



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Stakeholder-based Strategies for Sharing and Applying Healthcare Big Data: A Case Study on a Medical University and Its Affiliated Hospitals

FAN Geng

Doctor of Management

Supervisors:

PhD Pedro Fontes Falcão, Associate Professor,
ISCTE University Institute of Lisbon

PhD Jiang Hong, Professor,
Southern University of Science and Technology

April, 2024



BUSINESS
SCHOOL

Marketing, Operations and General Management Department

Stakeholder-based Strategies for Sharing and Applying Healthcare Big Data: A Case Study on a Medical University and Its Affiliated Hospitals

FAN Geng

Doctor of Management

Supervisors:

PhD Pedro Fontes Falcão, Associate Professor,
ISCTE University Institute of Lisbon

PhD Jiang Hong, Professor,
Southern University of Science and Technology

April, 2024

Marketing, Operations and General Management Department

Stakeholder-based Strategies for Sharing and Applying Healthcare Big Data: A Case Study on a Medical University and Its Affiliated Hospitals

FAN Geng

Doctor of Management

Jury:

PhD Inna Choban de Sousa Paiva, Associate Professor,
Iscte - Instituto Universitário de Lisboa

PhD Qian Yi, Professor,
Southern Medical University (China)

PhD António Rui Carvalho Moreira de Carvalho, Associate Professor,
ISG – Instituto Superior de Gestão

PhD Nelson Jorge Campos Ramalho, Associate Professor,
Iscte – Instituto Universitário de Lisboa

PhD Pedro Miguel Ribeiro de Almeida Fontes Falcão, Associate Professor with
Habilitation,
Iscte – Instituto Universitário de Lisboa

April, 2024

[This page is deliberately left blank.]

Abstract

Healthcare big data plays an important role in medical scientific research, medical decision-making, chronic disease prevention and control, infectious disease monitoring and early warning, medical administration, and medical and health policy formulation and evaluation, so it has become a basic consensus of all countries to actively develop and innovate the application of healthcare big data. However, healthcare big data comes from a wide range of sources with various forms and varieties, and involves multiple levels and a wide range of stakeholders. Stakeholders connect and interact with each other through competition, cooperation, conflict and confrontation, which promotes or hinders the effective sharing and application of healthcare big data.

Based on the Stakeholder Theory, with the combination of empirical and qualitative research, this research systematically explores the key influencing factors and implementation strategies of the sharing and application of healthcare big data through expert consultation, questionnaire survey, in-depth interviews and other methods. The research found that the National Health Commission, the government, the medical insurance department and hospital administrators are the dominant stakeholders in the application and sharing of healthcare big data. Also, to promote data sharing, it is necessary to start from the top-level design, improve the construction of data sharing and application systems, establish industry standards and unified platforms, ensure the investment of funds and resources, strengthen data network security barriers, and ensure data security and personal privacy. The study has thus theoretical and practical implications for the stakeholder approach to healthcare big data.

Keywords: big data, healthcare, stakeholders, data sharing

JEL: I18, M10

[This page is deliberately left blank.]

Resumo

Os megadados de saúde desempenham um papel importante na investigação científica médica, na tomada de decisões médicas, na prevenção e controlo de doenças crónicas, na monitorização e alerta precoce de doenças infecciosas, na administração médica e na formulação e avaliação de políticas médicas e de saúde, pelo que se tornou um consenso básico de todos os países a desenvolverem e inovarem ativamente na aplicação de megadados em saúde. No entanto, os megadados de saúde provêm de uma ampla gama de fontes, com diversas formas e variedades, e envolvem múltiplos níveis e uma ampla gama de partes interessadas. As partes interessadas ligam-se e interagem entre si através da competição, cooperação, conflito e confronto, o que promove ou dificulta a partilha e aplicação eficazes de megadados em saúde.

Com base na Teoria das Partes Interessadas, com a combinação de investigação empírica e qualitativa, esta investigação explora sistematicamente os principais fatores de influência e estratégias de implementação da partilha e aplicação de megadados de saúde através de consulta a especialistas, inquéritos, entrevistas aprofundadas e outros métodos. A investigação concluiu que a Comissão Nacional de Saúde, o governo, o departamento de seguros médicos e os administradores hospitalares são as partes interessadas dominantes na aplicação e partilha de megadados de saúde. Além disso, para promover a partilha de dados, é necessário começar pelo desenho de nível superior, melhorar a construção de sistemas de partilha e aplicação de dados, estabelecer padrões industriais e plataformas unificadas, garantir o investimento de fundos e recursos, fortalecer as barreiras de segurança da rede de dados, e garantir a segurança dos dados e a privacidade pessoal. O estudo tem, portanto, implicações teóricas e práticas para a abordagem das partes interessadas à temática de megadados de saúde.

Palavras-chave: Megadados, gestão da saúde, partes interessadas, partilha de dados

JEL: I18, M10

[This page is deliberately left blank.]

摘 要

健康医疗大数据在医学科学研究、医疗决策、慢性病防控、传染病监测预警、医疗行政管理、医药卫生政策制定与评估等方面具有重要作用，积极发展与创新应用健康医疗大数据已经成为各国的基本共识。然而，健康医疗大数据来源广泛、形式多样、种类繁多，涉及的利益相关者层次多、范围广。利益相关者彼此之间通过竞争、合作、冲突和对峙等方式相互联系、相互作用，促进或阻碍健康医疗大数据的有效共享与应用。

本研究以实证研究和质性研究相结合的方式，基于利益相关者视角，通过专家咨询、问卷调查、深度访谈等方法，系统性地探索了健康医疗大数据共享与应用的关键影响因素和实现策略。研究发现，卫健委、政府、医保部门和医院管理者是健康医疗大数据应用与共享的关键利益相关者。促进数据共享应从顶层设计出发，完善数据共享与应用制度体系建设、建立行业标准和统一平台、保障经费与资源投入、加强数据网络安全屏障、保证数据安全与个人隐私等。

关键词：大数据，健康医疗，利益相关者，数据共享

JEL: I18, M10

[This page is deliberately left blank.]

Acknowledgements

Time flies. After three and a half years of unremitting efforts, my dissertation writing is finally coming to an end. The bits and pieces of my doctoral career since September 2020 are tough but touching. From the decision to pursue a doctoral degree in management, to the in-depth study of various management theories and courses during the acquisition, then to the design of the thesis, the implementation of the research topic, and finally to the writing of the thesis, I have gone through many hardships, and now I finally complete my doctoral thesis. I put theories into practice and determined the research direction of my thesis by reading a large number of domestic and foreign literature under the guidance of my supervisors. From topic selection to conception, from research ideas to thesis framework, from expert consultation, questionnaire survey to in-depth interviews, from condensed views to final revision, I spent a lot of time and energy in each process, and it is inseparable from the careful guidance of my supervisors and the enthusiastic help of colleagues and students. I would like to express my heartfelt thanks to all the teachers, relatives and friends who cared, loved, helped and supported me.

First of all, I would like to express my sincere thanks to my supervisors, Professor Pedro Fontes Falcão from the University College of Lisbon, Portugal, and Professor Jiang Hong from Southern University of Science and Technology. Professor Pedro Fontes Falcão and Professor Jiang Hong have given me a lot of guidance and help in course learning, academic research and project practice, and these influences will benefit me for a lifetime. Professor Pedro Fontes Falcão and Professor Jiang Hong are knowledgeable and have profound academic foundation. Their spirit of excellence in scientific research, practical and rigorous working attitude, and interesting and enthusiastic educational style have deeply impressed me.

I would like to express my sincere thanks to Professor Virginia and Professor Nelson and other teachers, who have taught me many classic and cutting-edge medical policies and health management knowledge in and beyond China, which expanded my academic vision and enriched my knowledge system.

I would like to express my heartfelt thanks to Dean Mao Chen and Dr Zhang Xiru for their great support of this research. It is because of your dedication to help me that I was able to overcome all the difficulties and finally complete my thesis.

I would like to thank Vice president Wang Dong and Deputy Dean Zhang Chichen from the School of Health Management, Southern Medical University, Director Ou Weiyan, and the whole team for their thoughtful care during my study and valuable guidance in the process of the writing the thesis.

I would like to thank my colleagues Huang Chen, Zhao Li, Gao Tingting, Wang Chunmei, and my fellow classmates for their unselfish help during my study, and for sharing every success and joy in my work and study.

Thank my family and classmates for their love, understanding and tolerance to me as always. Your support and expectations are my inexhaustible motivation to move forward.

I would also like to thank all the experts and professors who participated in the thesis review and thesis defense for their careful guidance in their busy schedules.

致 谢

时光荏苒，岁月如梭。经过三年半的不懈努力，我的学位论文写作也终于接近了尾声。回顾自2020年9月开始的博士研究生生涯中的点点滴滴，不禁感慨万千。从决定攻读管理学博士学位，到攻取期间深入学习各种管理理论与各科课程，再到毕业论文的设计、研究课题的实施、最后到学位论文的撰写，历经千辛万苦，如今终于交上了一份答卷。将所学理论付诸实践，通过阅读大量国内外文献并在导师指导下确定毕业论文的研究方向，从选题到构思，从研究思路到论文框架，从专家咨询、问卷调查到深度访谈，从凝练观点到修改定稿，每一个过程都凝聚了许多的时间和精力，都离不开导师们的悉心指导和同事、同学们的热情帮助。谨在此衷心地向所有关心、爱护、帮助、支持我的师长和亲友表示最诚挚的感谢

首先，衷心感谢我的导师——葡萄牙里斯本大学学院Pedro Fontes Falcão教授、南方科技大学姜虹教授。Pedro Fontes Falcão教授和姜虹教授无论在课程学习上，还是在学术研究上，抑或是项目实践中都给予了我很多的指导与帮助，这些影响都将会使我受益终身。Pedro Fontes Falcão教授和姜虹教授博学多识，学术功底深厚，他们精益求精的科研精神，踏实严谨的工作态度，风趣热情的教育风格，令我由衷敬佩与感动。

衷心感谢Viginia、Nelson等授课教师，他们传授了国内外许多经典的和前沿的医疗政策与卫生事业管理方面的知识，扩展了我的学术视野，丰富了我的知识体系。

衷心感谢毛琛院长、张希如博士对本课题研究给与的大力支持，正是因为你们倾力相助，才能让我克服重重困难，最后完成论文。

感谢南方医科大学王冬副校长、张持晨副院长、欧玮艳主任以及整个项目团队在我课程学习期间给予的照顾，在论文撰写过程中给予的非常重要的指导。

感谢我的同事黄晨、赵莉、高婷婷、王春梅，还有博士班的同学们在我学习期间给予的无私帮助，感谢大家分享着我工作和学习过程中的每一份成功与喜悦，

感谢我的家人对我一如既往的关爱、理解与包容，你们的支持与期盼是我前进的不竭动力。

感谢参加论文评阅和论文答辩的各位专家教授在百忙之中给予我的精心指导。

[This page is deliberately left blank.]

Contents

Chapter 1: Introduction	1
1.1 Research background	1
1.1.1 Concept of healthcare big data	1
1.1.2 Main application fields and potential value of healthcare big data	1
1.1.3 Application potential and current situation of healthcare big data in a university and affiliated hospitals.....	5
1.2 Research problems (pain points)	7
1.2.1 Possible constraints in sharing and application of healthcare big data	8
1.2.2 Difficulties in healthcare big data sharing and application	10
1.3 Research questions	11
1.4 Research purpose and significance	13
1.4.1 Research purpose.....	13
1.4.2 Research significance	14
Chapter 2: Literature Review	17
2.1 The origin and development of Stakeholder Theory	17
2.2 The classification method of stakeholders	21
2.2.1 Multidimensional segmentation	22
2.2.2 Mitchell three-dimensional scoring approach	29
2.2.3 Research on stakeholder classification by Chinese scholars.....	35
2.3 Analytical method of stakeholder research	37
2.4 The application of Stakeholder Theory in the sharing and application of healthcare big data	41
2.4.1 The application of Stakeholder Theory in the sharing and application of genomic and bio-specimen data	41
2.4.2 The application of Stakeholder Theory in the sharing and application of medical and public health data.....	42
2.4.3 The application of Stakeholder Theory in the sharing and application of clinical trial data.....	43
2.4.4 The application of Stakeholder Theory in the sharing and application of scientific research data	43

2.5 Candidate stakeholders and interest demands of the sharing and application of healthcare big data.....	44
2.5.1 Candidate stakeholders of the sharing and application of healthcare big data	44
2.5.2 The interest appeals of candidate stakeholders of the sharing and application of healthcare big data.....	47
2.6 Concept, research process, and application of in-depth interviews	49
2.6.1 Concept of in-depth interviews	49
2.6.2 Research process of in-depth interviews	49
2.6.3 Application of in-depth interviews	57
Chapter 3: Research Method	61
3.1 Research idea.....	61
3.2 Research content	61
3.2.1 Theoretical research	62
3.2.2 The identification of key stakeholders involved in the sharing and application of healthcare big data	62
3.2.3 Research on influencing factors of the sharing and application of healthcare big data from the perspective of key stakeholders	62
3.2.4 Strategy construction of the sharing and application of healthcare big data...	63
3.3 Research methods.....	63
3.3.1 Literature review and theoretical research method	63
3.3.2 Case study method	64
3.3.3 Expert consultation.....	64
3.3.4 Questionnaire survey and expert grading method.....	65
3.3.5 In-depth interview method	65
3.4 Technical Route	66
Chapter 4: Stakeholder Identification in the Sharing and Application of Healthcare Big Data	69
4.1 Analysis of stakeholders engaging in healthcare big data sharing and application ...	70
4.2 Selection of key stakeholders in the sharing and application of healthcare big data .	71
4.3 Mechanisms of key stakeholders in promoting healthcare big data sharing and application	73
4.4 Identification of key stakeholders in the sharing and application of healthcare big data	76
4.4.1 Approach to identification.....	76
4.4.2 Identification method	77

4.4.3 Questionnaire	79
4.4.4 Statistical analysis	80
4.4.5 Survey results	81
Chapter 5: Interview Analysis of Healthcare Big Data Sharing and Application	91
5.1 Selection of interview subjects.....	91
5.2 In-Depth interview research process	93
5.2.1 Preparation for in-depth interviews.....	94
5.2.2 Implementation of in-depth interviews	95
5.2.3 Data compilation and entry	96
5.2.4 Data analysis and results synthesis.....	96
5.3 Key findings from interviews.....	97
5.3.1 Foundation and current state of healthcare big data sharing.....	97
5.3.2 Healthcare big data sharing benefits and risks	100
5.3.3 Attitudes of interviewees towards healthcare big data sharing	103
5.3.4 Problems of and recommendations for healthcare big data sharing.....	104
5.3.5 Legal guarantee for healthcare big data sharing.....	108
5.3.6 Security and privacy protection of healthcare big data sharing	109
Chapter 6: Strategies for Improving Healthcare Big Data Sharing and Application	113
6.1 Establishing a comprehensive system for healthcare data guided by policy	113
6.2 Establishing industry standards and unified platforms to promote data sharing.....	114
6.3 Ensuring funding and resource investment to support data platform construction..	115
6.4 Nurturing specialized and interdisciplinary talents to support healthcare data sharing and applications.....	116
6.5 Enhancing network and data security defense capabilities, strengthening data security	116
6.6 Advancing healthcare big data applications for sustainable development.....	117
Chapter 7: Conclusion.....	119
7.1 Main Conclusions.....	119
7.2 Research limitations and future research	121
Bibliography.....	123
Webliography	131
Annex A: Expert Consultation Form for Identifying Key Stakeholders in the Sharing and Application of Healthcare Big Data in Medical Colleges and Their Affiliated Hospitals	133
Annex B: Questionnaire for Identifying the Three-dimensional Attributes of Mitchell Stakeholders in the Sharing and Application of Healthcare Big Data	135

Annex C: Interview Outline for Healthcare Big Data Sharing and Application in a Medical College and Its Affiliated Hospitals (Government Department Manager).....	137
Annex D: Interview Outline for Healthcare Big Data Sharing and Application in a Medical College and Its Affiliated Hospitals (University or Hospital Manager).....	139
Annex E: Information Gathered During the Interviews on Healthcare Data Sharing	141

List of Tables

Table 2.1 Details of threat potential and cooperation potential proposed by Savage	27
Table 4.1 Analysis of stakeholders engaging in healthcare big data application and sharing..	70
Table 4.2 Expert consultation results on stakeholders in healthcare big data sharing and application	72
Table 4.3 Basic characteristics of e-survey participants.....	81
Table 4.4 Legitimacy dimension scores for 11 major stakeholders in healthcare data	83
Table 4.5 Descriptive statistical analysis results for power dimension scores of 11 major stakeholders in healthcare data.....	83
Table 4.6 Descriptive statistical analysis results for urgency dimension scores of 11 major stakeholders in healthcare data.....	84
Table 4.7 Paired-sample T-test results for legitimacy dimension score differences	86
Table 4.8 Paired-sample T-test results for differences in power dimension scores.....	88
Table 4.9 Paired-sample T-test results for differences in urgency dimension scores.....	89
Table 4.10 Classification results of 11 major stakeholders in healthcare big data.....	90
Table 5.1 Basic characteristics of interview subjects	92
Table 5.2 Interview of the status quo of healthcare big data sharing	100
Table 5.3 Potential benefits of sharing and application of healthcare big data	101
Table 5.4 Potential risks of sharing and application of healthcare big data	103
Table 5.5 Attitudes of healthcare big data sharing respondents	104
Table 5.6 Recommendations for healthcare big data sharing.....	106
Table 5.7 Relevant legal regulations for healthcare big data sharing.....	108
Table 5.8 Legal guarantee for healthcare big data sharing.....	109
Table 5.9 Privacy data protection and protective measures in healthcare big data sharing and application	110

[This page is deliberately left blank.]

List of Figures

Figure 2.1 Stakeholder classification method proposed by Frederick	24
Figure 2.2 Stakeholder classification method proposed by Wheeler	26
Figure 2.3 Results of stakeholder qualitative classification based on Mitchell scoring method	33
Figure 3.1 Technical route of this research	67
Figure 5.1 Composition of healthcare big data	98
Figure 5.2 Keyword analysis of information system for healthcare big data sharing and funds sharing	99
Figure 5.3 Keyword analysis of problems of and recommendations for healthcare big data sharing	104
Figure 5.4 Problems of healthcare big data sharing	105

[This page is deliberately left blank.]

List of Abbreviations

Hospital Information System, HIS

Laboratory Information System, LIS

Electronic Medical Record System, EMRS

Artificial Intelligence, AI

Coronavirus Pandemic Epidemiology, COPE

Personal Health Data, PHD

[This page is deliberately left blank.]

