharge in the middle. The doctor suggested I continue my hospitalization by spending at my own expense one day every month." (Patient 10)

4.2.6 Supervision and regulation

For the internal relationship between supervision and medical insurance violations, this interview mainly figures out whether the diagnosis upload of outpatient and inpatient diagnosis is standardized, whether the use of drugs is reasonable, and whether the insured person's medical behavior and drug purchase are reasonable;

Some research subjects described issues related to: whether hospital charges are reasonable in the past diagnosis and treatment process, whether outpatient and inpatient diagnosis upload is standardized, whether doctors will prescribe some unnecessary examinations and laboratory tests for the sake of profit, and whether patients will choose hospitalization to reduce their burden. The main manifestations include patients reducing the standard of hospitalization, project inspection, outpatient upload registration, hospital fee rationality, health examination, and common diseases. The main descriptions are:

"There are indeed some patients who would like to be hospitalized from time to time. Or in every year, they want to be hospitalized for some examinations, treatments." (Dr. 2)

"For the insured patient not retired, the basic out-of-pocket expenses are 20-30%, and if they are retired, the general out-of-pocket expenses ratio is 10%. That means more drugs prescription and less self-payment in hospitalization. The proportion of private payments in the outpatient clinic is relatively high." (Dr. 3)

"The examination needs to be self-paid, but you can get medical insurance reimbursement if you are hospitalized for a comprehensive examination." (Dr. 4)

"There are many professional things that we, ordinary patients, do not understand, but it does feel as if there are many items to be tested for a disease, and I do not understand what this project is for; I can only follow the doctor instruction that it is necessary. The doctor is very busy, so he will not explain carefully what is checked on and why it is necessary." (Patient 1)

"The frequency of medical insurance cards placed in the doctor's prescription is quite high. Because the elderly at home are afraid of queuing trouble, the community clinic offers some fixed medicines packages." (Patient 5)

"There have been times when the doctor said to let me do me a favor, put the health insurance card in his time for a while, dispense the medicine or give some appropriate compensation." (Patient 6)

It is described whether the use of drugs by research subjects is reasonable, specifically repeated medication, more than 3 antibiotics prescribed in a single outpatient clinic, and assistant drugs exceed the standard. Given the role of medical institutions as authorities, it is difficult for patients and non-stakeholders to question them for sound reasons. However, in interviews, some patients, alongside their trust in doctors, also questioned medical equipment and examinations that could not be explained and seen. It shows that the lack of openness, transparency, and supervision in this area must be improved. The main descriptions are:

"Many patients come to see something that does not fit their condition. For example, with the less commonly used painkillers and sleeping pills, we judge that he does not in line with the condition, and we will not agree to his demands." (Dr. 2)

"Now our medical insurance is networked; for example, the patient has got drugs in our hospital and then did the same in other hospitals. When he comes to visit us, whn we prescribe drugs to him online, on the computer it will pop up a reminder of what drugs he has been prescribed in a certain hospital, which needs you to pay attention." (Dr. 3)

"This kind of patient does have such a need - for example, if he is older and has a long-term chronic disease, it is not convenient for him to come to the hospital, and sometimes he will be prescribed a large prescription for several months. But some people use other people's health insurance cards to get drugs here. Under such conditions, generally hospital prescription system will send a warning; we will also pay attention and give them verbal warnings." (Dr. 4)

"In the face of the patient's condition not meeting this drug, as long as he proposes, we will prescribe it to him, without understanding how the patient would deal with these drugs, because it is true that each patient has very short time slot on the work schedule." (Dr. 5)

"Normal dosage is generally applied to some chronic disease patients who hope to come less often, do not have to register repeatedly, and hope to prescribe a long prescription. This situation happened. But we do not recommend prescribing long drugs for the requirement of health insurance, including considerations for his illness and health. Therefore, for this kind of patient, we recommend that he come within 3 months, for instance, and then follow up with a consultation. Others beyond the scope of medical insurance are also basically refused." (Dr. 6)

"As I went to see the doctor in the private medical center, I had more unnecessary items. For a small cold, they prescribed many drugs or healthcare drugs attached and examinations: such as CT and other imaging exams. All these added a lot and led to a quite high final payment. But in the public hospital, it is better." (Patient 1)

"In the past consultation and treatment, there were sometimes some doubts about the examination programs offered by the hospital, and I think it is quite common. There has been

a fever before that is a low-grade fever, which should also be a cold, and then the doctor will prescribe a lot of tests and will not explain the reason for each test. Prescribing drugs like some anti-antibiotics, which we know are anti-inflammatory. In addition, there are a lot of Chinese medicines that are accompanied by colds and fevers. Sometimes doctors are impatient and don't explain in detail the evidence of each medicine." (Patient 2)

"In the past medical treatment, I feel that the hospital has set up a lot of diagnoses and treatment items. There are more common in public hospitals that only communicate with patients for a little while; the doctor will advise examinations from head to toe, even if it is irrelevant. As far as I know, because the equipment used is expensive, he is asking you to share the cost of some of the machines." (Family member 2)

The research subjects described whether the insured person's medical behavior and drug purchase were reasonable. It is mainly manifested in the number of drug purchases in individual pharmacies greater than that of hospitals, the total amount of drugs purchased greater than 10,000, the set of drugs brushed, suspected impersonation of medical treatment, and the number of single monthly visits exceeding 20 times. The main descriptions are:

"Generally, when I want to buy some medicine for my parents from the doctor, and I come in a lot or I buy a lot of medicine, I put my medical insurance card there and I use the medical insurance card to write the medicine." (Patient 4)

"I remember that once I took my grandmother to the Chinese medicine to prescribe some drug, and the doctor did say so, put the card here, maybe the dose at one time cannot be so much, so you need to take the card away after a while. It is a private clinic, and friends also recommend the doctor, and it is quite famous." (Family member 2)

4.2.7 Conclusion

1. The awareness of the seriousness of health insurance is insufficient. With the full implementation of China's basic medical insurance system, the implementation of medical insurance gives different patients different benefits. Because public-funded medical care is completely different from the policy and operation mode, medical insurance will have a far-reaching and extensive impact on both the supply and demand sides of medical services (doctors and patients). The results of this survey show that most patients have not yet realized the importance of basic medical insurance, and the current medical insurance policy has a certain impact on patients, which will affect the rationality of medical expenses and the relationship between diseases and medical insurance expenses.

2. Medicare fraud is widespread. Some doctors get performance standards to increase the hospital revenue, driven by the economic profits, a very small number of hospitals have adopted illegal operations such as "decomposition hospitalization", such as when the hospitalization cost of medical insurance patients reaches the "hospital quota" and the condition does not meet the discharge conditions, individual hospitals require patients to go through the procedures for re-admission after paying their expenses, and private clinics will choose to give gifts in order to promote medical insurance consumption to carry out a series of fraudulent insurance behaviors, so the medical insurance management department must strengthen the supervision of policy implementation.

For insured patients and their families who can enjoy publicly-funded medical care, with the increase in the number and number of hospitalizations, most of the insured people in order to reduce the burden of medical expenses they bear are mostly out of simple concepts such as fear of trouble, chronic diseases do not want to run more, family members deserve to use my card, but still, choose to be hospitalized, and even some patients are premeditated illegal profits through private channels. Therefore, in order to make the reasonable treatment of patients more effective and economical in the use of health resources, it is necessary to strengthen the publicity of medical insurance policies, so that the insured can fully understand their rights and choices, and also reduce the economic sharing of diseases of medical insurance patients.

3. Medicare fraud is subtle and systematic. Due to the lack of awareness of health insurance laws, taking advantage of the policy loopholes as much as possible is regarded as their advantage for taking. From the perspective of malpractice, each step is operated in accordance with the regulations, so the specific steps and the fraud implementation are very hard to spot under the standard operation. As the supervision of medical insurance fraud is increasingly strict, simple behavior in one part of the fraud is not easy to succeed. The fraud behavior in the form of collaboration between upstream and downstream and medical insurance fraud with systematic characteristics approved by all relevant parties has become the main part of medical insurance fraud, which is highly harmful to the healthy development of the overall medical insurance system.

Chapter 5: Medical Insurance Fraud Identification Based on Big Data Model

Through the investigation and interview analysis in Chapter Four, we find that the risk of medical insurance funds does exist. Because medical insurance fraud involves various stakeholders, in the specific implementation, doctors deliberately neglect their duties, and patients take a chance to defraud, making the fraud behavior systematic, hidden and complex. This chapter will examine the source of existing health insurance risk problems, that is, how to identify health insurance risks. At first, this study builds a process model for healthcare fraud identification and then conducts healthcare fraud identification based on scenarios and collected data.

5.1 Construct a process model for medical insurance fraud identification

5.1.1 Thought and basic theory

This chapter studies the identification of medical insurance fraud, and its basic process ideas follow the logic of "data-algorithm-process-strategy", referring to the practice of (T. Zhang et al., 2022). The primary theoretical basis of the process design is: (1) Risk society theory; that is, through interview research, it is found that the risk characteristics of medical insurance fraud problems do exist. And it mainly analyzes the complex correlation interaction between these problems, artificiality, uncertainty, and the double-edged sword effect to classify this problem as a comprehensive systemic problem; (2) Regulatory sandbox theory, the process design of three algorithms, provides a pre-emptive basis for policy introduction. And for all interested parties, it is only necessary to focus on the output of this study and treat the identification of medical insurance fraud risks as a data-driven applied scientific algorithm. In this safe space, the administration reduces certain regulations for enterprises to allow communication and experiment with innovative solutions before entering the market; (3) Welfare economic theory and externalities. The fundamental purpose of health insurance fraud identification is to establish more equitable mechanisms and processes. If all parties involved (including fund managers) have the same need to bend the rules, initiatives should be reformulated to meet the expectations of all parties. This is another use of the results of health insurance fraud

identification in this study, which constitutes one of the practical contributions of this study.

5.1.2 Targeting the medical insurance fraud process model

Based on risk society theory and regulatory sandbox theory, the healthcare fraud process model mainly addresses the following social issues:

(1) The top-level design of the policy system

The high leverage of medical insurance has caused the continuous deterioration of fiscal expenditure. The design of the medical insurance system and the formulation of medical insurance policies and regulations are not reasonable, violating the principles of insurance and medical laws, the relevant laws, regulations, and measures are not perfect, and the supporting reform measures are lagging behind, thus causing risks to the safety of the medical insurance fund or failing to protect the basic medical needs of the insured.

Therefore, it is necessary to take advantage of various data in the primary insurance business and use the basic principles and methods of actuarial science to calculate the social medical insurance program, including the calculation of liability reserves and solvency analysis involved in medical insurance operation and management in addition to the calculation of insurance premiums.

(2) Expense risk caused by changes in the disease spectrum

In recent years, with the aging of the population, the medical insurance fund expenditure on severe diseases and chronic diseases occupies a large share of the total. Chronic non-communicable disease treatment costs are much higher than traditional infectious diseases in medical treatment. Therefore, the change in the disease spectrum inevitably leads to increased medical costs. A detailed management plan for the expenditure on serious and chronic diseases is necessary.

(3) Risks caused by backward informatization

The business operation and management of the medical insurance fund are highly dependent on the information system or the information provided by it; if the system fails or is damaged, resulting in business failure or information error, it will have very serious consequences. Insurance defrauds, cheating behaviors or erroneous payment of funds due to incorrect information, illegal alteration or forgery of information, and inaccurate information provided result in the failure of risk control measures and information leakage.

The theory of information asymmetry means that in market economic activities, there are differences in the understanding of relevant information among all types of personnel, and those with sufficient information are often in a good position. In contrast, those with poor information

are in a relatively unfavorable position. In medical behavior, some personnel will take advantage of the information asymmetry between regulatory authorities and medical institutions to carry out insurance fraud. The reimbursement information obtained by the regulatory authorities occurs afterward rather than being monitored during real-time patient visits, which is an important reason for the use of information-based risk fraud.

From the perspective of improving the management level of medical insurance funds, the medical insurance system is a closed-loop information system. In this closed management process, medical insurance fund regulators can introduce the theory of regulatory sandbox, which refers to the regulator establishing a specific framework after adopting appropriate security measures and introducing new risk identification technology to improve the refined management.

The above risks are an important cause of fraud, waste, and abuse of medical insurance funds, resulting in loopholes in the management of medical insurance funds. Medical insurance fund managers must find problems through existing medical data and optimize relevant management systems and processes.

(4) Problems caused by medical violations

Based on the risk society theory, the scope and depth of human intervention in social life and nature have expanded. Decision-making and repeated behaviors have become the main sources of risk, and consequently, man-made risks have surpassed natural risks becoming the leading content of risk structure. With the help of modern governance mechanisms and various governance tools, human beings have improved their ability to cope with risks, but at the same time, they are faced with new types of risks brought about by governance, namely institutionalized risks (including market risks) and technical risks.

In the medical insurance fund management, specific performance of the risk society is violations to defraud reimbursement fund, that is, violate medical insurance management regulations and policies, use fabricating the evidence, concealing the truth and other methods to defraud medical insurance funds or medical insurance treatment from medical insurance management institutions, resulting in medical insurance fund risks. The fraud, abuse and waste faced by medical insurance funds are particularly prominent, posing a serious threat to the medical insurance funds

5.1.3 Medical insurance fraud identification model

This research focuses on risk society theory and regulatory sandbox theory, while information asymmetry, welfare economic theory and externality theory explain data collection and policy

formulation. By taking the "data-algorithm-process-strategy" as guidance, the overall medical insurance fraud risk identification model is established, which uses the integration of local medical data to carry out research through the following three steps of data aggregation, data screening and data analysis of various tools, in order to provide reference ideas for the risk identification path in medical insurance management (See Figure C1 in Annex C).

Firstly, at the policy formulation risk level, through a hospital's data research, the overall impact of the top-level design of the local basic medical insurance policy on the reimbursement behavior of hospitals and patients is analyzed. Also, they are questioned about what problems are brought to the medical insurance fund when formulating the serious illness treatment policy.

The second step is to study what personal fraud and insurance fraud behaviors the medical insurance fund currently faces due to problems such as disease spectrum and backward medical insurance system. These behaviors include hospitals, patients and other entities.

Thirdly, due to the limitations of the rule engine, this method can only detect relatively simple individual medical insurance fraud and cannot profoundly identify new types of organized fraud. Therefore, this study introduces the persona portrait to make up for the shortcomings, which mainly studies organized fraud and fraud in specific scenarios.

5.2 Data sample selection

5.2.1 Data selection

This study takes data from Huzhou City, Zhejiang Province, as an example, and obtains data through daily business cooperation, through the municipal medical insurance bureau, hospitals and various pharmacies, and obtains the consent of all parties. There are 2.4 million people in Huzhou, and this analysis project covers outpatient, inpatient and pharmacy data from January 2018 to April 2019.

5.2.2 Descriptive presentation

In this dataset, from January 2018 to April 2019, the total number of outpatient visits was 50,365,776, and the number of hospital admissions was 598204, as shown in Table B1 in AnnexB.

Regarding the number of medical visits, the total outpatient visits were 5,291,775,942.14954 yuan, and the total number of inpatients was 6301992436.29 yuan, as shown in Table B2.

5.3 Basic medical insurance-related policy risk mining model based on association rule analysis

There are many kinds of medical insurance-related policies. Most relevant policies in various places are formulated using historical data analysis, which relevant data may not verify, and the risks within the policies cannot be known. Therefore, introducing big data analysis methods and analyzing the implemented policies is an important part of risk control.

5.3.1 Analysis content

The medical record's front sheet is a summary of the patient's basic situation and diagnosis and treatment process, including the patient's basic information, outpatient medical information, inpatient diagnosis information, and inpatient costs. This study uses the information on the first page of inpatient cases of HZ People's Hospital of HZ City, Zhejiang Province, in 2017, with a total of 7382 cases and nearly 160 indicators. Combined with the significance of the indicators and the research purpose of this dissertation, the indicators such as gender, age, household registration, occupation, marital status, discharge department, length of stay, hospitalization cost, payment method, number of surgeries, and hospitalization were selected as the research objects.

5.3.2 Analysis procedure

5.3.2.1 Medical record data cleaning

The data on the first page of the medical record involves personal privacy; first, the data is privacy filtered, and the relevant indicators that can infer the patient's identity are removed to protect the rights and interests of the hospital and patients. Secondly, according to the actual meaning of the variables and the requirements of the association rules, the following data processing is carried out:

1. Anomaly handling

There is an outlier of "gender=3" in the gender (male=1, female=2) indicator, which does not match the actual situation, and a total of 25 cases of "gender=3" are deleted.

2. Extreme-value handling

According to the national average daily inpatient cost data from 2015 to 2020, the average daily inpatient cost of China's tertiary hospitals and secondary hospitals represents the average daily inpatient cost level in China, as shown in Figure C2. Among all the data, the

hospitalization cost of one case exceeded 1 million yuan, the hospitalization cost of some cases was less than 1 million yuan, and some numbers of hospitalization days exceeded 1,000 days or less than 1 day. The above cases are relatively special; if such cases are included for analysis, the analysis results will inevitably deviate from the general situation, so seven cases containing more than 1 million yuan, eight cases with more than 1,000 days of hospitalization, and 151 cases with hospitalization expenses of less than 100 yuan were deleted, as shown in Figure C2.

3. Handle medical cases with missing key information

Some key indicators of medical cases were missing, such as the discharge department and payment method, and could not participate in the analysis, so 73 medical cases were deleted.

5.3.2.2 Descriptive analysis and grouping

According to the mining process, after selecting the indicators and cleaning the data, make a descriptive statistical analysis of the selected indicators (see Table B3). Then analyze and group the indicators (see Table B4). And apply the Apriori algorithm to generate correlation rules based on R language version 3.1.1.

5.3.2.3 Apply the Apriori algorithm to generate association rules

The R rules package was used to generate association rules, and the function "apriori()" in this package can perform association rule analysis of Apriori's algorithm.

1. Generate association rules. The preprocessed data is imported into R software through the "read.table()" function, and then the “apriori()” function is used to generate association rules.

2. Check and review rules. Use “inspect()” to check the returned association rules, and use the summary() command to view the summary of the association rules.

3. Extract valuable rules. According to the degree of support, confidence, or promotion, the sort() and inspect() functions are used to sort and view the obtained association rules, and valuable association rules are extracted and analyzed based on the actual meaning of variables.

5.3.2.4 Result analysis

According to the mining results, combined with relevant medical knowledge and the actual situation of the hospital, the following conclusions were drawn:

1. HZ City's serious illness medical insurance covers a wide range of diseases, which can reduce the pressure of hospitalization costs for patients to a certain extent;

2. The settlement procedures of inpatient expenses and medical insurance in different places are relatively complicated, which increases the burden of inpatient expenses for patients in different places to a certain extent.

3. There is a strong correlation between the discharge department and inpatient cost, and different factors lead to differences in inpatient expenses in oncology, interventional medicine, respiratory, otolaryngology and neurosurgery.

4. The high-quality medical resources in short supply are unevenly distributed.

The expanded analysis of the conclusions above are as follows:

(1) Huzhou City's serious illness medical insurance covers a wide range, partially reducing pressure on patients' hospitalization costs.

In order to reflect the relationship between hospitalization expenses and medical payment methods, the rules are shown in Table B5 for linking hospitalization expenses with medical payment methods.

The inpatient expenses of 80,000-150,000 yuan, 150,000-300,000 yuan and more than 300,000 yuan and the medical payment method of basic medical insurance for urban employees are 0.54, 0.50 and 0.55, respectively. Still, the improvement degree is less than or equal to 1, and the rules are unavailable. The hospitalization expenses were 5000-10,000 yuan, 10,000-30,000 yuan, 30,000-80,000 yuan, and the support of the basic medical insurance for urban employees was 0.15, 0.13 and 0.14, respectively. It can be seen that patients with more than 5,000 yuan hospitalization choose basic medical insurance payment methods for urban employees with a higher degree of support. The proportion of patients with hospitalization expenses in the above three ranges was 84.5%, reflecting the broad coverage of serious illness medical insurance in HZ, which can effectively reduce the medical burden of most patients.

(2) The settlement procedures of inpatient expenses and remote medical insurance are relatively complicated, which increases the burden of hospitalization expenses to a certain extent

In order to reflect the impact of medical treatment in different places on medical payment methods, household registration is linked to medical payment methods, and the rules are shown in Table B6.

The increase in the payment method of basic medical insurance for urban employees in WX district, NX district, DQ county, CX county, AJ county and other cities is less than 1, and the rule is not available, indicating that patients in the above cities are excluded from choosing the payment method of basic medical insurance for urban employees. The improvement degree of patients at the same level in Huzhou City in choosing the payment method of basic medical insurance for urban employees was greater than 1, and the support degree and confidence degree were 0.32 and 0.74, respectively, indicating that patients at the same level in Huzhou City chose the payment method of basic medical insurance for urban employees.

(3) The discharging department has a strong correlation with the amount of hospitalization cost

After the frequently set screening, it was found that the association rule between the discharge department and hospitalization cost was highly confident, as shown in Table B7.

The confidence and improvement of Ear, nose& throat department (ENT) and hospitalization costs of 0-5,000 yuan were the highest, 0.91 and 7.06, respectively; the confidence of respiratory and hospitalization expenses of 5000-10,000 yuan was 0.76 and 0.58, respectively, the improvement was 2.81 and 2.15; the confidence and improvement of neurosurgery and hospitalization expenses of 10,000-30,000 yuan were 0.57 and 2.59, respectively; the confidence and improvement of oncology, intensive care medicine and inpatient expenses of 50,000-80,000 yuan were 0.48 and 3.72, respectively. The above five association rules have a higher improvement degree than 1, and the rules are available and have a high confidence level, which is a strong association rule. This study takes the above five subjects as an example and uses multi-dimensional correlation rules to further analyze the reasons for the differences in hospitalization costs.

(4) The scarce medical resources are unevenly distributed

The relationship between hospital expenses and length of hospital stay can reflect the scarcity and uneven distribution of medical resources. The normal length of hospitalization is generally 10-20 days, and compared with surgery, conservative treatment has a slower recovery time, so that the hospital stay will be extended.

As seen from Table B8, the number of hospitalization days for first-time inpatients with hospitalization costs of less than 10,000 yuan and no surgery is less than 10, which is lower than the normal hospitalization days. Even if the hospitalization cost is 10,000-80,000 yuan, the number of hospitalization days is mainly 10-20, which is still in the normal range. Then we make associations between the hospitalization cost, the number of hospitalization days, and the number of surgeries.

As seen from Table B9, the hospitalization days of first-time inpatients with one operation, and hospitalization expenses of 5000-10,000 yuan, 10,000-30,000 yuan, and 30,000-80,000 yuan, respectively, are mostly within 10 days.

It can be seen from the above two rules that regardless of the hospitalization cost and whether surgery is performed, the hospital shortens the patient's hospital stay as much as possible to improve the bed turnover rate, reflecting the current uneven distribution of medical resources.

5.3.3 Summary

This section presented a hospital's settlement data mining multi-dimensional attribute association rule, finding out the relevant factors of hospitalization costs to understand the impact of local medical insurance policies on actual medical costs. Also, it provided strong support through big data analysis to understand the impact of the region's policy level on the insured; in the HZ area, the existing policy risks are:

(1) The distribution of medical resources in the region is uneven, and large hospitals are concentrated at the city level, resulting in an increase in the burden of medical treatment for patients outside the urban area.

(2) At present, the government's treatment of serious diseases covers a wide range, effectively alleviating the burden of medical treatment for residents to a certain extent.

(3) Critical diseases such as tumors, neurosurgical, and respiratory diseases. These are still important factors affecting medical expenses in the region, suggesting that managers need to adjust reimbursement policies and appropriately tilt toward diseases with higher medical expenses.

5.4 Medical insurance expenses risk identification model based on rule engine

5.4.1 Data source

In order to verify the risk control model, data from January 2018 to April 2019 in HZ City, Zhejiang Province, was used for verification. The drug model, hospital admission and exit inspection model, insured medication screening model, inspection item control model, and price model, were used for real settlement data screening.

This analysis related to the risk identification model covers multiple perspectives regarding the insured, medical institutions, and designated pharmacies. It also included an analysis of the insured's medical trajectory, and sets of drugs, the distribution of the insured's disease diagnosis, a comparative analysis of the basic information of the insured and medical visits, an analysis of the collusion items of medical institutions, rationality analysis of hospital admission, analysis of hospital medication repeatability, analysis of hospital auxiliary drug analysis, analysis of the limited scope of hospital directory, analysis of hospital medical insurance reimbursement scope and charging items, analysis of pharmacy violations.

5.4.2 Risk model modeling thinking

At present, the medical insurance policy of HZ City implements various forms of comprehensive payment methods such as payment by disease, payment by project, prepayment by total, per capitation, and according to a fixed amount. Therefore, the analysis will start from related business scenarios such as disease scores and projects.

5.4.2.1 Diagnosis-intervention packet variation method (Single department)

(1) The total cost has increased, and the total score value is also on an upward trend The specific score of the disease score has not changed since 2018, such as coronary heart disease 87 points have not moved. The cash value of the observation score has not shown a clear upward trend, which indicates that the increase in the total cost is related to the increase in the number of hospital cases.

(2) From another perspective, behind the upward trend of a hospital's total cost (P) is the increase in the number of admitted cases (N). Since the score (C) of a single disease is constant, and the amount of money behind each score is stable, it can be concluded that: P=N*C.

(3) Deduction 1: In the case of a significant increase in the number of medical visits, if the number of medical staff in hospitals/departments does not increase and the efficiency does not increase, it will inevitably lead to a decline in the quality of medical services.

(4) Deduction 2: If there are cities that have recently undergone payment method reform, the proportion of cases diagnosed by disease score in each hospital before and after the change can be counted to further analyze the change in the chain score value. For example, in a hospital department, the previous diagnosis of coronary heart disease accounted for 13%, and unstable angina for 7%. Still, after the reform of disease score, coronary heart disease accounted for 3%, and unstable angina accounted for 17%, which is the obvious high-score behavior. Cases with this part of the change are singled out for key audits.

5.4.2.2 The similarity of diagnosis and treatment items

(1) Standardize the diseases of the same department corresponding to a particular type of diagnosis and treatment by focusing on relevant drugs and examinations.

(2) Compare the similarity of the documented treatment items with the standard one.

(3) If the diagnosis and treatment items of the document are higher than the standard similarity (>80%, the threshold is adjustable), and the diagnosis score of the document is high, consider the possibility of swiping cards.

5.4.2.3 Decomposition of hospitalization period

(1) Eliminate emergency hospitalization, rescue hospitalization, and hospitalization in different departments

(2) The time of admission for the second hospitalization event, less than three days between the discharge from the previous hospitalization event

(3) The treatment plan had evident continuity during the two hospitalizations, and the treatment items were similar

5.4.2.4 Lowering the standard admission/unreasonable admission

(1) Compared with the average hospital stay for the same disease, less than 60% of the average hospital stay (threshold adjustable)

(2) Compare the average cost of the same disease and age, less than 60% of the average cost (threshold adjustable)

(3) Compare the nursing grade of the same disease and age, and mark the documents below the average nursing level.

(4) The proportion of examination costs is abnormal, or prescribe medicine for admission.

(5) The proportion of infusion times is abnormal. The drip time for the same disease may reach N, but patients admitted to the hospital with low standards may not have an intravenous drip in one composing period.