Chapter 1: Introduction

1.1 Research background

1.1.1 Concept of healthcare big data

Healthcare big data refers to all the information related to health status, disease and diagnosis and treatment generated throughout the life course, which constitutes a complete, reliable, comprehensive and systematic information base, mainly divided into clinical data generated in the process of hospitalization, outpatient and physical examination and a series of non-clinical data generated in the process of daily activities, sleep, exercise and other processes (Song et al., 2021; Yu et al., 2014; L. X. Zhang et al., 2018).

The rapid development of emerging information technologies such as the Internet of Things, the Internet, medical informatics, cloud computing, mobile intelligence and bioinformatics has provided unprecedented opportunities for the accumulation and application of healthcare big data.

1.1.2 Main application fields and potential value of healthcare big data

The huge application value behind healthcare big data is being widely accepted and recognized by all walks of life, and the active development and innovative application of healthcare big data has become an important consensus of governments and relevant departments around the world (General Office of Guangdong Provincial People's Government, 2021; General Office of the State Council of the People's Republic of China, 2016; H. Q. Li et al., 2019; S. S. Ma et al., 2018; Martin et al., 2014; Packer, 2018; Pisani et al., 2016).

Healthcare big data has shown an important role and a prospective future in medical scientific research, medical decision-making, prevention and control of chronic diseases, monitoring and early warning of infectious diseases, medical administration, formulation and implementation of medical and health policies, and “Internet plus healthcare” services (Armstrong, 2017; Berg, 2017; Damiani et al., 2018; Elliott et al., 2015; Han et al., 2020; Kazanjian, 1998; Lusher et al., 2014; Martin et al., 2014; Patrick, 2016; Pisani et al., 2016; L. R. Wu & Zeng, 2019; T. Xu & Yu, 2020).

1.1.2.1 Application and value of healthcare big data in clinical treatment

(1) Clinical research

Information is integrated by hospital information system (HIS), laboratory information system (LIS), electronic medical record (EMR) system and other systems of different medical institutions. Besides, major scientific research information databases such as single-disease databases and biological sample databases are linked to carry out clinical effect evaluation of diagnosis and treatment plans, and explore disease occurrence and development, new drug R&D, vaccine research and medical equipment R&D. That can provide strong support for the comprehensive treatment plans for major intractable diseases and frequently-occurring chronic diseases, the formulation of diagnostic guidelines, promoting the development of key characteristic medical technologies and the rapid validation and clinical promotion of cutting-edge medical scientific and technological achievements (Elliott et al., 2015; Martin et al., 2014; Packer, 2018; Song et al., 2021; L. X. Zhang et al., 2018).

(2) Precision medicine

According to the living environment, lifestyle, clinical symptoms and signs, pathophysiology features, medical imaging atlas and other patient information, accurate diagnosis and treatment services are provided. The application of genome, proteome, transcriptome and other multi-omics technologies can accurately determine the potential intervention targets of various diseases and the response degree of populations with different characteristics to the same intervention measures, so as to realize the individualization of disease prevention and treatment programs, and meet the diversified and multi-level medical and health needs of different populations due to differences in genetic and environmental factors (Armstrong, 2017; Song et al., 2021; L. X. Zhang et al., 2018).

(3) Support for clinical decision-making

The new generation of Artificial Intelligence (AI) technology, such as natural language information processing and machine learning, is used to extract the common characteristics, diagnosis basis and treatment plan of patients with the same type of disease from the imaging atlas, inspection information and electronic medical record data generated by previous medical practice, develop AI-assisted diagnosis technology, and establish clinical decision-making support system, so as to guide clinical decision-making accurately, timely and effectively (Elliott et al., 2015; Malykh & Rudetskiy, 2018). The clinical decision-making and treatment process is no longer solely dependent on the doctor's clinical experience and knowledge reserve, thus improving the speed and accuracy of clinical decision-making in different regions,

especially in remote areas (Song et al., 2021; L. X. Zhang et al., 2018).

(4) Hospital digital services

It is suggested that we develop multi-center smart healthcare services to benefit people, and promote the construction and optimization of online appointment and triage, electronic payment, medical insurance instant settlement, online inquiry and download of inspection and examination reports, online health consultation and prescription issuance, online inquiry of medical files, online follow-up of patients and clinical trial participants, and other services through mini programs and apps, so as to form a standardized and efficient disease diagnosis and treatment and health management with mutual sharing and trust (Fan et al., 2016; S. S. Ma et al., 2018).

1.1.2.2 Application and value of healthcare big data in public health

(1) Infectious disease surveillance and emergency response

In the face of the COVID-19 pandemic, American scholars have established Coronavirus Pandemic Epidemiology (COPE) Consortium for ordinary people to obtain information on epidemic prevention and control (Chan et al., 2020). China uses health codes, pass codes and information reporting systems of infectious diseases to provide strong support for outbreak surveillance and precise prevention and control. Besides, the information sharing and business collaboration among institutions in centers for disease control and prevention at all levels, import and export ports, inspection and quarantine departments, and hospitals contribute to timely access to distribution characteristics and changing trends of both major and emerging infectious diseases, helping to improve the level of emergency response and comprehensive handling capacity of relevant organizations and social groups to major public health emergencies (Berg, 2017; Song et al., 2021).

(2) Discussion on etiological mechanism of chronic non-infectious disease and evaluation of prevention and treatment effects

The collaborative sharing of multiple data resources such as the common chronic disease monitoring system, the cause of death monitoring system, the diagnosis and treatment system of hospitals and community health service centers at all levels, special disease cohorts, specific population cohorts, and general population cohorts, works together with molecular epidemiological analysis results and meteorological environmental data to contribute to comprehensive discussion on the influence of biological, genetic, and environmental factors on the occurrence and development process of major chronic non-infectious diseases, which provides high-quality data support for the formulation of various disease prevention and

intervention measures and the evaluation of prevention and control effects (Z. M. Chen et al., 2011; Littlejohns et al., 2020; Sullivan et al., 2019; Yu et al., 2014).

1.1.2.3 Comprehensive application and value of healthcare big data

(1) Assisted administration and management decision-making

The information interaction and sharing system of healthcare big data has functions including communication, business management, and timely supervision and therefore optimizes work processes and improves the efficiency of daily work and management. In terms of medical and health policies, the system can provide a scientific and effective basis for decision-makers in the medical industry to make policies and supervise the implementation by systematically combing, gathering and mining healthcare data resources covering different medical levels and different specific business fields as well as online public opinion information of major events, and using multidimensional mathematical models and algorithms for comprehensive analysis, thus helping optimize the allocation of public medical resources, improve the framework design of the medical security system, and safeguard the life and health of the public (S. S. Ma et al., 2018T.; Song et al., 2021; T. Xu & Yu, 2020).

(2) R&D of telemedicine and intelligent consultation system

We should use AI as a tool to encourage higher-level medical institutions to carry out innovative services such as remote consultation, surgical demonstration, training and teaching for primary medical service institutions, in a bid to the rapid extension and sinking of the strong diagnosis and treatment capabilities, scientific research capabilities and teaching capabilities of top institutions to primary institutions, and therefore enhance the comprehensive strength of primary medical institutions. Based on massive electronic medical record data and AI technology, we can develop a smart doctor assistant and an intelligent consultation platform to provide a more convenient and practical electronic service system for disease screening, diagnosis, treatment and follow-up of patients in remote areas with shortage of medical resources (Elliott et al., 2015; Han et al., 2020; Song et al., 2021).

(3) Cultivating new formats for healthcare industry

Based on the current situation of healthcare and the demand for high-quality medical services of Chinese residents, we can build a complete medical service industry chain that integrates modern medical technology with various health elements such as health maintenance, elderly care, community nursing, family medical services, leisure, entertainment, and sports, so as to promote the rapid growth of social economy and the continuous improvement of residents' happiness and satisfaction.

1.1.3 Application potential and current situation of healthcare big data in a university and affiliated hospitals

1.1.3.1 Application potential of healthcare big data in a university and affiliated hospitals

The medical university and its affiliated hospital involved in this research is located in Guangdong Province which has strong economic power, complete support policies for healthcare big data research and application (General Office of Guangdong Provincial People's Government, 2017, 2020, 2021), the most 5G base stations in China, well established medical systems, rich medical resources, numerous scientific research platforms and a good ecological environment. As the largest provincial economy, Guangdong has ranked first in China in terms of economic aggregate for 31 consecutive years (General Office of Guangdong Provincial People's Government, 2021; X. R. Ma, 2021), with three special economic zones and the Economic Open Zone of Pearl River Delta. Guangdong's innovative economy has a large scale, and its growth rate is among the highest in China. In 2019, high-end electronic information manufacturing grew by 8.8%, high-performance medical devices and biomedicine by 7.2%, and computer, electronic communication and other electronic equipment manufacturing by 7.4% in Guangdong. Driven by the Guangdong-Hong Kong-Macao Greater Bay Area, a world-class bay area, and the Shenzhen pilot Demonstration Zone, Guangdong has outstanding advantages in industry, technology, platform, talent and capital accumulation.

In Guangdong, this university is the only medical university that was co-built by China's Ministry of Education, National Health Commission and the province, and also a high-level university with key construction. After nearly 70 years of development, the university has formed a discipline system with medical science as the main body, and other coordinated disciplines including science, engineering, literature, management, law and economics. With strong talents, the university has over 100 national and provincial research platforms including national key laboratories, key laboratories of Ministry of Education and national clinical research centers. In terms of medical big data and digital healthcare, it has a national institute of health data science and a digital diagnostic and therapeutic equipment engineering center of the Ministry of Education, an electronic data evidence service engineering research center of Guangdong, a deep mining engineering research center of medical image big data, a medical big data integration and application engineering research center and other platforms.

The university has 13 directly affiliated hospitals located in Guangzhou, Shenzhen, Foshan, and other regions. They form as a medical group pattern with the Pearl River Delta as the core, radiating the east, west and north of Guangdong and serving the GBA. In those hospitals, a total

of 14,000 beds are deployed, the annual diagnosis and treatment number exceeds 17 million, the annual number of discharged patients is nearly 530,000, and the number of operations for discharged patients exceeds 210,000. There are 27 national key clinical specialties, 35 high-level key clinical specialties of Guangdong, 79 key clinical specialties of Guangdong, 15 medical quality control centers of Guangdong, and eight national specialized training pilot bases of standardized training for specialists. In addition, the university has 48 non-directly affiliated hospitals, of which more than 87% are tertiary first-class hospitals. It also has many educational practice bases represented by centers for disease control and prevention at all levels and occupational disease prevention and control institutes of Guangdong.

With advanced big data integration and analysis technology, the in-depth mining of the rich healthcare big data plays an important role in effectively promoting medical research, improving diagnosis and treatment capabilities and medical services, and stimulating medical innovation and scientific achievement transformation.

1.1.3.2 Current situation of healthcare big data resource sharing and application in a university and affiliated hospitals

Although the university has accumulated a large amount of multi-dimensional healthcare big data, the data is basically still distributed in different units, departments, systems, and platforms, failing to realize data sharing or effective integration. Meanwhile, since different teams have adopted advanced technologies in in-depth mining of the data accumulated so far, and have made certain achievements (Gao et al., 2022; Z. H. Li et al., 2020; Z. Li et al., 2023; Liu et al., 2022; P. D. Zhang et al., 2022), but it is far from playing the full value and potential of the data.

Specifically, there are few high-impact achievements based on multi-center randomized controlled clinical trials with large scale, different types of population cohorts, and multi-omics sequencing techniques. The application of AI based on high-quality images and pathological data has not fully achieved breakthroughs and substantial progress in the building of auxiliary diagnostic systems. Monitoring data of various major infectious diseases and chronic non-infectious diseases, data of occupational disease prevention and control, and data of key specialties such as orthopedics and stomatology are still scattered in the internal departments of related units. The industrialization of healthcare big data and the transformation of research results are still at a low degree. Currently the intelligent diagnosis and treatment system successively established by affiliated hospitals has only realized online appointment and triage, inspection and examination report inquiry, electronic payment and real-time medical insurance settlement within the hospital, but the intelligent diagnosis and treatment systems of different

affiliated hospitals have not been effectively linked, resulting in the inability to communicate and share the inspection and examination results and diagnosis and treatment plans of the same patient among different affiliated hospitals. The follow-up visits of patients in affiliated hospitals, especially those with high risk of recurrence such as stroke and tumor and those with chronic diseases accompanied by serious complications, have not basically shown an intelligent and standardized pattern.

1.2 Research problems (pain points)

In recent years, the academic researches on big data sharing and application of healthcare mainly focus on the establishment of data sharing systems, the formulation of data sharing and application standards, and the research and development of sharing and application technologies. Researchers in related fields have tried to start from the underlying design and facilitate the aggregation of data from different sources by formulating standardized data structure standards, semantic standards, and information exchange standards (Costeloe et al., 2018; Halamka, 2006; Koutkias, 2019). The research and development of distributed storage, blockchain technology and other technologies aim to realize the sharing and application of electronic medical data through technology, while focusing on data security and patient privacy protection (Mackey et al., 2019; Rowhani-Farid et al., 2017; Zheng et al., 2019). The development and application of AI-based data mining technology and algorithm, and the development of visual analysis technology are expected to maximize the mining and presentation of the core value and extended value of healthcare big data. To a certain extent, the development and progress of technologies has promoted data sharing and application (Nahar et al., 2013; Obenshain, 2004). However, the sharing and application of healthcare big data still faces severe constraints and huge challenges.

Currently, medical data across different hospitals nationwide are relatively independent, lacking integration and sharing among hospitals. Even within the same medical institution, achieving data interoperability and sharing poses a significant challenge. For example, the information systems of the 13 affiliated hospitals under Southern Medical University operate independently, making it difficult to achieve data interoperability and sharing. This results in increased costs for patients seeking treatment across multiple hospitals and complicates collaboration among these hospitals.

1.2.1 Possible constraints in sharing and application of healthcare big data

The author has systematically reviewed and sorted out relevant literature on the sharing and application of healthcare big data, and summarized several possible factors for low effectiveness in that respect.

1.2.1.1 Cost returns and incentive mechanism of sharing

The construction process of sharing and application of healthcare big data needs to configure professional and complete information software and hardware systems and other infrastructure, and the relevant infrastructure needs to be regularly maintained and upgraded in the future (S. S. Ma et al., 2018). It means that in the sharing and application process of healthcare big data, a large number of human resources, material resources and financial resources are continuously invested. However, there is currently no effective solution to protect the interests and returns of data providers and system builders of data sharing and application. The high input cost of healthcare data sharing and application, the long duration and the slow return pace, as well as the lack of perfect incentives, compensation, and guarantee mechanisms for healthcare big data sharing and application may lead to low enthusiasm for data sharing and application entities (Elsayed & Saleh, 2018; Zhou et al., 2021).

1.2.1.2 Data ownership

The issue of data ownership also hinders data sharing and application to a certain extent. Clinical data such as inspection and examination results, electronic medical records and diagnosis and treatment plans recorded in the hospital information system are generated by patients in the process of medical treatment or physical examination, and these data often involve patients' personal privacy, which should be owned by individuals and protected, and should not be used and spread without the informed consent of patients. However, these data are manufactured, managed, and stored by relevant medical personnel, and to a certain extent, it can also be considered to belong to the hospital. When third-party institutions such as scientific research institutes and data sharing system providers standardize and integrate all kinds of medical information, they naturally become new data producers, and it seems that they should also have the ownership of such non-original data. At present, there is no clear quantification and judgment standard for what kind of subjects should be given the attribution right of healthcare big data (Elsayed & Saleh, 2018; S. S. Ma et al., 2018; K. Wu & Ren, 2017; J. N. Zhang et al., 2020; Zhou et al., 2021). In addition, with the development of science and technology, operators of wearable devices, sports products, diet and health-related apps have

all recorded a large amount of information related to individual health or disease (Elliott et al., 2015). The ownership and attribution of these data are currently unclear and there are certain legal implications. The ownership relationship of the data is not clear, which can easily lead to a series of ownership disputes in the subsequent development and utilization. What is even more worrying is that even if disputes arise when the concept of healthcare big data ownership is still vague, it is difficult to reasonably define the rights and responsibilities of different participants, which also means that the personal privacy and data security of the collection is difficult to be effectively protected (S. S. Ma et al., 2018; J. N. Zhang et al., 2020).

1.2.1.3 Interest competitions among sharing entities

The main entities of healthcare big data sharing and application may often have significant interest competitions. On the one hand, managers in some medical institutions, especially the medical institutions with strong comprehensive strength, believe that healthcare big data is generated in the process of daily business, clinical diagnosis and treatment and scientific research within their units, and it should be “privately owned by institutions”. They regard the sharing of healthcare big data, especially the core data related to daily business and departmental functions, as the abandonment of their own power or the loss of key resources, so they will negatively treat or even resist the sharing and application of healthcare big data. On the other hand, the diagnosis and treatment plans contained in the electronic files of medical institutions are often related to the core competitiveness of hospitals, and may also involve the personal privacy information of patients. In order to safeguard the overall interests of medical institutions and reduce the potential disputes caused by the disclosure of patient privacy, some managers are unwilling or afraid to share data (S. S. Ma et al., 2018; Yu et al., 2014).

In addition, the diagnosis and treatment data and research data accumulated by different hospitals, colleges and clinical departments, as well as the monitoring data of cooperative units, are often involved in the signing of result publishing, guideline formulation, and patent authorization, as well as authorship and benefit distribution in the transformation of research achievements. The lack of relevant policies and systems and clear distribution schemes with rationalization and standardization has made data owners often “work by themselves” and fail to achieve effective data sharing and joint application.

1.2.1.4 Patient privacy and willingness to share

In the process of sharing and application of healthcare big data, the widespread dissemination and use of data has led to a significant increase in the possibility of privacy disclosure of the people whose data are collected (S. S. Ma et al., 2018; L. X. Zhang et al., 2018). According to

the guidelines of medical ethics, in order to protect the privacy of patients, different organizations should obtain the informed consent of patients before using the relevant data. In previous research models, informed consent was usually signed for the conduct of a scientific study without informing participants of what data might be collected and reused in the future (S. Bull & Bhagwandin, 2020; McKeown et al., 2021; Parasidis et al., 2019; Room, 2004; L. X. Zhang et al., 2018). However, with the development of informatization and AI technology, the mining of the potential value of healthcare big data is increasingly inclined to aggregate the data of different specific populations to obtain representative large data sets and produce more convincing conclusions, or use AI algorithms to generate corresponding conclusions based on the relationship among different data sets, or use datasets from different sources for cross-validation to evaluate the extrapolation of research results. In this case, the patient data collected by the researcher is no longer just for a single scientific study, but can be reassembled and reused at any time, but the researcher cannot predict which data in the dataset may be reused in subsequent studies (Mittelstadt & Floridi, 2016), which challenges the previous informed consent model to protect patient privacy.

In addition, multi-omics data such as sociodemographic characteristics, family history, past diseases and health information, as well as proteomic, transcriptomic, and genomic data based on biological samples, inevitably involve multiple levels of privacy information to a large extent (S. S. Ma et al., 2018; Yu et al., 2014). The disclosure of privacy information such as past diseases, health status, and family history may lead to a series of problems such as discrimination in daily work and social life, inability to purchase medical insurance, and even unemployment, which may have a potential adverse impact on the patient's life and property, while the disclosure of multi-omics information with ethnic health characteristics may harm the interests of the public and even endanger national security. In the process of sharing and integrating data collected by different institutions to form large-scale healthcare big data, the contact group of data has increased. The academic value, commercial value and social value of the big data generated by the convergence increase, and the convergence of multi-party data may lead to the failure of the previously set privacy protection model such as data desensitization and anonymization, all of which greatly increase the risk of privacy disclosure of the collected data (S. S. Ma et al., 2018; H. H. Wang et al., 2017).

1.2.2 Difficulties in healthcare big data sharing and application

To sum up, the author has found that the problems such as high input cost, slow return pace, unclear data ownership, interest competitions among sharing entities, insufficient patient

privacy protection and lack of willingness to share among data holders exist in the sharing and application of healthcare big data, indicating that in the process of healthcare big data sharing and application, there are many stakeholders with different interest demands, and the interest demands of different stakeholders have not been effectively balanced and satisfied, and the driving force for data sharing and application remains weak.

1.3 Research questions

The Stakeholder Theory is developed in the process of exploring the transformation of enterprise management mode. This theory holds that the operation and development of enterprises are jointly participated and promoted by various stakeholder groups, and stakeholders can directly or indirectly affect the realization of the established strategic goals and business benefits of enterprises, and can also be affected by the operation and development process of enterprises. The Stakeholder Theory emphasizes that the ultimate goal of enterprise operation and development should be to provide specific benefits for various types of stakeholder groups within the enterprise, rather than simply to guarantee and safeguard the interests of shareholders (R. E. Freeman, 1984; R. E. Freeman & Reed, 1983; Jawahar & McLaughlin, 2001; Zinkhan & Balazs, 2004).

In the process of establishment, development, operation and decision-making, enterprises need to fully consider the interests and requirements of different types of stakeholders, and try to coordinate and meet their demands. Different types of stakeholders have different or even conflicting interest demands, and their degree of realization of interest demands and satisfaction with enterprise operation and management are also different, which will have a very important impact on the realization of enterprise strategic objectives and the improvement of business performance (M. B. E. Clarkson, 1995; Costa & Menichini, 2013; Kenny, 2013).

The sharing and application process of healthcare big data are often influenced by many stakeholders. Different types of stakeholders have different interest demands, and there are interest competition and conflict among them. Different types of stakeholders also have significant differences in the realization of interest demands and satisfaction in the process of data sharing and application. Without satisfying and balancing the interests of stakeholders, it is difficult or even impossible to generate effective driving forces to carry out the sharing and application of healthcare big data (Mitchell et al., 1997).

There are many stakeholders in the sharing and application of healthcare big data, who connect and interact with each other through competition, cooperation, conflict and

confrontation, thus promoting or hindering the sharing and application of healthcare big data. If the exploration is based on all stakeholders, it can often be difficult or even chaotic to coordinate the interest demands of all stakeholders (R. E. Freeman et al., 2010; R. E. Freeman & Evan, 1990; X. L. Xu, 2018).

Different types of stakeholders vary in such aspects as status, value, role and rights while participating in the sharing and application of healthcare big data. Some stakeholders can play a key and decisive role in the effective sharing and application of healthcare big data, yet some stakeholders have a minor impact (Luo & Jiang, 2011; Mitchell et al., 1997; J. H. Wu et al., 2019).

In different times, the value, status, role and characteristics of the same type of stakeholders are not unchanged, but change dynamically with the progress of the practice process. The effective driving for the sharing and application of healthcare big data depends on the interaction and synergy among dominant stakeholders in the current era and healthcare context (Mitchell et al., 1997; L. Wu & He, 2005).

In the current practice of the sharing and application of healthcare big data, the accurately identifying dominant stakeholders that affect data sharing and application, eliminating possible conflicts and frictions between dominant stakeholders, integrating dominant stakeholders together organically, satisfying and coordinating the interests of dominant stakeholders as well as pushing all dominant stakeholders to actively communicate and collaborate in order to realize the common goal of sharing and application of healthcare big data will all effectively driving the sharing and application of healthcare big data.

This study proposes the following assumptions:

(1) The insufficient driving force for the sharing and application of data caused by the unsatisfied and unbalanced interest demands of stakeholders of healthcare big data is the key factor that makes it difficult to realize effective data sharing and application.

(2) The core of improving the driving force of data sharing and application lies in identifying the dominant stakeholders of healthcare big data, exploring the core demands of dominant stakeholders through in-depth interviews, and formulating corresponding strategies to satisfy and balance the interest demands of the dominant stakeholders.

Based on that, the researchers raised the following questions:

(1) Who are the key stakeholders in the sharing and application of healthcare big data, and what are their attitudes towards it?

(2) From the perspective of key stakeholders, what are the barriers to the sharing and application of healthcare big data?

(3) What strategies should be adopted to overcome the current challenges in order to achieve data sharing and application, and realize the intrinsic value of healthcare big data?

1.4 Research purpose and significance

1.4.1 Research purpose

Taking a medical university and its affiliated hospital in Guangdong Province as an example, this research constructs a strategy for the sharing and application of healthcare big data in the context of current medical practice from the perspective of stakeholders in the sharing and application of healthcare big data, aiming to provide support for the effective sharing and in-depth mining of healthcare big data. The research focuses on the following purposes:

(1) To identify dominant stakeholders in the sharing and application of healthcare big data. There are many stakeholders involved in the sharing and application of healthcare big data, and only the behavior strategies of dominant stakeholders will have an important impact on the sharing and application of healthcare big data. In view of the current status of the sharing and application of healthcare big data, this thesis sorts out the stakeholders and their roles that affect the sharing and application of healthcare big data, and identifies the dominant stakeholders, so as to accurately grasp which stakeholders play a key role in hindering and promoting the sharing and application of healthcare big data, thus laying the foundation for the determination of the following in-depth interviews.

(2) To clarify the key influencing factors of the sharing and application of healthcare big data from the perspective of dominant stakeholders. Different dominant stakeholder entities may have the same interests that overlap with each other, or they may have different interests. Stakeholders often compete, collaborate, and compromise to meet their core interests, at the expense of those that are not as important to them. Through in-depth interviews, the core interests of dominant stakeholders and the degree of realization of their interests are maximized to promote or hinder the sharing and application of healthcare big data, and the decisive factors affecting the sharing and application of healthcare big data are deeply analyzed, so as to provide a basis for formulating strategies for the sharing and application of healthcare big data based on the core interests of dominant stakeholders.

(3) To propose effective strategies to promote the sharing and application of healthcare big data. Through in-depth interviews with key stakeholders, this study proposes strategies for the sharing and application of healthcare big data based on the core needs of these key stakeholders,

thus an ecosystem for the sharing and application of healthcare big data with active participation, coordination and mutual assistance of various dominant stakeholders is formed. In this way, this study aims to obtain an in-depth understanding of relevant data, and give full play to the intrinsic value of healthcare big data, thereby helping the medical university and its affiliated hospital to make major breakthroughs in medical research, continuously improve the quality of diagnosis and treatment as well as medical services, and promote innovation and transformation of research results in medical science. At the same time, this thesis also intends to provide reference for other medical colleges and relevant institutions in the sharing and application of healthcare big data.

1.4.2 Research significance

First, this research identifies the dominant stakeholders that affect the sharing and application of healthcare big data, which is conducive to promoting the enrichment and application of the Stakeholder Theory. Based on the perspective of stakeholders, on the basis of comprehensively combing, analyzing, concluding and summarizing the stakeholders involved in the sharing and application of healthcare big data, the researchers use scientific methods to identify the dominant stakeholders of the sharing and application of healthcare big data, and propose strategies for the sharing and application of healthcare big data based on the perspective of dominant stakeholders, which enriches the significance and connotation of the Stakeholder Theory.

Second, it will help promote the participation of dominant stakeholders in the sharing and application of healthcare big data. This thesis embeds the Stakeholder Theory into the process of sharing and application of healthcare big data, and advocates that different types of dominant stakeholders can jointly build symbiotic, win-win, and mutually beneficial behavior measures through coordination and mutual assistance, so as to concretize and orient their responsibilities and rights and interests in the process of the sharing and application of healthcare big data, thus promoting the active participation of dominant stakeholders in the sharing and application of healthcare big data.

Third, it will help improve the prevention, diagnosis, treatment and prognosis of diseases, promote medical scientific research and innovative applications, optimize the allocation of public medical resources, improve the utilization efficiency of medical resources, improve the emergency response level and comprehensive handling capacity of major public health emergencies, and provide more high-quality, efficient, fair, accessible, convenient and affordable medical services for people.

The structure of the thesis includes the Literature Review, followed by the Research Method. Then, there is the Stakeholder Identification in the Sharing and Application of Healthcare Big Data, followed by the Interview Analysis of Healthcare Big Data Sharing and Application. After, strategies for Improving Healthcare Big Data Sharing and Application are presented, and the Conclusions are presented at the end.

[This page is deliberately left blank.]

Chapter 2: Literature Review

The literature review is composed of six parts. (1) The origin and development of Stakeholder Theory. (2) The classification method of stakeholders. (3) The analysis method of stakeholder research. (4) The application of Stakeholder Theory in the field of healthcare big data sharing. (5) The candidate stakeholders and their interest demand in the sharing and application of healthcare big data. (6) Concept, research process, and application of in-depth interviews.

2.1 The origin and development of Stakeholder Theory

The Stakeholder Theory originated in the 1960s, and was gradually developed and improved in the process of questioning the long-term promotion and implementation of external control mode in the process of corporate governance in United Kingdom, the United States and other countries (S. H. Jia & Chen, 2002). In 1963, scholars from the Stanford Research Institute in the United States were inspired by the word “shareholder” and first proposed the concept of stakeholders. They defined stakeholders as “all groups of people who are intimately involved in the survival of the enterprise, without whom the enterprise or organization could not survive”. Different from the traditional enterprise management concept with the interests of shareholders as the core, the Stakeholder Theory pointed out in a pioneering way that the survival and development of any enterprise is inseparable from various types of stakeholders, such as shareholders, government departments, different forms of creditors, employees of different ranks, distributors, customers who purchase the enterprise’s products and services, and others, continuously investing some resources or keeping long-term participation. Although this concept only defines stakeholders from the perspective of influencing the enterprise, and limits the scope of stakeholders to a small group that affects the survival of the enterprise, it makes the academic and enterprise management circles begin to realize that not only shareholders, but also many social and organizational forces will have an impact on the operation and development of enterprises. The goal of enterprise management should be the overall interests of different types of stakeholders, rather than the single interests of a certain type of stakeholders such as shareholders (Yang & Zhou, 2000).

Ansoff (1965), an American economist and management scientist, was the first to try to formally apply the concept of “stakeholder” to academic research and the management practice

of organizations such as enterprises. Ansoff put forward that “in order to develop a more ideal strategic goal for the enterprise, it is necessary to comprehensively consider and balance the conflicting interests of multiple stakeholder groups such as government departments, shareholders of the enterprise, employees of different ranks, suppliers, distributors and consumers”

In the 1970s, the Stakeholder Theory gradually began to be accepted and adopted by experts, scholars and enterprise managers in Europe and the United States, and people’s understanding of stakeholders also expanded from “as an external factor affecting the survival of enterprises” to “directly or indirectly participating in the survival and development of enterprises”. In 1977, Wharton Business School, located in Pennsylvania, the United States, first opened the “stakeholder management” course, the purpose of which is to apply the concept of “stakeholder” to the strategy formulation and management practice of enterprises and other organizations, initially forming a relatively complete theoretical and analytical framework.

Since 1980, with the acceleration of economic globalization, the competition among different enterprises has become increasingly fierce. Enterprise managers and scholars in related fields have gradually realized that the one-sided definition of whether certain groups are stakeholders of an enterprise only based on “whether they can affect the survival of the enterprise” proposed by economists in the early days has great limitations. The academic and management circles tried to define stakeholders from different perspectives, but so far, no definition has been fully agreed by the enterprise management circles and the academic circles.

Mitchell and Wood (1997) have summarized 27 representative stakeholder concepts since 1963, when the concept of stakeholder was first formally proposed by Stanford Research Institute. Among nearly 30 stakeholder concepts, the most representative and influential one is the definition proposed by R. E. Freeman and M. B. E. Clarkson. They described the stakeholders in broad and narrow sense respectively.

The American economist R. E. Freeman has conducted a detailed and in-depth study of Stakeholder Theory and has given a very broad definition of stakeholders. In 1984, in his classic and landmark book *Strategic Management: A Stakeholder Analysis Approach*, stakeholders were defined as “any group or individual who can affect or is affected by the achievement of the organization’s objectives” (R. E. Freeman, 1984). The publication of this book also marked the formal formation of the Stakeholder Theory. In the broad definition of stakeholder given by R. E. Freeman, the basis of stakeholder’s interest with the organization can be either one-way or two-way, that is, “can actively influence or be influenced”. In R. E. Freeman’s view, stakeholders who actively influence the enterprise’s strategic goals and business performance

can be regarded as stakeholders of the enterprise, and individuals and groups that are passively affected by the actions taken in the process of achieving the enterprise's strategic goals can also be regarded as stakeholders of the enterprise (R. E. Freeman & Evan, 1990; R. E. Freeman & Reed, 1983). According to the above concept, government departments, shareholders of an enterprise, creditors, employees of different ranks, upstream suppliers, downstream distributors, final consumers and other stakeholders closely related to the survival of the enterprise, as well as relevant social organizations and social groups, the public around the enterprise, the community environment where the enterprise is located, and other groups or individuals who have a direct or indirect interest relationship with the enterprise to a large or small extent can be included in the list of stakeholders of the enterprise. Only those who cannot influence and are not influenced by the survival and development of the organization are excluded from the scope of the organization's stakeholders. This concept has greatly expanded the scope and connotation of stakeholder research.

R. E. Freeman's understanding of stakeholders coincided with the concept of corporate social responsibility, which was booming in Western countries at that time, and was recognized by many management scientists and economists. Corporate social responsibility emphasizes that in the process of production and operation, enterprises should not only bear the corresponding legal and contractual responsibilities to shareholders and employees, but also bear corresponding responsibilities to consumers, the environment and the community, rather than just creating profits as the only goal. R. E. Freeman's views were increasingly accepted by the business management community and became a mainstream reference and standard paradigm for defining stakeholders in different fields in the early 1990s. However, an overly broad definition will lead to too many stakeholders who need to be concerned and satisfied in practice, and the identity uncertainty is strong, which may make researchers and managers feel unable to start.

In the narrow sense of stakeholders, the most representative is the expression of voluntary or involuntary risk-bearers proposed by M. B. E. Clarkson (1995). He thinks that "voluntary stakeholders are those who have incorporated certain means of production, such as physical capital, human capital, financial capital or other valuable things in the development process of enterprise for its better survival and development and the continuous progress of strategic objectives, and thus bear some forms of risks for the smooth development of enterprise production and business activities; voluntary stakeholders are at risk due to enterprise activities, and if there is no risk factor, there is no stake among them and enterprises". In short, M. B. E. Clarkson (1994, 1995) believes that stakeholders are "individuals or groups who have invested

some different types of special capital in the operation and development of an enterprise, and therefore bear the corresponding risks”. In other words, stakeholders directly or indirectly participate in the production and operation activities of the enterprise for some benefits, promote the enterprise to achieve performance objectives, and therefore bear risks and share profits of the enterprise. They have ownership, claim and interest requirements for the enterprise. The more different forms of capital stakeholders invest in the enterprise and the greater the risk they take, the closer their relationship with the enterprise will be. This concept highlights the risk linkage between stakeholders and the enterprise, and illustrates the role of dedicated investment, narrowing the scope of stakeholders to a small group of people who have invested some form of dedicated capital in the enterprise. According to this definition, groups or individuals such as the media that may influence the development of the enterprise but have not invested special capital in the enterprise are no longer included in the list of stakeholders of the enterprise.

The difference between the narrow definition and the broad definition of a stakeholder is that the narrow definition is based on realities such as limited resources, limited time, limited attention and limited patience of managers in dealing with external constraints. In general, from a narrow perspective of stakeholders, researchers attempt to define stakeholder groups based on their direct relevance to the enterprise’s core economic interests, such as, the necessity for the enterprise’s survival (R. E. Freeman et al., 2010; Luper-foy, 1988), contractors or participants who have some kind of quid pro quo relationship with the enterprise (Cornell & Shapiro, 1987; R. E. Freeman & Evan, 1990; Hill & Jones, 1992), and the corresponding risks due to certain investment relationships with the enterprise (M. B. E. Clarkson, 1995). In contrast, the definition of stakeholders in the broad sense is based on empirical reality and encompasses all groups and individuals who can indeed have a significant impact on the enterprise, or who can indeed be influenced by the enterprise to a large extent. Regardless of whether the concept is broad or narrow, researchers generally agree to define stakeholders in terms of “influencing and/or being influenced” and “whether there are certain relevant interests or requirements to bear risks for the enterprise’s activities”.

In the 1990s, Chinese scholars’ research on stakeholders mainly received and relayed the concepts and methods put forward by western scholars. Since entering the 21st century, Chinese researchers and enterprise managers have begun to conduct in-depth research and application exploration on the Stakeholder Theory.