5.4.2.5 Other scenarios

(1) Frequent medical visit: The average number of medical visits for the same disease, age, and period in a specific area exceeds the number of medical visits. It can be distinguished by outpatient or by hospitals, pharmacies, and the threshold of the relevant number of times can be adjusted.

(2) Excessive medication: within a period, the scope and quantity of drugs used by a hospital/department exceed the number of drugs used by hospitals of the same level to treat the same disease; In a single document, the amount of relevant drugs used exceeds the maximum use limit in the instructions; In the documents, there are more than two auxiliary drugs for the same symptom, such as the abuse of blood circulation and stasis drugs.

(3) Excessive examination and treatment: within a period, the prescription inspection volume of a hospital/department is abnormal, exceeding the prescription examination volume of the same disease treatment in the same level of the hospital; For some individual examinations, such as glycated hemoglobin, bone density, blood type, tumor markers., set a

limit on the frequency of examinations per unit time, and more than this number is positive; Related examinations are not associated with diagnosis and treatment, such as bronchitis, secondary care patients, head MRI and other examinations are done during the treatment.

(4) Screening concept for non-compliant behaviors in the limited categories of medical insurance catalog library such as drug, diagnosis and treatment items, materials, and diseases:

a) Drug-related: drug contraindications, hitchhiking prescription, single flavor tablets are not paid, single compound refusal, drug excess, brush drugs, second-line drugs, over-course medication, contraindications to injection mixing, limited children, special control of brain calm, excessive testing, repeated medication, hitchhiking prescription in the same group, more than three kinds of antibiotics, over-limited drug purchase, serial drug exchange, abnormal medication in the population, auxiliary medication

b) Diagnosis and treatment items: serial exchange items, one hospitalization beyond the limit of the project, payment limit of the project, non-integer number of surgeries, excess of the project, bilateral addition, large-scale imaging examination review

c) Materials: Material and associated item analysis

(5) Screening concept for unreasonable irregularities of price charges:

a) False billing: In the medical details, there is a logical relationship between some items, such as nursing days and bed fees, air conditioning and heating fees, and oxygen intakes up to 24 hours a day. If the relevant medical treatment item data exceeds the normal logical value, it is considered false billing.

b) Duplicate claims: Documents are reviewed to see if they coincide with previous document dates.

c) Repeated reimbursement: Drugs and related diagnosis and treatment items, such as CT examination, and large dressing changes, exceed the appropriate hospital charges.

d) Double charges: For the same diagnosis and treatment items, multiple charges are judged according to the frequency requirements in the treatment interval of the relevant project unit.

5.4.3 Analysis procedure

5.4.3.1 Data warehousing construction

The data warehouse model is a window that directly reflects the business logic in the data warehouse. The data warehouse model is formulated according to the business development plan and strategy. The data warehouse in this topic is based on the medical insurance data

mining platform and the Apache Hadoop big data platform, using Spark distributed real-time computing. The results of the algorithm's model are presented to various anti-fraud services in the application layer.

5.4.3.2 ETL process

The ETL process is data extraction, transformation, cleaning and loading. It is necessary to select data according to the topics needed to mine and establish proper ETL decisions. For example, what time and frequency to extract from the data source, how to clean the data after extraction, delete redundant data, vacant attributes, redundant attributes and conversions between attributes, join between some tables, establish related views, and so on.

5.4.3.3 Modeling and computing

(1) Risk feature extraction: For the cleaned data, extract the required features of the model and perform the necessary feature transformations;

(2) Machine learning models: The model is trained to identify various risk scenarios based on the generated features. The schematic diagram of the modeling and training process see Figure C3.

5.4.3.4 Details of the extracted data field

The data must be stored in the corresponding field in the warehousing process. The settlement data storage method in this topic is as follows: Set the corresponding field name and field example, determine whether it is necessary, and draw a medical receipt table and a medical visit details table, series of operations seen in Tables B10, Table B11, and Table B12.

In Table B12 we make the corresponding example of the code meaning and display the values of the code involved in the above table.

5.4.4 Risk control model results

5.4.4.1 The insured over swiping cards for drug

The drug model is calculated by the algorithm that the insured gets the same drug on the same day, and the basic logic is to filter the entire data of the insured through data screening to find out the problematic data, as the rankings of drugs cost and amount of HZ civilians over swiping card cases shown in Figure C4 and C5. As seen from the following Table B13, the two insured people have the largest amount of drugs, and it gives detailed examples of the analysis of the patient's medication time and the visited hospital.

5.4.4.2 Exchange of medicines

Drug exchange is a kind of pharmacy-based risk scenario. In some pharmacies, in order to pursue profits, some daily necessities are exchanged for drugs; in this scenario, the abnormal behavior of pharmacies includes the following characteristics.

(1) The time of card swiping is abnormal

The abnormal number of card swipes includes two meanings; one is that the number of people who swipe cards in a pharmacy is more concentrated, and the second is that the number of individual card swipes is large. Through model analysis, it was found that in a pharmacy, more people are using several types of cards that meet the above two characteristics, as shown in Table B14.

(2) Different individuals swipe the card to get a similar prescription

The content of individual card prescriptions is similar, meaning that different individuals have similar drug combinations and amounts. As can be seen from the table below, the drug combinations of the two groups of different insured persons are similar, as seen in Table B15.

(3) The package of prescription drugs is not practical, and the amount is similar

As seen from Table B16, the daily prescription amount of an insured person is around 110 yuan, and the amount is similar; in such cases, there is a suspicion of drug exchange.

5.4.4.3 Decomposition hospitalization period

In order to reduce the average cost of hospitalization, some departments will decompose the length of hospitalization for patients with more prolonged or higher amounts into two or more hospitalizations. It is a means to control expenditure and reduce the average hospitalization cost and length of stay in the department. It is a means for hospital departments to prevent overspending. Some details can be founded in Tables B17 and B18, where it can be found that some hospitals had repeated admissions.

5.4.4.4 Lower the standard for admission

Low-standard admission refers to (1) Admitting patients who can be treated on an outpatient basis or for outpatient observation treatment in violation of the hospitalization standards set by the health department. (2) The admitted patients suffer from mild diseases and stable vital signs, and long-term medical advice is mainly oral medication. (3) The patient's hospital stay after admission is short, mainly based on inspection items, and there are no corresponding treatment measures, as shown in Table B19.

5.4.4.5 Off-label adjuvant medication

Adjuvant medication refers to drugs that help increase the effect of the main therapeutic drug or increase its efficacy by affecting the absorption, mechanism of action, and metabolism of the primary therapeutic drug. According to medical insurance policies, complementary medication for off-label medication means that certain adjuvant drugs must be used under a proper disease diagnosis. According to the Notice of Implementing and Improving the Guiding Opinions on the Centralized Bulk-buy of Drugs in Public Hospitals, it has been clearly stated: "Establish a prescription review and physician interview system, focus on tracking and monitoring auxiliary drugs and drugs used by hospitals beyond normal use, clarify the authority of physicians to prescribe, and when prescriptions involve valuable drugs, they should take the initiative to communicate with patients, standardize dosage, and strive to reduce the burden of drug costs on patients with acute and long-term medication." Through data mining, the use of adjuvant drugs was screened, and the inpatients with adjuvant drugs exceeding 50% of the total drug cost were reserved, presenting a total of 91547.3 yuan and a total of 135 inpatient records. See Table B20, where the screened information is summarized.

According to local medical insurance policies, Xingnaojing injection can only be used by patients who need shock or rescue, see Figure C6.

Through data mining, the use of auxiliary drugs exceeding the standard in Changxing Hezhou Hospital was accurately located, as shown in Table B21.

5.4.4.6 Frequent medical visit

Frequent medical visits refer to the average number of medical visits in a particular area, the same disease, age, and time period, which can be distinguished by outpatient, inpatient and other types, as well as by hospitals, and pharmacies. The screening results are shown in Table B22.

Due to the large amount of data, only the top 5 insured persons were taken as an example, and the time span is from January 2018 to September 2019. As seen from the following table, this type of risk indicates that the insured frequently takes too many drugs, and the amount of drugs taken exceeds the regular dosage.

5.4.4.7 Medication beyond the applicable scope

According to the provisions of the national medical insurance catalog, certain drugs must be used within the scope of their application, and the risk warning is as follows by using the catalog modeling:

(1) Medication beyond the hospital level

Individual hospitals do not fully grasp the range of indications for medication, and secondary hospitals prescribe drugs that can only be used in tertiary hospitals, and the data are shown in Table B23.

(2) Rescue drugs used for non-rescue patients

According to the provisions of the national medical insurance catalog, Ticagrelor tablets are used for acute coronary syndrome rescue. Still, according to the model settings, the diagnosis related to the medical records of the drug is checked, and many diagnoses are not acute rescue.

This is illustrated in Figure C7. As shown in the figure, many patients diagnosed with healthy physical examination have used Ticagrelor, which is unreasonable and seriously impacts health, insurance fraud, and costs.

5.4.4.8 Overdiagnosis

Overdiagnosis refers to applying unreasonable and unnecessary diagnostic measures to the insured person during hospitalization. This time, glycated hemoglobin is taken as an example.

Glycated hemoglobin (GHb) is the product of combining hemoglobin in red blood cells with carbohydrates in serum. It is formed by a slow, continuous and irreversible glycation reaction. Its content depends on the blood glucose concentration and the contact time between blood glucose and hemoglobin, independent of factors such as blood draw time, whether the patient is fasting, and whether insulin is used. Therefore, HbAlc (GHb,TypeAlc) can effectively reflect the blood glucose control of diabetic patients in the past one to two months. Therefore, this test has no clinical significance if repeated during a short hospital stay. This time, it was found that some hospitals had the phenomenon of repeated examination during one hospitalization. It can be seen from Table B24 that DQ County has the most significant number of hospitalized patients who overuse glycated hemoglobin.

Taking the specific medical records of the hospital as an example, most patients tested HbAlc twice during a single hospitalization, and for the same patient in short-term hospitalization, it makes no sense to have a two-time examination, which is a waste of medical insurance funds, see Table B25.

5.4.4.9 Doctor repeat medication

Duplicate drugs refer to a doctor prescribing two drugs simultaneously with the same effect to a patient, which is a typical waste. Here are some examples produced by the data mining process presented in Table B26.

Below is an analysis based on the actual data of HZ No.1 Hospital. Prescribing drugs with the same effect to the same patient on the same day is a waste of the medical insurance fund, see Table B27.

5.4.4.10 False billing

False billing refers to items that exceed the number of days of hospitalization, such as bed fees, nursing fees, and other items billed daily. As shown in Table B28, nursing fees and bed fees exceed the length of hospital stay, which is a typical unreasonable billing.

5.4.4.11 Private hospitals overcharge

Some private hospitals charge a high price for adjuvant medication, seriously exceeding the standard, and a high price for related diagnosis and treatment items, such as CT examination, large dressing changes, and exceeding the appropriate level of hospital fees. See examples in Table B29.

5.5 Construction of medical insurance anti-fraud model based on entity portrait technology

Entity portrait (or user portrait) is one of the core data technologies of the Internet, that drives the company's sales, product optimization, operation optimization, and risk prevention. The basic principle of entity portrait is to collect the user's contact points in all aspects (such as visiting the women's sneaker page at 11:12 pm), starting from its behavior trajectory, according to the specific business scenario usage rules, machine learning or deep learning models for labeling the user, and build a panorama of the user at various business levels. Taking an e-commerce marketing scenario as an example, the user group who purchased women's sneakers can be used as a positive sample for constructing the label classification model of "will you buy women's shoes", for labeling the customer group. Then potential customer groups will be targeted in subsequent promotional activities.

Entity portrait technology can also greatly facilitate the supervision of medical insurance funds through the data profiling of participants, hospitals, pharmacies, doctors, and other entities of the medical insurance system to achieve overall supervision of medical insurance funds (see Chapter Four for details). By using the insured as an example, it is possible to record his historical medical trajectory, medication habits, and other related behaviors. Regarding the insured identified in the system abusing medical insurance, it can extract the frequency of

medical treatment, reimbursement expenses, disease types, and other features from the user portrait to find similar insured groups for identification

Figure C8 and C9 illustrates the Entity portrait for analysis of purchasing process and decision of women's shoes as an exemple.

5.5.1 Proactive defense self-learning monitoring network based on entity portraits

This chapter explains how to build an intelligent data-driven medical insurance fund monitoring network driven by physical portraits and based on deep learning and machine learning. We innovatively introduce the physical portrait technology commonly used by Internet technology companies into the field of medical insurance anti-fraud, accurately depict the data behavior trajectory of participants, doctors, and hospitals, and comprehensively apply different algorithms in the field of a rule engine, machine learning, deep learning and graph link analysis to build a set of active defense and self-learning medical insurance fund monitoring network.

This research plan is divided into three main steps. First, we create physical profiles of different entities of health insurance entities, such as enrollees, doctors, and pharmacies. Next, based on the physical portraits, we use different modeling techniques to detect and warn of medical insurance risks, and finally, we use data visualization to display and process risk results. Below we introduce the functional modules and technical selection of each step, see Figure C10.

5.5.2.1 Build an entity portrait of medical insurance

Entity portrait technology starts from the entities of concern, such as insured, doctors, and hospitals, integrates relevant data, and profoundly analyzes entity attributes and the relationship between different entities from various dimensions. Entity portraits allow us to easily obtain various entity attributes and entity association relationships, while constantly precipitating new analysis results. The physical portrait provides strong support for various downstream anti-fraud function designs and risk control models, and is the data foundation of the medical insurance intelligent monitoring network. Constructing a perfect physical portrait requires a deep understanding of the risk control business and a long-term iterative process. Let's take the portrait of the insured as an example to briefly explain how to gradually create a physical portrait of the medical insurance risk control system.

(1) Integrate data of medical insurance participants

Medicare data is rich in information: Each reimbursement document contains information such as specific time, enrollee, medical service provider, doctor, disease, medical program, drug cost, and medical treatment (outpatient or inpatient). All the documents can be grouped

according to the disease or arranged chronologically for a given insured. Preliminary statistics can be made, such as: basic information about age group, address, gender, medical insurance status of the insured.

Types of diseases of the insured, classification of expenses, everyday medical items, number of outpatient visits, number of hospital stays, common medical service institutions

Information on the insured's medical trajectory, such as the maximum monthly fee, the maximum number of outpatient visits in a single month

(2) Based on the requirement of risk control, we construct a labeling and index system

The construction of the label system needs to start from the business scenario, which is a certain correlation description of the entity within the scope of the scene. It is necessary to communicate in-depth with medical insurance audit experts, grasp the principal contradictions, and build an index system for insured persons according to actual needs. For example, suppose we want to build a risk control model for chronic disease people to overtake drugs. In that case, we should consider the insured's previous diseases, age group, frequency of visits, commonly used drugs and other relevant information. These needs are our guidelines for building a labeling system. After preliminary analysis, we can establish the following participant labels:

a) According to the medical division of diseases, the insured is labeled with "gynecological diseases", "chronic diseases," and other labels so that the subsequent index can focus on analysis.
b) The insured occupies medical insurance expenses and the frequency of medical treatment, and labels such as "high cost" and "high-frequency diagnosis and treatment" are marked for focused analysis
c) The medical expenses and frequency of diagnosis and treatment of participants suffering from the same disease - were longitudinally compared, and the index score of excessive medical tendency was obtained.
d) The drug use of insured with the same disease is monitored, and the probability score of suspected drug application is given to the population who deviates from the normal value
e) For insured persons with hospitalization behavior, analyze the proportion of drugs and hospitalization rationality score.

These labels describe the risk behavior of the insured from different dimensions and have good generality. By improving the label system of entity portraits, we can quickly analyze entities from different angles. Moreover, these indicators or labels can be used for feature engineering, which significantly facilitates the construction of downstream risk control models,

see Figure C11.

(3) Entity association map

An important attribute of an entity is the association with other entities (such as pharmacies and doctors frequently visited by the insured), and digging deep into the entity association relationship is the key to medical insurance fraud prevention for the following reasons:

a) A single case appears legal and compliant in many medical insurance risk scenarios. An in-depth analysis of entity associations is necessary to discover cluster-type violations and doctors and insured persons colluding to falsely prescribe drugs to obtain premiums. Although each prescription is Medicare compliant, we observe a persistent call relationship between the insured group and a certain doctor in the entity association. The graph link analysis found that these insured nodes are centered on fraudulent doctors and present community characteristics (see Chapter Six for details).

b) Risk transmission is a core feature of fraud, i.e., other entities closely linked to high-risk entities are also at high risk. By mining entity link relationships, we can automatically discover potentially high-risk entities.

c) We use the graph method to represent and describe the association relationship between entities from the data level: take the entity as the node, and establish a graph model when the association relationship between the entities is the edge. By talking to healthcare experts and taking a closer look at the healthcare data structure, we propose virtual entities for "medical events." A medical event represents an outpatient or inpatient visit to a healthcare facility on a specific date. Through medical events, we can integrate the relationship between the insured, doctor, hospital, disease, and medical program, as shown in Figure C12.

By establishing an entity graph according to the above graph data model in a graph database (such as Neo4j), we can use various link analysis algorithms to achieve community detection and risk transmission functions.

5.5.2.2 Detection and early warning of risks

This section describes how to effectively detect and warn of unknown risks based on physical portraits. Built-in rules and risk models are mainly adopted.

Previous expert experience is invaluable in the regulation of health insurance funds. We collectively refer to the common fraud, violations and abuse patterns found by experts in actual operations as risk scenarios. As a high-level summary of expert experience, risk scenarios are the cornerstone of medical insurance fund supervision. Here we start from the risk scenario and

combine rules and models to accumulate past expert experience. The relevant concepts are explained as follows in Table B30.

The advantages of the rules are clear logic and strong interpretability. Still, the disadvantages are obvious: the logic is rigid, the coverage is narrow, and it cannot adapt to frauds that change from moment to moment. Machine learning models complement the lack of rules quite well. The model is trained on a large amount of historical data to generate complex decision logic and patterns, and is not constrained by artificially agreed thresholds. Through extensive healthcare anti-fraud modeling work, the general modeling process is shown below:

(1) Experts summarize and define risk control scenarios through rich experience, such as insured people hoards lots of medical insurance drugs.

(2) Experts define the expected behaviors in this risk scenario, such as if the insured prescribes more than 10 different medical insurance drugs in one week, it is suspected of over purchase drugs.

(3) Experts verify the results generated by the rules, continuously adjust the rule thresholds, accumulate positive cases, and provide annotated data for subsequent modeling. During this process, some of the rules that have been validated are rolled up into a rule set.

(4) When the label (positive case) accumulates to a certain extent, the modeling process can begin. In general, the risk model is a classification problem in supervised learning. In this process, experts will provide model engineers with guidance on feature engineering, such as the insured who generally prescribes drugs with high frequency, and these features can be obtained through physical portraits.

(5) Before the model is officially launched, experts judge the results of modeling risk control to ensure the accuracy and recall of the model.

(6) The risk control results can be fed back to the entity portrait for subsequent graph link analysis of the risk control model.

Starting from the fraudulent motives of entities such as insureds, pharmacies, doctors, and hospitals, we gradually build a library of rules and models for risk scenarios. Every day, the risk scenario library is used to comprehensively sort out and explore the entities with data updates and aggregate the risk results to show to audit experts for review.

5.5.2.3 Proactive alerts based on anomaly detection

Because healthcare fraud is highly evolutionary, the existing expert experience cannot fully protect against unknown risks, and anomaly detection is a common means of protecting against unknown risks in anti-fraud methods. The rationale is that the small number of fraud cases is

somewhat different from a large number of expected cases at the data level. So that we can model normal behavior, and any sample that deviates from normal behavior can be considered to have a higher risk. There are three types of exceptions:

(1) Point-wise anomaly: refers to the fact that a sample has a significant attribute difference from other samples. For example, suppose the general diabetic type II patient has a monthly outpatient frequency of one to four times and costs 300 to 500 yuan. In that case, a patient's monthly frequency of 10 times and spending 2,000 yuan will be a single-point anomaly.

(2) Contextual anomaly: refers to the behavior anomaly under a specific situation. For example, suppose the daily sales curve of a drug in a pharmacy usually is seasonal and cyclical but suddenly increases on a certain day. In that case, the date change should be considered abnormal.

(3) Group anomaly: refers to the population anomaly that appears when a set of normal samples appears as a cluster. For example, dozens of hospitalization documents handled by a doctor are entirely worth the same in terms of cost and should be regarded as anomalies.

Starting from the physical portrait, we model the normal behavior of many insured persons, doctors, and medical institutions, build an anomaly detection model for the above three abnormal situations, and present the detected samples to experts for review. In general, the Isoforest, DBscan algorithm in classical machine learning, or the AutoEncoder algorithm in deep learning can be used for anomaly detection.

5.5.2.4 Self-learning modeling

Anomaly detection is a very effective way to detect unknown risks. However, anomaly detection tends to produce more false positives than supervised classification models, because not all samples that deviate from normal are fraudulent. Another problem is that anomaly detection models often fail to explain the cause of anomalies. Therefore, anomaly detection is more suitable as a detector for unknown risks, and after discovering new risk scenarios, we still use supervised classification models to detect risks.

As shown in Figure C13, modeling a new risk scenario requires immediate assistance from business experts and model engineers, and a long development cycle. Below we introduce a self-learning modeling process based on active learning in partial-supervised learning, which can quickly precipitate newly discovered risk scenarios into high-precision supervised classification models.

(1) The cases detected by anomaly detection are judged using the known risk model library, and if they do not belong to any known risks, they are prompted for expert review.

(2) Experts analyze abnormal cases to determine whether they are fraudulently abused or not. If so, create a new risk scenario and summarize the case.

(3) Find a batch of positive cases in the risk scenario (you can find similar cases by using the K-nearest neighbor algorithm, and establish an initial classification model. Deep learning and transfer learning (see section 4.2.6) can be used for automatic feature extraction during the modeling process.

(4) As mentioned in Chapter Three, the performance improvement of supervised classification models requires a lot of labeled data, and the accumulation of labeled data in medical insurance anti-fraud is very time-consuming and labor-intensive. Active learning can optimize the labeling strategy by using existing models, picking out the most uncertain samples in the data, and passing them to experts for review and annotation, which can improve the performance of the model through a small number of annotation datasets, significantly reducing labor costs.

5.5.2.5 Graph link analysis

We have established a relationship diagram of the insured-doctor-service organization in the physical portrait, which mainly includes the application in the following scenarios:

(1) Specific risk scenarios: In specific risk scenarios involving the cooperation of multiple entities to commit fraud, graph link analysis plays an irreplaceable role. Taking the Multiple cards collected for fraudulent medical treatment risk scenarios as an example, several doctors collect the medical insurance cards of multiple insured persons and collude to collect premiums. We need to identify this group fraud accurately and to identify all the members of the fraud gang (see Chapter Six Multiple cards collected for fraudulent medical treatment scenario for details). We can detect suspected doctor-patient groups through the community detection algorithm in the graph algorithm, and combine it with the supervised classification model to more accurately screen out the doctor-patient groups that conspire to set drugs.

(2) Risk transmission: A reasonable assumption in health care monitoring is that other entities closely associated with high-risk entities are themselves high-risk. For example, a doctor frequently visited by an insured person who determines that there is a drug set fraud is also highly suspected. According to this assumption, we can use the label Propagation algorithm in the graph algorithm to transfer the risk to the associated entity from the identified high-risk entity, and focus on the entities with a high-risk score that has not yet been confirmed.

(3) Graph-based feature engineering: Previous studies have shown that entity-relationship graphs can provide powerful features for anti-fraud models. Taking the model of insured people

taking medical insurance drugs as an example, the number of hops in the entity relationship diagram of current insured persons and high-risk doctors and pharmacies is a strong feature.

5.5.2 Risk identification results based on entity portrait technology

5.5.2.1 Fraud risk identification in D.R.G. payments

We entirely model the medical records of Huzhou City from January 2018 to April 2019, including 50365776 outpatient visits and 598204 hospital admissions, then figure out the fraudulent use. For example, the patients' cards used in other pharmacies during their hospitalization, the falsified hospitalization medical record, and the men's medical insurance card used to purchase gynecological medicine.

(1) Risk description

Decomposition hospitalization refers to the act of the hospital handling multiple discharge and admission procedures for the patient when the inpatient has not yet recovered. In order to reduce losses or earn medical insurance funds, hospitals often break down a patient's hospitalization into multiple hospitalizations to reduce the cost of a single hospitalization, thereby earning medical insurance funds, as shown in Figure C14.

(2) Modeling process

Each step of the modeling process is analyzed as follows

a) Define risk scenarios: Under DRGs, hospitals may break down hospitalization reimbursements and collect higher premiums.

b) Analysis strategy: This risk scenario is a typical classification model scenario, and some samples suspected of having decomposed hospitalization can be determined by training a supervised classification model to determine whether some samples suspected of decomposition hospitalization belong to decomposition hospitalization.

c) Labeling strategy: Screen two or more hospitalization documents of a certain insured person in a short period of time (with an interval of less than 1 month) as a sample set in the data of hospitals that implement DRGs reimbursement. After interviews with audit experts, it was learned that the rules for decomposing hospitalization risk behaviors generally include: the cost of hospitalization for a single disease is lower than the cost of the MDC group for the disease, and the length of a single hospitalization is shorter. Two or more adjacent hospitalizations with shorter intervals, similar hospitalization and discharge diagnoses, or their hospitalizations tend to be in the same department. Based on the above behavior, we use the rules to select

potentially positive samples for expert labeling.

d) Algorithm selection: Use the XGBOOST algorithm, which has many optimizations in processing small data volume, unbalanced data sets, and complex features and is widely used by the industry.

e) Feature engineering: time interval between two hospitalizations, cosine similarity of disease diagnosis vector (according to 3.2.6 Med2Vec), the ratio of diagnosis and treatment cost/MDC cost, length of stay, hospital level, department, hospital history of previous violations

(3) Analysis of sample results

A hospital admitted Jia, a patient with heart failure, on December 3rd. On December 17th (the 14th day of Jia's hospitalization), Jia was discharged on the grounds stipulated by the hospital when Jia did not improve and agreed with Jia to be hospitalized the next day again. On December 18, Jia was hospitalized again, and after another 10 days of treatment, his condition improved, and he was discharged. It is a significant decomposition of hospitalization, which divides patients with a long course of treatment into two courses to reduce the average hospitalization cost and obtain two DRGs settlement costs, seen in Table B31.

5.5.2.2 Detection of diagnosis exaggeration

(1) Risk Description

Low-standard condition and high-standard treatment disease group (diagnosis exaggeration) means that after the patient completes treatment, the hospital does not fill in the discharge according to the patient diagnosis, choosing a more serious medical diagnosis so as to form a low-code high-number. In order to pursue greater interests, hospitals defraud the medical insurance fund by placing the low-standard treatment disease group in the high-standard treatment disease group in the process of diagnosis and treatment, as shown in Figure C15.

The main risk behavior factors are: the single disease is lower than the overall cost of the disease in the MDC group, the distribution of disease items is inconsistent with the distribution of the disease items, the distribution of fees for each charge category is inconsistent, the disease items are inconsistent with the disease items, the treatment relevance is low or does not contain disease-specific diagnosis and treatment items, the degree of treatment is inconsistent with the severity of the disease, the timing distribution of various costs is different from the disease-related distribution, and the treatment plan is inconsistent with the clinical pathway.