In China, the economist Wan et al. (1998) was the first to define stakeholders. In his book *Stakeholder Management*, he and his colleagues defined stakeholders as individuals, groups

and institutions that enjoy one or more interests in an enterprise. H. H. Chen and Jia (2004) defined stakeholders as “groups or individuals who have invested some valuable and dedicated capital in the process of survival and development of the enterprise, and thus bear certain risks accordingly, and their behavior can affect the process of achieving the strategic goals of the enterprise, or they are affected in the process of achieving the management goals of the enterprise”. The concept of stakeholders proposed by H. H. Chen and Jia emphasizes not only the role of dedicated investment, but also the two-way interest relationship between stakeholders and enterprises. This view is more consistent with M. B. E. Clarkson’s (1995) and Starik’s (1995) definition of stakeholders, that is, to consider whether a group of people are stakeholders in the organization in terms of dedicated investment and two-way impact. Yang and Yi (2003) defined stakeholders as people who have a contract or contractual relationship with the enterprise, such as corporate shareholders, corporate managers, general employees, suppliers, and others when exploring the paradigm shift from shareholder profit to stakeholder win-win. Stakeholders bear different levels of risk as a result of enterprise activities. L. Wu and He (2005) believe that stakeholders are those individuals, groups and institutions that can influence the realization of strategic goals and their processes to varying degrees. In recent years, based on the views of M. B. E. Clarkson and H. H. Chen, as well as realities such as limited resources, time and attention, Chinese scholars have defined groups and individuals who have certain relationships and interests with an organization, have invested certain resources into the organization, have assumed corresponding responsibilities and risks, and have a great interaction with the organization as stakeholders of the organization (Luo & Jiang, 2011; S. Zhang, 2021).

2.2 The classification method of stakeholders

With the development and application of Stakeholder Theory, practitioners and theoretical researchers of enterprise management in developed countries such as the United States and Japan have soon realized that only defining enterprise stakeholders is far from enough to comprehensively assess and deeply analyze the extent of their impact on the enterprise’s business performance and strategic objectives. Defining the stakeholders of an enterprise does not mean that the characteristics and importance of stakeholders are fully grasped, which just mixes together stakeholders with different purposes and requirements. After accurately defining the stakeholders of the enterprise, it is also necessary to classify a large number of stakeholders according to some standards or rules, so that different measures can be taken to classify and

manage stakeholders with different characteristics and importance. As Jones (1980) proposed, the Stakeholder Theory needs to answer the question “What are the stakeholder groups? Which of these people need to be served? Which of their interests are the most important? How should their interests be balanced?”

In the late 1980s, western academic circles began to focus on exploring how to divide the many stakeholders of enterprises into different categories according to one or some attributes. Researchers have tried to divide stakeholder groups according to different attributes or characteristics, which has deepened people’s understanding of stakeholders. Similarly, different scholars have different opinions on the classification of stakeholders. The main classification methods of stakeholders using time clues can be concluded and summarized as the “multidimensional segmentation” and “Mitchell score-based approach”. These two methods are the most remarkable achievements of western scholars in the exploration of classification of stakeholder, and they are also the methods that Chinese scholars often use for reference.

2.2.1 Multidimensional segmentation

“Multidimensional segmentation”, also known as “conceptual segmentation”, is a speculative analysis of stakeholder groups from different characteristic dimensions at the conceptual level, in order to classify stakeholders from multiple dimensions to find out the differences in certain characteristics of different types of stakeholders. Before 1995, enterprise managers and academic researchers mainly used multidimensional segmentation to classify stakeholders, represented by R. E. Freeman, Frederick, Charkham, M. B. E. Clarkson, Wheeler, Savage, Carroll and Walker.

(1) R. E. Freeman (1984) creatively subdivided stakeholders from three different dimensions of ownership, economic dependence and social interests. The first category is the stakeholders who have ownership of enterprise assets. The classification of such stakeholders is based on holding company’s stock, mainly the directors of the enterprise, shareholders, technical personnel and management personnel who hold the shares of the enterprise, and other individuals or groups who hold shares. The second category is the stakeholders economically dependent on the enterprise, including corporate bond holders, professional managers who are paid in the process of enterprise operation, employees of different positions, consumers, suppliers and competitors of enterprises, and the surrounding communities of enterprises. The third category is the stakeholders who have a close relationship with the enterprise in various social interests, mainly including governments at all levels, various types of media and other special groups that are closely related to the social interests of enterprises.

(2) Frederick (1988) believed that stakeholder refers to “all groups that can influence the strategic objectives and business decisions of the enterprise”. In order to deeply study the interest relationship between stakeholders and enterprises and their impact on enterprise strategic objectives and business decisions, Frederick divided stakeholders into direct interest groups and indirect interest groups. Direct stakeholders refer to individuals or groups directly related to the enterprise in the process of market transactions, mainly including employees in different positions in the enterprise, internal and external shareholders of the enterprise, various forms of creditors, suppliers, agents, existing customers and potential consumers, peers and potential competitors with market competition. Indirect stakeholders are those who have non-market relationship with the enterprise’s operation management and strategic decision-making process. It includes government departments at different administrative levels that supervise and manage enterprises, foreign government departments that have friendly or hostile relations with enterprises due to international cooperation and competition, social groups such as consumer associations and environmental protection associations that put pressure on enterprises, various media such as news, radio, film and television, and the general public. The stakeholder classification method proposed by Frederick is shown in Figure 2.1.

(3) According to whether there is a market-oriented contractual relationship between stakeholders and enterprises (i.e., the nature of contractual relationship with enterprises), Charkham (1992) divided stakeholders into contractual stakeholders and community stakeholders.

Contractual stakeholders include internal shareholders and external shareholders related to market contracts, employees of different positions, customers, suppliers, distributors, suppliers and various forms of creditors, while community stakeholders include all consumers, organizations that supervise and manage enterprises, government departments, groups that put pressure on enterprises, news media, and communities around enterprises.

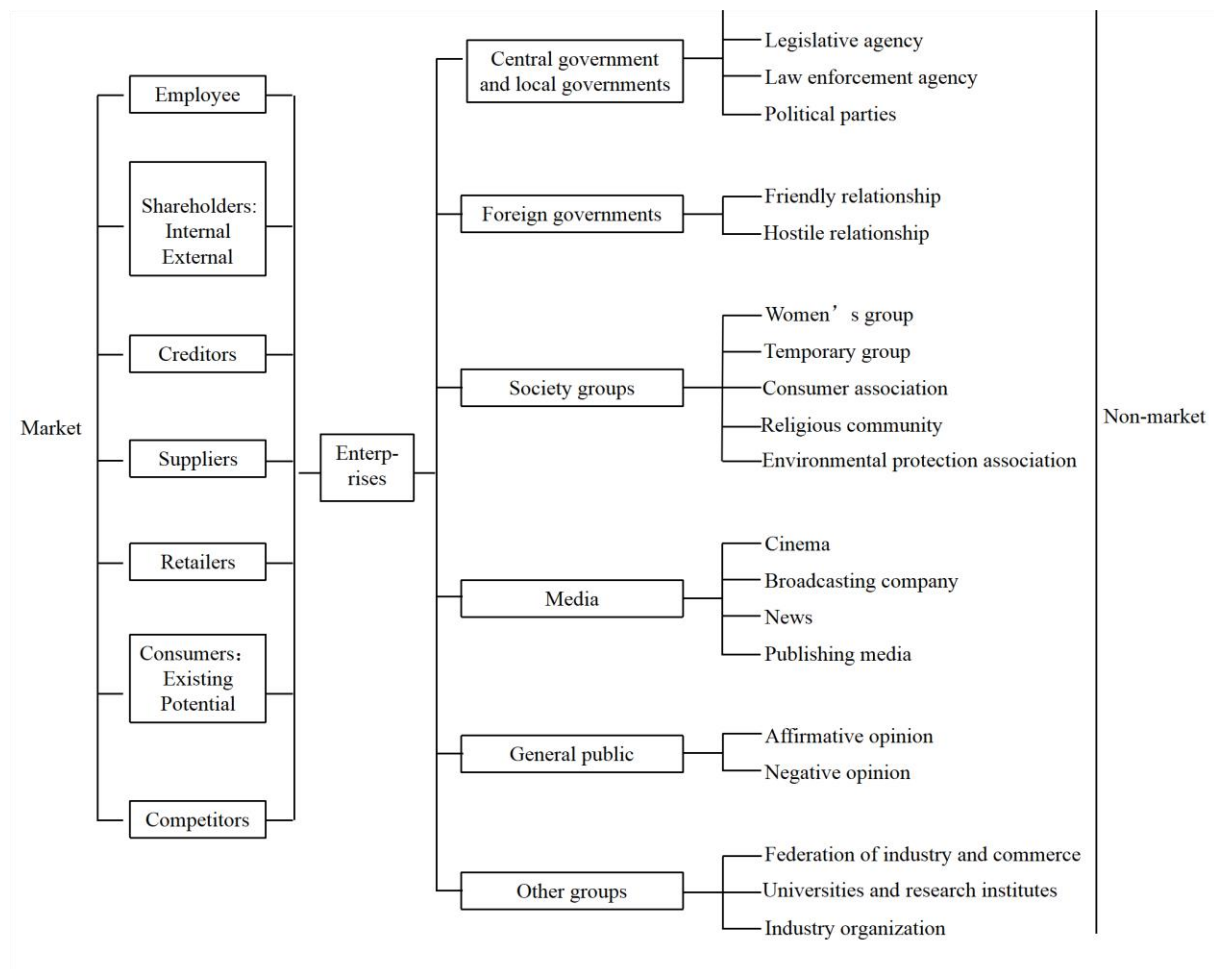


Figure 2.1 Stakeholder classification method proposed by Frederick

Source: Frederick (1988)

(4) When studying the management of enterprise stakeholders, M. B. E. Clarkson (1995) put forward two typical classification methods according to the attitude of stakeholders towards the enterprise or the importance of stakeholder behavior.

One divides stakeholders into voluntary stakeholders and involuntary stakeholders, based on the types and ways of taking risks by stakeholders in enterprise operation, management or decision-making activities. Voluntary stakeholders refer to those individuals or groups who have actively invested certain property capital, human capital, material capital or other valuable capital in the operation, management and decision-making activities of enterprises. They actively and voluntarily bear some forms of risk that business activities may bring to them. The latter means the individuals or groups that passively bear some forms of risks due to the operation, management and decision-making activities of enterprises. In other words, as explained in M. B. E. Clarkson's definition of stakeholders, "the stakeholders of an enterprise are individuals or groups who bear different forms of risks in the operation and development of the enterprise due to the investment of some special capital."

The other divides stakeholders into primary stakeholders and secondary stakeholders according to the closeness of interests between relevant groups and enterprises. Primary stakeholders are the individuals and groups that have a significant impact on the business activities and development strategies of the enterprise. Without their continuous participation or investment of resources, the enterprise cannot be continuously maintained and survived. Such stakeholder groups mainly include internal and external shareholders of the enterprise, creditors of different forms, investors of the enterprise, employees and managers of different positions, consumers who buy the products and services of the enterprise, suppliers of the enterprise, and public stakeholder groups that provide the market and infrastructure such as government departments and the communities where the enterprise is located. Secondary stakeholders are the individuals or groups who indirectly affect the production and business activities of enterprises or are indirectly affected by the production and business activities of enterprises, but do not conduct direct transactions with enterprises, and do not play a fundamental role in its survival, so they are not essential to the survival and development of the enterprise, and without them, the enterprise can still survive. Such stakeholder groups mainly include media such as news, radio, television and film, experts and scholars, and a series of less important stakeholder groups around enterprises.

(5) The British scholars Wheeler and Maria (1998) introduced the social dimension into the classification and definition of stakeholders based on M. B. E. Clarkson's research, which had a profound impact. The introduction of the social dimension also has a profound impact on the research of stakeholder classification methods and the practice of stakeholder classification.

Wheeler believes that some stakeholders have a direct interest relationship in the enterprise through "people who exist in reality and can be concrete", so they are social. Some stakeholders do not have an interest relationship in the enterprise through "people who exist in reality and can be concrete", such as the natural environment that is continuously deteriorating or being improved, some non-human species, and others, which are not social. Combined with M. B. E. Clarkson's proposal for the closeness of the stakes with the enterprise, and considering whether the two dimensions of sociality introduced by Wheeler exist respectively, Wheeler proposed to divide all types of stakeholders of an enterprise into the following four types:

First-level social stakeholders, who have a direct relationship with the enterprise and directly participate in the production and operation activities of the enterprise as a person who exists in reality, including customers, investors, managers and general employees, suppliers of enterprises, distributors and other business partners, and local communities.

Secondary social stakeholders, who form indirect contact with enterprises through a series

of social activities in which people who exist in reality and can be concrete participate. They may also have a very big impact on the enterprise, especially in terms of corporate reputation and corporate credibility, but this impact is only indirectly caused, rather than directly formed. Such stakeholder groups mainly include government departments, consumer associations and other groups that put pressure on enterprises, trade unions, social news media, academic critics, as well as some trade groups and competitors that do not directly interact with enterprises.

First-level non-social stakeholders, who are directly related to the enterprise and have a direct impact on the enterprise, but they are not specific people in the real society, mainly including non-human species such as natural environment, plants and animals.

Secondary non-social stakeholders, who have an indirect impact on the enterprise and are not specific people in the real society, including environmental pressure groups and animal interest groups. Wheeler's classification of stakeholders according to the two dimensions of closeness and sociality with the enterprise is shown in Figure 2.2.

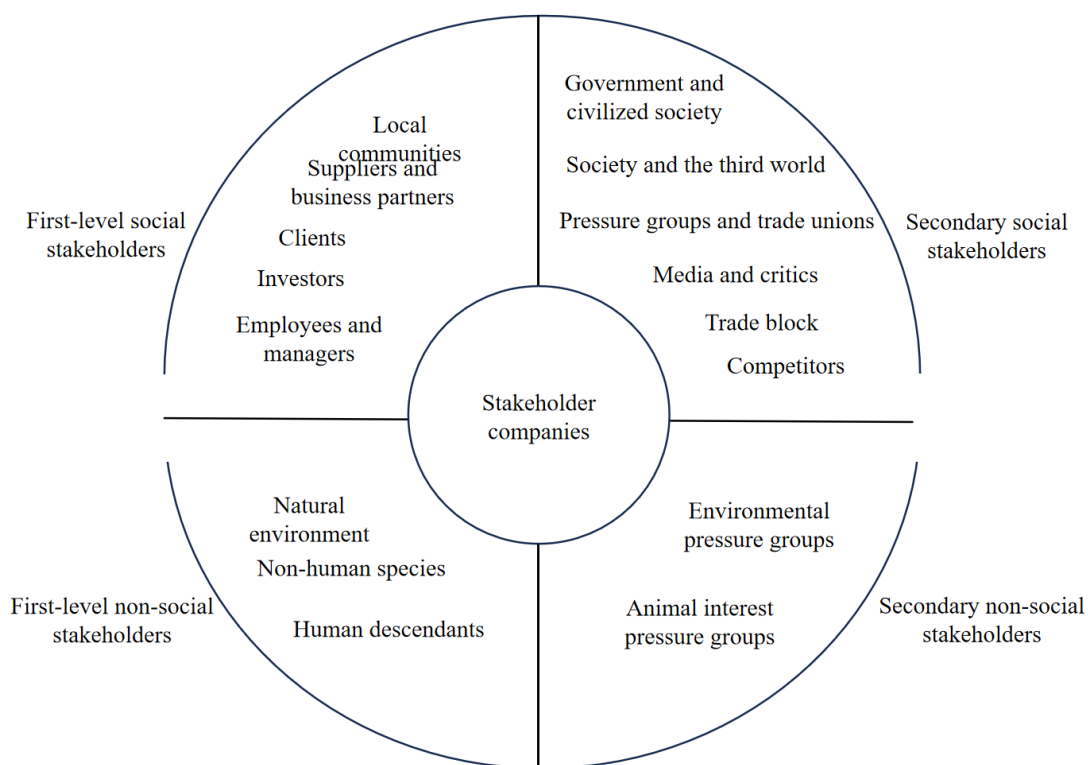


Figure 2.2 Stakeholder classification method proposed by Wheeler

Source: Wheeler and Maria (1998)

(6) In addition, from the perspectives of the potential of threat and cooperation, Savage (1991) and his colleagues divided the enterprise's stakeholders into the supportive, marginal, mixed blessing, and non-supportive stakeholder.

According to Savage, the following four factors affect the threat potential and cooperation potential of these stakeholders to enterprises: (1) Whether they have mastered the resources that

are critical to the survival and development of certain enterprises; (2) Whether their own strength is stronger or weaker than the enterprise; (3) Are they likely to take action for or against it? Or do not act; (4) What other groups will they form alliances with?

Enterprises can judge the level of threat potential and cooperation potential according to the realistic situation of stakeholders in the above four types of factors. If the number of “increased” in the threat potential of a certain stakeholder group is greater than that of “decreased”, the threat potential of that stakeholder group is judged to be high, and vice versa, the threat potential is judged to be low. Similarly, if the number of “increased” in the cooperation potential of a certain stakeholder group is greater than that of “decreased”, the cooperation potential of that stakeholder group is judged to be high, otherwise, the cooperation potential of that stakeholder group is judged to be low. When the total number of increases and decreases of a certain potential is equal, it is impossible to judge. Table 2.1 shows the details of the threat potential and cooperation potential posed by stakeholders to the enterprise.

Table 2.1 Details of threat potential and cooperation potential proposed by Savage

Factors		Threat potential	Cooperation potential
Whether it has mastered the key resources (needed by enterprises)	Stakeholders master the key resources of enterprises	Increased	Increased
	Stakeholders do not master the key resources of enterprises	Decreased	Unchanged
	Stakeholders are more powerful than enterprises (dependent)	Increased	Unchanged
How does its own strength compare with the enterprise	Stakeholders and enterprises are equally powerful	Unchanged	Unchanged
	Stakeholders are not as powerful as the enterprise	Decreased	Increased
	Stakeholders may take supportive actions	Decreased	Increased
What actions might be taken	Stakeholders may take non-supportive action	Increased	Decreased
	No action will be taken by the stakeholders	Decreased	Decreased
Which groups are likely to be allied with	Stakeholders may be allied with other stakeholders	Increased	Unchanged
	Stakeholders may be allied with enterprises	Decreased	Increased
	Stakeholders do not participate in any alliances	Decreased	Decreased

Source: Savage et al. (1991)

Stakeholders with low threat potential and high cooperation potential with the enterprise are defined as supportive stakeholders, mainly including board members, managers, employees in different positions, parent companies, and possibly suppliers of the enterprise. They are the ideal type of stakeholders that are most sought-after by enterprises.

Marginal stakeholders neither pose a high threat to enterprises nor are particularly willing to cooperate with enterprises. This type of stakeholders mainly included consumer interest groups, some shareholders and employee associations. Although they have some potential interests with enterprises in production and operation activities, most of the time they do not pay special attention to and care about these interests and problems.

The stakeholders with high threat potential and cooperation potential to the enterprise are called mixed blessing stakeholders, which mainly include employees currently in short supply of the enterprise, customers and agents of the enterprise, and organizations that can provide supplementary products or services to the enterprise. Enterprise managers should maximize their support and minimize their threat potential.

Non-supportive stakeholders are groups or individuals with high threat potential but low cooperation potential to the enterprise, which may mainly include competitors, trade unions, and news media. This kind of stakeholders is the most troublesome for enterprise managers.

The same group may belong to different types of stakeholders as specific circumstances and scenarios change. For example, employee federations may in general be marginal stakeholders who do not attract much attention from enterprise managers. However, when the interests of employees are seriously threatened or persecuted, employee federations may turn into the most troublesome type of non-supportive stakeholders for enterprise managers.

(7) Thereafter, Carroll (1996) proposed two classification methods. One is direct stakeholders and indirect stakeholders based on the formality of the relationship between stakeholders and the enterprise. The former refers to an individual or group that has a direct claim on the enterprise due to a certain contractual relationship or other types of formal relationship recognized and protected by law, while indirect stakeholder refers to an interest group or individual that has an informal relationship with the enterprise. There is no formal contractual relationship or other legally recognized and protected relationship between indirect stakeholders and the enterprise, so their influence on the enterprise is not so important. This classification method makes it clear to a certain extent that if there is a conflict among stakeholders, the priority should be paid to the interests of the direct stakeholder groups that have formal relations with the enterprise. According to the participation in the survival and development of the enterprise, the second classification method proposed by Carroll is to divide stakeholders into ① core stakeholders that are decisive for the survival of the enterprise, ② strategic stakeholders that are particularly important when the enterprise has specific opportunities or face specific threats, and ③ environmental stakeholders who are outside the

enterprise.

(8) According to Dick and Dasu's definition of loyalty in the process of enterprise operation and management, Walker and Hampson (2003) evaluated stakeholders according to different levels of enterprise commitment. Combined with the stakeholders' attitudes to the enterprise and their loyalty of behavior, stakeholders can be divided into truly loyal stakeholders, accessible stakeholders, trapped stakeholders, and high-risk stakeholders.

In addition, Walker Information proposes a stakeholder attribute classification method that is different from the above classification model. They believe that the criterion of whether a certain stakeholder group should be given priority is whether this group is mentioned in the mission and core values of the enterprise and whether it is extremely important to the survival and development of the enterprise. In their analysis, they found that the stakeholders mentioned in the core mission and values of the enterprise were, in descending order, customers (82%), employees (70%), shareholders (57%), suppliers (33%), the communities (40%) and government departments (14%). This means that they believe that the relationship with the stakeholder group at the highest level should be with the customer.

The introduction of the method of conceptual segmentation of stakeholders according to different dimensions into corporate governance and strategic management can help to expand the thinking of managers and researchers. But there is a significant practical problem with conceptual speculative analysis, that is, it makes people's understanding and cognition of stakeholders only stay in the academic theoretical research level, and the operability in practical application is low, which restricts the practical application of the Stakeholder Theory in enterprise management and decision-making and other fields to a certain extent.

2.2.2 Mitchell three-dimensional scoring approach

In the late 1990s, American scholars Mitchell et al. (1997) proposed a score-based approach to comprehensively define and classify stakeholders. They clearly pointed out that "there are two main problems in the research of enterprise stakeholder theory to solve. The first problem is the definition of stakeholders. Who are the stakeholders, or who are the stakeholders of the enterprise? And the second, what are the basic characteristics or attributes to classify these stakeholders? Under the premise of limited resources and energy, what standards or rules should enterprise managers rely on to give different degrees and levels of attention to the stakeholders of the enterprise?"

Based on the in-depth discussion of the research content of the previously published Stakeholder Theory, and combined with the actual situation, Mitchell and Wood proposed that

possible stakeholders can be comprehensively scored and analyzed from the three dimensions of legitimacy, power and urgency, and then determine

Legitimacy is a universal perception that the actions of an individual or group are desirable, right, or appropriate within a certain system of social norms, values, beliefs, and rules. In enterprise management, legitimacy refers to whether an individual or group has been given the right to construct or claim the right to participate in the operation of the enterprise in a social, moral, legal or other specific type. The members of the board of directors of the enterprise, the customers, the employees of different positions in the enterprise, and other groups closely related to the enterprise have the legal right to demand the enterprise to protect their rights and interests and claim from the enterprise, which also means that the legitimacy of their demands is larger, while the social groups, competitors and other stakeholders are not so closely related to the enterprise, so their demands and their claims have lower legitimacy.

Power is a relationship among social actors. In such a relationship, one party has or has access to certain coercive, utilitarian or normative powers or means and is therefore able to impose its will on the other party. In other words, social actor A can ask another social actor B to do something that B would otherwise be unwilling to do, which means that A has some special powers. In enterprise management, power refers to whether a certain type of stakeholder group has or can obtain the status, ability or means that can influence the strategic objectives and business decisions of the enterprise.

Legitimacy and power are regarded as independent variables in the relationship between stakeholders and enterprise managers, which promotes the identification and attribute determination of stakeholders, but fails to capture the interaction dynamics between stakeholders and enterprise managers. Mitchell added the attribute of urgency to the stakeholder classification model, which helped move the classification model from static to dynamic. *Merriam-webster* defines urgency as “demanding immediate attention” or “pressing”, Mitchell argues that urgency (synonyms including “compelling”, “driving”, and “imperative”) exists only when two conditions are met. (1) The relationship or requirement has time sensitivity. It is unacceptable for the manager of the enterprise to delay the processing of the requirement or relationship with the stakeholders. (2) Criticality. The requirement or relationship is very important to the stakeholders. Thus, according to Mitchell’s definition, urgency indicates the degree to which a stakeholder’s requirement can immediately attract the attention of enterprise managers and regulators, or the degree to which public opinion immediately supports it.

Mitchell and Wood proposed that a stakeholder in an enterprise can only be identified with at least one of these three attributes. In other words, the stakeholders of the enterprise, at least,

are either endowed with social, moral, legal or other specific types of construction rights to participate in the operation of the enterprise or the right to claim from the enterprise; either it can arouse urgent attention of enterprise managers or regulators or immediately widespread support of public opinion; either have access to sufficient power or means to have a significant impact on enterprise performance and strategic decisions. After scoring the possible stakeholder groups of an enterprise on three dimensions according to the specific situation, they can be divided into stakeholders and non-stakeholders. Stakeholders can be further subdivided into definitive stakeholders, expectant stakeholders and latent stakeholders.

Firstly, definitive stakeholders, namely, core stakeholders, who possess the three attributes proposed by Mitchell, that is, the legitimacy of claiming in the process of enterprise survival and development, the power, and the urgency that the demand will be paid immediate attention. In order to achieve the strategic objectives of the enterprise and obtain higher business performance, the managers of the enterprise must always pay attention to the wishes and interest requirements of the definitive stakeholders, and try to meet their interest requirements, so as to improve their satisfaction and ensure the sustainable development of the enterprise. Generally speaking, definitive stakeholders include internal and external shareholders, employees in different positions, and consumers who purchase the products and services of the enterprise.

Secondly, expectant stakeholders, who have close relationships with the enterprise and two of the above three attributes. This kind of stakeholders can be divided into the following three situations.

① Dominant stakeholder, the group with both legitimacy and power, such as the government departments. They are eager to get the close attention of the company's senior management, and often can achieve their expected goals with the legitimacy and power. In some cases, they will formally participate in the enterprise's decision-making and management process.

② Dependent stakeholder. They have legitimacy and urgency to the enterprise, but lack the corresponding power to help realize their interest requirements. Such groups cannot realize their own interests independently, but need to depend on other stakeholder groups. If they want to achieve its goal, it needs to win the support of other stronger stakeholders with rights or rely on the benevolence of enterprise managers. Therefore, they often promote their goals by forming alliances with other stakeholder groups, participating in the political activities of government departments with power, and morally appealing to the conscience of corporate governance.

③ Dangerous stakeholder. They have urgency and power to the enterprise, but lack legitimacy to claim rights from enterprises. Such stakeholders are very dangerous for the operation, management and decision-making activities of enterprises. Due to the combination of the power to make demands on the enterprise and the urgency of the interest claim that needs to be paid attention to, if the interest demand in the daily business activities of the enterprise is not met in time, they may often choose to take behaviors, means and measures that are a great threat to the enterprise or even seriously destructive. For example, when the internal contradictions of the enterprise continue to intensify, some employees who are dissatisfied with the current policies of the enterprise will launch strikes and demonstrations to disrupt the production order of the enterprise, and environmentalists may take actions such as demonstrations to counter the damage and threat that the enterprise may bring to the ecological environment.

Thirdly, latent stakeholders, the group with only one of the three characteristics of legitimacy, power and urgency. With limited asset, time, energy and other resources, latent stakeholders are unlikely to receive attention. Similarly, they are unlikely to give more attention or recognition to the company. What's more, latent stakeholders have three categories based on their individual attributes.

① Discretionary stakeholder, the group that only has legitimacy but lacks power and urgency. They depend on the operation of the enterprise to decide whether to play its role. There is absolutely no pressure for managers to establish positive relationships with such stakeholders, although managers can choose to do so. Discretionary stakeholders are a particular focus of scholars who study corporate social responsibility and performance, because they are the most likely recipients of corporate philanthropy.

② Dormant stakeholder, a group only with power but no legitimacy and urgency. Their power is not used and is in a dormant state, with little or no interaction with enterprises. However, such stakeholders have the potential to obtain the second attribute. For example, they may become dangerous stakeholders once they gain urgency, or they may become dominant stakeholders after gaining legitimacy through alliances with certain stakeholder groups, requiring adequate attention from enterprise managers.

③ Demanding stakeholder, a group that only has urgency but lacks legitimacy and power. In Mitchell's view, they are like "mosquitoes buzzing in the ears of managers, annoying but not dangerous, troublesome but not requiring much attention". Unless there are indications demonstrating that their demands and goals have some forms of legitimacy, or that they have

gained some power, the management does not need or has little enthusiasm to pay attention to them (Mitchell et al., 1997).

Figure 2.3 shows the stakeholder classification results based on Mitchell's three-dimensional attribute classification method. Latent stakeholders: 1, 2, 3; expectant stakeholders: 4, 5, 6; definitive stakeholders, namely core stakeholders: 7. 8 indicates that a certain group does not have any dimension of legitimacy, urgency and power, and is not a stakeholder of the enterprise.

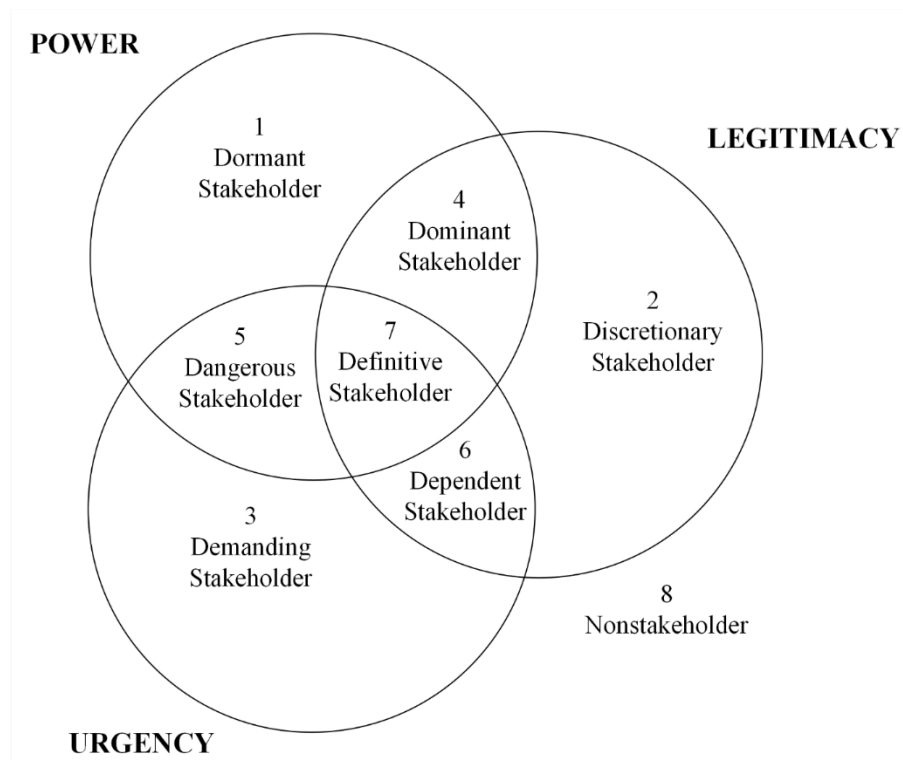


Figure 2.3 Results of stakeholder qualitative classification based on Mitchell scoring method

Source: Mitchell et al. (1997)

It is worth noting that Mitchell's stakeholder classification model is not static and fixed, but dynamic and variable. At different stages of the survival and development of an enterprise, if an individual or group gains or loses certain attributes due to external factors or changes in its own state, it will jump from one type to another with more or less importance. For example, if a political or social factor makes a dominant stakeholder urgent, it will jump up to become a definitive stakeholder, or a dependent stakeholder who lacks power will also jump to a definitive stakeholder if they gain power through an alliance with other rights institutions or a change in some political factors. Another example is that a demanding stakeholder, as a latent stakeholder of the enterprise, does not receive much attention from enterprise managers, but once it acquires some kind of power and becomes a dangerous stakeholder, or it gains legitimacy due to changes in policy and political factors and becomes a dependent stakeholder, the enterprise managers need to pay extra attention to it.

The Mitchell scoring and classification approaches are clear, simple and feasible, making the stakeholder more easy to be classified. More importantly, the enterprise managers can quantitatively subdivide the stakeholders according to the score, extending the stakeholder research from the academic speculative mode to the empirical research, which greatly promotes the application of Stakeholder Theory in management practice and is widely respected by the academic and enterprise management circles. It has gradually become the most commonly used method to define and classify stakeholders in different fields of empirical research. According to the specific situation of enterprises or organizations, many researchers and managers use Mitchell's three-dimensional attribute scoring method to define and classify stakeholders, so as to provide a reasonable basis for enterprises or organizations to make management decisions.

Norwegian scholars Knut and Sven (2001) studied in detail the stakeholders in Norwegian fishery enterprises based on Mitchell's three-dimensional scoring approach. The research shows that many stakeholders in fishery enterprises are not static, but in a state of dynamic change. After the changes of certain political factors, as well as internal and external environment of the enterprise, latent stakeholders such as citizens, news and radio, and municipalities may gain attributes that they otherwise lack and become expectant stakeholders, while expectant stakeholders such as local residents and environmental protection groups may acquire one or more attributes that they would otherwise lack and thus become definitive stakeholders. Similarly, definitive stakeholders may lose certain attributes due to changes in the political and socio-economic environment and become expectant stakeholders or even latent stakeholders. Only after clearly understanding the current characteristics of different stakeholders and fully grasping the further development trend, can the enterprise management formulate and implement corresponding strategies to ensure the sustainable healthy development of the enterprise.

Rachel and Baskerville (2004) found that the stakeholders in the crisis period of accounting professional institutions are not fixed but dynamic. Stakeholders who acquire some of their missing features will promote to become other type of stakeholder. Later, Magness (2008) conducted a study on dynamics of Mitchell's three-dimensional attribute scoring model against the background of two major mining disasters and environmental accidents, supporting that the state of stakeholders is temporary and dynamic. Therefore, the identification and classification of stakeholders of an organization should be based on the current social and economic reality and specific practical process.

2.2.3 Research on stakeholder classification by Chinese scholars

After the mid-1990s, Chinese scholars began to pay attention to the stakeholder classification, and realized that the definition and classification of stakeholders is “a very critical issue” (Yang & Zhou, 2000). Chinese scholars’ research on the classification of stakeholders is based on western research. In the early stage, stakeholders were mainly classified from the perspective of multidimensional segmentation by Chinese scholars, such as Wan et al. (1998) and X. H. Li (2001).

According to whether stakeholders have formal and contractual official contracts with enterprises, Wan et al. (1998) divided stakeholders into primary stakeholders and secondary stakeholders. Primary stakeholders include owners of financial capital and human capital, governments with management and supervision powers, suppliers, and customers who purchase the company’s products and services, while secondary stakeholders include consumer rights protection organizations, environmental protection organizations, news and radio media. This classification method is similar to Charkham’s classification.

X. H. Li (2001) drew on the pairwise attribute combination classification model proposed by Savage and divided stakeholders into four categories from the two dimensions of cooperation and threat: (1) Supportive stakeholders, mainly including internal and external shareholders of the enterprise, different forms of creditors, managers and employees of the enterprise, suppliers and distributors. They benefit from the economic interests of the enterprise and naturally have close ties with the enterprise, so they will cooperate with the enterprise in a friendly way. For such stakeholders, enterprises can choose to actively encourage them to participate in the operation and management of enterprises according to the principle of decentralization of rights and participatory management, making joint efforts to promote the efficiency of enterprises. (2) Marginal stakeholders, including enterprise professional associations, consumer interest protection groups, unorganized minority shareholders and others. It is characterized by a low likelihood of both threat and cooperation, so reasonable “monitoring” should be carried out to grasp the change of its characteristics in time. (3) Non-supportive stakeholders, including groups, trade unions, news and radio media that have competitive relations with enterprises. It is characterized by high potential threat to enterprises and low likelihood of cooperation. They are often in a non-cooperative relationship with the enterprise and should be “guarded” against it, and their relationship with the enterprise should be handled carefully to minimize the threat they may pose to the enterprise. (4) Mixed stakeholder. including employees in short supply. It is characterized by high potential threat and

cooperation to enterprises. The enterprise managers should cooperate with them as much as possible to increase their support level and reduce their potential threat to the enterprise.

These studies are mainly normative analysis, and make logical analysis and reasoning based on conceptual contents, and the classification results are not tested and supported by empirical research. At the same time, the previous research is often static at a certain moment, does not conduct dynamic analysis in combination with the specific internal environment and external environment of the enterprise in different operating stages, and does not pay attention to the differences among enterprises with different forms, so that managers are at a loss in the face of many classification results formed by numerous dimensions, and do not know how to use these research results to effectively guide enterprise strategic objectives and management practice.