(2) Modeling Process

a) Define risk scenarios: the act of hospitals falsely selecting more serious medical

diagnoses under DRGs to obtain more reimbursement amounts.

b) Analysis strategy: adopt model plus rules. First, the classifier is trained with the hospitalization details as input and the disease diagnosis as the label. Then, compare the speculative disease diagnosis output by the classifier for each hospitalization document with the reported disease diagnosis. According to the MDC grouping rule, if the reported disease diagnosis group is higher than the estimate, there is a suspicion of low code and high editing.

c) Labeling strategy: The hospitalization receipt of medical insurance itself contains hospitalization details and diagnostic information, and does not need to be manually marked

d) Algorithm selection: Considering that Med2Vec can accurately capture the correlation between diseases and drugs through word vector embedding, this model adopts DNN deep learning model

e) Feature engineering

   i. Drug features: Med2Vec as a pharmaceutical word vector embedding
   ii. Cost characteristics: proportion of the surgical cost, the proportion of medicine, the total cost
   iii. Characteristics of hospitalization: length of stay, hospital level
   iv. Patient characteristics: age, sex, past medical history
   v. Hospital characteristics: previous high set history

(3) Analysis of sample results

A people's hospital admitted patient Chen on the grounds of heart failure, and the entire treatment process was carried out with heart failure standards. The diagnosis of heart failure was changed to a better group for the diagnosis of coronary heart disease when the discharge diagnosis was reported. This case is a low-code high-number, and the patient group with a lower treatment degree and cost is set to a higher patient group to obtain a higher settlement cost, as seen in Table B32.

5.5.2.3 Overtreatment

(1) Scenario description

Overtreatment refers to doctors' excessive diagnosis and treatment of patients due to their interests (drug rebate, hospital income generation) in the treatment process.

The main risk factors are: 1. The use of drugs with the same pathological effect is significantly higher than that of other patients diagnosed with the same diagnosis. 2. The

average number of adjunct medications (e.g., Chinese herbal medicines) and ancillary treatments (e.g., acupuncture massage) is greater than the specified limit.

(2) Modeling Process

a) Define risk scenarios: Doctors abuse medical insurance funds to generate income for excessive diagnosis and treatment.

b) Analysis strategy: This risk scenario is a typical classification model scenario, and whether some documents are over-diagnosis and treatment can be determined by training a supervised classification model. However, it is worth noting that the abuse of medical insurance resources by doctors in violation of regulations is often a long-term, continuous behavior. For outpatient clinics, it is necessary to observe the doctor's medication for a certain patient for a while. Therefore, we train classification models for outpatient and inpatient visits.

c) Labeling strategy: For each pair of doctors-patients, screen out documents with excessive outpatient services, high total amounts, and an excessive number of auxiliary drugs, and let business experts mark them.

d) Algorithm selection: XGBOOST algorithm

e) Feature engineering

i. Outpatient classifier

Outpatient characteristics: outpatient frequency, the proportion of adjuvant medication, the total cost

Statistical characteristics: the frequency of outpatient clinics for the same disease, z-score

Historical characteristics: History of misuse by doctors

ii. Hospitalization classifier

Characteristics of hospitalization: length of hospitalization, the proportion of auxiliary items

Statistical characteristics: length of hospitalization for the same type of disease, z-score

Historical characteristics: History of misuse by doctors

(3) Analysis of sample results

A patient, Hao, was diagnosed with auricle chondritis, a hospitalization time of 6 days, a total cost of 3541 yuan, monosialic acid tetrahexose ganglion glyceryl ester used 1950 yuan, accounting for more than 55% of the total cost, which is excessive use of auxiliary drugs.

5.5.2.4 Illegal swiping of medical insurance cards

(1) Scenario description

Designated pharmacies or health service stations collude with drug dealers who collect medical insurance cards to use medical insurance funds in large quantities to exchange drugs. Pharmacies make illegal revenues, and drug dealers profit by taking daily necessities or quickly reselling drugs. When the pharmacy settles with medical insurance, the item is exchanged for items covered by medical insurance for settlement, defrauding the medical insurance fund, as illustrated in Figure C16.

The main risk behaviors are:

a) Drug purchase documents often appear in a certain combination of drugs, and different price ranges will appear in different drug combinations. Regionally, this combination is rarely found in other pharmacies. There is no pharmacological correlation between the combined drugs.

b) The same pharmacy has been swiped frequently in a short period of time. Further, when analyzing the transactions of pharmacies, it can be found that the peak and low peak of their trading volume on the day do not match the situation in the region.

(2) Modeling process

a) Define risk scenarios: bulk drug exchange carried out by pharmacies to obtain medical insurance funds.

b) Analysis strategy: Triggered by the evolution of the daily behavior of pharmacies, as shown in Figure C17, the number of daily insured persons, the sales volume of different drug types, and the document volume of a pharmacy in the past period can be represented by a multidimensional time series. Here, we use a multidimensional time series data classifier to determine whether pharmacies are suspected of batch-swiping over a certain period.

c) Labeling strategy: LSTM-AutoEncoder is used to detect anomalies in the multi-dimensional time series data of each pharmacy's sales volume, allowing experts to label the time points of pharmacy anomalies.

d) Algorithm selection: LSTM classifier.

e) Feature engineering: the sales behavior of the pharmacy in the past week, that is, the total quantity and total amount of different types of drugs per day; the number of documents, the number of insured persons, the average number of drugs per order, and the average amount per order.

(3) Analysis of sample results

Guo, a drug dealer, collected many medical insurance cards employing inducement and replacement and collaborated with Liu Yuan, a cardiologist at the People's Hospital, to illegally swipe medical insurance cards and obtain a large number of chronic disease drugs.

5.5.2.5 Collecting medical insurance cards for fraudulent healthcare

(1) Introduction to the model

The so-called model for card collection for medical treatment use algorithms to find out the population gathering for medical treatment, our FPGrowth frequent itemset algorithm, strongly connected component algorithm, and Bron-Kerbosch algorithm of undirected graphs. Establish insured people who meet the configuration conditions (the same day and place to visit the doctor, the items used are similar or even completely consistent, and the time of treatment is consistent), establish a complex undirected map, and then find out the suspicious people who meet the conditions according to the combination of three algorithms, see Figure C18.

(2) Screening purposes

Currently, most fraudulent insured people use clusters, similar treatment processes, and fake hospitalized cases. For example, in the Shenyang fraud case, we have sorted out several characteristics, the same length of hospitalization, the same diagnosis, and the same frequency. The local medical insurance management department found that the hospital falsified the medical records of these so-called patients and prescribed medications, but did not actually give the patients medications, and only did simple physiotherapy or no treatment. The stay for these so-called patients is generally four days. Afterward, they all received a cash commission of 300 yuan. In order to capture such patients, the traditional rules no longer work, and a single medical record is in line with medical and medical insurance regulations in the rules. Therefore, only by establishing the corresponding model in the early stage and analyzing it from the perspective of multiple medical record similarities and aggregation degrees can these behaviors be nowhere to hide. A firewall can be better established so managers can have clues and data to patrol.

(3) Brief introduction of the results

In this examination, we found that 24 medical insurance cards were gathered in a kidney disease hospital limited company, totaling more than 1.66 million. This part of the population is fixed, the interval between visits is the same, and they appear at the hospital almost every Tuesday, Thursday, and Saturday for dialysis. The number of visits was similar; no patients with such fixed dialysis were found in other hospitals. The hospital's trajectory deviated significantly from the average dialysis visit in HZ City. The clustering is more pronounced, 22

patients throughout the year, no other groups of people enter, or some patients are out of the group. Strong fixation.

(4) Conclusion analysis

According to the model, clustered visits can be found, and the relevant data model can provide managers with relevant clues, if traditional rules, single-dimensional screening, can not find similar clues, it must be achieved through mathematical modeling. At the same time, the results of this screening need to be verified on-site using clues to determine whether the hospital has retained medical insurance cards.

5.5.2.6 Diagnosis-related groups screening

In order to understand through data analysis whether public hospitals strictly reported the relevant documents of disease payment in 2018, we developed a data-driven model for payment by disease. The model is briefly presented:

(1) Modeling principles

a) The model serves the business, and the business understanding is carried out throughout
b) It is assumed that its behavioral characteristics come from the analysis of fraud
c) The same fraud scenario can correspond to multiple models
d) Model landing finds a balance between accuracy, explanation, and input and output

(2) Modeling method

a) Similarity method

In data analysis, data mining and search engines, we often need to know the magnitude of differences between individuals, and then evaluate the similarities and categories of individuals. Common examples include correlation analysis, classification clustering (K-Means.) algorithms in data mining, and search engines when recommending items. The similarity is to compare the parity of two things. Generally, by calculating the distance between the characteristics of data objects, if the distance is small, then the similarity is large; If the distance is large, then the similarity is small. For example, two fruits will compare similarities in characteristics such as color, size, and vitamin content. Problem definition: There are two objects X and Y, both containing N-dimensional features, X=(x1, x2, x3,.....,xn), Y=(y1, y2, y3,.....,yn), calculate the similarity of X and Y. The data analysis of HZ City adopts the Euclidean distance method, one of the most

commonly used distance calculation formulas, which measures the absolute distance between points in multidimensional space and is suitable when the data is dense and continuous.

b) Topic Model (LDA)

LDA is an unsupervised machine learning technique to identify latent subject information in a large-scale document collection or corpus. It uses the Bag of words approach, which treats each document as a word frequency vector, transforming textual information into digital information that is easy to model. However, the bag of word approach does not consider the order between words, simplifying the problem's complexity and providing an opportunity for model improvement. Each document represents a probability distribution of topics, and each topic represents a probability distribution of many words.

i. Modeling objectives

For documents paid by disease groups, it is judged whether the details are consistent with the corresponding diagnosis, and if not, the payment for a single disease is not established

Documents that are not paid by disease groups should be judged by comparing the details to determine whether they are consistent with the payment for a certain disease, and if they are met, they should be paid according to the disease

ii. Introduction on logic

Only review hospitalization documents

According to the patient's discharge diagnosis, summarize the list of disease diagnoses and treatment items

iii. Analysis of results

This data screening involved all cases in 2018, and a total of 7415 documents were screened out by LDA model, similarity analysis, druid import, and other processes (mainly operated in the specified catalog of diseases, without related serious complications, but not uploaded according to the diagnosis of single disease). The total amount of documents is 166 million, the reimbursement amount is 87 million, and the top 20 medical institutions in terms of the number of documents are as follows in Table B33.

The top 20 reimbursement amounts involved are shown in Table B34.

Taking HZ Maternal and Child Health Care Hospital as an example, according to the document, if a singleton is delivered spontaneously, the upload code of the disease should be

BZ0097, and the name of the disease should be a natural vaginal delivery, as shown in Table 5.1.

Table 5.1 Spontaneous single delivery

BZ0097	Spontaneous vaginal delivery	080000 and Z37.000	Spontaneous single normal delivery	5000	Newborn screening

The analysis found that 909 documents used spontaneous singleton delivery, but the upload diagnosis was spontaneous delivery, and the amount exceeded the standard.

Furthermore, statistics show that HZ Maternal and Child Health Care Hospital uses 995 documents for the singleton spontaneous delivery project throughout the year, and 909 documents exceed 5,000 yuan. The diagnostic code is not the prescribed BZ0097. It is preliminarily judged that 99% of the cases may not be settled according to a single disease. So, it is necessary to verify the settlement list of each hospital in HZ City in the previous year according to the single disease to clarify whether there is any under-reporting. (see Table B35).

In this analysis, we found that after using the relevant items, the diagnosis in the data differs from the diagnosis required to be uploaded. Since we did not have the actual settlement list of hospitals by single disease in 2018, it was not possible to verify the upload of a single disease, but through data processing, it was suggested that there might be relevant risks. (It is recommended that the management department sort out the payment for diseases and truly implement the payment for diseases under the spirit of the document.)

5.5.2.7 Pharmacy exchanging

(1) Modeling ideas

Characteristics of the swapping behavior: First, exchanging self-paid items or other non-reimbursable items for the medical insurance reimbursement catalog is a waste of the medical insurance fund. Second, for pharmacies, some drugs and daily necessities will be replaced with medical insurance items, so that items not reimbursed will enter the reimbursement catalog.

The method of natural semantic recognition and entity alignment is adopted to verify the correspondence between hospitals and medical insurance items. This method supports structured data, such as medical data in JSON format, or semi-structured or unstructured data, such as medical text after OCR recognition. The Name Entity Resolution (NER) model, belonging to natural language processing, was trained from a large amount of medical text data accumulated by business experts through annotation tools, independently developed by business experts. NER can identify and extract entities such as diseases, medical institutions, and items in the text.

For pharmacy swaps, the anomaly detection model is used for verification. For a given

dataset x(1),x(2),..,x(m), we calculate estimates for the mean and variance(σ2) for each feature. Once we have an estimate of the mean and variance, given a new training instance, calculate P(x) according to the model: when P(x) <ε, it is an anomaly.

(2) Analysis of hospital swap results

a) Analysis of hospital swap results:

This time, by setting the analysis model, it was found that the visit fee of Shanghe County People's Hospital corresponds to the pre-hospital emergency fee. The total amount is 34120 yuan. A total of 241 documents were involved. This is the only hospital in the city, and according to the medical insurance reimbursement regulations, the "home visit fee" should not be included in the scope of reimbursement. Some examples can be observed in Table B36.

b) Pharmacy swap analysis

Pharmacy swap has the following characteristics: 1) abnormal pharmacy data, 2) abnormal trajectory of insured persons. This is shown in Table B37.

In the data set of abnormal prescription volume, some insured people have similar drug varieties. This is shown in Table B38.

Some participants have a similar daily combined amount. According to the courseware in the table below, some insured people have similar drug combinations, and the amount remains around 110 yuan, as seen in Table B39.

5.5.3 Summary

This chapter establishes the medical insurance fraud identification process model, screens a large number of sample data, analyzes the medical insurance policy risk mining model by using association rules, formulates the medical insurance expense risk identification model by using the rule engine, and finally constructs the medical insurance anti-fraud model based on the materialization technology, which plays a crucial role in the identification of medical insurance fraud.

[This page is deliberately left blank.]

Chapter 6: Conclusions

Based on the analysis of the current situation of medical insurance research at home and abroad, this chapter uses the literature review, research output, big data mining, and other methodologies, and takes the medical insurance data of Huzhou City from January 2018 to April 2019 as the data source. Then it explains the reasons for the occurrence of medical insurance fraud, qualitatively discovers the characteristics of various medical insurance fraud, constructs the data model, and proposes the prediction identification. Finally, management suggestions based on the research findings are extracted to get conclusions on the academic and management significance of the research. The limitations of the research and new direction for future investigation are also presented.

6.1 Conclusions

6.1.1 Construct a medical insurance fraud risk identification model

The study uses a data-algorithm-process-strategy risk identification model. Through the data research of a hospital, the study analyzes local basic medical insurance policy, specifically, its overall impact on critical illness reimbursement of hospitals and patients and the associated problems in the medical insurance fund.

We use the association rule to analyze the influencing factors of total hospitalization cost and explore the feasibility of studying these factors through a mathematical model. In modeling and analyzing actual settlement data, specific analysis of unreasonable medical expenditure is carried out by establishing risk scenarios. We get information on the relevant risks in the region by setting multiple risk scenarios, such as those defined by the insured, pharmacies, and medical institutions.

The results show that big data modeling technology on medical insurance can be widely applied. Strengthening medical services and supervision based on credit means a promising direction in the future. Medical insurance involves many parties such as participants, physicians, medical institutions, and pharmacies and, according to the credit score of the behavioral characteristics of historical subjects, realize tiered management and services (rapid service blocklist, monitoring, and suspension of settlement), which improve the convenient services

and reduce management costs.

The employed research found that the optimization and adjustment of medical insurance policies rely largely on medical insurance data. Based on medical insurance settlement data, prescriptions, and detailed diagnosis and treatment data, through big data methods, we can effectively support medical insurance payment methods such as single disease and disease grouping and reasonably divide the payment level of various diseases, to support the adjustment of the catalog of medical insurance drugs, diagnosis and treatment items, manage to transfer drugs with high demand and good efficacy to the catalog; to support the pricing and negotiation of medical insurance programs, assist in the formulation of drug payment standards, achieve differentiated pricing of different drugs, and determine the price of special drugs according to usage. In a word, using big data methodologies, we will continue to optimize policies of the medical insurance settlement and facilitate the reform of medical insurance payment methods. At the same time, medical insurance data plays an excellent supporting role in other social security businesses and larger livelihood security fields. For example, in social security welfare certification, we can use health insurance data to know the health status of the elderly. If there are recent medical records, we can not perform the one-to-one authentication with the elderly through data comparison. If the participant’s original medical treatment was frequent but no longer sought medical treatment after a long time, we could conduct a one-on-one work certification. In this way, combined with data, management is integrated into the service, and the convenience of the masses is maximized. Another example is providing public health management service support through health profiles or health scores, providing individuals with convenient inquiries and directions such as the nearest designated medical institution, designated pharmacy, and doctors and hospital classification ranking. Another example is to provide relevant verification services for commercial insurance verification and claim settlement, provide personalized commercial insurance product customization services for different insured groups, support the establishment of a multi-level social security system, support product research and development of pharmaceutical and medical device manufacturers, and allocate social resources according to market demand.

6.1.2 Design three algorithmic closed-loop identification modes

The study designs and verifies three algorithmic close-loop identification modes to uplift medical insurance fraud identification accuracy.

This study uses the rule engine's association rules and risk identification methods to study the personal fraud faced by medical insurance funds and caused by the backwardness of the

existing medical insurance system and the disease spectrum. However, the rule engine has the limitation that it can only detect individual fraud but cannot identify new group fraud in depth. Therefore, this study introduces the overall scheme of entity portrait to compensate for the shortcomings of the rule engine.

The data warehouse model is a window that directly reflects the business logic in the data warehouse. The data warehouse model is formulated according to the business development plan and strategy. The data warehouse in this study is based on the medical insurance data mining platform and built on the Apache Hadoop big data platform, using Spark distributed real-time computing. The results of the algorithm's model are presented to various anti-fraud services in the application layer.

By starting from the actual medical insurance situation, this study discusses the necessity of new data technologies, such as artificial intelligence and big data, in supervising medical insurance funds. Since 2010, real-time settlement of basic medical insurance cards for medical treatment has been in place. In 2011 the audit rule engine was launched, which set up monitoring indicators such as excessive inspection, excessive treatment and use of expensive materials, and the cost of key monitoring groups for the extraction of transaction information such as excessive prescription of drugs, high frequency of medical treatment, and large amounts transactions. But the vast majority of abnormal transaction types screened have a low rate of manual chargebacks, and the commissioning rate of audit manpower is low. Using the closed-loop identification of three algorithms, the following problems are solved in a targeted manner:

(1) Poor flexibility in the audit process: The existing audit rules are compiled by the experience of medical insurance audit experts, and the update and transformation of their rules require a long development cycle. At the same time, due to the huge differences between different coordination areas, medical institutions and populations, the current audit system generally has the characteristics of insufficient flexibility and poor adaptation.

(2) Narrow capture area: Most rules only impose simple behavioral restrictions on common fraud and abuse, such as primary institutions cannot prescribe more than 5 kinds of Chinese herbal medicines at a time, which is not necessarily reasonable in complex medical practice and can be easily avoided. Some macro indicators, such as the proportion of drugs, although they have a certain effect on the overall standardization of medical institutions, are not effective in identifying fraudulent subjects.

(3) Unable to resist the evolution of fraud: Since the subject of medical insurance fraud is often the practitioner of medical institutions, it has a robust professional nature, can quickly detect the logic of medical insurance regulatory rules, and quickly adjust its behavior to avoid

supervision.

(4) Unilateral audit granularity: At present, audit rules mainly restrict a single medical behavior (outpatient, drug prescription, hospitalization), while medical insurance fraud, violations and abuse often have collective, continuous, serial, rapid evolution and other behavioral characteristics, such as some doctors in a health service station in series with patients, collect a large number of medical insurance cards, and falsely open drugs to cash out. Currently, the audit system is unable to identify such collective violations.

Through this study, it can be seen that entity portrait and big data mining technology are essential ways to make up for the shortcomings of audit rules and identify new fraud methods.

6.1.3 Experimental test

This study identified the risk of medical insurance fraud among 2.4 million people in Huzhou City, Zhejiang Province, and provided a scientific empirical basis for local policy formulation.

(1) HZ City's serious illness medical insurance has wide coverage, which can reduce the pressure of hospitalization costs for patients to a certain extent;

(2) The settlement procedures of inpatient expenses and medical insurance in different places are relatively complicated, which increases the burden of hospitalization expenses of patients in different places to a certain extent.

(3) There is a strong correlation between discharge department and inpatient cost, and different factors lead to differences in inpatient expenses in oncology, interventional medicine, respiratory medicine, otolaryngology and neurosurgery.

(4) The scarcity of high-quality medical resources is unevenly distributed.

6.1.4 Institutional mechanism proposition

This study puts forward the design of the medical insurance fraud risk management system and operation mechanism based on the conclusions of this study.

(1) Through new technologies, medical insurance fraud identification efficiency has been improved. This study finds that it is necessary to build a health insurance fraud identification process, introduce new technologies, and collect massive amounts of data to perfect our processes and algorithms.

(2) Optimizing and adjusting medical insurance policies depend largely on medical insurance data. Based on medical insurance settlement data, prescriptions, diagnosis, and treatment data, combined with big data, can effectively support the scientific calculation of

medical insurance payment methods such as single disease type and group according to disease diagnosis and reasonably divide the payment grade of various diseases. It can support the adjustment of medical insurance drugs, diagnosis and treatment items and transfer drugs with large demand and good efficacy into the catalog.

(3) At the same time, medical insurance data has a good supporting effect on other social security businesses and people's livelihood protection in larger fields. For example, for the qualification certification of social security benefits, the health status of the elderly can be learned through medical insurance medical data, and those who had recent medical records after data comparison can no longer carry out one-to-one certification with the elderly. For cases where had often sought medical treatment but have not sought medical treatment for a long time, it can be carried out one-on-one certification work. This way, combined with data resources, the management will be embedded in the service and the convenience of the masses to the greatest extent. Another example is to provide public health management service support for individuals through health portraits or scoring and provide individuals with convenient inquiries and services such as the nearest designated medical institutions, designated pharmacies, and physicians and hospital classification ranking queries. Another example is to provide relevant verification services for commercial insurance underwriting and claims settlement, provide personalized commercial insurance product customization services for different insured groups, support the establishment of a multi-level social security system, support product research of pharmaceutical and medical device manufacturers, and allocate social resources according to market demand.

6.2 Contributions

6.2.1 Academic contribution

First, according to the conclusions and characteristics of big data mining applications and referring to the advanced risk management concepts of medical funds and medical insurance industries, the medical insurance fraud risk identification model is designed, and the medical insurance cost control model is used to test and verify the data application.

Second, starting from the theory of quasi-public goods and medical insurance-related theory, the progressive mode of medical insurance fraud identification based on big data mining, association rules, and entity portraits is analyzed. This study applies medical insurance big data mining to the practical application and analysis of medical fraud to realize medical insurance

cost control. This process collected a batch of real-world data, confirming that the three analysis algorithms had good practical results in medical insurance cost control management.

Third, the model proposed in this study is verified using big data samples. In this study, a large number of data samples were collected for three years, which have regional comprehensiveness and representative population characteristics, and have theoretical solid research value. The desensitized data can be made available to researchers who need it, facilitating related research efforts. This is one of the main contributions of this research process to this topic and related research.

6.2.2 Management contribution

The problem of fraud and supervision, prevention, and control of medical insurance funds is not a new topic but a common social problem that needs to be gradually improved with new technology, new management theory, and new regulatory means in the process of social development. This study achieves good results in analysis and monitoring from big data mining technology and entity portrait in-depth crowd tagging method. Through the research and analysis of medical insurance fraud, this study has made the following three outstanding contributions:

First, this study constructs a medical insurance fraud management flow chart to manage medical insurance funds to prevent illegal and fraudulent actions, reduce medical insurance fund expenditure, improve the reasonable and compliant use of medical insurance funds, and reduce the economic burden of patients.

Second, this study introduces various new technologies, including the application practice of physical portrait technology, which can manage the rational use of medical insurance funds, control the payment of medical insurance fund expenses, provide cost control management methods for medical insurance fund expenditure, and management units, and provide a reference for the use of medical insurance funds and units that formulate policies.

Third, the conclusions obtained in this study can be directly applied to the identification and management of medical insurance fraud problems in the sample collection area. The research conclusions can also be applied to other places, using the same process and inserting a three-fold algorithm to solve similar problems to reduce the control of intentional and unintentional fraud in the medical insurance process.

6.3 Solutions and advice

The medical security fund is the people's life-saving money, which has the nature of special funds, and the fraud and deception that occurs by treating the medical security fund as Tang monk meat will ultimately damage the vital interests of every insured person. In the long run, it is necessary to change the governance concept in the medical insurance field, use healthcare big data to identify appropriate and necessary medical behaviors, identify fraudulent behaviors, reduce insurance fraud, and watch for life-saving money. It is necessary to use information technology to fully use the insight of big data monitoring to achieve intelligent supervision and precise crackdowns. Use big data to capture, analyze, and screen for doubts. Using big data monitoring to drive targeted regulation can help keep insurance fraud in the dark.

6.3.1 Focus on the factors of health insurance risk and identify health insurance fraud

First, it is caused by the characteristics of China's medical insurance system. China's existing medical insurance system is gradually transformed from the previous public-funded medical care. The opacity of pharmaceutical consumption makes it difficult for medical regulators to accurately define the value of pharmaceutical consumption and the rationality of consumption content, which causes great difficulties for the standard and accurate payment of medical expenses by the pooled fund.

Second, under the current situation, the financial investment obtained by each designated medical institution remains unchanged or even reduced. Its economic benefits mainly rely on its ability to generate income, which inevitably makes medical institutions take the acquisition of economic benefits as an extremely important pursuit goal. Even some medical institutions do not hesitate to take various means in order to obtain maximum economic benefits. The Harbin sky-high medical fee incident that made a lot of noise some time ago is the best example.

Third, with the strengthening of medical and health conditions, the problem of population aging is becoming more and more prominent, and the medical consumption needs of this part of the population are also increasing, which is bound to cause a significant increase in medical costs.

6.3.2 Establish and improve the medical insurance information platform

At the beginning of the system construction, due to insufficient demand analysis, the height of the station is not enough, and the software cannot adapt to the changing requirements of the

medical insurance business, resulting in many systems falling into the embarrassing situation of not dying or living. In some places, design and planning are only carried out from the perspective of medical insurance, and the integration requirements of social insurance business development are not fully considered. System fragmentation increases operating costs, and duplicate construction causes huge waste. There are also some systems, due to various reasons, the application effect is not ideal, and long-term shelving is not used, resulting in insufficient development of the system.

China's basic medical security system is diverse, but the system design and security provision has problems such as multiple management, multiple government departments, inconsistent policies, and other practical problems, so their respective information is not being shared, resulting in a waste of resources. Public services are inefficient. The overall level of basic medical insurance protection needs to be further improved, and the overall settlement is carried out on a regional and municipal basis. The level of basic medical security is unbalanced between urban and rural areas and between regions. Urban medical insurance treatment is significantly higher than that in rural areas, and in the same area, there are also significant differences in the level of medical security for different groups of people.

As urban employees change their work units or move to work in other provinces and cities, their social insurance and medical insurance will naturally follow the transfer, and their contributions will be deducted proportionally from their wages. Currently, the central government adopts follow-up financial subsidies, and the mobility of migrant workers has also increased the difficulty of fundraising. Therefore, the lack of information platforms has become the biggest obstacle to long-distance reimbursement and transfer continuation.

Through the Internet Plus initiative and the construction of a public data service platform, realize the informatization project of big data, accelerate the establishment of a national basic medical insurance information sharing platform, realize nationwide information sharing on population fertility and health, disease diagnosis and treatment, chronic disease rehabilitation and management, establish a new support system for promoting medical reform, reform medical insurance payment methods, and accelerate the promotion of national networking of basic medical insurance and settlement of medical treatment in different places. The platform is jointly built and shared, forming a big data and large platform, providing objective data for government departments to promote the integration of three guarantees, provide medical insurance handling units with residents' insurance information, reasonably and effectively control costs, avoid duplicate examinations and medications, and provide support for solving medical reimbursements in different places; provide assistance to public security departments

in grasping information on population birth, death, and investigation of injuries involved, and thoroughly realize data sharing and three-doctor linkage.

To do an excellent job in medical insurance, the construction of information systems is the key. Under the spirit of the *Guiding Opinions on Carrying out Domestic Medical Insurance Work* issued by six ministries and commissions, medical insurance should strengthen the connection with urban residents' medical insurance and new rural cooperative services and provide one-stop instant settlement services. To this end, the Insurance Regulatory Bureau carefully investigates, actively plans, and guides insurance institutions to continuously strengthen the construction of medical insurance information systems. The one-stop instant settlement service of residents' serious illness medical insurance has taken a key step, which will further facilitate the reimbursement of people's medical expenses. Medical insurance service windows have been set up in designated medical institutions in the region, staff has been stationed, and information collection, settlement and payment, statistical analysis, and other work have been carried out, which has greatly strengthened the effective connection between basic medical care and severe illness medical insurance.

The top-level design realizes the national unification of basic medical insurance, reflecting fairness, and commercial insurance and social assistance are adapted to local conditions and reflect characteristics. The framework of China's medical security system includes three levels: basic medical security, commercial health insurance, and medical social assistance, making it clear that basic medical security is the main body and is sponsored by the government; Commercial insurance and social assistance are supplements, and commercial insurance is sponsored by enterprises and reflects the market attributes. Since the government sponsors basic medical insurance, this requires that it should be unified throughout the country, and citizens should enjoy fair state benefits. Only by accelerating the construction of the basic medical security system and realizing the integration of the three security systems of urban workers' medical insurance, urban residents' medical insurance, and new rural cooperative medical care can we lay the foundation for the sharing and joint construction of various types of information. The state initiates project investment, led by the health department, to realize the co-construction and sharing of health examination, diagnosis and treatment information, and basic medical security data, clearly establish and improve the data sharing and co-management mechanism between social security departments, medical institutions and health authorities, and realize data information sharing, solve referral channels, realize hierarchical diagnosis and treatment, improve medical treatment links, reform medical insurance payment methods, accelerate the promotion of national networking of basic medical insurance and settlement of

medical treatment in different places, and solve problems such as difficulty in reimbursement in different places. Achieve the goal of a healthy and well-off China for all.

6.3.3 Complete medical insurance deductible and reimbursement ratio settings

Under the premise that the total amount of medical insurance funds is certain, if the policy threshold is low, the number of people who benefit is relatively large, and the people with high expenses will receive relatively little compensation. On the contrary, if the policy threshold is high, theoretically, the people with high fees will receive more compensation, but the number of people who enter the compensation will also be reduced. The study believes that Japan's measures against high costs can be used as a reference (Wang, 2022), and different policy thresholds can be set according to the income and age of the insured, especially taking into account vulnerable groups such as children and the elderly. The results of multiple regression in this study show that the older the age, the lower the reimbursement growth rate and China has entered an aging society. More and more elderly groups will face a high burden of medical expenses, so it is necessary to set specific policy thresholds for the elderly to reduce their economic burden.

Secondly, the new rural cooperative medical insurance policy is still in the trial stage, the financing level, operation, and guarantee capacity are still weak, and patients hope to get more compensation through the medical insurance policy to reduce the economic burden of family diseases. At present, due to the policy's limitations, there may be a low compensation ratio due to high medical expenses outside the scope of the policy, so the medical insurance policy can not meet the needs of a small number of patients. Implement hierarchical diagnosis and treatment among medical institutions and differentiated payment policies for medical insurance. The outpatient reimbursement ratio of employee medical insurance in primary medical institutions is higher than that of high-level medical institutions, and the minimum payment standard for hospitalization is lower than that of high-level medical institutions. Adjust the threshold for outpatient reimbursement, and if the outpatient expenses in the current year have not reached the threshold, the threshold fee will be reduced in the next year. That is if employees and insured persons of resident medical insurance have continuously participated in the insurance and paid premiums or enjoyed medical insurance treatment for one year (natural year), and the outpatient medical expenses incurred within the policy scope in the current year do not exceed the minimum payment standard, the minimum payment standard will be reduced by 100 yuan based on the prescribed standard when the following year is treated for outpatient treatment; If the minimum payment standard has not been exceeded for two consecutive years,

the starting payment standard for the following year will be reduced by 200 yuan; If the minimum payment standard has not been exceeded for three consecutive years or more, the starting payment standard will be reduced by 300 yuan for the following year. The payment threshold is calculated continuously for outpatient, outpatient-specific diseases, inpatient referrals, and transfers. The scope of drug reimbursement for primary medical institutions has been extended to the basic medical insurance drug list specifications. Doctors in primary medical institutions issue long-term prescriptions, and basic medical insurance pays for them. Outpatient reimbursement was extended to secondary hospitals. The general outpatient medical expenses incurred by insured urban and rural residents seeking medical treatment in secondary hospitals that have implemented the reform of public hospitals are included in the scope of reimbursement by resident medical insurance, with a minimum payment standard of 500 yuan and a maximum payment standard of 3,000 yuan within the year, with a reimbursement rate of 50%.

In addition, due to the lack of policy publicity, patients' expectations of the reimbursement ratio are too high, and there is dissatisfaction with the policy. It is recommended to increase the financing level of the New Agricultural Cooperation Fund, appropriately expand the scope of protection based on not exceeding the funds' ability to pay, reasonably increase the compensation ratio of medical insurance, and increase the compensation level. Through reforming medical insurance payment methods, we will strive to build a joint negotiation and negotiation mechanism between hospitals and medical insurance.

6.3.4 Reform compensation methods, explore and establish a prepayment system

According to the 2015 serious illness medical insurance compensation policy, if the cumulative out-of-pocket eligible inpatient medical expenses reach 5,000 yuan, you can obtain serious illness medical insurance, which is lower than farmers' annual per capita net income. However, relevant scholars have disputed this standard.

Some scholars believe that because of the uneven distribution of income, the per capita net income as the standard means that a large proportion of patients will face economic risks of disease. They suggest further increase the actual compensation ratio of serious illness medical insurance. At the same time, the development of new technologies and the economic evaluation of drugs and materials should be carried out to reflect the real needs of patients as a guide to carry out a cost-benefit analysis of drugs and materials; so as to provide a theoretical basis for doctors to weigh the basic needs and special needs of disease treatment, so that medical insurance funds can be used on the cutting edge.

6.3.5 Governments at all levels will enhance new agricultural cooperatives financing

The new agricultural cooperative subsidies should be mainly based on the state, and the financing of new agricultural cooperation in the eastern developed provinces should be mainly based on local financial subsidies. While ensuring rapid economic growth, local governments should continuously increase subsidies for new agricultural cooperatives, improve the utilization rate of funds, and actively explore new agricultural cooperatives systems that adapt to local development to expand the level of protection for new cooperatives. In addition, under the premise of ensuring the safety of the operation of the New Agricultural Cooperation Fund, ensure a high utilization rate of the fund or carry out recompensation to avoid a large amount of waste of the fund.

6.3.6 Intensify the publicity and strengthen the direction of public opinion

It is essential to increase the publicity, raise the awareness and popularization level, promote the work of residents of serious illness medical insurance for urban residents, and improve the residents' serious illness medical insurance system. It is necessary to make more insured objects receive timely assistance when suffering from serious diseases, enjoy the fruits of policy development in a timely and effective manner, and promote harmonious social and economic development.

6.3.7 Innovate the approach and improve service quality

Combining urban residents' medical and basic medical insurance is suggested for reimbursement. When the insured person receives the basic medical reimbursement expenses, he also receives the serious illness reimbursement expenses, and then the commercial insurance and the social security department settle the settlement. At the same time, the online direct settlement business of designated medical institutions in the city will be opened as soon as possible to shorten the reimbursement cycle of inpatient expenses and facilitate insured patients.

6.3.8 Innovate ways to improve the effectiveness of rescue

First, continue to do a good job in sharing information and information on medical reimbursement for serious illnesses of urban residents. By providing study copies and system resource sharing, the insurance company will be provided with accurate and timely information about qualified rescuers to ensure that sick residents can enjoy serious illness assistance the first time and gradually alleviate the problem of difficulty and high cost of medical treatment for

urban residents in difficulty. The second is to inform promptly. At the same time that residents provide the original hospitalization receipts, inform the relevant policies on the secondary reimbursement of serious illnesses to ensure that the hospitalized residents are aware of the rescue information the first time. The two working methods can not only facilitate the rapid reimbursement of huge medical expenses by hospitalized residents but also simplify the procedures and improve efficiency.

6.4 Shortcomings and limitations

First, the data variables of the health insurance system are limited. The more complete the variable information collected, the results of this study will be more representative, making the inpatient cost control strategy more effective for insured patients.

Second, the data in this study came directly from the medical insurance database system, and some important variables that may affect hospitalization costs were not included in this study due to the content of the database information system. For example, information on the patient's education, financial status, admission, and discharge outcome, was not collected from the cases.

Third, this study only discusses the impact of index variables such as patients' social and clinical characteristics on cost. The lack of information on factors such as hospital care (such as hospital infection) and disease severity and outcome are often reported as important factors affecting inpatient medical costs, so considering them as potential variables to the model needs further exploration.

Fourth, the data source has certain limitations; medical insurance data is limited by hospital upload, many fields are missing, which has a certain impact on the modeling of risk control models, and the accuracy of results needs to be further improved.

[This page is deliberately left blank.]

Bibliography

Albanese, E., Liu, Z., Acosta, D., Guerra, M., Huang, Y., Jacob, K. S., Jimenez-Velazquez, I. Z., Llibre Rodriguez, J. J., Salas, A., Sosa, A. L., Uwakwe, R., Williams, J. D., Borges, G., Jotheeswaran, A. T., Klibanski, M. G., McCrone, P., Ferri, C. P., & Prince, M. J. (2011). Equity in the delivery of community healthcare to older people: Findings from 10/66 Dementia Research Group cross-sectional surveys in Latin America, China, India and Nigeria. *Bmc Health Services Research, 11*(1), 153.

Arrow, K. J. (2001). Uncertainty and the welfare economics of medical care. 1963. *J Health Polit Policy Law, 26*(5), 851-883.

Ataguba, J. E., & McIntyre, D. (2012). Paying for and receiving benefits from health services in South Africa: Is the health system equitable? *Health Policy Plan, 27 Suppl 1*, i35-i45.

Banderier, S. C. (2000, July 31st). *Continued fractions, comparison algorithms and fine structure constants*. Canadian Mathematical Society, Limoges, Région Limousin, France.

Baumol, W., & Ordover, J. A. (1988). Antitrust policy and high-technology industries. *Oxford Review of Economic Policy, 4*(4), 13-34.

Blank, F. (2020). The state of the German social insurance state: Reform and resilience. *Social Policy & Administration, 54*(3), 505-524.

Blesch, G. (2008). Halting medical fraud. New coalition aims to stop various insurance scams. *Mod Healthc, 38*(26), 12.

Bloom, S. W. (2002). Professionalism in the practice of medicine. *Mount Sinai Journal of Medicine A Journal of Translational and Personalized Medicine, 69*(6), 398-403.

Bommier, A., & Stecklov, G. (2002). Defining health inequality: Why Rawls succeeds where social welfare theory fails. *Journal of Health Economics, 21*(3), 497-513.

Brockett, W., R., & Liberzon, D. (2000). Quantized feedback stabilization of linear systems. *Ieee Transactions On Automatic Control, 45*(7), 1279-1289.

Bundes Ministerium Für Familie, S. F. U. J. (2013). Politischer Bericht zur Gesamtevaluation der ehe- und familienbezogenen Leistungen. *Fachlexikon der sozialen Arbeit*(3 Auflage), 939.

Caudill, S. B., Ayuso, M., & Guillén, M. (2005). Fraud detection using a multinomial logit model with missing information. *Journal of Risk and Insurance, 72*(4), 539-550.

Chen, Z., Xu, Y., Tian, Z., & Meng, X. (2022). The Impact of Social Insurance Law on Corporate Innovation: Evidence from a quasi-natural experiment. *Economic Modelling, C*(111), 5-10.

Cleeton, G. C., & Knight, F. B. (1924). Validity of Character Judgments Based on External Criteria. *Journal of Applied Psychology, 8*(2), 215-231.

Coase, R. H. (1992). The problem of social cost. *Systems Research, 9*(1), 79-81.

Coleman, M. P. (2014). Cancer survival: Global surveillance will stimulate health policy and improve equity. *Lancet, 383*(9916), 564-573.

Cooter, R., & Ulen, T. (2008). *Law & economics*. Pearson/Addison Wesley.

Davies, J. (2017). "It's a really grey area": An exploratory case study into the impact of the Jackson Reforms on organised insurance fraud. *International Journal of Law, Crime and Justice, 51*, 45-57.

Dawson, W. H. (1912). Social Insurance in Germany, 1883-1911. *The Economic Journal, 22*(87), 457-460.

Dionne, G. (2000). The empirical measure of information problems with emphasis on insurance

fraud. In G. Dionne (Ed.), *Handbook of insurance* (pp. 395-419). Springer Netherlands.

Einav, L., & Finkelstein, A. (2018). Moral hazard in health insurance: What we know and how we know it. *Journal of the European Economic Association*, *16*(4), 957-982.

Evans, R. G. (1974). Supplier-Induced demand: Some empirical evidence and implications. In M. Perlman (Ed.), *The economics of health and medical care: proceedings of a conference held by the International Economic Association at Tokyo* (pp. 162-173). Palgrave Macmillan.

Feldstein, M. (1970). The rising price of physicians' services, *52*(2), 60-88.

Fernandes Antunes, A., Jacobs, B., de Groot, R., Thin, K., Hanvoravongchai, P., & Flessa, S. (2018). Equality in financial access to healthcare in Cambodia from 2004 to 2014. *Health Policy and Planning*, *33*(8), 906-919.

Fuchs, V. R. (1978). The supply of surgeons and the demand for operations. *Journal of Human Resources*, *13 Suppl*, 35-56.

Gholamreza, Haddad, Keshavarz, & Amirkhanloo, M. (2011). Test of asymmetric information in the automobile insurance industry of IRAN. *The Journal of Economic Studies and Policies*, *0*(20), 130-158.

Glied, S., & Stabile, M. (2001). Avoiding health insurance crowd-out: Evidence from the medicare as secondary payer legislation. *Journal of Health Economics*, *20*(2), 239-260.

Greenger, T., Schimke, R., Ackermann, K., & Su, J. (2013). 德国医疗保险体系的渐进式制度变迁——渐行渐远的“俾斯麦模式” [A long farewell to the Bismark system: Incremental change in the German health insurance system]. *Jianghai Academic Journal*, (05), 31-36.

Grobler, L., Marais, B. J., & Mabunda, S. (2015). Interventions for increasing the proportion of health professionals practising in rural and other underserved areas. *Cochrane Database Syst Rev*, *2015*(6), D5314.

Hajizadeh, M., & Nghiem, H. S. (2011). Out-of-pocket expenditures for hospital care in Iran: Who is at risk of incurring catastrophic payments? *Int J Health Care Finance Econ*, *11*(4), 267-285.

Hall, G., Hong, G., & Terrell, T. (1996, Oct 18th). *Prediction by back-propagation neural network for lossless image compression*. Proceedings of Third International Conference on Signal Processing (ICSP'96).

Hong, Z., Wang, H., & Chen, W. (2000). Simultaneously applying multiple mutation operators in genetic algorithms. *Journal of Heuristics*, *6*(4), 439-455.

Jehu-Appiah, C., Aryeetey, G., Spaan, E., de Hoop, T., Agyepong, I., & Baltussen, R. (2011). Equity aspects of the National Health Insurance Scheme in Ghana: Who is enrolling, who is not and why? *Social Science & Medicine*, *72*(2), 157-165.

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Improving fraud and abuse detection in general physician claims: A data mining study. *Int J Health Policy Manag*, *5*(3), 165-172.

Kim, M. J., Kam, S., & Lee, W. K. (2017). Inequality status of household medical expenditure distribution by household income quartile and associated factors with household catastrophic medical expenditure in korea. *J Health Info Stat*, *42*(3), 232-240.

Komisar, J. B., & Buchanan, J. M. (1969). Buchanan, James M., The Demand and Supply of Public Goods, Chicago, Rand McNally and Company, 1968, ix + 214 pp. ($5.00). *American Journal of Agricultural Economics*, *51*(1), 213.

Kou, Z., Rey, P., & Wang, T. (2013). Non-Obviousness and screening. *The Journal of Industrial Economics*, *61*(3), 700-732.

Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the craft of qualitative research interviewing, 2nd ed.* Sage Publications, Inc.

Lagomarsino, G., Garabrant, A., Adyas, A., Muga, R., & Otoo, N. (2012). Moving towards

universal health coverage: Health insurance reforms in nine developing countries in Africa and Asia. *Lancet*, *380*(9845), 933-943.

Liou, Shwu-Ru, Grobe, & J, S. (2008). Perception of practice environment, organizational commitment, and intention to leave among asian nurses working in u.s. Hospitals. *Journal for Nurses in Staff Development (JNSD)*, *24*, 276-282.

Manchester, W. (1974). *光荣与梦想:1932-1972年美国社会实录* [The Glory and the Dream: A Narrative History of America, 1932-1972]. Rosetta Books LLC.

Meade, J. E. (1952). A geometry of international trade. *Acta Paediatrica*.

Menon, N. M. (2018). Information spillover and semi-collaborative networks in insurer fraud detection. *MIS Q.*, *42*(2), 407-426.

Mills, A., Ataguba, J. E., Akazili, J., Borghi, J., Garshong, B., Makawia, S., Mtei, G., Harris, B., Macha, J., Meheus, F., & McIntyre, D. (2012). Equity in financing and use of health care in Ghana, South Africa, and Tanzania: Implications for paths to universal coverage. *The Lancet*, *380*(9837), 126-133.

Munge, K., & Briggs, A. H. (2014). The progressivity of health-care financing in Kenya. *Health Policy and Planning*, *29*(7), 912-920.

Ortega, P., Figueroa, C., & Ruz, G. (2006, June 26-29). *A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile*. Proceedings of the 2006 International Conference on Data Mining, Las Vegas, Nevada, USA.

Pauly, M. V. (1986). Taxation, health insurance, and market failure in the medical economy. *Journal of Economic Literature*, *24*(2), 629-675.

Roll, J. M., Kennedy, J., Tran, M., & Howell, D. (2013). Disparities in unmet need for mental health services in the United States, 1997-2010. *Psychiatr Serv*, *64*(1), 80-82.

Rudman, W. J., Eberhardt, J. R., Pierce, W., & Hart-Hester, S. (2009). Healthcare fraud and abuse. *Perspect Health Inf Manag*, *6*(Fall), 1g.

Sokol, L., B., G., M., W., J., R., & K., J. (2001, January 3-6). *Precursory steps to mining HCFA health care claims*. Proceedings of the 34th Annual Hawaii International Conference on System Sciences, Hawaii.

Stefan, G., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talon, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420-3435.

Stiglitz, J. E. (1988). *Economics of the public sector*. W.W. Norton.

Vera-Hernandez, M. (2003). [Evaluating health interventions without experiments]. *Gaceta Sanitaria*, *17*(3), 238-248.

Viaene, S., Derrig, R. A., & Dedene, G. (2004). Cost-sensitive learning and decision making for massachusetts pip claim fraud data. *International Journal of Intelligent Systems*, *19*(12), 1197-1215.

Wagstaff, A., Lindelow, M., Jun, G., Ling, X., & Juncheng, Q. (2009). Extending health insurance to the rural population: An impact evaluation of China's new cooperative medical scheme. *Journal of Health Economics*, *28*(1), 1-19.

Wagstaff, A., van Doorslaer, E., van der Burg, H., Calonge, S., Christiansen, T., Citoni, G., Gerdtham, U., Gerfin, M., Gross, L., Häkinnen, U., Johnson, P., John, J., Klavus, J., Lachaud, C., Lauritsen, J., Leu, R., Nolan, B., Perán, E., Pereira, J.,... Winkelhake, O. (1999). Equity in the finance of health care: Some further international comparisons. *Journal of Health Economics*, *18*(3), 263-290.

Wan, L., Wu, B., Du, N., Ye, Q., & Chen, P. (2006, August 14-16). *A new algorithm for enumerating all maximal cliques in complex network*. International Conference on Advanced Data Mining and Applications, Xi'an, China.

Watt, Moreno, I., Vázquez, F. J., & Richard. (2006). Can Bonus-Malus allieviate insurance fraud? *The Journal of Risk and Insurance*, *73*(1), 123-151.

Yip, W. (2010). Impacts of social and demographic factors on residents' subjective well-being in rural china. *Chinese Journal of Population Resources and Environment (English Version), 8*(1), 7.

An, K. (2020). *Research on the investigation of social medical insurance fraud crimes*. [Master's thesis]. Chinese People's Public Security University.

Bao, L. (2010). Teradata 洞悉地理空间数据零售, 汽车保险行业身先士卒 [Teradata reads the geospatial data retailing and lead the auto insurance industry]. *Telecom Market*, (4), 12.

Pigou. (1920). *福利经济学* [Welfare Economics]. Huaxia Publishing House.

Cao, P., & Gu, C. (2012). 我国城镇居民基本医疗保险制度存在的问题及对策 [Problems and countermeasures of the basic medical insurance system for urban residents in China]. *Reform & Opening*, (04), 140-142.

Cui, F. (2016). 基本医保待遇支付与筹资平衡实证分析——基于广元市的实践 [Empirical analysis on the balance between basic social insurance payments and financing: Based on the practice of Guangyuan City]. *China Medical Insurance*, (05), 45-48.

Deng, C. (2012). *Research on social insurance anti-fraud legal system in China*. [Master's thesis]. Liao Ning University.

Deng, W. (2014). 整合各种力量构建多层次医疗保障体系 [On the consolidation of various forces oto constuct a multi-level medical security system]. *Journal of Social Science of Hunan Normal University*, *43*(5), 5.

Di, X. (2021). *Medical insurance fraud identification system based on isolated and random forests algorithm*. [Master's thesis]. Nanjing University of Posts and Telecommunications.

Du, Q., Liu, H., & Hu, Q. (2020). 社会医疗保险基金欺诈行为的扎根理论研究——基于58个医保欺诈刑事案件分析 [The grounded theory research of social medical insurance fund fraud: Based on the analysis of 58 criminal cases]. *Legal and Economics*, (08), 70-71.

Du, Y., & Wen, J. (2010). 从福利为本到资产为本:社会政策发展的新趋向 [From welfare-based to asset-based: New trend in social policy development]. *Hebei Academic Journal*, *30*(04), 120-123.

Fei, P. (2021). 《社会保险法》实施十周年经办服务发展情况 [Administration Development for the tenth anniversary of the implementation of the Social Insurance Law]. *China Labor*, (05), 77-84.

Feng, X. (2022). 社会研究:方法、能力与关键 [Social Research: Methods, Competencies and Keys]. *Journal of China Women's University*, *34*(05), 5-12.

Fu, M., Chen, D., & Xiao, M. (2019). 关于构建我国基本医疗保险参保人诚信体系的思考 [Reflections on the construction of a credit system for basic medical insurance participants in China]. *Laboratory Medicine and Clinic*, *16*(06), 857-860.

Gao, J. (2007). *Improve the government supply mechanism of social medical insurance in Chinese western countryside*. [Doctoral dissertation]. Northwest University.

Gao, N. (2010). 浅议中国城镇职工基本医疗保险制度的建立与发展 [Brief discussion on the establishment and development of the basic medical insurance system for urban workers in China]. *Economic Research Guide*, (10), 63-65.

Gao, X. (2006). 浅论我国城镇职工基本医疗保险制度的运行与完善 [Brief discussion on the operation and perfection of the basic medical insurance system for urban workers in China]. *Journal of West Anhui University*, (06), 61-64.

Ge, T., & Wu, D. (2013). 宪政视阈下山地型新农村建设中养老保险政策执行的困扰与对策——以q区a镇为例 [The difficulties and solutions of the implementation of pension insurance policy in the construction of new mountainous countryside from the constitutional perspective: A case study on A county of Q district]. *Journal of Shandong Agricultural University (Social Science Edition), 15*(04), 6-11.

Gu, X. (2017). *Research on the improvement of anti-fraud mechanism of social insurance in Shanghai*. [Master's thesis]. Shanghai Normal University.

He, P. (2017). 德国社会保险的探索——以长期照护保险制度为例 [Exploration of social insurance in Germany: A case study of the long-term care insurance system]. *Theory Monthly*, (09), 177-183.

Huang, D., Qin, J., & Mao, Y. (2017). 发挥医保数据监测和警示作用的创新之路——基于成都市医保大数据挖掘的实践 [Seeking an innovative road of playing the role of monitoring and warning in healthcare insurance data- based on the practice of big data mining of medical insurance in Chengdu]. *China Health Insurance*, (05), 38-41.

Huang, H. (2020). *Research on medical insurance anomaly based on improved fuzzy clustering algorithm*. [Master's thesis]. Beijing Jiaotong University.

Huang, K., & Liu, H. (2017). 中华人民共和国成立以来中国历史人口数据的比对分析——基于《世界人口展望》2017年修订版的深度解读 [Comparative analysis of the historical population data since the founding of the People's Republic of China: Based on the interpretation 2017 revision of the Prospect of World Population]. *Population and Family Planning*, (08), 15-18.

Huang, Y., & Zhang, Q. (2016). 福祉经济学 [Welfare Economics]. *Dongyue Tribune*, *37*(01), 5-14.

Huang, Z. (2020). *Research on medical insurance supervision of violations by designated medical institutions in Y City*. [Master's thesis]. Ji'nan University.

Huang, Z. (2016). 浅谈法律监管下医保基金反损害 [Brief analysis on the countermeasures on damage of healthcare insurance fund under the law]. *Legal and Economy*, (10), 116-118.

Kuang, C., & Chai, Y. (2000). 我国城镇职工基本医疗保险制度推行过程中的问题及法律对策 [Problems and legal measures in the implementation of the basic medical insurance system for urban workers in China]. *China Health Economics*, (08), 20-21.

Jiang, X. (2018). 关于加强企业社会保险管理工作的思考 [Reflections on strengthening the management of social insurance in enterprises Translation]. *China Economist*, (02), 292-294.

Li, C. (2015). *Theoretical analysis and empirical study of health insurance fraud in China*. [Doctoral dissertation]. Qing Dao University.