H. H. Chen and Jia (2004) comprehensively used Mitchell's three-dimensional attribute scoring method and multidimensional segmentation for conceptual dialectical analysis, and conducted empirical research from three dimensions of initiative, power and urgency through questionnaire. First of all, they believe that some stakeholders of an enterprise actively and spontaneously participate in the business process and bear the corresponding risks, while the other part of all stakeholders passively participate in the enterprise and bear different risks because of some factors (M. B. E. Clarkson, 1995). Different types of stakeholders differ in how they influence enterprise activities and risk-taking initiatives. In addition, different types of stakeholders have significant differences in the degree of influence on the survival and development of enterprises (M. B. E. Clarkson, 1995; R. E. Freeman & Evan, 1990). Finally, there are significant differences in the degree to which different stakeholders can get immediate attention (Mitchell et al., 1997). Therefore, H. H. Chen and Jia (2004) conducted a questionnaire in nine provinces in China, asking respondents to rank ten types of stakeholders from three perspectives: initiative, power and urgency. According to the ranking results collected from 423 valid questionnaires, the ten types of stakeholders identified in the early stage were divided into core stakeholders, dormant stakeholders and marginal stakeholders.

Core stakeholders score less than 4 points in at least two dimensions of the above three dimensions (stakeholder groups are scored on a dimension based on the order of the forward ranking. The higher the ranking of stakeholder groups in a dimension is, the lower their score in that dimension will be), or they need to be in the top 4 in at least two dimensions. They are one or more groups that are indispensable in the process of enterprise survival and development, and have a very deep and close interest relationship with the enterprise, and can even be directly related to the survival and long-term development of the enterprise, mainly including shareholders, managers and general employees of the enterprise.

Dormant stakeholders, who score more than 4 points and less than 6 points in at least two dimensions, mainly including consumers who buy products or services of enterprises, different forms of creditors, government departments, upstream suppliers of enterprises and downstream distributors. Such groups themselves have often formed a relatively close dependence relationship with enterprises, and the special capital they invest makes them bear certain risks because of the production and operation activities of enterprises. In the case of a good business situation, they seem to be just an obvious legal contractor of the enterprise. Once some of their actual interests are not fully met or they feel that their own interests are damaged and threatened, they may immediately turn into an active state, and the degree of active state may also be very strong, thus affecting the healthy development of the enterprise.

Marginal stakeholders are mainly special interest groups and communities. They score more than 6 points in at least two dimensions, with the forward ranking in the bottom 40 percent. They have low initiative and are often passively affected by the enterprise. In the eyes of enterprises, their importance is very low, and their urgency to realize interest requirements is not strong. H. H. Chen's empirical research on stakeholders has certain representativeness and practical significance. However, the three dimensions of initiative, urgency and power are not completely independent attributes, and their meanings and connotations overlap.

L. Wu and He (2005) studied the power change of stakeholders from the perspective of enterprise life cycle, analyzed the changes in the importance of stakeholders faced by enterprises in different life cycles from the aspects of enterprise life cycle characteristics, crises faced by enterprises at various stages and the characteristics of core stakeholders, and divided them into critical stakeholders, non-critical stakeholders and marginal stakeholders, and this classification is verified by empirical research. The classification study conducted by L. Wu and He takes into account the differences in the different life cycles of entrepreneurship, growth and maturity of enterprises, and therefore has more profound practical significance than the previously published empirical research on stakeholders.

In the context of the sharing and application of healthcare big data in a medical university and its affiliated hospitals in Guangdong Province, the thesis will adopt Michell three-dimensional attribute scoring method to determine and categorize stakeholders through questionnaire survey and expert scoring.

2.3 Analytical method of stakeholder research

At the beginning of the study and application of Stakeholder Theory, enterprises had been the

absolute subject of research. Before the classic book *Strategic Management: An Analytical Approach to Stakeholder Management* by R. E. Freeman, an American economist, was published in 1984, the focus of stakeholder research was mainly to analyze which subjects were the stakeholders of enterprises and the rationality of their participation in corporate governance. R. E. Freeman introduced the stakeholder research method into enterprise strategic management, emphasizing the influence of enterprise stakeholders on the formulation and implementation of enterprise strategy. It was not until the late 1980s and early 1990s that the research based on the perspective of stakeholders began to expand to government departments, surrounding communities, urban construction and development, medical and health undertakings, different types of social groups, and related economic and social environments.

American scholars J. D. Blair and Whitehead (1988) were the first to introduce the research and analysis of Stakeholder Theory into the health field and proposed a practical approach and research steps for stakeholder analysis. J. D. Blair found that hospital managers must effectively manage stakeholders in order to cope with the environmental factors and various uncertainties faced by hospitals, and proposed analytical methods and steps for the diagnosis and management of hospital stakeholders. According to J. D. Blair, the research on diagnosis and management of hospital stakeholders can be broadly divided into two steps: the first step is to identify the dominant stakeholders of the organization (hospital) under research; the second step is to draw the stakeholder diagram. The information and content to be conveyed in the diagram varies in according with the purpose of the analysis. For example, if it is used to analyze the position and attitude of the stakeholders on the organization's objectives or a specific project, the specific stance of each stakeholder should be manifested and appropriately categorized in the diagram, such as support, opposite, neutral, etc.

Dymond et al (1995) evaluated and empirically studied stakeholders in pharmaceutical distribution networks based on organizational control, alliance formation, resource control and relative rights. Topping and Fottler (1990) studied the management improvement and performance improvement of health maintenance organizations from the perspective of stakeholders. J. D. Blair et al (1990) proposed a six-step strategic approach and corresponding maps and models to improve the success rate of cooperation between hospital managers and medical staff: (1) Identifying key stakeholders and their links to each other; (2) Using problem-oriented maps to reveal conflicts among stakeholders; (3) Diagnosing cooperation from two dimensions that promote successful cooperation (cooperation and finance); (4) Classifying cooperation using two dimensions that promote successful cooperation; (5) Selecting a strategy to optimize the potential for success of the current collaboration; (6) Choosing an approach with

limited potential for success to improve the current cooperation model.

The above researches mainly focus on the role of stakeholders in management decision-making or their attitude towards organizational decision-making, but rarely systematically analyze their specific interest content and influence. However, through integrating theories and methods of social science, public health policy and medical ethics into stakeholder studies by these scholars, the application scope of stakeholders in the field of medical and healthcare has continuously expanded, and the research framework and analysis methods have been continuously improved.

In addition, Ruairi and Zsuzsa, from a stakeholder perspective, draw on the experience of Varvasovszky and colleagues in analyzing Hungary's alcohol restriction policy to systematically illustrate how stakeholder analysis can be used in conducting health policy analysis, forecasting policy development, and implementing specific policies or programs. They outlined the factors that should be considered before conducting a stakeholder analysis, including the stakeholder's purpose and time dimension, the time frame and the context in which the analysis is performed. Ruairi described how to identify and approach stakeholders and considered the use of qualitative or quantitative data collection methods to assess the position, status, level of interest and degree of influence of stakeholders. A key message is that the data collection and analysis process require interactive iteration; when new data is obtained, analysts need to revise and deepen the earlier analysis results (Brugha & Varvasovszky, 2000; Varvasovszky & Brugha, 2000).

In the field of healthcare institution supervision and management, based on the stakeholder analysis method in the health policy analysis by Nandraj et al.(2001), Sunil and colleagues, through empirical research, found that when introducing the certification system, a mechanism to ensure the service quality of medical groups, into the supervision of medical groups, especially private hospitals, it is necessary to fully consider the wishes and demands of major stakeholders such as hospital owners, professional groups, government officials and third-party payment institutions. In the field of health policy implementation process, Javanparast et al. (2009) analyzed and evaluated the perceptions and behaviors of stakeholders at different levels in promoting and implementing primary health care policies to improve child malnutrition in Iran.

Chinese scholars have also increasingly applied stakeholder analysis methods to researches in the health field in recent years. Luo and Jiang (2011) have systematically reviewed the progress of stakeholder theory and its analytical methods in the health field, and proposed the following six steps for the study of stakeholders in the healthcare industry. First, the

stakeholders in the field or organization under research are defined, usually using the broad definition of stakeholders proposed by R. E. Freeman (1984) in strategic management, i.e., stakeholders are any group or individual who can influence the achievement of organizational goals or are affected by the achievement of organizational goals. Second, possible stakeholders are initially described, mainly in terms of their job content, mechanism of action, the role played in the operation of the institution and the status. Third, Mitchell's three-dimensional attribute is used to invite experts in the industry and related fields to rate stakeholders sorted out in three dimensions, namely legitimacy, power and urgency with a five-point scoring method, and to classify stakeholders according to the research results to determine the core stakeholders and expectant stakeholders. Fourth, the core (definitive) stakeholders are characterized in terms of resources, power, position, and knowledge. Fifth, according to the research purposes, the results of stakeholder classification and analysis are used in a targeted and selective manner to formulate corresponding strategies to solve real-world dilemmas. Sixth, the proposed strategies are evaluated. If necessary, a final continued evaluation of the possible impact of the implementation of the developed strategy on the core stakeholders will be carried out.

Based on the analysis of Luo, X. Na et al. (2015) of the Chinese Academy of Medical Sciences conducted a study on the influencing factors of residents' electronic health records sharing based on stakeholder analysis. First, through literature review and questionnaire, the needs of different stakeholders were clarified; then, through the Mitchell three-dimensional scoring method and expert consultation, the stakeholders, the classification of stakeholders and the interests of core stakeholders of residents' electronic health records sharing were defined, analyzing the factors affecting the sharing of residents' electronic health records; finally, policy recommendations for improving residents' electronic health records sharing were put forward from the legal, management, and technical aspects.

J. H. Wu and colleagues (2019) explored ways to optimize stakeholder management from the perspective of the healthcare services ecosystem. Firstly, the latent stakeholders of healthcare service and management are identified by constructing a stakeholder map in the healthcare ecosystem. Then, Mitchell's three-dimensional attribute scoring method is used to score the potential stakeholders identified. Finally, according to the scoring results, stakeholders are divided into three categories: leading group, core group and supporting group, which formed the medical service ecosystem in the context of China. Wu's research presents a complete framework for stakeholder positioning, management strategy formulation, and interest balance mechanism construction in the medical service ecosystem, and also provides certain theoretical support and practical guidance for the optimization of medical services.

The successful application of the previously published stakeholder research method by different scholars shows that the method has high feasibility and practical operability.

2.4 The application of Stakeholder Theory in the sharing and application of healthcare big data

In 2001, Love et al. (2001) introduced the concept of stakeholders when studying data sharing and dissemination strategies in the field of healthcare, and proposed that the general principles of the promotion of healthcare data for the health data initiative include the establishment of standardized databases, the support of relevant policies, links inside datasets and across multiple datasets, continuous development of collection, integration and dissemination technologies, stakeholder support and others. Subsequently, scholars have applied the idea of stakeholders to research related to the sharing and application of healthcare big data.

2.4.1 The application of Stakeholder Theory in the sharing and application of genomic and bio-specimen data

Foster and Sharp (2007) discussed how to distribute the scientific and social benefits of genomic data, and argued that when formulating genomic data sharing policies, we should take into account stakeholders and their interests that are complexly intertwined, rather than developing strategies just base on a single issue and the perspective of a single stakeholder. International Charter of principles for sharing bio-specimens and data stated that shared data and bio-specimens are essential for discovering and creating new knowledge, and for applying the results of various biomedical research to improving the power of diagnosis, treatment, patient care, and health service planning and the health of the general population (Mascalzoni et al., 2015). It also prospectively reviewed emerging issues in this field from the perspectives of patients, healthy volunteers, researchers, biobanks, funders and other stakeholder groups, and provided possible solutions to these issues; Husedzinovic and his colleagues found that the success of biobank-based genome research largely depends on people's willingness to donate tissue specimens. When best practice guidelines for research plan and the recruitment of participants are formulated, the opinions and demands of stakeholders should be considered. Meanwhile, Husedzinovic et al (2015) conducted a systematic review of the views of patients, the public, professionals and other stakeholders on biobank-based genome research; Harza et al. (2018) asserted that the key factors behind the successful sharing of data and specimens from

the National Institute of Child Health and Human Development include the formulation of guiding principles, leadership support and commitment, close coordination between technical and non-technical teams, and continuous participation of dominant stakeholders, and he particularly emphasized the importance of making early contact with stakeholders; in addition, in order to clarify the basic principles, technical challenges, as well as social and cultural challenges of data sharing in biomedical research, Byrd et al. (2020) analyzed in detail the characteristics of stakeholders and their roles in scientific enterprises.

2.4.2 The application of Stakeholder Theory in the sharing and application of medical and public health data

Szirbik et al. (2006) proposed six methodological steps to build medical data warehouse for research. This method emphasizes current trends, including early identification of key needs, data modeling, timely and close interaction with stakeholders, ontology construction of the data warehouse, quality management and exception handling. In the 2012 European Summit on Trustworthy Reuse of Health Data, representatives (Geissbuhler et al., 2013) pointed out that the availability of electronic health record (EHR) triggers and amplifies privacy issues of the person being collected, but shared data is extremely crucial for public health, longitudinal patient care and clinical research. If various stakeholders fail to cooperate on a common policy framework, they will lose important opportunities to promote key European markets (including pharmaceuticals, health technologies and equipment, as well as e-health solutions) and to respond to increasing global competition. S. Bull et al. (2015) systematically reviewed the views of stakeholders on data sharing in medical and public health research, summarized the interests and demands of junior researchers, research participants, research project funders and other stakeholders in data sharing. S. Bull and his research team recommended prioritizing research development and implementation on the opinions of research participants, community representatives, researchers, research ethics committees, and data managers, so as to provide useful information for current policy making and major initiatives. Based on the perspectives and experience of different research stakeholders, Denny and his colleagues (2015) elaborated on the ethics to be followed for public health research data sharing in South Africa. Since then, Carr and Littler (2015) have also carried out similar research from the perspectives of different stakeholders, with the hope to share research data and improve guidelines and practices for public health.

2.4.3 The application of Stakeholder Theory in the sharing and application of clinical trial data

In 2015, Naci et al. (2015) proposed in the guidelines for data sharing in institute of medical research that efforts to strengthen multi-stakeholder collaboration should be made to lead and manage the challenging process of enabling data sharing to become the new norms. In 2019, Miller and his colleagues (2019) published an article entitled *Sharing of Clinical Trial Data and Results Reporting Practices among Large Pharmaceutical Companies*, in which the data sharing measures formulated by researchers were adapted from the 10 important data sharing guidelines written by professional institutions, and were reviewed and further perfected by patients, industry, academia, regulatory agencies, and other stakeholders.

2.4.4 The application of Stakeholder Theory in the sharing and application of scientific research data

Y. Y Chen and Wang (2020) adopted Wheeler model to analyze the role positioning and mechanism of stakeholders that may be involved in the sharing process of data generated in scientific research, systematically sorted out the interest demands of different types of stakeholders in the sharing of scientific research data, and vividly draw the interaction relationship diagram between different types of stakeholders in the process of data opening and sharing, so as to put forward strategies to effectively promote the openness and sharing of scientific research data from the perspective of stakeholders. Sheng and Wang (2019) analyzed the responsibilities and roles of governments, researchers, research institutions, libraries, professional associations and other stakeholders in the process of data openness and sharing based on the open and shared policies of scientific data of international organizations, and proposed that data openness and sharing policies should be formulated based on the above stakeholders to improve the level of data sharing and application.

In addition, Bietz et al. (2016) analyzed the opportunities and challenges of using personal health data (PHD) collected on mobile devices to conduct health research from the perspective of three types of stakeholder groups; Dexheimer et al. (2019) discussed issues related to the sharing of PHD data under protective surveillance from the perspective of stakeholders; Kalkman et al. (2019) systematically reviewed relevant literature and ethical guidelines on the principles and norms of international health research data sharing, finding that there is a convergence among a large number of principles and norms on four themes: social benefits and value; distribution of risks, benefits and burdens; respect for related individuals and groups;

public trust and participation. Costeloe et al. (2018), in their article about data sharing to promote medical development and newborn care, proposed that the steps of newborn data sharing are (1) developing data standards and sharing standards; (2) strengthening interactions with major stakeholders in data sharing; (3) promoting the collection of high-quality datasets. A paper published in the *New England Journal of Medicine* directly pointed out concerns about the current use of data and required stakeholders to publish ethical guidelines for health-related data use (Parasidis et al., 2019).

2.5 Candidate stakeholders and interest demands of the sharing and application of healthcare big data

2.5.1 Candidate stakeholders of the sharing and application of healthcare big data

Accurately defining the stakeholders of the sharing and application of healthcare big data is a key issue in the in-depth study of the strategy of the sharing and application of healthcare big data. Only by correctly defining the concept of stakeholders can we fully understand their interest appeals and their impact on the sharing and application of healthcare big data. The author sorted out the definitions of stakeholders in healthcare big data put forward by Chinese and foreign scholars:

From the aspect of EHR and PHD, Bahga and Madiseti (2013) proposed that stakeholders of EHR include patients, providers, and payers; Leyens et al. (2017) divided the stakeholders of big data used in drug development and personal health care into: manufacturers, regulatory agencies, payers, healthcare providers, decision makers, and researchers; X. Na (2016) defined the stakeholders of residents' EHR sharing as: government departments (National Health Commission, Ministry of Finance, Ministry of Human Resources and Social Security, primary health administrative departments, centers for disease control and prevention), medical institutions (primary medical institutions, hospitals, maternity and child healthcare hospitals), medical staff, residents, medical enterprises (such as information product providers, and medical insurance enterprises); in the *2021 Bulletin of the World Health Organization* pointed out that to address the ethical issue related to PHD collection, use, and sharing, the current and future value of data should be taken into consideration (Dulhanty, 2021). The stakeholders involved in this process include individuals, healthcare providers, public health authorities, government decision makers, and commercial organizations that collect and use medical information. At the same time, the Bulletin also stated that in the medical industry,

PHD should be freely shared among all stakeholders who may benefit from the data. Given that the value of PHD is uncertain but may continue to grow, it is recommended to inform individuals of the potential value before they agree to provide their data. People who collect and use medical data, such as medical practitioners and researchers, public health departments, governments, and commercial entities, can make vulnerable groups aware of the potential uses of their data. Further communication between stakeholders helps individuals to be treated fairly regarding the value of their information.

From the aspect of clinical research, Kim et al. (2014), when discussing the data governance requirements of a distributed clinical research network, proposed that the network often entails complex collaboration, involving interactions among many stakeholders such as federal and state governments, research centers and universities, commercial entities, healthcare organizations and patient groups; the conference report of the American Heart Association Data Summit in 2015 divided stakeholders of data acquisition, analysis and sharing in the field of cardiovascular and stroke science into: patients, basic scientists, clinical researchers, public health researchers, clinicians and healthcare system managers, as well as industry and regulatory agencies (Antman et al., 2015); Mazor et al. (2017) divided the stakeholders of multicenter research into five categories: patients, researchers, institutional review boards and supervisors, multi-center research governance experts, and healthcare system leaders.