Li, F. (2013). *Research on the reform on basic medical insurance system for urban employee*. [Master's thesis]. Inner Mongolia University.

Li, J., & Xu, L. (2007). 我国城镇居民医疗补偿机制的完善 [The perfection of medical compensation mechanism for urban residents in China]. *Jiangxi Social Sciences*, (12), 164-168.

Li, L., & Li, L. (2011). 社会医疗保险欺诈行为实证研究——基于成本—收益的视角 [Empirical research on social health insurance fraud—based on cost-benefit perspective]. *Hunan Normal University Journal*, *40*(01), 56-59.

Li, X. (2016). *Research on the classification ensemble algorithem for medical insurance anomaly detection*. [Master's thesis]. University of Electric Science and Technology of China.

Li, Z. (2002). 医保、医院和患者之间的关系及处理初探 [A preliminary study of the relationship between patients, hospitals and medical insurance]. *Journal of Shandong Medical College*, (05), 368-370.

Lin, Y. (2015). 基于bp神经网络的新农合欺诈识别实证研究——以定点医疗机构欺诈滥用为中心 [An empirical study on fraud identification of new rural cooperative medical insurance based on BP neural network: Focusing on fraud abuse in designated medical institutions]. *Journal of Yunnan Normal University*, *47*(03), 117-128.

Liu, C., & Zhu, X. (2018). 基于bp神经网络的医保欺诈识别 [Medical insurance fraud detection based on BP neural networl]. *Computer system application*, *27*(06), 34-39.

Liu, H., He, L., & Li, J. (2017). 深度学习进展及其在图像处理领域的应用 [Progress of deep learning and its application in image processing]. *Zte Communications Journal*, *23*(04), 36-40.

Liu, J. (2019). 中国社会医疗保险制度40年的历史经验、结构困境与改革方向 [Structural dilemma and reform direction of China's social medical insurance system in the past 40 years]. *The Journal of Humanities*, (03), 20-29.

Liu, L. (2019). *Research on healthcare fraud identification model and application based on big data mining*. [Master's thesis]. Hebei University of Technology.

Liu, S. (1995). 关于社会安全立法的比较研究 [Comparative study of social security legislation]. *Peking University Law Journal*, (03), 44-47.

Liu, W. (2004). 萨缪尔森的效率与公平观探析 [Analysis of Samuelson's thoughts on equality and efficiency]. *Research on Economics and Management*, (06), 65-68.

Liu, X. (2008). 社会医疗保险基金不同支付模式下的保险欺诈行为分析 [Analysis of insurance fraud under different payment models of social medical insurance funds]. *Population & Economics*, (02), 64-70.

Liu, Y. (2009). *Study on current Chinese new rural cooperative medical system development*. [Doctoral dissertation]. Shandong University.

Liu, Y. (2016). *Research on regional inequanlity of the new rural cooperative medical care system based on two-stage nested Theil decomposition method*. [Master's thesis]. Shandong University.

Liu, Y. (2017). 基于logistic模型的保险欺诈分析 [Analysis on the insurance fraud based on the logistic model Financial Economy]. *Finance & Economy*, (14), 140-141.

Liu, Y. (2020). *Analysis and research on abnormal deep learning algorithm of medical insurance fraud data*. [Master's thesis]. Chengdu University Of Technology.

Liu, Q. (2013). 初探养老金缺口的弥补 [Preliminary Exploration on the pension gap filling]. *Modern Enterprise Culture*, (06), 92-93.

Lu, P. (2010). *Data module design and implementation of Medicare fund prevention and control platform*. [Master's thesis]. Donghua University.

Luo, K., & Dai, M. (2016). 数据挖掘技术在医疗大数据中的应用研究 [A study on application of data mining technology in medical field]. *China Computer and Communication*, (06), 45-47.

Luo, Z. (2016). 社会医疗保险诈骗及其刑法规制探讨 [Discussion on social medical insurance fraud and its criminal law system]. *Medicine and Society*, *29*(12), 67-69.

Ma, J. (2011). *Research on improving basic medical insurance system for urban resident of Shanghai*. [Master's thesis]. Shanghai University of Engineering Science.

Ma, X., Sun, R., & Song, X. (2018). 浅析医疗大数据对医学研究领域的创新进展 [Analysis of the innovation progress of medical big data in the field of medical research]. *Inner Mongolia Science Technology& Economy*, (07), 93-96.

Mao, Z. (2003). 新型合作医疗的特征及其涵义 [Characteristics and Definition of the new rural cooperative health care system]. *Health Economics*, (8), 2.

Mao, Z., & Jiang, J. (2005). 新型农村合作医疗制度的特征及目前面临的挑战 [Characteristics and challenges of the new rural cooperative health care system]. *China Health Economics*, *24*(1), 2.

Mei, L. (2016). “聪明监管”:基本医疗保险监管的模式和路径选择 [Smart supervision: Pattern and Path Selection of the Supervision on the Basic Medical Insurance]. *Journal of Hubei Administration Institute*, *35*(06), 13-18.

Miao, X. (2015). *A study of fairness and efficiency of Samuelson's thoughts*. [Master's thesis]. Nanjing Normal University.

Pei, G. (2011). 反欺诈监管已成为各国保险监管的重点内容 [Anti-fraud supervision has become the priority in insurance supervision in countries]. *Financial Times*, *24*(8), 101.

Qu, G. (2014). *Research on management problems of utilization and supervision of medical insurance fund: Taking management of medical insurance fund in Dalian as example*. [Doctoral dissertation]. Northeastern University.

Ran, M. (2009). 完善城镇居民基本医疗保险制度的研究——基于试点城市实证调研的比较分析 [Research on Improving the Basic Medical Insurance System for Urban Residents: A Comparative Analysis Based on Empirical Research in Pilot Cities]. *Journal of China University of Labor Relations*, *23*(05), 120-122.

Ren, H. (2012). 城镇职工基本医疗保险制度中的道德风险及控制对策研究 [Research on moral hazard and control measures in the basic medical insurance system for urban workers]. *Modern Marketing (Academic Edition)*, (12), 128-129.

Ren, Y., He, X., Lei, J., Feng, Y., & Pei, Y. (2018). 基于医保大数据的我国急诊医保患者急（抢）救药品使用现状 [Present situation about the medical resources used for emergency patients with medical insurance in China based on analysis of large numbers of data]. *27*(8), 949-954.

Shen, C. (2021). 参保人医保欺诈的识别——基于国家医保局公开曝光案例的分析 [Identification of medical insurance fraud by participants: An analysis based on the public exposed cases of the National Medical Insurance Administration]. *China Social Security*, (11), 54-55.

Shi, Z. (2011). *Research on fraudulent behavior of social medical insurance designated hospitals based on externality theory*. [Master's thesis]. Hunan University.

Sun, M. (2020). *Research on the social security insurance fraud*. [Master's thesis]. Yangzhou University.

Tian, C. (2014). *Research on the legal issues of anti-fraud in China's social medical*. [Master's thesis]. Liaoning University.

Tian, Y. (2010). 社会政策体系构建的社会理论基础 [The social theoretical basis for the construction of social policy systems]. *Hebei Academic Journal*, *30*(04), 117-120.

Wan, Q. (2020). *A study on the differences in China's basic medical insurance system*. [Doctoral disseratation]. Henan University of Economics and Law.

Wang, G. (2004). 新旧合作医疗比较 [Comparison of old and new cooperative healthcare]. *Qinghai Social Science*, (6), 3.

Wang, W. (2020). 在医疗保险信息化建设中应用大数据分析的研究 [Study on the data analysis in medical insurance information construction]. *Computer Measurement and Control*, *28*(05), 160-164.

Wang, X. (2008). 城镇居民医疗保险付费制度研究——兼析镇江市居民医保付费方式的完善 [Research on the payment system for urban residents' medical insurance and analysis of the improvement of the payment method of residents' medical insurance in Zhenjiang City]. *Chinese Health Economics*, *27*(12), 23-26.

Wang, X. (2005). 城镇职工基本医疗保险制度缺陷分析 [Analysis of the defects of the basic medical insurance system for urban workers]. *Health Economics Research*, (06), 23-24.

Wang, Z., Wang, J., Song, F., Zhang, J., & Yan, Y. (2007). 我国城镇居民基本医疗保险制度初探 [Preliminary Study on urban residential basic medical insurance system in China]. *Chinese Health Service Research*, (10), 677-678.

Wu, J. (2013). 医保新政下医疗机构住院费用不合理增长的原因与控制措施研究 [Research on reasons and measures for the unreasonable increase in hospitalization

expenses of medical institutions under the new medical insurance policy]. *Journal of Frontiers of Medicine, 14*(000), 10-11.

Xiang, D. (2010). 发展型社会政策及其在中国的建构 [Developmental social policy and its construction in China]. *Hebei Academic Journal, 30*(04), 114-117.

Xiang, D. (2013). 发展型社会政策及其构建 [Developmental social policy and its construction in China]. *Basic Theory Research on Philosophy*, (00), 149-157.

Xiao, Z., & Shen, H. (2010). 萨缪尔森经济伦理思想述评 [Review on Economics Ethics Thoughts of Samuelsson]. *Financial Teaching and Research*, (01), 57-63.

Xie, H., & Wan, R. (2014). 社会保险权的公平保障研究 [Research on the fairness and security of social security insurance right]. *The Science of Leadership Forum*, (05), 58-60.

Xie, J. (2013). *The asymmetric information of insurance brokerage market in China*. [Master's thesis]. Southeastern University of Finance and Economics.

Xu, J. (2007). *Application of data mining technology in CRM system of supermarket*. [Master's thesis]. Nanjing University of Science & Technology.

Yang, X. (2016). 医保变革进行时 [The progressive medical insurance change]. *China Hospital CEO*, (02), 44-53.

Yao, X. (2013). 城镇职工基本医疗保险制度缺陷分析 [Analysis of the defects of the basic medical insurance system for urban workers]. *Assets and Finances in Administration and Institution*, (08), 85.

Yao, Y., Sun, Q., Lin, S., & Fan, Q. (2020). 健康保险道德风险防范研究——以商业健康保险欺诈为例 [Alleviating moral hazard in health insurance:evidence on commercial medical expence insurance fraud in China]. *Economic Research Journal, 55*(06), 189-206.

Yin, X., & Liu, X. (2020). 商业医疗保险逆向选择与道德风险的分离检验 [Separation test on adverse selection and moral hazard of commercial Medicare]. *Journal of Finance and Economics*, (07), 33-41.

Yu, F. (2018). *Research on the problem and countermeasure of medical insurance fraud based on big data*. [Master's thesis]. Shandong University of Finance and Economics.

Yu, R., Su, Y., & Xu, A. (2003). 新型农村合作医疗制度新在哪里 [What's new in the new rural cooperative medical system]. *Chinese Health Economics*, (04), 10-11.

Yu, T. (2005). 德国医改增强国民自我责任——从免费就诊改为个人承担部分费用 [German healthcare reform to strengthen national self-responsibility - from free treatment to part of personal expense]. *Health Horizon*, (11), 52.

Yu, L. (2016). *A research of the intelligent audit system of medical insurance based on rule*. [Master's Thesis]. University of Electronic Science and Technology of China.

Zhang, C. (2004). 新型农村合作医疗制度相关因素分析 [Analysis of related factors of new rural cooperative medical system]. *Chinese Health Economics*, (08), 12-14.

Zhang, H. (2014). *Study on the problems and solutions on the social security insurance system*. [Doctoral Dissertation]. Northwest Agriculture and Forestry University.

Zhang, H., & Xu, L. (2007). 城镇居民医疗保险制度的风险及规避 [Risks and avoidances of the urban residents' medical insurance system]. *Health Economics Research*, (12), 17-19.

Zhang, T., Wang, C., & Wu, N. (2022). 以业务风险防控为核心的全面风险管理体系建设的探索和实践 [Exploration and practice of building a comprehensive risk management system centered on business risk prevention and control]. *China Management Magazine*, (11), 50-60.

Zhao, F. (2005). 1935年美国《社会保障法案》的出台及启示 [The introduction and revelation of the Social Security Act of 1935]. *Sub National Fiscal Research*, (04), 53-56.

Zhao, G., & Wu, H. (2011). 保险市场道德风险实证研究评述 [Review of empirical research on moral hazard in the insurance market]. *Insurance Studies*, (04), 116-123.

Zhao, Y. (2020). *Analysis of the crime and punishment of social insurance fund fraud*. [Master's thesis]. Shanxi University.

Zheng, Y. (2011). *Theoretical analysis on the correlation between social security and fairness and efficiency*. [Doctoral dissertation]. Liaoning University.

Zhou, J., Zhang, X., & Cao, Q. (2015). 城镇居民基本医疗保险制度对医疗及预防保健服务利用的影响研究 [Study on the Impact of the Basic Medical Insurance System for Urban Residents on the Utilization of Medical and Preventive Healthcare Services]. *Chinese Journal of Health Policy*, *8*(12), 36-40.

Zhu, L., Zuo, M., Wan, X., Cui, J., & Liu, Y. (2015). 医院临床研究中的大数据应用需求与策略 [The Requirement and Strategy of Big Data Technology in Hospital Clinical Research]. *China Digital Medicine*, *10*(12), 14-15.

Zhu, Y., Sun, L., & Lu, J. (2002). 城镇职工基本医疗保险制度相关主体行为的经济学分析 [Economic analysis of the behavior of relevant entities in the basic medical insurance system for urban workers]. *Journal of Shanxi University of Technology*, (01), 51-55.

Zhu, Z. (2017). *A study of the social medical insurance fraud behavior*. [Master's thesis]. Shanghai Normal University.

Zong, J. (2018). *Research on the existing problems and regulatory countermeasures of basic medical insurance for urban employees in Shuozhou City*. [Master's thesis]. Capital University Of Economics And Business.

[This page is deliberately left blank.]

Annex A: Interview Outline

Interview Outline

Date: 2022. .
Interviewer:

Part One Interview Guide
Interview Topic: Whether the Fraudulent Use of Medical Insurance Funds is Widespread
Interviewee: Hospital physicians, patient and family members
Interview form: Individual interview
Circumstance and Language: in relatively quiet and private space in hospital, speak local Chinese
Interview Purpose: Through interviews with doctors, patients, and their families, we can investigate from both personal and institutional perspective- whether there are medical insurance violations in different channels and methods, the frequency, scope and scale of the problem.
Part Two Opening Remarks
Background information: Since the 1950s, China has gradually established a medical insurance system in line with the times. The medical insurance has played an important role in ensuring people's health and reducing the burden on families, especially in recent years, China has achieved rapid economic growth, people's well-being, stable social order and a harmonious socialist society. The medical insurance system is an effective financing mechanism for public healthcare, a progressive design that makes up for a social security system, and a health expense management model that is widely used in the world so far. However, in recent years, medical insurance fraud cases have spread frequently, involving a large amount of money, complicated cases, many and complex personnel, and hidden criminal processes. Today, I will interview doctors, patients, and their families to find out if there are common violations in the issue, and whether such violations are widespread.
Part Three Interview Questions (to Doctor, patient & family member)
To doctor: 1 In medical practice and prescription, can patients fully understand the use of medical insurance and its protection role? 2 Have any patients or family members been asked to prescribe some drugs for their relatives and family members? For example, there may be a husband who asks for some items for his lover, nutrition for children, and so on. 3 Given a natural year as the cycle, are there any patients who prefer to visit the clinic or are impatient? It is likely that their problems are not as serious as described, but they are still keen to prescribe medication and be hospitalized. Does this often happen? 4 During the patient's visit, have you found that the patient's identity does not match the medical insurance card? Maybe someone changes their identity or bribes a doctor to turn a blind eye, or he says he forgot to take the card, or took the wrong card. What would you do if that case? 5 Are there any patients asked to prescribe a large number of drugs that are not fit for the condition? Have you learned how they would handle these drugs? **To patient and family members:** 1 In your visit, did any doctor say that there are other channels to prescribe lower-priced drugs? 2 Have you ever met someone who borrowed your insurance card for reimbursement? Will you lend your card to someone else? Or get paid to borrow your card? 3 Has the hospital held a such activity on purpose? For example, exchange daily necessities for a

	certain designated drug.
4	Can you use the balance in your medical insurance card to cash out via the doctor or get other compensation?
5	Does the hospital provide services for decomposing the hospitalization period or hospitalization for minor illnesses to facilitate reimbursement?
6	Do you think the hospital's charges were reasonable in the past treatment? Have you ever had any unnecessary concerns about some medical practices and items? For example, have you ever felt deliberate repetition? Are there some ridiculous, unexplained, and undone charges in your medical bills and charge items?
7	Has the doctor ever asked you to do a little favor, such as collecting some bills, keeping your medical insurance card in him for a period of time, or reminding a more economical way to prescribe drugs? Does this often happen?
Part Four Interview Conclusion	
In response to insurance fraud, the State Council officially issued the Regulations on the Practical Supervision and Administration of Medical Security Funds on February 19, 2021, focusing on medical insurance fraud. The regulations specify that if an institution or individual is involved in medical insurance fraud, the online settlement of medical expenses shall be suspended for 6 months (3 months) to 12 months, and a fine of not more than 2-5 times the amount of fraud shall be imposed. This seemingly ordinary news has a deep connotation. As we all know that medical insurance concerns the interests of every insured person, and every penny defrauded is out of the civilians pocket. There is still a lot to be done to protect life-saving money. Moreover, due to the lack of medical insurance common sense and basic knowledge of the law, we do not realize that many behaviors in daily life are insurance fraud. Many of us may accidentally become accomplices in such behaviors, or fail to identify them. In this regard, we should always be in awe, vigilant and disciplined ourselves. Do not indulge ourselves for petty advantages, or try to challenge the law.	

Annex B: List of Other Tables

Table B1 Table of outpatient and inpatient visits by region

	Outpatient Visit					Inpatient Visit				
Month	DQ	HZ	CX	ANJ	Total	ANJ	CX	HZ	DQ	Total
2018/01	711,775.00	1,139,157.00	674,413.00	558,721.00	3,084,066.00	6,370.00	9,240.00	14,533.00	5,610.00	35,753.00
2018/02	680,446.00	1,071,302.00	656,692.00	546,169.00	2,954,609.00	5,640.00	8,270.00	12,785.00	5,073.00	31,768.00
2018/03	788,297.00	1,252,901.00	752,383.00	682,813.00	3,476,394.00	6,985.00	10,530.00	15,909.00	6,465.00	39,889.00
2018/04	736,624.00	1,165,085.00	693,080.00	587,467.00	3,182,256.00	6,337.00	9,411.00	14,830.00	5,927.00	36,505.00
2018/05	751,948.00	1,172,046.00	724,266.00	627,235.00	3,275,495.00	6,972.00	10,377.00	15,371.00	6,291.00	39,011.00
2018/06	725,428.00	1,199,628.00	704,182.00	590,603.00	3,219,841.00	6,458.00	10,392.00	16,107.00	6,347.00	39,304.00
2018/07	743,178.00	1,105,048.00	709,693.00	589,251.00	3,147,170.00	6,740.00	10,537.00	15,942.00	6,110.00	39,329.00
2018/08	786,535.00	1,175,068.00	693,400.00	622,145.00	3,277,148.00	7,238.00	10,521.00	16,338.00	6,437.00	40,534.00
2018/09	761,545.00	1,177,383.00	700,501.00	609,167.00	3,248,596.00	7,451.00	11,129.00	16,356.00	6,246.00	41,182.00
2018/10	745,079.00	1,171,297.00	710,856.00	601,385.00	3,228,617.00	6,749.00	10,309.00	14,789.00	5,632.00	37,479.00
2018/11	743,295.00	1,199,371.00	687,216.00	605,864.00	3,235,746.00	7,128.00	10,469.00	16,640.00	6,206.00	40,443.00
2018/12	797,717.00	1,339,929.00	684,704.00	646,444.00	3,468,794.00	8,654.00	9,574.00	16,952.00	6,437.00	41,617.00
2019/01	889,402.00	1,379,822.00	755,171.00	680,253.00	3,704,648.00	7,664.00	10,925.00	17,180.00	6,547.00	42,316.00
2019/02	722,218.00	1,162,911.00	679,709.00	562,360.00	3,127,198.00	6,323.00	9,423.00	13,912.00	5,286.00	34,944.00
2019/03	807,137.00	1,313,118.00	787,725.00	680,848.00	3,588,828.00	8,287.00	12,246.00	17,455.00	7,010.00	44,998.00
2019/04	274,208.00	438,904.00	212,471.00	220,786.00	1,146,369.00	2,419.00	3,020.00	5,433.00	2,260.00	13,132.00
Average	729,052.00	1,153,935.63	676,653.88	588,219.44	3,147,860.94	6,713.44	9,773.31	15,033.25	5,867.75	37,387.75
Total	11,664,832.00	18,462,970.00	10,826,462.00	9,411,511.00	50,365,775.00	107,415.00	156,373.00	240,532.00	93,884.00	598,204.00

The city name is indicated only by the first letter, such as HZ on behalf of Huzhou

Table B2 Table of outpatient and inpatient amounts by region

	Outpatient Amount					Inpatient Amount				
Month	ANJ	CX	HZ	DQ	Total	ANJ	CX	HZ	DQ	Total
2018/01	48,657,079.70	67,604,455.79	131,250,910.85	58,893,256.93	306,405,703.27	56,368,200.54	83,358,065.77	162,973,486.41	50,319,009.24	353,018,761.96
2018/02	49,684,715.81	66,173,735.80	123,804,059.13	58,249,621.14	297,912,131.88	57,414,330.85	71,261,462.97	144,843,448.72	50,163,752.08	323,682,994.62
2018/03	61,288,849.71	77,305,779.38	145,791,721.54	67,974,109.02	352,360,459.65	71,830,906.38	94,445,270.60	180,951,723.52	68,934,000.07	416,161,900.57
2018/04	52,300,191.33	71,225,313.48	135,107,444.67	63,232,928.83	321,865,878.31	62,787,470.37	82,533,037.79	162,687,823.96	61,908,055.02	369,916,387.14
2018/05	57,270,597.36	75,483,091.60	140,818,142.08	65,559,309.56	339,131,140.60	69,480,791.85	93,306,654.13	170,570,593.73	65,153,154.02	398,511,193.73
2018/06	53,750,227.16	76,740,643.37	156,232,397.42	66,418,252.39	353,141,520.34	67,583,145.94	94,625,138.10	182,132,625.86	69,336,217.63	413,677,127.53
2018/07	53,446,338.54	74,530,983.92	133,058,411.34	65,490,154.45	326,525,888.25	69,821,444.33	95,517,571.36	179,785,508.54	65,490,057.67	410,614,581.90
2018/08	58,053,909.21	75,342,173.74	143,088,487.53	69,422,235.81	345,906,806.29	71,629,374.14	95,428,245.91	190,233,759.18	68,438,360.29	425,729,739.52
2018/09	57,113,966.88	75,501,872.93	144,902,757.65	66,482,523.76	344,001,121.22	77,448,422.25	100,405,219.81	186,775,498.12	67,820,873.26	432,450,013.44
2018/10	55,256,141.09	76,123,031.21	143,364,545.44	63,957,482.97	338,701,200.71	65,349,310.24	90,150,409.35	174,896,832.66	61,015,293.12	391,411,845.37
2018/11	55,734,411.74	76,711,554.67	153,784,109.20	66,836,583.05	353,066,658.66	71,411,155.70	98,531,317.91	200,180,989.86	73,257,026.52	443,380,489.99
2018/12	60,020,902.31	71,393,816.45	166,316,387.49	70,763,620.29	368,494,726.54	92,549,358.55	88,755,561.23	207,563,629.30	73,466,090.01	462,334,639.09
2019/01	65,898,924.41	83,182,132.22	168,440,433.73	77,186,437.89	394,707,928.25	66,766,685.67	108,615,977.72	207,159,478.02	71,637,066.40	454,179,207.81
2019/02	54,010,490.40	71,076,826.66	140,541,870.89	63,795,643.61	329,424,831.56	63,678,503.06	88,141,745.28	170,283,193.52	58,411,659.46	380,515,101.32
2019/03	67,007,839.14	86,174,667.28	166,456,757.29	78,011,133.12	397,650,396.83	79,729,202.25	111,319,196.65	207,706,178.87	78,552,719.42	477,307,297.19
2019/04	20,847,099.61	22,878,211.12	54,241,042.81	24,513,166.37	122,479,519.91	22,636,253.32	31,932,101.87	67,956,510.31	26,576,289.61	149,101,155.11
Total	870,341,684.40	1,147,448,289.62	2,247,199,479.06	1,026,786,459.19	5,291,775,912.27	1,066,484,555.44	1,428,326,976.45	2,796,701,280.58	1,010,479,623.82	6,301,992,436.29
Average	54,396,355.27	71,715,518.10	140,449,967.44	64,174,153.70	330,735,994.52	66,655,284.72	89,270,436.03	174,793,830.04	63,154,976.49	393,874,527.27

The city name is indicated only by the first letter, such as HZ on behalf of Huzhou

Table B3 Frequency descriptions of some variables

Variable	Category	Case number	Proportion%	Variable	Category	Case number	Proportion%
Gender	Male	3212	45.23	Household register condition	Local Municipal	3554	49.93
	Female	3906	54.87		DQCounty	603	8.5
	Total	7118	100		CXCounty	887	12.46
Occupation	Public servant	234	3.29		AJCounty	717	10.07
	Technique professional	543	7.63		NX Dist.	803	11.28
	Office clerk	523	7.35		WX Dist.	554	7.78
	Worker	354	4.97		Total	7118	100
	Peasant	1322	18.57	Payment method	Urban employee	3225	45.3
	Student	121	1.7		Urban resident	3112	43.72
	Active duty Military	23	0.32		Serious disease	32	0.45
	Self-employed	277	3.89		Business Insurance	175	2.469385
	Unemployed	553	7.77		All public expense	349	4.9
	Retired	2779	39.04		All self-payment	221	3.1
	Others	389	5.47		Other insurance	0	0
	Total	7118	100		Others	4	0.056
					Total	7118	100

Table B4 Variable grouping and its meaning

Factor	Group	Assignment
Hospitalizati-on expense	7	1=[0-5000)，2=[5000,10000),3=[10000,30000), 4=[30000,80000),5=[80000,150000)，6=[150000,300000)，7=[300000,-) Unit:CNY
Gender	2	1=male,2=female
Age	10	1=[0,6）,2=[6,17）,3=[17,28）,4=[28,40）,5=[40,48）, 6=[48,55）,7=[55,65）,8=[65,72）,9=[73,84）,10=[85,-) Unit:years old
Hospitalizati-on number	5	1=[0,10),2=[10,20),3=[20,40),4=[40,80),5=[80,-) Unit:day
Sugery time	8	0-7 Unit:time
Hospitaliz-ation time	53	1-53 Unit:time
Payment method	9	1=urban employee basic medical insurance,2=urban residence pension,3=serious disease pension，4=business insurance,5=all public expense,6=all self-payment,7=other insurance,8=others
Discharge department	1	1=pediatrics,2=E.N.T.,3=oncology，4=rheumatology,5=O.B. 6=infectious disease,7=proctology,8=orthopedics，9=pneumology 10=intervention,11=oral surgery,12=geriatrics,13=urology surgery 14=endocrinology,15=dermatology,16=general surgery,17=galactophore，18=neurology, 19=neurosurgery, 20 = intrarenal, 21 = gastroenterology, 22 = digestive endoscopy,23 = extracardial, 24 = intracardiac, 25 extrathoracic, 26 = blood, 27 = extravascular,28 = ophthalmology, 29 = Traditional Chinese Medicine, 30 = Intensive Care Ward
Residence	12	Local municipal 330501 WXDist. 330502 NXDist.330503 DQCounty 330521 CXCounty 330522 AJCounty 330523
Occupation	11	11=public servant;13=technique professional;17=office clerk;24=wrker;27=peasant;31=student;37=active duty military;54=self-employee;70=unemployed;80=retiree;90=others;
Treatment Drug Consumable	7	1=[0,500),2=[500,1000),3=[1000,5000),4=[5000,10000) 5=[10000,30000).6=[30000,80000),7=[80000,-)

Table B5 Association rules on hospitalization expense and payment method

Antecedent	Consequent	support	confidence	lift
{hospitalization expense =3}	=> {payment method=1}	0.13102929	0.6012482	1.0907077
{hospitalization expense =4}	=> {payment method =1}	0.13591101	0.6006870	1.0896897
{hospitalization expense =1}	=> {payment method =7}	0.07650459	0.5937795	1.3639205
{hospitalization expense =2}	=> {payment method =1}	0.15007043	0.5537731	1.0045844
{hospitalization expense=7}	=> {payment method =1}	0.55124593	0.5512459	1.0000000
{hospitalization expense =5}	=> { payment method =1}	0.07079710	0.5449617	0.9885999
{hospitalization expense =6}	=> {payment method =1}	0.01061349	0.5046189	0.9154153

Table B6 Rules of household register and basic medical insurance for urban employee payment method

Antecedent Consequent	support	confidence	lift
{ household register =330523}=> {payment method=1}	0.0492544	0.424979	0.7709427
{ household register =330502}=> { payment method =1}	0.0250887	0.4240558	0.769268
{ household register =330521}=> { payment method =1}	0.0473843	0.4237619	0.7687348
{ household register =330522}=> { payment method =1}	0.0480643	0.3939092	0.7145799
{ household register =330501}=> { payment method =1}	0.3245009	0.7428556	1.3475937

Table B7 Rules of discharge department and hospitalization expense

Antecedent Consequent	support	confidence	lift
{ discharge department =2} => { hospitalization expense =1}	0.03859232	0.90955924	7.0594142
{ discharge department =9} => { hospitalization expense =2}	0.05860494	0.76143894	2.8097766
{ discharge department =20} => { hospitalization expense =2}	0.01622383	0.58289703	2.1509412
{ discharge department =30} => { hospitalization expense =3}	0.01406227	0.56542969	2.5945617
{ discharge department =3} => { hospitalization expense =5}	0.0261330	0.48272768	3.7158029

Table B8 Rules of hospitalization expenses and length of stay (for first-time hospitalization and no surgery)

Antecedent Consequent	support	confidence	lift
{which hospitalization time =1,surgery number=0, expense=1} => {hospital stay=1}	0.0682955	0.9398396	1.6223098
{which hospitalization time =1,surgery number=0, expense=2} => {hospital stay=1}	0.0591635	0.7483871	1.2918329
{which hospitalization time =1,surgery number=0, expense=3} => {hospital stay=2}	0.0343663	0.5099099	1.6950614
{which hospitalization time =1,surgery number=0, expense=4} => {hospital stay=2}	0.0355564	0.6365217	2.1159491

Table B9 Rules of hospitalization expense and hospital stay（for initial hospitalization and one-time surgery）

Antecedent Consequent	support	confidence	lift
{ which hospitalization time =1, surgery time=1, hospitalization expense=2}=> { hospital stay =1}	0.07461019	0.9029982	1.5587159
{ which hospitalization time =1, surgery time=1, hospitalization expense=3}=> { hospital stay =1}	0.05114878	0.7908374	1.3651088
{ which hospitalization time =1, surgery time=1, hospitalization expense=4}=> { hospital stay =1}	0.03239909	0.8007203	1.1402385

Table B10 Settlement information——Medical receipt table

Field name	Field example	Yes/no	remark
Patient No.	1000012345	Y	
billing status mark	Normal settlement, refund record, write-off record, cancellation record	Y	For encoding, provide the dictionary value
Hospital No.	H1001	Y	
Hospital name	People‘s Hospital	N	
Department No.	S10	N	
Department Name	Internal medicine	N	
Doctor No.	H1001-55	N	
Doctor Name	Li Yi	N	
Participant No.	100335-1	Y	
Gender	Female	Y	For encoding, provide the dictionary value
Age	25	Y	
Type of treatment	1-hospitalization	N	For encoding, provide the dictionary value
Medical Category	21-general hospitalization	Y	For encoding, provide the dictionary value
Admission Date	2018-05-12 08:10	Y	
Discharge Date	2018-05-30 10:30	Y	
Settlement Date	2018-05-31 12:20	Y	
Admission Diagnosis No.	S10.310	N	
Admission diagnosis Name	Accidental injury	N	
Discharge diagnosis No.	S10.310	Y	
Discharge diagnosis name	Accidental injury	N	
Discharge secondary diagnosis No.	E14.910	N	
Discharge secondary diagnosis name	Diabetic nephropathy	N	
Total	3100.00	Y	
Compensable amount	2700.00	N	
Actual compensation amount	2100.00	Y	
Severe illness compensation amount	100.00	N	
Other compensation amounts	20.00	N	
Sel-payment Amount	1000	Y	

Table B11 Settlement information——medical visit details

Field name	Field example	Remark
Patient No,	1000012345	
Item Coding	100335	
Settlement status mark	Normal settlement, refund record, write-off record, cancellation record	For encoding, provide the dictionary value
Expense Occurrence Time/BillingTime/Prescription/Advice Time	2015-05-30 10:30	
Hospital item name	Glucose injection *500ml	
Category item name	Glucose injection	
Cost type	1-drug cost	For encoding, provide the dictionary value
Invoice type	11-Western medicine cost	For encoding, provide the dictionary value
Specification and unit	500ml	
Drug form	Injection	
Unit price	5.00	
Quantity	2	
Total amount	10.00	
Compensable amount	10.00	

Table B12 The following is an example of code meanings, values of codes in the above table are provided.

Table name	Field code	Field name	Value of code
KA02 drug category			
KA02	AKA060	Drug No.	No value
KA02	AKA061	Drug name	
KA02	AKA070	Drug form	101-tablet 102-capsule 201-injection
KA03 treatment category			
KA03	AKC065	Invoice type	11-nursing care 12-consultation 13-surgery 14-examination 99-other

Table B13 Patient 228940 medication data in a certain period

Medical visit Date	Medical insurance drugs	Quantity	Amount (CNY)	Hospital Name	Personal No.	Gender	Age
2018/2/23	Valsartan	5	171.35	HZ T.C.M. Hospital	228940	M	64
2018/2/23	Finasteride	3	169.11		228940	M	64
2018/2/23	Atorvastatin	4	222		228940	M	64
2018/2/23	Ursodeoxycholic acid	1	211.86		228940	M	64
2018/2/23	Clopidogrel	4	433.16	A hospital of the P.L.A.	228940	M	64
2018/2/23	Atorvastatin	4	222		228940	M	64
2018/2/23	Finasteride	3	169.11		228940	M	64
2018/2/23	Valsartan	4	137.08		228940	M	64
2018/2/23	Ursodeoxycholic acid	2	423.72		228940	M	64
2018/3/6	Ursodeoxycholic acid	1	211.86	HZ T.C.M. Hospital	228940	M	64
2018/3/6	Valsartan	5	171.35		228940	M	64
2018/3/6	Finasteride	3	169.11		228940	M	64
2018/3/6	Atorvastatin	4	221.96		228940	M	64
2018/3/6	Atorvastatin	4	221.96	DQ People's Hospital	228940	M	64
2018/3/6	Clopidogrel	4	433.16		228940	M	64
2018/3/6	Finasteride	3	169.11		228940	M	64
2018/3/6	Donepezil	4	534.76		228940	M	64
2018/3/6	Valsartan	5	171.35		228940	M	64
2018/3/6	Esomeprazole	4	262.76		228940	M	64
2018/3/6	Ursodeoxycholic acid	2	423.72		228940	M	64

Table B14 Medication data of a pharmacy in a town in DQ City for a certain time period

Personal No.	Number of visits
12080809	231
Renyi pharmacy of one town in DQ	231
12078314	229
Renyi pharmacy of one town in DQ	229
13105268	225
Renyi pharmacy of one town in DQ	225
13008781	223
Renyi pharmacy of one town in DQ	223
14564051	219
Renyi pharmacy of one town in DQ	219
13013637	213
Renyi pharmacy of one town in DQ	213
12031426	211
Renyi pharmacy of one town in DQ	211
13482432	211
Renyi pharmacy of one town in DQ	211
22114432	211
Renyi pharmacy of one town in DQ	211
13461106	204
Renyi pharmacy of one town in DQ	204
12045565	198
Renyi pharmacy of one town in DQ	197
12037613	197
Renyi pharmacy of one town in DQ	197
13050618	195
Renyi pharmacy of one town in DQ	195

Table B15 An example of a pharmacy with different insured persons' drug package

Insured person 120***00		Insured person 120***37	
2018/1/5	Banlangen Granules	2018/1/3	Bezoar detoxification tablets
2018/1/7	Glycerine Enema	2018/1/5	Glycerine Enema
2018/1/10	Tendon relaxing and blood activating tablet	2018/1/7	Bezoar detoxification tablets
2018/1/12	Paracetamol capsule	2018/1/13	Paracetamol capsule
2018/1/15	Anti skeletal degeneration tablet	2018/1/15	Glycerine Enema
2018/1/17	Glycerine Enema	2018/1/17	Paracetamol capsule
2018/1/19	Paracetamol capsule	2018/1/19	Bezoar detoxification tablets
2018/1/22	Bezoar detoxification tablets	2018/1/21	Glycerine Enema
2018/1/25	Glycerine Enema	2018/1/24	Paracetamol capsule
2018/1/27	Paracetamol capsule	2018/1/26	Bezoar detoxification tablets
		2018/1/27	Glycerine Enema
Insured person 120***63		Insured person 130***53	
2018/1/11	Strong loquat dew	2018/1/3	Bezoar detoxification tablets
2018/1/11	Colloidal pectin bismuth capsules	2018/1/5	Glycerine Enema
2018/1/11	Baixian Xiatare Tablet	2018/1/7	Glycerine Enema
2018/1/13	Liuwei Dihuang capsule	2018/1/8	Glycerine Enema
2018/1/15	Neck and Back Health capsule	2018/1/13	Duyiwei soft capsule
2018/1/17	Sao Niu Tablets	2018/1/15	Strong loquat dew
2018/1/17	Compound bean root aminophenaminetablet	2018/1/17	Glycerine Enema
2018/1/19	Duyiwei soft capsule	2018/1/19	Paracetamol capsule
2018/1/21	Glycerine Enema	2018/1/21	Paracetamol capsule
2018/1/23	Paracetamol capsule	2018/1/24	Glycerine Enema
2018/1/25	Bezoar detoxification tablets	2018/1/26	Bezoar detoxification capsule
2018/1/27	Glycerine Enema	2018/1/27	Glycerine Enema
2018/1/31	Paracetamol capsule		

Table B16 Prescription details of the insured person

Date of visit	Drug name	Amoun（CNY）
2018/8/11	Finasteride tablet	75
2018/8/11	Biantong tablet	39
Total		114
2018/7/31	Qingkailing Granule	20.6
2018/7/31	Tongxinluo capsule	44
2018/7/31	Celeoxib capsule	44
Total		108.6
2018/7/29	Compound Kangling Granule	24
2018/7/29	Compound Hypotensive Tablets	8.58
2018/7/29	Finasteride tablet	75
Total		107.56
2018/7/27	Gukang Capsule	66
2018/7/27	Liuwei Dihuang capsule	53.8
Total		119.8
2018/7/11	Levamlodipine Besylate Tablet	53
2018/7/11	Neck pain tablets	36
2018/7/11	Anshen Bunao Syrup	23.5
Total		112.5

Table B17 The typical diagnosis and risk amount of hospitalization period decomposition

Hospital name	Amount (CNY)
Huzhou Evergreen Elderly Rehabilitation Nursing Home	4317508.23
Changxing Jizhou Hospital	288959.12
Changxing County Jizhou Hospital	220695.44
Deqing Aixia Hospital Co., Ltd	210996.29
Nanxun Jiu'an Geriatric Hospital	79630.66
Huzhou Mingzhou Hospital	77522.47
Anji Perdue Bone Injury Hospital of T.C.M.	72605.58
Deqing Aimjia Hospital Co., Ltd	44908.37
Huzhou Banger Orthopedic Hospital	44349.96
Deqing Aimujia Hospital	40094.59
Huzhou Tai Lake Mingde Hospital	26281.42
Huzhou Aier Eye Hospital	18313.89
Huzhou Maria Obstetrics and Gynecology Hospital	12842.05
Deqing County Wuzhou Hospital	2014.92
Total	5456722.99

Table B17 Diseases groups and the amount involved in hospitalization period decomposition in private medical institutions

Private medical institution	Disease group	Amount (CNY)
Huzhou Evergreen Elderly Rehabilitation Nursing Home	Sequelae of cerebral infarction	1020384.17
	Alzheimers	944022.51
	Hypertention	757877.38
	CAHD	362082.82
	COPD	194080.35
	Cerebral infarction	139909.99
	Cerebral hemorrhage sequel	115133.21
	Coronary artery occlusion	111057.37
	Lumber spine disc herniation	105205.49
	Fracture	96441.02
	Parkinson's disease	90599.54
	Diabetes	89667.1
	AECOPD	62677.96
	Cerebral thrombosis sequel.	51938.76
	Type II diabetes	47770.68
	Brain atrophy	47408.53
	Level 3 hypertension	37025.54
	Intertrochanter fracture of femur	17452.85
	Rheumatoid Arthritis	16619.6
	COPD with acute lower respiratory infection	5831.65
	Lumbar disc herniation with sciatica	4321.71
	Total	4317508.23
Changxing Jizhou Hospital	Anthracosilicosis	161676.45
	Sequelae of cerebral infarction	43402.17
	Cerebral infarction	22960.91
	CHD	14411.02
	Fever	14115.09
	Hypertensive heart disease	13833.87
	Lumbar spine disc herniation	9526.4
	Chronic heart failure	9033.21
	Total	288959.12
Changxing County Jizhou Hospital	CHD	98088.87
	COPD	60904.73
	Sequelae of cerebral infarction	39433.87
	COPD	15643.31
	Ileus	6624.66
	Total	220695.44
Deqing Aimujia Hospital	Arrhythmia of CHD	71243.61
	Sequelae of cerebral infarction	49853.84
	COPD	19492.15
	Alzheimer’s	14744.71
	Diabetes	12726.85
	Level 3 hypertension	12472.54
	Hypertention	10272.61
	Lung infection	9686.76
	Multiple nerve damage	6181.13
	Chronic Bronchitis	4322.09
	Total	210996.29

Table B18 Diseases groups and the amount involved in hospitalization period decomposition in private medical institution (Continue)

Private medical institution	Disease group	Amount (CNY)
Nanxun Jiu'an Geriatric Hospital	Cerebral hemorrhage	29985.37
	Vasculitis	17093.27
	Senile hypertension	9610.64
	Type II hypertention	8257.93
	Hypertention	6870.19
	Cerebral infarction	4808.48
	Cerebral hemorrhage sequel	3004.78
	Total	79630.66
Huzhou Mingzhou Hospital of North Zhejiang Province	Intracranial hemorrhage	40254.73
	Sleep disorder	37267.74
	Total	77522.47
Anji Perdue Bone Injury Hospital of T.C.M.	Lumbar spine disc herniation	63540.44
	Knee joint pain	6762.65
	Neck spine disc herniation	2302.49
	Total	72605.58
Deqing Aimujia Hospital	Sequelae of cerebral infarction	11598.71
	Diabetes	9961.81
	Arrhythmia of CHD	8964.72
	Multiple nerve damage	7957.59
	Hypertention	6425.54
	Total	44908.37
Huzhou Banger Orthopedic Hospital	Gout	22354.8
	Lung infection	15886.25
	Knee Joint Pain	6108.91
	Total	44349.96
Huzhou Tai Lake Mingde Hospital	Lung infection	18058.15
	Senile osteoporosis	4230.57
	Pulmonary nodules	3992.7
	Total	26281.42
Huzhou Aiyer Eye Hospital	Cataract	18313.89
	Total	18313.89
Huzhou Maria Obstetrics and Gynecology Hospital	Acute mastitis	6521.07
	Pelvic inflammatory	5062.61
	Urinary incontinence	1258.37
	Total	12842.05
Deqing County Wuzhou Hospital	Chronic Prostatitis	2014.92
	Total	2014.92
Deqing Aimujia Hospital	COPD	12069.67
	Hypertension	11864.31
	Type II diabetes	11011
	Sequelae of cerebral infarction	5149.61
	Total	40094.59

Table B19 Examples of low-standard admission

Hospital	Personal No.	Gen.	Age	Type of Visit	Admission	Discharge	Time	Discharge diagnosis	Dept.	Total CNY	Reimbursable Amount	Admission No.
DQ People's Hospital	121***41	M	47	inpatient	2018/7/29	2018/7/30	1	Sleep apnea syndrome (SAS)	Internal medicine 3	575	457	32734901
	130***25	M	36	inpatient	2018/11/8	2018/11/9	1		Internal medicine 3	575	457	33812901
	120***05	M	47	inpatient	2018/1/15	2018/1/16	1	SAS	Internal medicine 3	573	455	30725901
	134***09	M	46	inpatient	2018/3/7	2018/3/8	1	SAS	Internal medicine 3	573	455	31222002
	131***10	M	47	inpatient	2018/5/20	2018/5/21	1	SAS	Internal medicine 3	573	455	32004501
	120***56	M	32	inpatient	2018/6/6	2018/6/7	1	SAS	Internal medicine 3	573	455	32175301
	145***30	M	49	inpatient	2018/6/16	2018/6/17	1	SAS	Internal medicine 3	573	455	32275801
	120***41	M	56	inpatient	2018/7/12	2018/7/13	1	SAS	Internal medicine 3	573	455	32560901
	121***15	F	52	inpatient	2018/7/19	2018/7/20	1	SAS	Internal medicine 3	573	455	32632401
	121***33	M	37	inpatient	2018/10/7	2018/10/8	1	SAS	Internal medicine 3	573	455	33461901
	133***18	M	55	Inpatient	2018/11/3	2018/11/4	1	SAS	Internal medicine 3	573	455	33752801

Table B20 The number of patients taking off-label adjuvant medication

Hospital name	The number of patients taking off-label adjuvant medication
Changxing Jizhou Hospital	35
Deqing County Hospital of Traditional Chinese Medicine	15
Anji County People's Hospital (city off-site)	9
(Designated) Changxing County People's Hospital	8
Huzhou Lianshi Hospital, Nanxun District	6
Huzhou Jiaotong Hospital	6
Deqing County People's Hospital (Wukang)	5
Huzhou Rehabilitation Hospital	5
Changxing County Guotai Rehabilitation Hospital	5
Deqing County Zhongguan Town Health Center	5
Huzhou Nanxun District Linghu People's Hospital	5
Nanxun Jiu'an Geriatric Hospital	4
Changxing Orthopedic Hospital	3
Zhejiang Anji Jianheng Diabetes Hospital Co., Ltd	3
Changxing County Hospital of Traditional Chinese Medicine	2
Shuanglin People's Hospital, Nanxun District, Huzhou City	2
Huzhou Nanxun District Hospital of Integrated Traditional Chinese and Western Medicine	2

Table B21 The use of exceeding adjunct drugs in Changxing Zhangzhou Hospital

Disease/Drug Name	Samm-ai inject i-on	Tanshinone IIA	CAMP	Alprostadil	Shuxuening injection (Ginkgo leaf injection)	Xuesa-itong inject i-on	Edarav-one	Vinpo--cedin	Total
Type II diabetes				369	434.4				803.4
Hypertention				1968	434.4				2402.4
CAHD	214.8			246	868.8			793.5	2123.1
Tenosynovitis	358			369					727
Unstable neck spine				492	289.6				781.6
Chronic Cholecystitis	358								358
Cerebral infarction	805.5					129			934.5
Sequelae of cerebral infarction	358			246		86	1230		1920
Abnormal kidney function	1254.6		104					621.6	1980.2
Diabetes	537			369					906
Heart Palpitation	143.2				289.6				432.8
Vertigo syndrome	155.6			1476	1013.6		615	1408.5	4668.7
Vertigo syndrome		303		492	434.4			634.8	1864.2
Spinal disc herniation	315			246	434.4				995.4
Total	4499.7	303	104	6273	4199.2	215	1845	3458.4	20897.3

Table B22 Top 5 insured persons with frequent medical visits

Participant No,	Total amount of drugs（CNY）	Total number（box）
A00800XXX	2270.79	276
Aspirin	66.4	62
Amlodipine	388.31	13
Irbesartan	428.7	45
Clopidogrel	1082.7	10
Telmisartan	76	48
L-amlodipine	228.68	98
A01331 XXX	268.52	212
Felidipine	154.8	6
Telmisartan	108	6
Niquendipine	5.72	200
A00157 XXX	389.01	194
Amlodipine	51	2
Felidipine	145.45	50
Candesartan Cilexetil	41.16	2
Telmisartan	151.4	140
A04761 XXX	1033.84	192
Telmisartan	181.68	168
Valsartan	852.16	24
A03511 XXX	133.74	153
Telmisartan	52.02	150
L-amlodipine	81.72	3

Table B23 Beyond hospital level medication

Hospital /Drugs	Amount（CNY）
AJ People’s Hospital	166498.34
Irinotecan	166498.34
DQ County People’s Hospital	145783.32
Fasudil	41341.6
Irinotecan	104441.72
DQ County T.C.M. Hospital	46795
Irinotecan	46795
CX County People’s Hospital（local）	14589.08
Fasudil	183.4
Memantine	14405.68
DQ Hospital	12410.76
Irinotecan	12410.76
HZ Jiaotong Hospital	5854.75
Entacapone	452.68
Memantine	5402.07
AJ County No.3 People’s Hospital	4187.31
Memantine	1080.39
Irinotecan	3106.92
AJ County T.C.M. Hospital	3132
Irinotecan	3132
AJ County No.2 People’s Hospital	905.4
Entacapone	905.4
DQ County new city YY town health center	720.34
Memantine	720.34
HZ City Jiaotong Hospital	360.17
Memantine	360.17
HZ City MY Health Center	360.13
Memantine	360.13

Table B24 Overuse of HbAlc in hospitals and the amount at risk

Hospital name	Sum / Amount
DQ County People’s Hospital	642960
HbAlc	642960
HZ City No.1 People’s Hospital (local)	486600
HbAlc	486600

Table B25 Diagnosis and time of admission in cases of HBA1C use

Document No.	Medical insurance item	Quant.	Amt.	Hospital	Admission Date	Discharge date	Discharge diagnosis
693044845	HbAlc	1	30	DQ County People's Hospital	2018/9/15	2018/9/17	Skin masses
693044845	HbAlc	1	30	DQ County People's Hospital	2018/9/15	2018/9/17	Skin masses
693034402	HbAlc	1	30	DQ County People's Hospital	2018/9/17	2018/9/17	Malignant tumors of the lung
693034402	HbAlc	1	30	DQ County People's Hospital	2018/9/17	2018/9/17	Malignant tumors of the lungs
693026680	HbAlc	1	30	DQ County People's Hospital	2018/8/30	2018/9/17	Malignant thyroid tumors
693026680	HbAlc	1	30	DQ County People's Hospital	2018/8/30	2018/9/17	Malignant thyroid tumors
693024286	HbAlc	1	30	DQ County People's Hospital	2018/9/15	2018/9/16	Lumber spine fracture
693024286	HbAlc	1	30	DQ County People's Hospital	2018/9/15	2018/9/16	Lumber spine fractrue
693015721	HbAlc	1	30	DQ County People's Hospital	2018/9/9	2018/9/13	Stomachache
693015721	HbAlc	1	30	DQ County People's Hospital	2018/9/9	2018/9/13	Stomachache
692975976	HbAlc	1	30	DQ County People's Hospital	2018/9/11	2018/9/16	TSAH
692975976	HbAlc	1	30	DQ County People's Hospital	2018/9/11	2018/9/16	TSAH
692940450	HbAlc	1	30	DQ County People's Hospital	2018/8/29	2018/9/15	TSAH
692940450	HbAlc	1	30	DQ County People's Hospital	2018/8/29	2018/9/15	TSAH

Table B26 Ranking of hospitals with repeated medications

HZ City No.1 People's Hospital (Local)	5252.15
Atorvastatin	132.51
Esomeprazole	522.96
Omeprazole	219.36
Lansoprazole	144.18
Rabeprazole	594.56
Clopidogrel	2844.36
Rosuvastatin	598.23
Simvastatin	195.99
HZ City Central Hospital (Local)	3034.56
Atorvastatin	498.3
Esomeprazole	402.84
Omeprazole	798.01
Lansoprazole	19.84
Rabeprazole	1012.5
Pantoprazole	303.07
HZ City NX District Town Central Health Center	2127.35
Acarbose	123.88
Atorvastatin	44.17
Omeprazole	879.33
Lansoprazole	397.78
Rabeprazole	358.86
Pantoprazole	225.08
Simvastatin	98.25

Table B27 Cases of Repeated Medication

Medical Insurance card No.	Receipt No.	Medical insurance No.	Price	Qt.	Amt.	Hospital	Dept.	Gen.	Date	Discharge diagnosis
A00124 xxx	990667629169	Omeprazole	54.84	1	54.84	HZ City No.1 People's Hospital (Local)	Gastroenterology	M	2018/8/30	Gastritis
A00124 xxx	990667629169	Esomeprazole	65.37	1	65.37			M	2018/8/30	Gastritis
B00880 xxx	659105878	Simvastatin	32.58	3	97.74		Cardiovascular hypertension	M	2018/5/30	Coronary heart disease
B00880 xxx	659105878	Rosuvastatin	48.5	5	242.5			M	2018/5/30	Coronary heart disease
C00389 xxx	649036577	Rabeprazole	48.38	5	241.9		Cardiology	M	2018/2/8	Coronary heart disease
C00389 xxx	649036577	Lansoprazole	40.39	2	80.78			M	2018/2/8	Coronary heart disease
C00752 xxx	658009681	Rabeprazole	32.06	4	128.24		E.N.T.	F	2018/5/17	Chronic pharyngitis
C00752 xxx	658009681	Esomeprazole	65.37	3	196.11		E.N.T.	F	2018/5/17	Chronic pharyngitis

Table B28 False billing cases

Hospital Item	Medical insurance item	Qt	Hospital	Personal No.	Gen.	Age	Admission	Discharge	Inpatient (Day)	Diagnosis	Inpatient No.
Nursing	Nursing (Tertiary hospital)	58	HZ City No.1 People's Hospital (Local)	108476	M	75	2018/6/22	2018/8/8	47	Stage 5 chronic kidney disease	4316030
Bed fee	General ward Bed fee	28		108476	M	75	2018/6/22	2018/8/8	47		4316030
Nursing	Nursing (Tertiary hospital)	60		810605	F	56	2018/7/10	2018/8/3	24	Personal history of gastric malignancy	4342000
Bed fee	General ward Bed fee	60		810605	F	56	2018/7/10	2018/8/3	24		4342001
Nursing	Nursing (Tertiary hospital)	28		917564	F	61	2018/7/13	2018/8/4	22	Lumbar spondylolisthesis	4346790
Bed fee	General ward Bed fee	28		917564	F	61	2018/7/13	2018/8/4	22		4346790