From the aspect of bio-specimen data and public health research, Lemke and Harris-Wai (2015) explored how to take advantage of stakeholders' participation to formulate policies and guidelines for specific genomics research and public health research, and proposed that the main stakeholders of genomics research include: different patient groups, research participants, the public, providers, researchers, advocacy groups, payers, and decision makers; Laird et al. (2020) divided stakeholders of healthcare big data in the field of public health research into five categories: academics, practitioners, policy-makers, knowledge brokers and a funder.

In addition, we sorted out the definitions of stakeholders in the medical field:

Jin et al. (2013) analyzed the stakeholders of China's medical treatment partnerships and defined them as government, core hospitals, non-member hospitals, other member hospitals, and patients.

Guo et al. (2014) analyzed and evaluated the stakeholders of the new rural cooperative medical scheme. He defined its main stakeholders as government, non-designated medical institutions, designated medical institutions, farmers, and pharmaceutical manufacturing and operation enterprises, and analyzed the interest appeals of various stakeholders.

Lei et al. (2015), by conducting an empirical study on the interests of stakeholders in two-

way referrals, and divided the stakeholders in two-way referrals in Dongguan city into three categories, that is, government, medical institutions, and individuals. At the government level, government, medical insurance departments, health administration departments, and hospital management centers are included. At the medical institution level, hospitals and community health service institutions are involved. At the individual level, the management of hospitals and community health service institutions, physicians of hospitals and community health service institutions and residents are covered.

X. L. Xu et al. (2017) combed the definitions of stakeholders in “two-way referral”, “medical service integration”, and “medical group” put forward by Chinese scholars from 2009 to 2016, and preliminarily assumed 16 categories of subjects as stakeholders who participate in the division of labor and collaboration of medical institutions, namely, government, health commissions, medical insurance departments, managers of urban general hospitals, physicians in urban general hospitals, nurses in urban general hospitals, managers of ordinary secondary hospitals, physicians in ordinary secondary hospital, nurses in common general secondary hospital nurses, managers of primary medical institutions, physicians in primary medical institutions, nurses in primary medical institutions, drug supply departments, equipment supply departments, and patients (residents). Subsequently, through a questionnaire survey using 50% as the selection criteria, he defined the main stakeholders involved in the division of labor and collaboration of medical institutions as government, health and family planning commissions, medical insurance departments, managers of urban general hospitals, physicians in urban general hospitals, managers of ordinary secondary hospitals, physicians in ordinary secondary hospital, managers of primary medical institutions, physicians in primary medical institutions, and patients (residents) (X. L. Xu, 2018).

This research mainly takes the sharing and application process of healthcare big data of a medical university and its affiliated hospitals in Guangdong as the example, and according to their different role and mechanism of action, preliminarily assumed that the stakeholders that may be involved in the sharing and application of healthcare big data mainly include source data subjects (such as patients receiving medical services, participants in clinical trials, large population cohorts and biobanks, individuals participating in physical examination and users of mobile medical equipment and software), derived data subjects (such as medical staff, scientific researchers, department directors, hospital information center managers, hospital and college managers, healthcare system, developers and operators of big data platform and application software, organizations providing biospecimen testing, and online storage operators), and data monitoring entities (governmental departments, health commissions, medical insurance

departments).

2.5.2 The interest appeals of candidate stakeholders of the sharing and application of healthcare big data

Patients, hospital medical practitioners performing physical examination, and volunteers of medical science research are not only the producers of healthcare big data, but also the most direct and main beneficiaries of healthcare big data. The source data subjects expect to use healthcare big data platform and software with the premise of fully protecting their legal rights and interests such as property and personality. In this way, they can realize online appointments, dynamic inquiry of diagnosis and treatment information and health conditions, online consultation, doctor-patient interaction, and mobile payment (Fan et al., 2016; S. S. Ma et al., 2018), enjoy more active, accessible, high-quality, continuous and full-cycle health management services, improve diagnosis and treatment efficiency, and reduce medical costs.

The interest appeals of **medical staff**: as the most direct collector and user of healthcare big data, medical staff formulate precise and individualized diagnosis and treatment plans to meet diversified and multi-level health demands, based on patients' living habits, disease history, clinical symptoms, laboratory tests, imaging inspections, device diagnosis results, as well as biological information such as genome and proteome; they take advantage of the sharing and application platforms of healthcare big data to promote the realization of multidisciplinary diagnosis and treatment models, and make follow-up visit to patients to provide patients with healthy behavior guidance and remote intervention to improve the efficiency and quality of diagnosis and treatment and increase patient satisfaction (Fan et al., 2016; S. S. Ma et al., 2018).

The interest appeals of **scientific researchers, department directors and information center managers**: through the integration of medical clinical databases and scientific research databases, they evaluate the clinical effects of different diagnosis and treatment plans, explore disease development mechanisms, and create new drugs, vaccines and medical devices so as to provide strong support for the formulation and update of diagnostic guidelines for infectious diseases and common chronic non-communicable diseases, the creation of comprehensive treatment plans, and the development of new diagnosis and treatment technologies, and to promote the verification and popularization of key medical technologies and medical scientific and technological achievements; they can use machine learning and natural language processing, deep learning and other technologies to build artificial intelligence-assisted diagnosis technology so as to scientifically and accurately guide clinical decision-making (Elliott et al., 2015; Song et al., 2021; L. X. Zhang et al., 2018).

The interest appeals of **hospital and college administrators**: healthcare big data in medical institutions covers all aspects of medical services and operations. The sharing and application of healthcare big data plays an important role in optimizing work processes, improving working efficiency, standardizing diagnosis and treatment behaviors, controlling and reducing diagnosis and treatment errors, and promoting refined management of medical services; meanwhile, it can help dynamically understand business volume and business expenses of hospitals, rationally allocate and integrate hospital resources, and improve scientific decision-making capabilities and overall management capabilities; it can also be used to carry out remote consultations, remote demonstration, remote surgery, and multi-center smart medical services to facilitate and benefit the public, promote the coordinated development of multiple medical research models, and bring good revenues and reputation to the hospital (Song et al., 2021; L. X. Zhang et al., 2018).

The interest appeals of developers and operators of healthcare platform and application software, organizations providing biospecimen testing, and online storage operators: such derivative stakeholders are stakeholders that mainly aim to make profits. Their main legitimate interest appeals is to obtain direct and objective economic benefits by providing technical services such as launching, operating and maintaining online systems or specimen testing; besides, they also expect to gain praise from users to increase their enterprises' reputation and popularity; in addition, some enterprises hope to have rights to control and benefit from data information, and to use data for production, research and development, as well as innovation, so as to obtain greater commercial benefits and increase the market value of their enterprises (H. Q. Li et al., 2019; Littlejohns et al., 2020).

The interest appeals of **governmental departments, health commissions, and medical insurance departments**: the data monitoring entity's reasonable development and application of healthcare big data can provide medical and health care decision-makers with a scientific and effective support for policy formulation and implementation under supervision, optimize the allocation of medical and health resources, improve medical security system, enhance governance capabilities and levels in the field of health and medical care, maintain public health, improve departmental performance, and increase social reputation and public satisfaction (Berg, 2017; Song et al., 2021; L. X. Zhang et al., 2018).

2.6 Concept, research process, and application of in-depth interviews

2.6.1 Concept of in-depth interviews

In-depth Interview is one of the most commonly used, important and effective methods in qualitative research. According to the research topic and the pre-set interview outline, researchers have in-depth and detailed conversations and interactions with interviewees, observe their expressions and reactions in the process, encourage their divergent thinking on the basis of understanding the interview questions, and based on their own experiences, provide in-depth and detailed descriptions and explanations of the questions raised by researchers. Through in-depth interviews, researchers can deeply understand the experience, focus, views and attitudes of certain stakeholder groups, explore the formation process of certain social phenomena and research pain points related to the research theme, and propose ideas and effective programs to solve research pain points. The outline of the in-depth interview is usually semi-structured, which means that the researcher can prepare some questions in advance according to the research theme for use, and can also flexibly adjust and questions according to the interviewee's response during the interview (Brinkmann & Kvale, 2018; Gerson & Damaske, 2020; Knott et al., 2022).

The main advantage and purpose of the in-depth interview is that it can provide detailed and rich information, which is conducive to in-depth and detailed exploration of complex topics, so as to clarify the reasons and solutions for the formation of research pain points from a deep and multi-angle perspective. Therefore, researchers of in-depth interviews are more concerned about the quality of interviews rather than the quantity of interviews.

To carry out an effective in-depth interview, attention should be paid to the interview design steps (especially the interview outline), the selection of the person in charge of the interview, the collation and analysis of interview data, ethical rules of in-depth interview, and quality control and evaluation of in-depth interview (Brinkmann & Kvale, 2018; Gerson & Damaske, 2020).

2.6.2 Research process of in-depth interviews

2.6.2.1 Research steps for in-depth interview

(1) Preparation before the interview

An interview research team should be set up, including main researchers, interviewers, interview recorders, results analysts, and report summarizers. After comprehensive

consideration of the research purpose, research topic, pain points and potential interviewees, the appropriate interview mode is selected and the interview schedule is formulated. An appointment with the interviewee is made about the time and location, and a suitable and relatively undisturbed space is selected as the interview venue. The informed consent form of the interview is drawn up to effectively inform the interviewee of the purpose of the interview and the scope of use of the interview data.

(2) Formulating the interview plan

The execution of the interview plan should mainly consider the following parts:

① The interview outline and plan should be formulated based on the research topic and purposes, and at the same time combined with the background of the interviewees. The questions set in the interview outline should have a certain logic and hierarchy. If necessary, experts and scholars in the research field can be consulted to improve the interview outline according to their opinions and suggestions. In addition, in the in-depth interview, some semi-structural focused questions can be designed and tested in combination with the means and strategies for building social theories. The interviewees are encouraged to make responses based on their own experiences and real views, and the interview outline and interview plan are added, deleted, revised and improved according to the feedback of the interviewees, so that the interview outline is relevant and relevant in content, thus improving the effect of the interview. The interview outline should not be a rigid list of questions, but should be carefully designed, and incorporate certain interview skills in the arrangement of interview questions. For example, at the beginning of the interview, it can provide some questions that can gain the emotional recognition of the interviewee, and then set some inspiring open questions, and gradually focus on the main content of the interview outline and research theme in the subsequent in-depth process, to improve the efficiency and quality of the interview (X. M. Chen, 2000; Z. M. Jia, 2015).

② The time, personnel and equipment of the interview. The determination of the time of the interview should not only consider the duration required to answer the questions listed in the interview outline, but also fully consider the accessibility and cooperation of the interviewees, to ensure that the interviewees have enough undisturbed time to for the interview and to obtain detailed and complete information. It should ensure that there are sufficient personnel in the project team to conduct interviews and have relatively complete interview skills. At the same time, it is also necessary to prepare the necessary interview tools and equipment, such as voice recorders and cameras.

③ The purpose, coverage and number of interviewees of the interview. The purpose and objectives of the in-depth interview should be clarified according to the research topic and pain points. The determination of the coverage of the interview and the number of respondents should not only consider the overall distribution of the concerned stakeholder groups, but also pay attention to the representativeness of the included respondents. Based on the principle of combining theoretical sampling and purposive sampling, it selects certain people who are relatively balanced in terms of age, nature of work unit, years of service, position and title, and who are familiar with the research field as the target group for interview. The number of respondents depends on the research objective, the researcher's resource scheduling ability and the principle of theoretical saturation. At the same time, it is also necessary to find the key informant in the interview.

(3) Training of interview skills

The improvement of interview quality and efficiency requires interviewers to have certain interview skills, such as establishing a harmonious and friendly interview atmosphere, giving interviewees sufficient freedom and relaxed communication space, focusing on the research theme and appropriately guiding interviewees to engage in divergent thinking during the interview process and open up to communicate and interact with interviewers to express their views truthfully and objectively. Therefore, it is necessary for experienced professionals to train the interview skills of the entire interview team, and improve the quality and efficiency of the interview by mastering enough interview skills, so as to obtain comprehensive, complete and useful information as much as possible (Gerson & Damaske, 2020).

(4) Ending the interview

The duration of the interview should be flexibly controlled according to the actual situation, to obtain effective and complete information, without hindering the normal work and life of the interviewee. Sometimes, interviewees only have a limited amount of time for the interview, so it is important to use this time to guide interviewees to provide meaningful information. After the interview, the interview record should be summarized as soon as possible, including the interview notes, and audio or video recordings, and the key points in the interview process should be reviewed and checked to ensure the accuracy and completeness of the interview information. At the same time, it is necessary to communicate the progress of the interview with the research team in time to ensure the consistency and reliability of the interview data. In addition, if it is found that the main points in the interview outline need to be further supplemented during the interview, it is necessary to communicate and discuss with the research team in time and determine whether additional interviews are needed.

2.6.2.2 Selection of the person in charge of in-depth interviews

The person in charge of the interview is critical to the organization of the interview, process control, and ultimately the overall quality of the interview. Interviewers need to be screened based on the following specific requirements:

(1) Background requirements

The interviewers should have certain interview experience and communication skills, cultural level and educational experience that match the research project, and be good at interpersonal communication and expression. A relatively close work, education background or similar work and life experience with the interviewee are preferred.

(2) Good at listening, appropriate probing, and timely questioning

A good interviewer is not only good at listening, but also gives positive feedback, and encourages interviewees to fully and freely express their thoughts and feelings. When encountering incomprehensible content, the interviewer should be decisive and promptly ask the interviewee to explain key terms or words to ensure that there is no misunderstanding or information bias. If there is ambiguity in meaning, the interviewee can be requested to express themselves clearly in a different way and provide further explanation and clarification on the content that may cause misunderstandings. When the interviewee finds it difficult to answer a question, if the question is only one that the interviewee has not previously considered but has a willingness to answer, appropriate guidance can be given and sufficient room for thinking can be provided; if it is a question that the interviewee avoids discussing, the interviewer can provide appropriate guidance or probing under the premise of grasping the sense of proportion. The interviewer also needs to follow the important principle of timely questioning. That is, the interviewer can organize the language to ask interviewees about their previous opinions or behaviors based on their cognition and insights, to confirm that there is no misunderstanding. At the same time, the interviewer can also know the interviewee's views on the research question, the formation process, and the reasons for its formation to a certain extent (X. M. Chen, 2015; Gerson & Damaske, 2020).

(3) Recording interviews

In the interview process, it is necessary to record important situational content and non-natural language information such as the body language, movement and expression in time. In addition to preparing the recording work in advance, it is best to have a full-time reporter to record the interview process in detail. This can not only ensure the interview data collation and information extraction based on the understanding of different members of the research team

in the interview process, but also ensure that the main interviewers can concentrate on listening to the interviewee's narration and lay the groundwork for the next proposal of interview questions, guidance and questioning (B. G. Glaser, 1998).

2.6.2.3 Data collation and analysis of in-depth interview

In qualitative research, coding is the core process of data analysis and information extraction. After the face-to-face interview, the researcher needs to classify, summarize, and explain the data obtained during the interview process with relatively simple names. In general, interviewers record the information obtained during the interview on data analysis cards or memos, and then classify and file them through coding. The coding process shows how the acquired data is selected, distinguished, classified, and analyzed (Charmaz, 2009).

(1) Data coding

The coding methods mainly include open coding, axial coding and selective coding (Strauss & Corbin, 1990). Open coding starts from scratch and encodes the collected information according to the properties and topics of the data itself. In the process of open coding, researchers need to think outside the original theoretical stereotypes, and understand and analyze materials with an open mind. Through open coding, researchers are able to discover new concepts and relationships that lead to a deeper understanding of the research object. Axial coding requires the researcher to encode and organize the scattered source data collected by discovering the internal connections between primary and secondary concepts. In axial coding, researchers first select a core concept or core topic, and then connect other related concepts and topics to this core to form an organic coding structure. Axial coding can help researchers organize and understand more complex data. In selective coding, the researcher needs to choose a unified topic as the focus of coding according to the research purpose and research problem, and then encode the collected original data under the unified theme framework (Creswell, 1998). Through selective coding, researchers can deeply analyze and understand the collected data in a more targeted and focused way, and extract key information closely related to the research theme and research question (X. M. Chen, 2000). The above three coding methods are usually used interchangeably during the research process. The researcher can start with open coding, then proceed to axial coding according to the discovered concepts and relationships, and finally proceed to selective coding. Through the systematic application of these coding methods, researchers can summarize, organize and analyze data, extract representative and important results, and construct a theoretical framework that reflects the research object and research question.

(2) Data classification

It refers to the classification of meaningful units on memos or data cards, such as ideas, words, phrases, and behavior patterns, according to different subject categories in order to conduct in-depth analysis and understanding of data. The subject categories of the classification consist of three levels: central themes, situated themes, and general themes. The central theme is the articulation of a single discourse through declarative sentences, which can be an important concept, idea or experience mentioned in the interview. By categorizing and analyzing the central theme, it is possible to gain a deeper understanding of the respondent's thoughts and experiences. A situated theme is a combination of several related central themes to form a situational or meaningful expression. It can provide a deeper and more comprehensive understanding, which helps to accurately grasp the situation of the respondent. A general theme is a combination of multiple contextual themes to reflect or express the essence of the problem. It plays a role in summarizing the overall research problem and research objective, and helps researchers to propose ideas and methods to solve the problem.

(3) Inductive analysis

Inductive analysis is the process of finding commonalities from specific data, integrating and theorizing them, and gradually abstracting from phenomena to concepts. The most common inductive analysis methods used in qualitative research include the funnel method and modification analysis method. The funnel method is a process of continuously narrowing down a large range of data for analysis. The first step is to comprehensively and completely collect information and data related to the research theme, and the analysis objectives are clarified in the process of data collection. The second step is to gradually screen the collected data according to the analysis objectives, so as to focus the collected data in a relatively narrow range that can reflect special phenomena. The third step is to conduct an in-depth analysis of the remaining data after screening and the specific problems and phenomena reflected, so as to make a general conclusion. Unlike the funnel method, which requires extensive data collection in advance, the modified analysis method first proposes an initial pattern based on a small amount of data, and then continuously revises it through the new data added. The specific steps are as follows. Firstly, starting from the existing data related to the problem of concern, it analyzes and summarizes a possible applicable model. Secondly, it adds new data and determines whether it conforms to the model summarized in the earlier stage. If there is any inconsistency, it amends the model summarized in the earlier stage. Then, it continuously adds new analytical data, and repeats the process of the second step until the summarized and generalized comprehensive model can reasonably and broadly explain the specific issues and

phenomena that the researcher is concerned about.