Table B29 Excessive use of auxiliary drugs in Changxing Jizhou Hospital

Item	Sum / Amount
General acupuncture (body) (⩽ 20 points)	434598
X-ray	252775
General acupuncture (rapid) (⩽ 20 points)	211040
Local infiltration anesthesia	145200
Endoscopic tissue biopsy and diagnosis	126906
intravenous injection	90254.1
Regular biochemical screening	87240
Color ultrasound regular examination (single part)	72000
Electroacupuncture (general electric)(20 points)	58338
Acupoint regulation therapy	53400
Ultra-short wave (low power)	46944
Intravenous fluids (outpatient/with infusion)	41850.8
Moxibustion (moxa bar)	37380
Intravenous fluids (outpatient/without infusion)	28129
Thermoelectric pulse therapy	20916
minor dressing change	16165
annular mixed hemorrhoidectomy	15360

Table B30 Concept of risk scenarios

Name	Description
Risk scenarios	Common fraud in a certain category of lines is the result of the experience of experts dealing with a large number of past cases
Rule	Conditional logic of risk behavior, e.g. suspected concealment of a history of diabetes if insulin use is detected
Model	Characterize complex conditions with supervised classification models

Table B31 Case group of hospitalization period decomposition

Bill quantity	Medical institution	Participant No.	Discharge diagnosis disease	Ad Admission and discharge times	Risk amount	Audit status
3	Huzhou Tuberculosis Hospital	665993525	--	--	41802.2	--
1. December, 3rd No.319482502	Huzhou Tuberculosis Hospital	665993525	ST elevation MI	2018.4.28-2018.5.10	25123.88	Pending
2. December, 17th No.319455924	Huzhou Tuberculosis Hospital	665993525	Heart failure	2018.5.13-2018.5.27	8315.32	Pending
3. December, 18th No.319453242	Huzhou Tuberculosis Hospital	665993525	Coronary disease	2018.5.30-2018.6.10	8363	Pending
	Huzhou Linghe District Hospital of T.C.M.	537491040	Diabetes	--	10232.41	--

Table B32 Medical insurance items

Prescription date	Medical insurance item No.	Medical insurance item	Hospital item
2018/08/08	x161101008960924	MRI scan	MRI scan
2018/08/08	x060102312512371	CT	CT
2018/08/08	x120942842910471	Multi-slice spiral CT	Multi-slice spiral CT
2018/08/08	x180500312361234	Alprostadil injection	Alprostadil injection
2018/08/08	x129481792872134	Oxiracetam for injection	Oxiracetam for injection
2018/08/08	x161101008960004	Atorvastatin	Atorvastatin calcium tablets
2018/08/08	x060102002771033	Vitamin B1	Vitamin B1
2018/08/08	x180500009920001	Clopidogrel	Clopidogrel bisulfate tablet
2018/08/08	x160600008550001	Terazosin	Terazosin hydrochloride tablet
2018/08/08	x03010307047d002	phenomepin	Phenol Mememine Tablets (Tylenol)

Table B33 Summary of abnormal cases of diseases charging

Hospital Name	Count of Bills
HZ City Maternal and Child Health Hospital	909
ZJ university QLhospital	881
ZJ provincial hospital	706
ZJ university QLChildren's Hospital(HZ City Children's Hospital)	603
ZJ Provincial Qianfoshan Hospital	589
ZJ University No.2 Hospital	571
ZJ affiliated Hospital of the University of Chinese Medicine	330
HZ City No.5 People's Hospital	322
ZJ Provincial E.N.T. Hospital（ZJ Provincial Hospital West hospital）	283
HZ City Central Hospital	242
ZJ Provincial Maternal and Child Health Hospital	215
HZ City No.4 People's Hospital	209
HZ City No.3 People's Hospital	190
ZJ Provicial Hospital（West hospital）	142
ZJ Provincial No.3 Hospital	135

Table B34 Summary of abnormal cases of diseases charging

Hospital Name	Sum of reimbursement amount
ZJ University QLHospital	13091871.68
ZJ Provicial No.2 Hospital	10923093.96
ZJ Privincial Hospital	10548761.66
ZJ affiliated Hospital of the University of Chinese Medicine	8875782.91
HZcity central hospital	5425441.2
ZJ provincial ENT Hospital	5033731.02
ZJ university No.2 Hospital	4530480.51
HZCity No.4 Hospital	2917521.64
ZJ university QL Children's Hospital (HZ Children's Hospital)	2772471.14
ZJ affiliated No.2 Hospital of the University of Chinese Medicine	2622442.57
ZJ provincial Hospital (West hospital)	2333238.14
ZJ provincial Hospital	2300545.8
General Hospital of XXX Military Region of the P.L.A.	2022590.97
HL city No.5 People's Hospital	1856949.24
Military General Hospital	1697790.15
HZCity No.3 People's Hospital	1510945.56
County People's hospital	1064186.5
HZCity Maternal and Child Health Hospital	966565.46
HZ City Jigang Hospital	709665.6

Table B35 Illustration of an abnormal case in one hospital

Medical Institution	Discharge diagnosis disease	Unit Price (CNY)	Hospital Class	Settlement date	Billing Amount	Discharge diagnosis disease code	Name	Total Amount (CNY)	Medical Insurance Item
HZ City Maternal and Child Health Hospital	Vaginal surgical delivery	550.0	Tertiary hospital	2018-01-02	9098.5	SY.002	Ma,JJ	9098.5	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	Vaginal surgical delivery	500.0	Tertiary hospital	2018-01-02	9098.5	SY.002	Ma,JJ	9098.5	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	normal delivery	660.0	Tertiary hospital	2018-01-02	7413.35	SY.001	Geng, J	7413.35	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	normal delivery	660.0	Tertiary hospital	2018-01-02	4847.76	SY.001	Liu,DM	4847.76	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	normal delivery	660.0	Tertiary hospital	2018-01-02	5954.27	SY.001	Zhang,XR	5954.27	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	normal delivery	660.0	Tertiary hospital	2018-01-02	7607.86	SY.001	Yu,MJ	7607.86	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	Vaginal surgical delivery	660.0	Tertiary hospital	2018-01-02	8571.75	SY.002	Li,WQ	8571.75	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	Vaginal surgical delivery	650.0	Tertiary hospital	2018-01-02	8571.75	SY.002	Li,WQ	8571.75	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	normal delivery	660.0	Tertiary hospital	2018-01-02	8690.99	SY.001	Zheng,LL	8690.99	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	normal delivery	660.0	Tertiary hospital	2018-01-03	4472.4	SY.001	Dang,Y	4472.4	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	Vaginal surgical delivery	650.0	Tertiary hospital	2018-01-03	6798.46	SY.002	Yi,ZhE	6798.46	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	Vaginal surgical delivery	660.0	Tertiary hospital	2018-01-03	5918.67	SY.002	Zhou, YG	5918.67	Spontaneous single delivery
HZ City Maternal and Child Health Hospital	Vaginal surgical delivery	650.0	Tertiary hospital	2018-01-03	5918.67	SY.002	Zhou, YG	5918.67	Spontaneous single delivery

Table B36 Examples of swapping items in medical institution

Consultation No.	Admission Date	Medical insurance item	Hospital item	Amount (CNY)
1180136489	2019-01-17	Pre-hospital emergency expenses	Visit fee	120
1180150975	2018-12-12	Pre-hospital emergency expenses	Visit fee	20
1180184877	2018-12-30	Pre-hospital emergency expenses	Visit fee	20
1180198416	2018-12-22	Pre-hospital emergency expenses	Visit fee	60
1180225472	2019-02-13	Pre-hospital emergency expenses	Visit fee	20
1180225868	2019-02-14	Pre-hospital emergency expenses	Visit fee	160
1180234568	2019-02-21	Pre-hospital emergency expenses	Visit fee	60
1180245140	2019-01-16	Pre-hospital emergency expenses	Visit fee	60
1180287201	2019-01-23	Pre-hospital emergency expenses	Visit fee	20
1180296409	2019-02-03	Pre-hospital emergency expenses	Visit fee	40
1180311617	2019-02-28	Pre-hospital emergency expenses	Visit fee	60
1180325581	2019-02-22	Pre-hospital emergency expenses	Visit fee	40
1180327652	2019-03-02	Pre-hospital emergency expenses	Visit fee	40
1180333659	2019-03-12	Pre-hospital emergency expenses	Visit fee	300
1180344616	2019-03-13	Pre-hospital emergency expenses	Visit fee	60

Table B37 Summary of swapping prescriptions in the pharmacy

Pharmacy name	Number of prescriptions
Renyi Pharmacy, Wukang Town, DQ County	16357
DQ Tianrun Pharmacy North Lake East Street Store	8732
DQ Tianrun Yinong joins the pharmacy	5292
DQ County New City Kangyi Pharmacy	2670
Tianbaokang Pharmacy, Xin'an Town, DQ County	2642
Tianrun Wuqian chain pharmacy	2416
people's big pharmacy Qianyuan store	2207
A pharmacy in a town of DQ County	2057
DQ New Anbao Ningtang Pharmacy	1899
DQ Zhengbentang Pharmacy	1752
PEOPLE'S BIG PHARMACY	1319
DQ County Renkang Pharmacy	1232
DQ Tangda Pharmacy TCM Sitting Clinic	1199
DQ Tianrun Pharmacy Co., Ltd. Store	1188
DQ Tianrun Pharmacy Co., Ltd. Store	1156
DQ County Wanmin Pharmacy Co., Ltd. store	1155
DQ County Ordinary People Pharmacy	1102
DQ Tianrun Pharmacy Co., Ltd. qingchuntanstore	1018
DQ Tianrun Chengxing Franchise stores	994
people big pharmacy Co., Ltd. Store	934
people’s big pharmacy chain Co., Ltd. DQ Wukang store	924
DQ County People’s Big Pharmacy Wukang Qianqiu Store	867

Table B38 Examples of swapping prescriptions in the pharmacy

Participant 120***00		Participant 120***37	
2018/1/5 0:00	Banlangen Granules	2018/1/3 0:00	Bezoar detoxification tablets
2018/1/7 0:00	Glycerine enema	2018/1/5 0:00	Glycerine enema
2018/1/10 0:00	Shujinhuoxue tablet	2018/1/7 0:00	Bezoar detoxification tablets
2018/1/12 0:00	Paracetamol capsule	2018/1/13 0:00	Paracetamol capsule
2018/1/15 0:00	Anti-skeletal degeneration tablet	2018/1/15 0:00	Glycerine enema
2018/1/17 0:00	Glycerine enema	2018/1/17 0:00	Paracetamol capsule
2018/1/19 0:00	Paracetamol capsule	2018/1/19 0:00	Bezoar detoxification tablets
2018/1/22 0:00	Bezoar detoxification tablets	2018/1/21 0:00	Glycerine enema
2018/1/25 0:00	Glycerine enema	2018/1/24 0:00	Paracetamol capsule
2018/1/27 0:00	Paracetamol	2018/1/26 0:00	Bezoar antidotal tablet
		2018/1/27 0:00	Glycerine enema

Table B39 Swapping prescription amount in pharmacy

Date of visit	Drug name	Amount(yuan)
2018/8/11 0:00	Finasteride tablets	75
2018/8/11 0:00	Laxative tablets	39
Total		114
2018/7/31 0:00	Qingkailing particles	20.6
2018/7/31 0:00	Tongxinluo capsules	44
2018/7/31 0:00	Celecoxib capsules	44
Total		108.6
2018/7/29 0:00	Compound cold spirit granules	24
2018/7/29 0:00	Compound reserpine triamferidine tablets	8.58
2018/7/29 0:00	Finasteride tablets	75
Total		107.56
2018/7/27 0:00	GuKang capsules	66
2018/7/27 0:00	Liuwei Dihuang capsules	53.8
Total		119.8
2018/7/11 0:00	L-amlodipine besylate tablets	53
2018/7/11 0:00	Neck pain tablets	36
2018/7/11 0:00	Anshenbunao Syrup	23.5
Total		112.5

Annex C: List of Other Figures

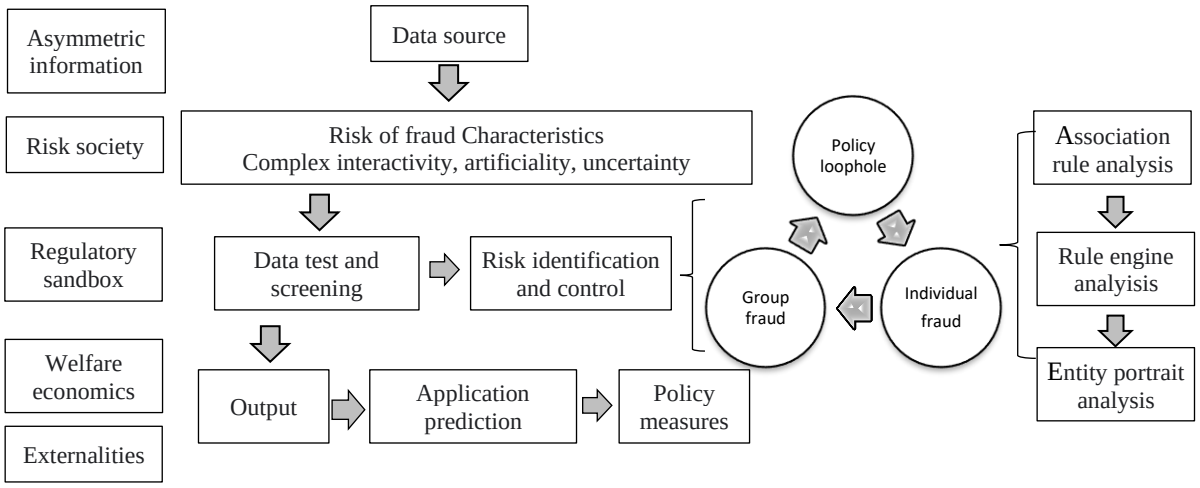
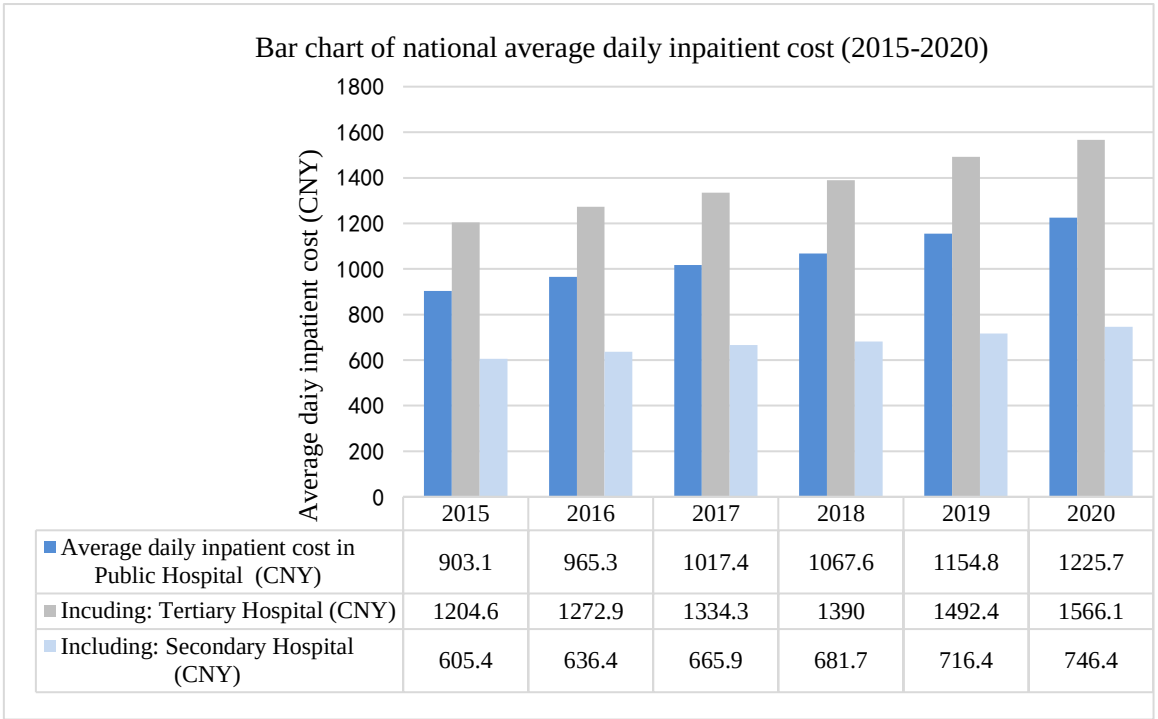


Figure C1 Medical insurance risk identification management flow chart



	2015	2016	2017	2018	2019	2020
Average daily inpatient cost in Public Hospital (CNY)	903.1	965.3	1017.4	1067.6	1154.8	1225.7
Incuding: Tertiary Hospital (CNY)	1204.6	1272.9	1334.3	1390	1492.4	1566.1
Including: Secondary Hospital (CNY)	605.4	636.4	665.9	681.7	716.4	746.4

Figure C2 Histogram of national average daily inpatient cost (2015-2020)

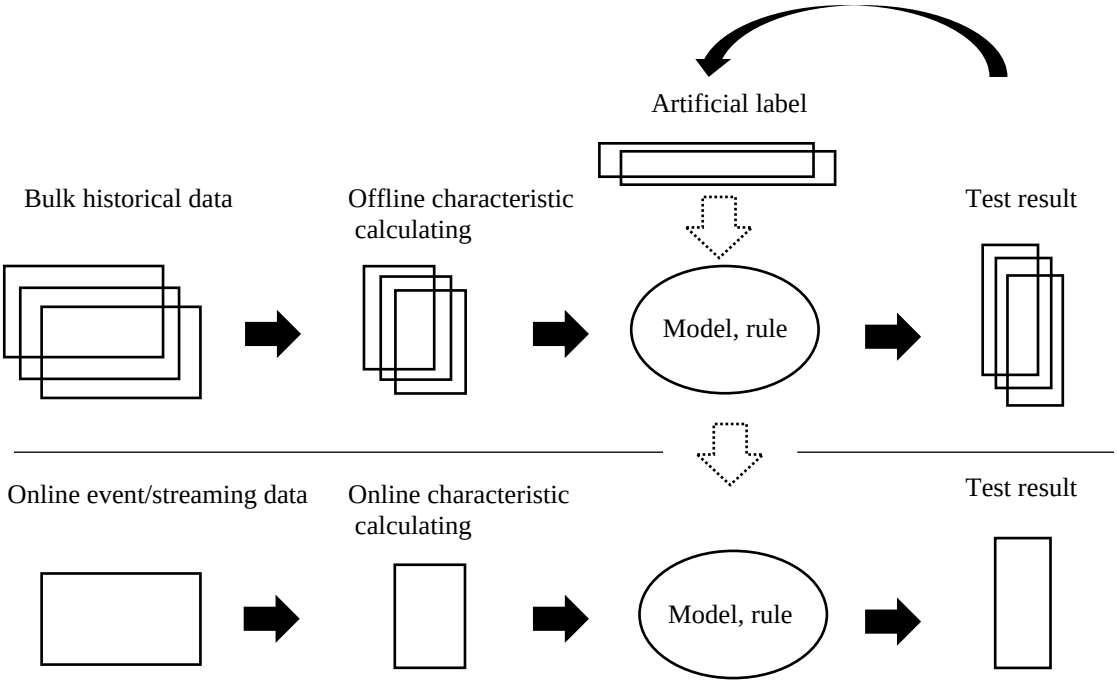


Figure C3 Basic modeling and training workflow

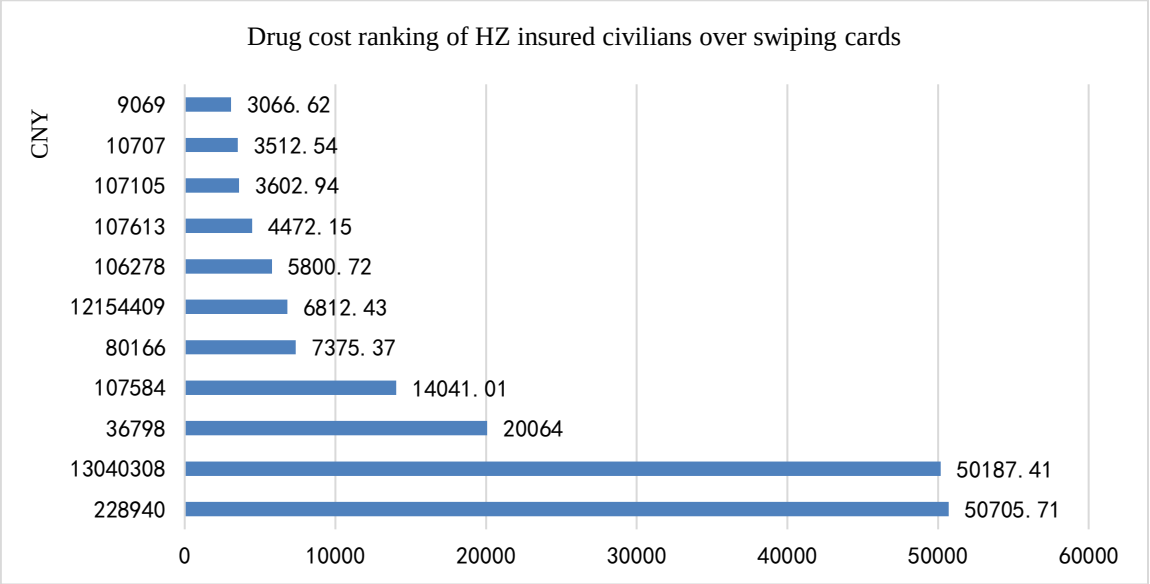


Figure C4 Drug cost ranking of HZ insured civilians over swiping cards

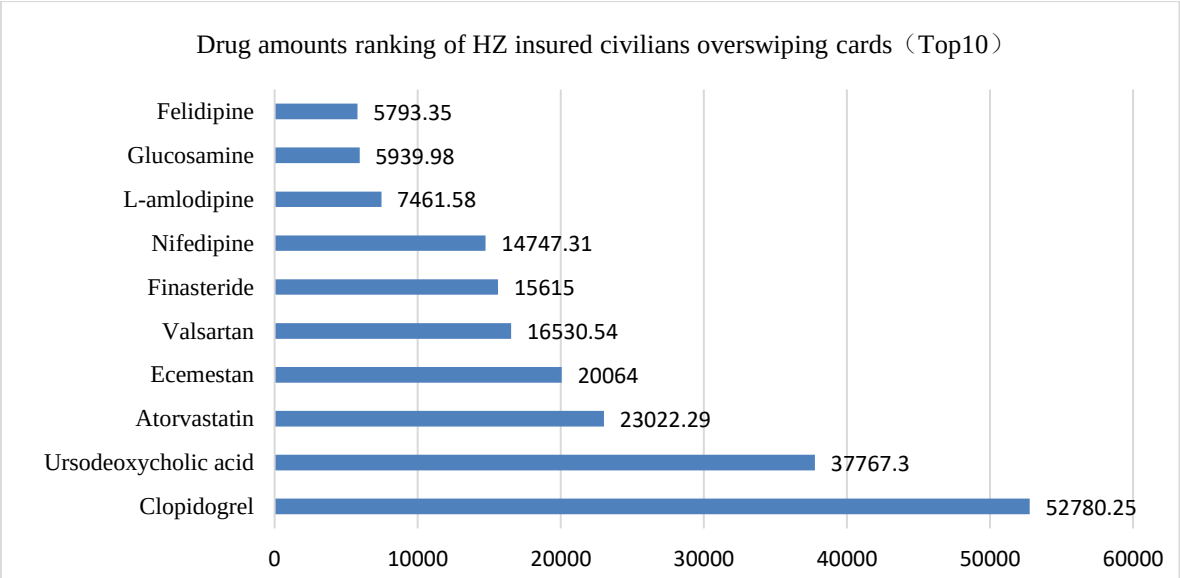


Figure C5 Drug amounts ranking of HZ insured civilians over swiping card

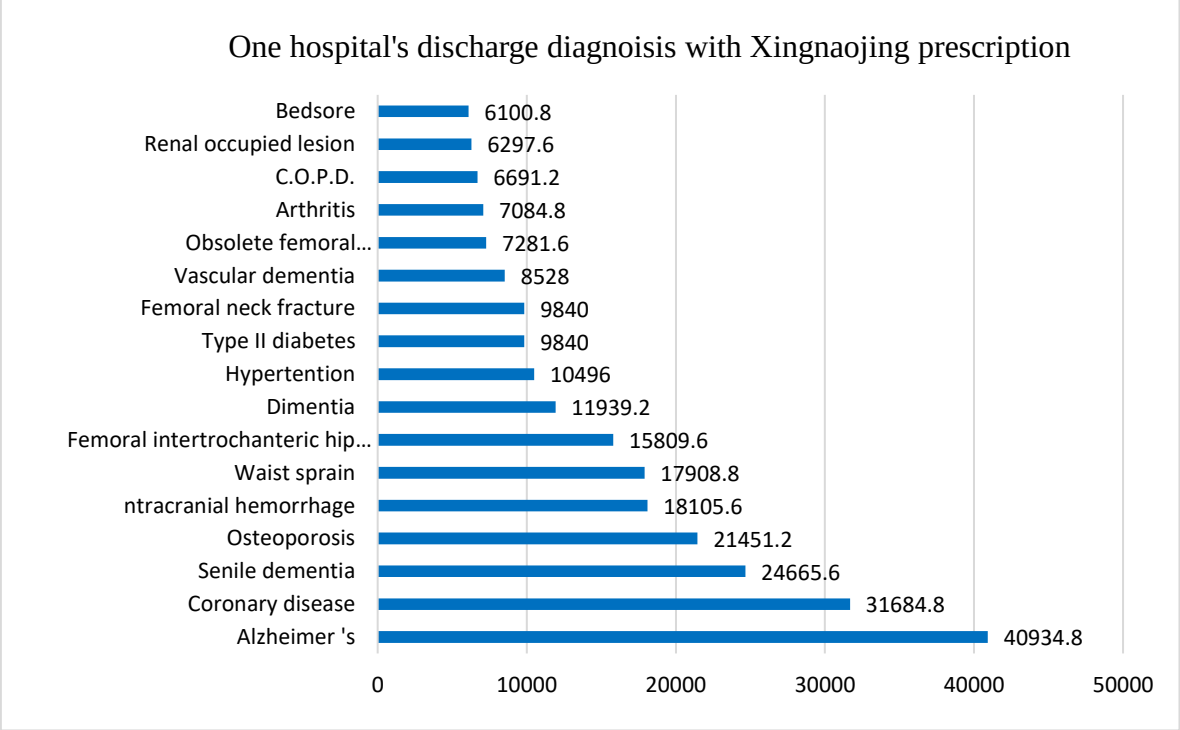


Figure C6 Adjuvant medication for off-label use

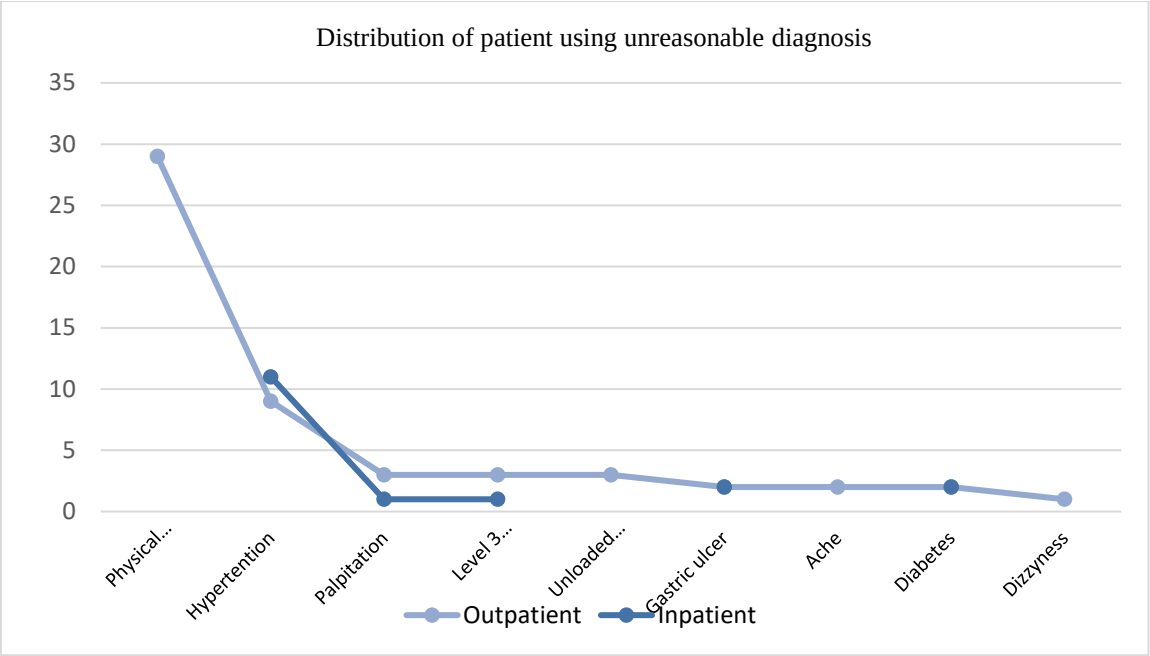


Figure C7 Off-label medication

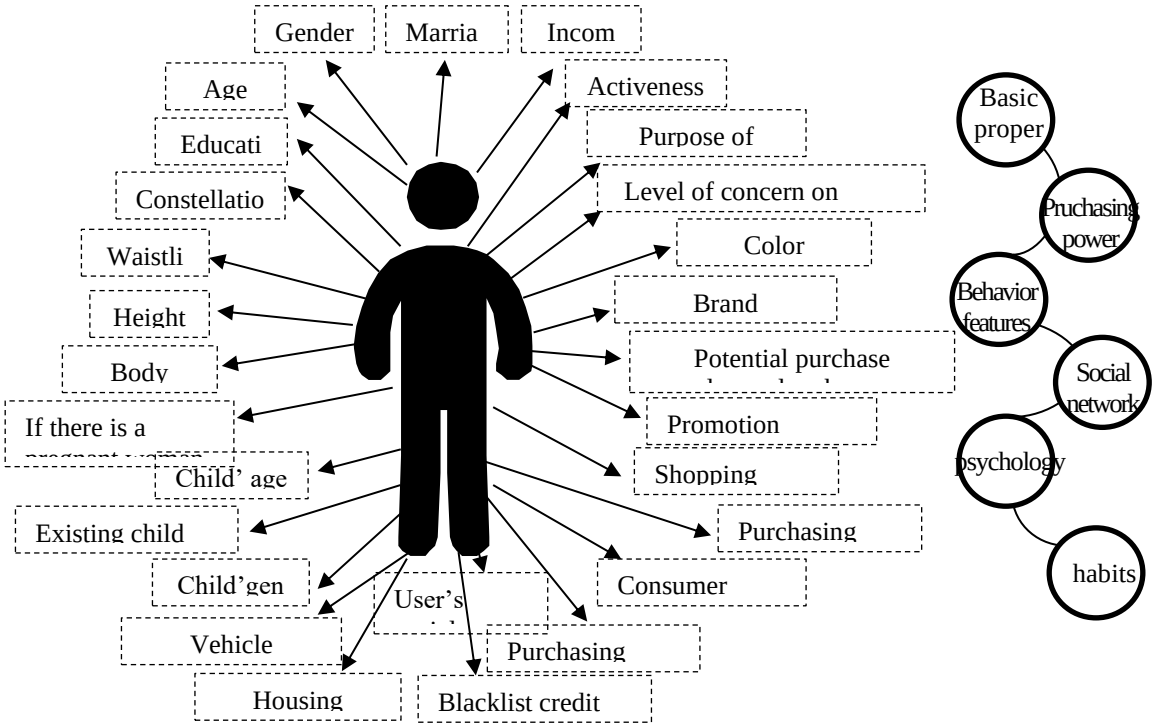


Figure C8 Entity portrait for analysis of purchasing women's shoes

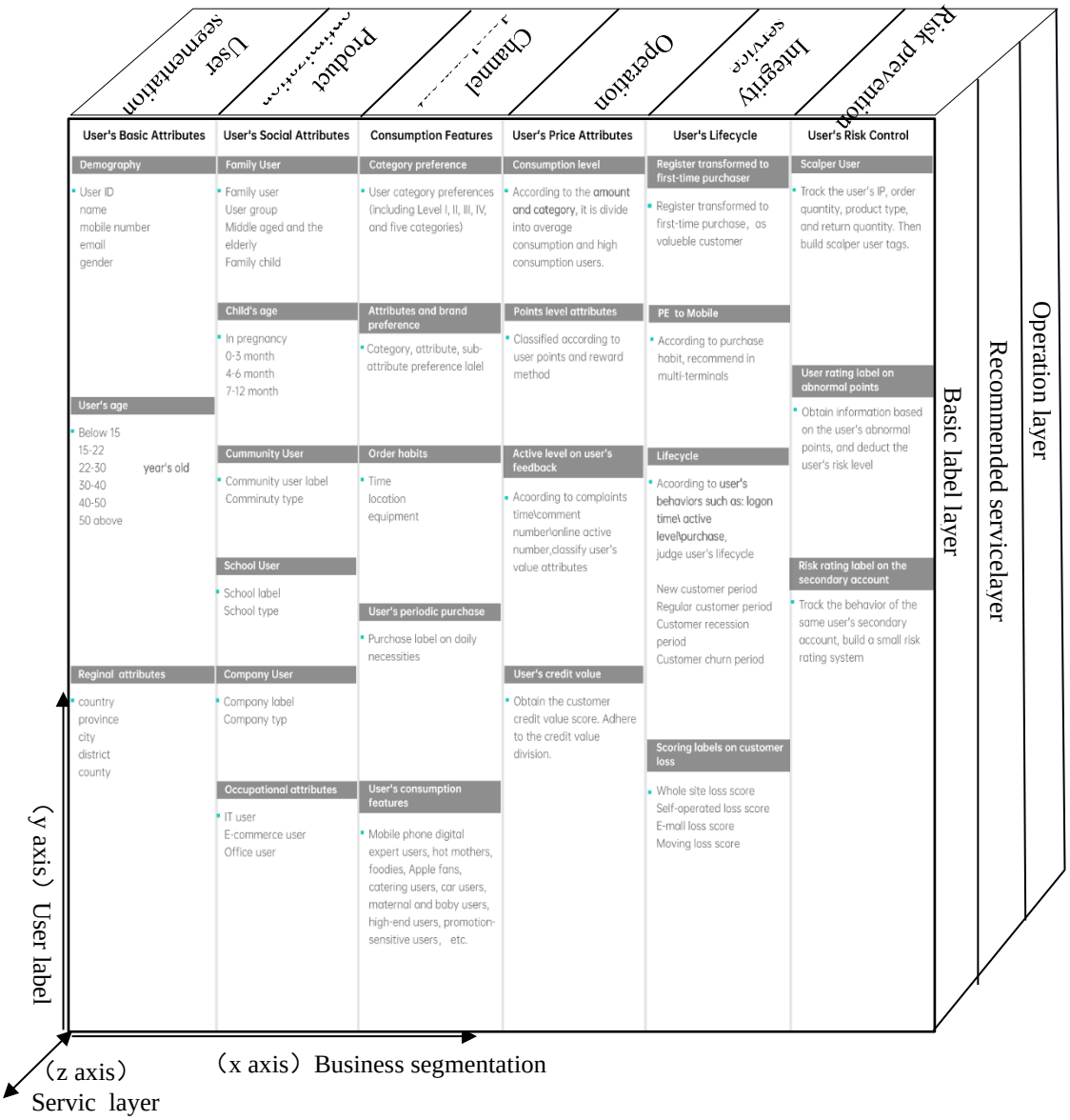


Figure C9 Entity portrait for analysis of the purchase decision of women's shoes

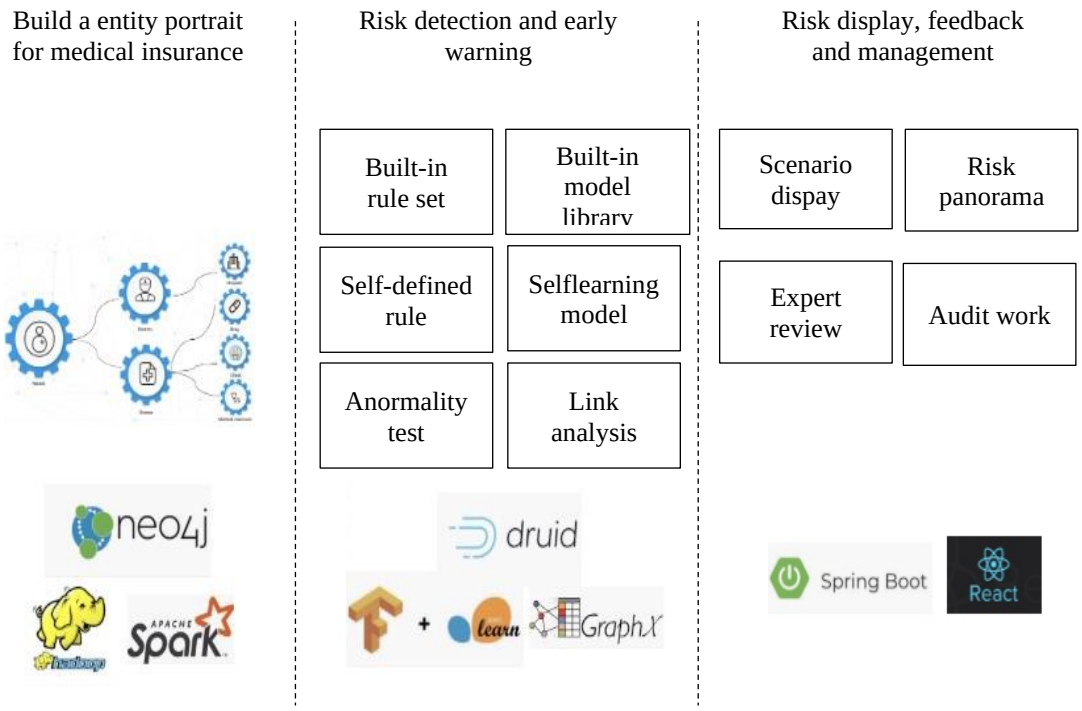


Figure C10 Entity portrait technology proactive defense self-learning model

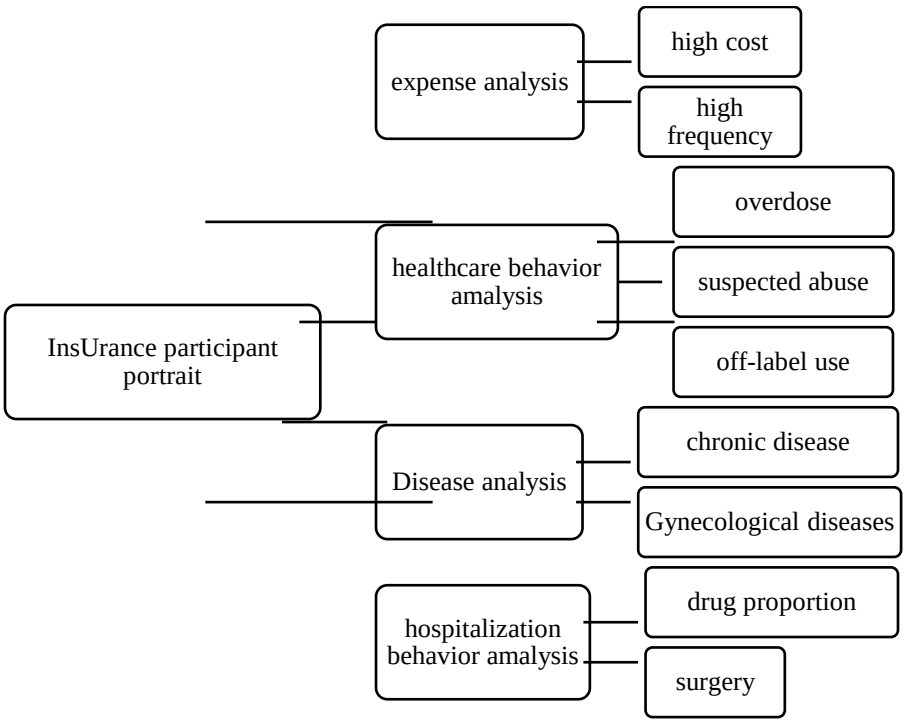


Figure C11 Label system for the insured person

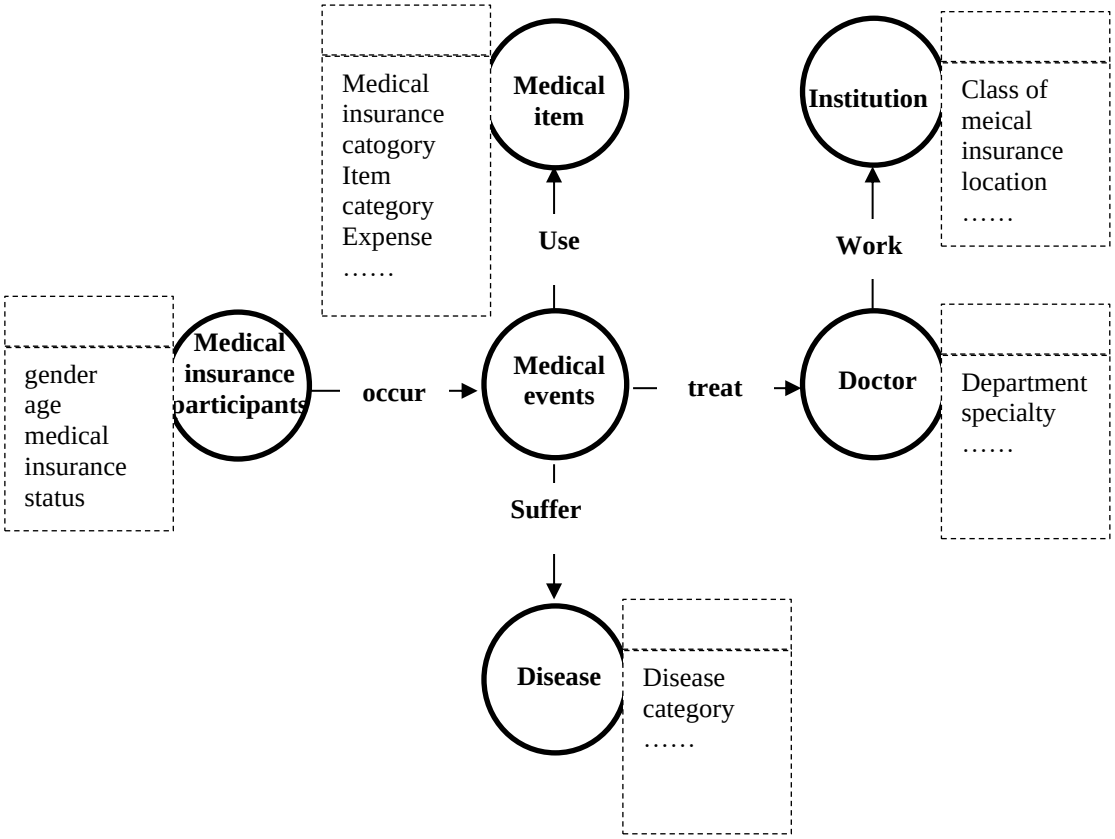


Figure C12 Entity association graph of the insured person

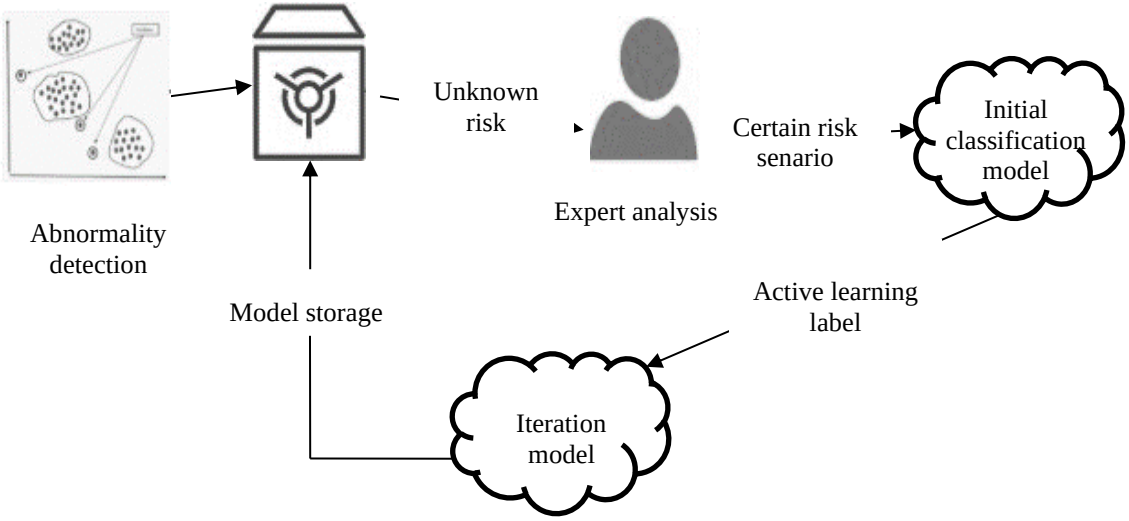


Figure C13 Modeling process of self-learning

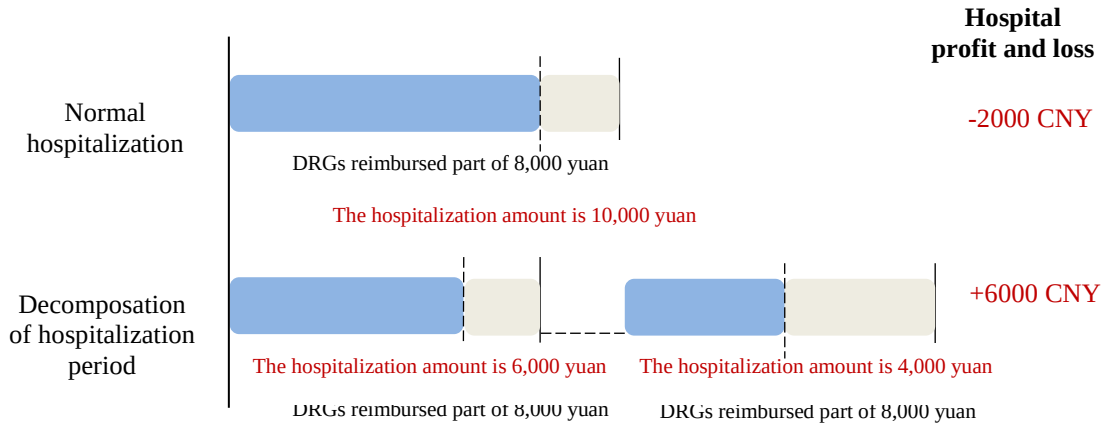


Figure C14 Illustration of decomposition of hospitalization period

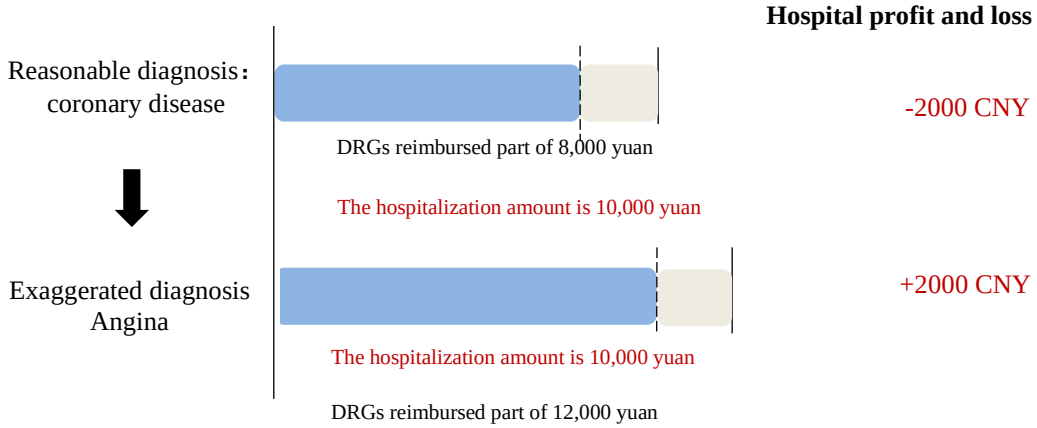


Figure C15 Illustration of low-standard condition and high-standard treatment disease group

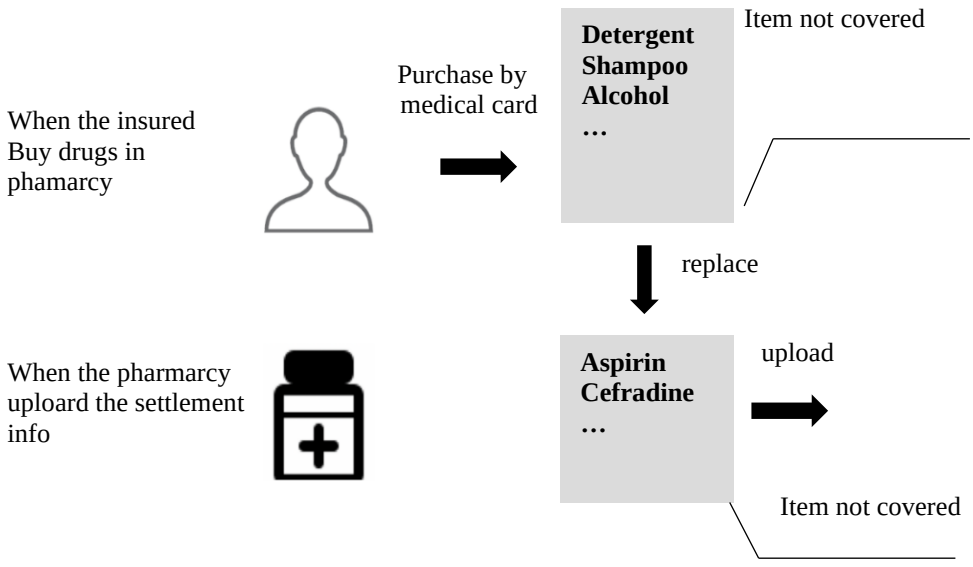


Figure C16 Swaping medicine in the pharmacy

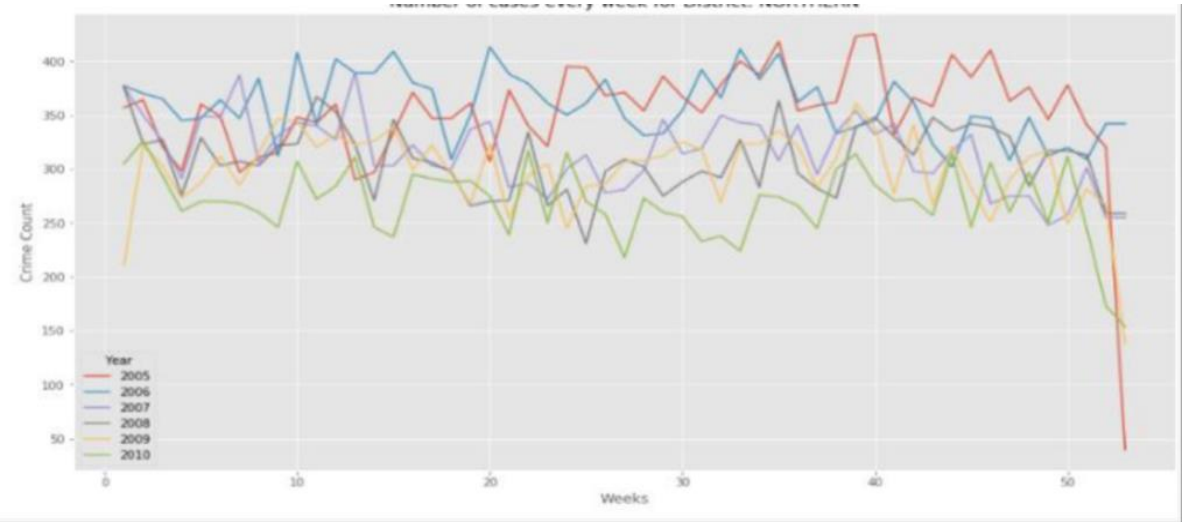


Figure C17 Frequent daily operation evolution in pharmacy

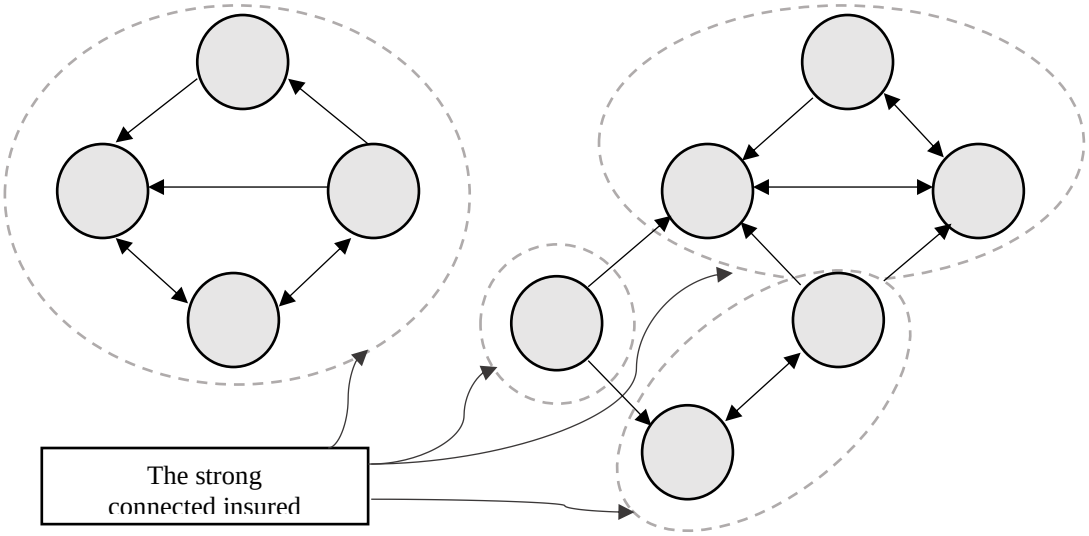


Figure C18 Complicated undirected graph of insured persons