2.6.2.4 Principles for in-depth interviews

In-depth interviews require researchers to fully respect and protect the privacy, security and legitimate rights and interests of the interviewees, and abide by some basic rules and norms, so as to mobilize the enthusiasm and cooperation of the interviewees and enhance their support for the research project and trust in the researchers. In-depth interviews should follow the following principles:

(1) Principle of informed consent

The informed consent principle is the first principle that research teams conducting in-depth interviews must follow. It requires that the interview process must fully respect the willingness of the interviewee. Before the formal start of the interview, with sufficient data collection, the interviewer should inform the interviewee of the research purpose and the scope of application of the collected information, to ensure the interviewee's right to know the research content and purpose to the maximum extent, so that they can understand the project rather than sharing their own experience without knowing or being deceived. Meanwhile, it is necessary to obtain the oral or written consent of interviewees to ensure that they voluntarily participate after fully understanding the study. Before the interview, the interviewer should also clearly inform the interviewee that he or she has the right to refuse the interview and can refuse to answer any questions he or she does not want to answer during the interview. After the survey results are formed, it is also advised to share the main findings and research results with the interviewee.

(2) The minimum harm principle

In-depth interviews should follow the minimum harm principle. Under normal circumstances, the execution of an in-depth interview will not cause physical harm to the interviewee. However, in some unexpected cases, due to some unexpected reasons or the interviewer's negligence or even aggressiveness in the process of questioning or probing, the interviewee may suffer mental or psychological harm, which is more likely to occur when interviewing some special and relatively vulnerable groups. Therefore, when conducting in-depth interviews, interviewers should abide by professional ethics, reasonably assess the possible negative effects of the interview process, and consider minimizing potential harm through skills or acceptable wording when designing the interview. During the interview, it is necessary to pay attention to the sensitivity, consider the situation to avoid stimulation, and take certain measures to eliminate possible potential negative effects after the interview.

(3) The anonymity and confidentiality principle

To avoid the potential harm to the interviewee caused by the accidental disclosure of interview information, researchers and research teams must always bear in mind and follow the anonymity and confidentiality principle during in-depth interviews, to protect the privacy, security and legal rights of respondents. The research established a corresponding index system of the interviewee's name and number, which authorized some researchers to use through document protection. In the process of in-depth interviews and data collation and analysis, the number was used to replace the sensitive information such as the interviewee's name. In terms of healthcare big data, it is also necessary to keep in mind the desensitization and privacy protection of respondents' personal information. For the sensitive information obtained in the interview process that cannot be effectively controlled in advance, the researcher should keep in mind the requirements of professional ethics, and take protective measures to communicate and use only within the scope of authority. Other members of the research team should abide by the confidentiality principle and must not disclose the interview information to irrelevant personnel, which may cause information diffusion and dissemination. The anonymity and confidentiality principle should also be implemented in subsequent data analysis, information extraction, result condensation, thesis writing and even after the end of the research.

2.6.2.5 Quality control and evaluation of in-depth interviews

In in-depth interviews, the researcher is the research tool himself (Guba & Lincoln, 1981). The key to the effectiveness of the research lies in the researcher's ability, research skills and rigorous execution in the operation process. Researchers and their team members will restructure the behavior and thinking of the subject. In general, quality control and evaluation of in-depth interviews can be carried out from the following aspects: ① research value; ② the clarity and specificity of the research problem; ③ the rationality and feasibility of the research design; ④ research background; ⑤ the sampling method; ⑥ data collection and analysis; ⑦ the introspection about the interpreted results; ⑧ the scope of reference.

Different from the random sampling used in mathematical research, in-depth interviews usually adopt nonrandom sampling to flexibly select interviewees according to research objectives and realistic factors. In in-depth interviews, common nonrandom sampling methods mainly include:

(1) Purposive sampling. Researchers conduct subjective analysis of potential interviewees and their coverage according to the research theme and objectives, so as to select and determine the interviewee. As mentioned above, the purpose of the in-depth interview is to have as much in-depth and detailed communication as possible with the interviewees, guide them to engage

in divergent thinking based on their own experience, and express their views and opinions fully, freely, comprehensively, completely and accurately. Therefore, in in-depth interviews, researchers focus more on the high correlation between the interviewees and the research theme and the quality of the interview execution, rather than excessively pursuing a sufficiently large sample size. In the process of in-depth interviews, the researcher will use the purposive sampling to include key interviewees who are highly related to the research theme and research objective and require special attention, to obtain useful information related to the research theme to the maximum extent.

(2) Diversity sampling. Compared with mathematical studies such as statistics, the sample size of in-depth interviews is generally smaller. Due to the limitation of sample size, after purposive sampling, researchers of in-depth interviews usually follow the principle of diversity to select interviewees. Thus, the interviewees cover different ages, occupations, education, work experience, educational background and other levels, which facilitates researchers to listen to different voices and opinions, improving the representativeness of the research objects and the reliability of the research results. In addition, by analyzing groups with different characteristics at multiple levels and dimensions, it can identify the diversified views and opinions of different groups on the same academic problem and social phenomenon, thereby providing a basis for reasonable solutions to problems based on different groups.

(3) Snowball sampling. After identifying some interview subjects through purposive sampling and diversity sampling, the researcher can request the included interviewees to recommend some familiar personnel who meet the overall requirements of the research objectives to be interviewed according to the characteristics and number of the included interviewees, combined with the research theme and research purpose. In this way, the sample continues to expand like a snowball. Snowball sampling allows researchers to collect rich and diverse information from a sufficient sample size. Snowball sampling is especially suitable for the investigation of a small number of specific groups, but the samples included by snowball sampling are often too homogenous. Therefore, in general, snowball sampling is not recommended to be used alone. Instead, it is recommended to be used when the sample size needs to be supplemented after purposive sampling and diversity sampling, to ensure that the interviewees have certain representativeness and wide coverage while ensuring their criticality and high correlation.

2.6.3 Application of in-depth interviews

After determining the key stakeholders in the sharing and application of healthcare big data,

this study adopts in-depth interviews to construct a strategy for the sharing and application of healthcare big data, starting from the research dilemma and the characteristics of the research problem itself. There are many key stakeholders that affect the sharing and application of healthcare big data, and different types of stakeholders have different interest demands, statuses, values and roles, and they promote or hinder healthcare big data sharing and application through competition, cooperation, conflict and confrontation. Under different historical backgrounds and practical states, the values, status, roles and characteristics of the same type of stakeholders are not static, but change dynamically with the practical process. This research problem involves a wide range of topics, with intertwined interests and a certain degree of complex derivative. Under the guidance of the *Healthy China 2030 Plan*, it actively promotes the development of “Internet + Medical health”, combines digital technology and digital thinking, and promotes healthcare big data sharing and application through the traction of digital reform.

Since most studies focus on the analysis of the factors that promote or hinder the sharing and application of healthcare big data from a macro perspective, there is a certain distance between the strategies and solutions proposed to promote data sharing and application and the practical application. An in-depth interview is a major qualitative research method, and the main conclusions obtained through it have strong interpretability. Besides, an in-depth interview takes into account the characteristics of focusing on the research theme and divergent questioning, often resulting in unexpected discoveries for researchers. In the field of healthcare big data research, the qualitative research method of in-depth interview helps to obtain stakeholders’ views and opinions on the current development status, the implementation degree of relevant policies and laws and regulations, and existing problems and constraints from multiple perspectives and at a deep level, which are usually difficult to obtain through quantitative methods, but need to be obtained through qualitative research. Through in-depth interviews, this study will conduct in-depth and detailed exchanges with key stakeholders in the sharing and application of healthcare big data, and guide them to comprehensively and truthfully express their views on issues related to the sharing and application of healthcare big data based on their work and life experiences, clarify the reasons and key influencing factors that hinder the sharing and application of healthcare big data, and put forward ideas and effective solutions for the research pain points.

This study will follow the normative requirements and research procedures of in-depth interviews. Firstly, it will delve into the key stakeholder groups that affect the sharing and application of healthcare big data, conduct in-depth interviews according to the preset interview outline to obtain the original data, and conduct a “deep description” of the event, to analyze the

essence and details of the event. Secondly, it encodes and transforms all kinds of data scientifically and systematically, and displays the characteristics of things, the intrinsic nature of events and the nature of behaviors contained in the data analysis as far as possible, so as to form an analytical framework for the events. Finally, this study combines the preliminary analytical framework to return to the data, and achieves theoretical saturation and a reasonable framework through continuous comparison, enrichment, corroboration and improvement of the theoretical model. On the basis of effectively identifying the key stakeholders involved in the sharing and application of healthcare big data in a medical university and its affiliated hospital, it is expected to reveal the fundamental causes and mechanisms affecting the sharing and application of healthcare big data through qualitative analysis at a deep level and from multiple perspectives, so as to make an effective strategy to remove the obstacles for the sharing and application of healthcare big data.

To sum up, the research steps to be adopted in this thesis are as follows: through literature research, the thesis systematically sorts out possible candidate stakeholders in the sharing and application of healthcare big data in a medical university and its affiliated hospital from a relatively broad perspective; through expert consultation, major stakeholders were selected according to the 50% election rate, and their roles and interest demands are sorted out; Mitchell three-dimensional attribute scoring method is used to identify the core stakeholders who play a key role in promoting or hindering the sharing and application of healthcare big data in the medical university and its affiliated hospitals through questionnaire survey results; conducting in-depth interviews with dominant stakeholders, exploring the core demands among dominant stakeholders and the key influencing factors and obstacles for the sharing and application of healthcare big data, and conducting effective strategies to promote the sharing and application of healthcare big data.

[This page is deliberately left blank.]

Chapter 3: Research Method

3.1 Research idea

Taking a medical university and its affiliated hospital in Guangdong Province as an example, this thesis takes effective promotion of the sharing and application of healthcare big data as the goal orientation, focuses on the key stakeholders of the sharing and application of healthcare big data, and follows the path of “identification of health care data sharing and application of major stakeholders → determination of the key stakeholders → in-depth interview with key stakeholders → health care data sharing and application of strategies construction around the core demands of key stakeholders and key influencing factors of the sharing and application of healthcare big data”. To this end, the research firstly expounds the related concepts, origin and development of stakeholder theory, and their application in the field of medicine and healthcare big data through theoretical research methods, and analyzes the feasibility and rationality of introducing the above theories into this study. Secondly, literature review, expert consultation, questionnaire survey and other methods are used to identify and determine the main stakeholders and key stakeholders that affect the sharing and application of healthcare big data. Thirdly, the in-depth interview is used to clarify the core demands of key stakeholders and analyze the decisive factors affecting the sharing and application of healthcare big data. On this basis, countermeasures and suggestions are proposed to promote the formation of healthcare big data sharing and application ecosystem with the active participation and coordination of all key stakeholders.

3.2 Research content

The researcher focuses on the stakeholders and their attribute identification in the process of the sharing and application of healthcare big data, the research on the influencing factors of the sharing and application of healthcare big data from the perspective of key stakeholders, and on the implementation strategy of the sharing and application of healthcare big data. Therefore, the main content of this research is as follows.

3.2.1 Theoretical research

The purpose of this research is to elaborate and analyze the definition, origin, development and application of relevant theories involved in this study, and to discuss the impact of stakeholders on the sharing and application of healthcare big data as well as the application of Evolutionary Game Theory in the sharing and application of healthcare big data. This thesis lays a theoretical foundation for the construction of the sharing and application strategy of healthcare big data through the evolutionary game of key stakeholders from the perspective of stakeholders.

3.2.2 The identification of key stakeholders involved in the sharing and application of healthcare big data

The research aims to sort out and analyze the key stakeholders involved in the sharing and application of healthcare big data, and identify the key stakeholders among numerous stakeholders with complex interactions and relationships, then takes the key stakeholders involved in the sharing and application of healthcare big data as the context throughout the research. Firstly, the candidate stakeholders participating in the sharing and application of healthcare big data are comprehensively found through literature review. Secondly, the main stakeholders that affect the sharing and application of healthcare big data are selected based on expert consultation with an enrollment rate of 50% as the standard, and the mechanism of the participation of each stakeholder in the sharing and application of healthcare big data is analyzed. Then, a questionnaire is set, according to Mitchell's three-dimensional attribute classification method and the 5-point Likert scoring method, the main stakeholders selected by the invited industry experts and leaders are scored and tested from the three attribute dimensions of legitimacy, power and urgency, so as to conduct statistical analysis on the scoring results and then identify the key stakeholders affecting the sharing and application of healthcare big data.

3.2.3 Research on influencing factors of the sharing and application of healthcare big data from the perspective of key stakeholders

This research aims to clarify the selection process and influencing factors of behavioral strategies for key stakeholders to participate in the sharing and application of healthcare big data. The interview outline and interview plan are designed from six aspects, including the foundation and status quo of the sharing and application of healthcare big data, benefits and risks, the attitudes of key stakeholders towards healthcare big data sharing, legal guarantee for the sharing and application of healthcare big data, data security and privacy protection, and

problems and suggestions in the process of the sharing and application of healthcare big data. In-depth interviews are conducted with key stakeholders involved in the sharing and application of healthcare big data.

Based on the perspective of key stakeholders, this thesis discusses in detail the status quo and key influencing factors of the sharing and application of healthcare big data in the context of current medical practice, as well as the roles and core needs of various key stakeholders in the process of data sharing and application, clarifies the obstacles and problems in the sharing and application of healthcare big data, which provides a basis for clearing the barriers of the sharing and application of healthcare big data and building an effective plan to promote the the sharing and application of healthcare big data.

3.2.4 Strategy construction of the sharing and application of healthcare big data

Based on the results of the above theoretical research, empirical research and qualitative research, the research aims to focus on the goal of the sharing and application of healthcare big data, propose specific strategies to promote the effective sharing and application of healthcare big data, form an ecosystem of the sharing and application of healthcare big data with the active participation, coordination and mutual assistance of various key stakeholders, give full play to the intrinsic value of healthcare big data, promote the medical university and its affiliated hospitals to make major breakthroughs in medical scientific research, constantly improve the diagnosis and treatment capacity and service level, and promote medical innovation and the transformation of scientific and technological achievements.

3.3 Research methods

3.3.1 Literature review and theoretical research method

The definition, origin and development of Stakeholder Theory involved in this research are reviewed and elaborated in detail. The development and application of Stakeholder Theory in the field of healthcare big data are comprehensively presented by generalizing, sorting and summarizing the present relevant research on the sharing and application of healthcare big data, Stakeholder Theory in the existing literature, which provides important theoretical support for in-depth interviews with key stakeholders that affect the sharing and application of healthcare big data, and the further discussion of the sharing and application strategies of healthcare big data from the perspective of stakeholders in the research.

3.3.2 Case study method

Taking a medical university and its affiliated hospitals in Guangdong Province as the research subject, this study first analyzes the potential stakeholders and their mechanisms of healthcare big data sharing within the research subject, and then combines purposive sampling and random sampling to take the stakeholders within the research subject as the main survey objects, and extends to government departments, health commissions and medical insurance departments. Through questionnaire surveys, expert consultations and in-depth interviews, the study determines the core demands of key stakeholders of the research subjects, and proposes strategies for effective sharing and application of healthcare big data of the research subject and similar organizations. During the research process, every step of the study should be recorded in detail as much as possible to make the whole research process repeatable and ensure the reliability of the research results (X. M. Chen, 2000; Yin, 2017).

3.3.3 Expert consultation

Aiming at all potential stakeholders who have an impact on the sharing and application of healthcare big data sorted out in the early stage, a structured electronic consultation questionnaire is designed according to the results of literature research and theoretical research. And the method of combining theoretical sampling and purpose sampling is adopted to 50-60 managers, experts and scholars with deep experience in relevant fields for consultation invited from Health Commission of Guangdong Province, Guangdong Provincial Medical Insurance Bureau, Guangdong Provincial Drug Administration, and a medical university and its affiliated hospitals in Guangdong Province, thus screening out the main stakeholders in the field of healthcare big data with the selection ratio of 50% as the standard.

The content of consultation consists of two parts: (1) basic personal information of the interviewed experts and managers, mainly including gender, age, employer, position, professional title, duration of employment and familiarity with stakeholders in the sharing and application of healthcare big data; (2) Screening of main stakeholders: the electronic questionnaire will include a Yes/No question “Do you think the following are stakeholders in the sharing and application of healthcare big data?”.

The interviewed experts should include at least officials of the National Health Commission of the PRC, presidents of public hospitals and directors of relevant administrative departments, senior medical workers, and senior experts and scholars in scientific research institutes or universities. According to the screening results of experts, scholars and managers, 50% is the

cut-off value. Those whose selected ratio is of 50% or above are the main stakeholders and included in the following questionnaire survey to determine the key stakeholders. See Table 1 for details (X. L. Xu, 2018).

3.3.4 Questionnaire survey and expert grading method

The state of stakeholders is dynamic. In different states, the role, characteristics and status of the same stakeholder will change. Based on the above theoretical research and expert consultation and Mitchell three-dimensional attribute classification method, industry experts and scholars and related practitioners will be widely invited to score stakeholders involved in the sharing and application of healthcare big data from three attribute dimensions of legitimacy, power and urgency with the 5-point Likert scoring method. The software R is used to conduct a descriptive analysis of the results of the questionnaire survey and a paired samples T-test (X. L. Xu, 2018) with statistical methods. Then, by referring to the score segmentation method of H. H. Chen and Jia (2004), stakeholders are divided into different segments in each dimension. Then statistical tests will be conducted to identify the key stakeholders involved in the sharing and application of healthcare big data and to lay the foundation of the in-depth interview with key stakeholders and the establishment of the sharing and application of healthcare big data.

The structured electronic questionnaire is composed of two parts: (1) the basic personal information of the respondents, which is consistent with that in the Expert Consultation Form; (2) score the main stakeholders of the sharing and application of healthcare big data identified above from the three attribute dimensions. The questionnaire adopts the 5-Point Likert scale, according to which the higher the score, the higher the legitimacy, power and urgency of stakeholders involved in the sharing and application of healthcare big data. See Table B.1 for details.

3.3.5 In-depth interview method

In-depth interview is the most commonly used and the most important effective method in qualitative research. In the interview process, it usually uses open questions to directly engage in face-to-face and detailed communication and interaction with the interviewee, and deeply understands the interviewee's views and opinions on specific issues through questioning and guidance, so as to more comprehensively and accurately reveal the interviewees' real thoughts, ways of thinking and behavioral motivations, and then to better understand the essence, internal mechanism and related influencing factors of the research problem (Brinkmann & Kvale, 2018;

Gerson & Damaske, 2020; Knott et al., 2022).

In the practice of the sharing and application of healthcare big data in the medical university, the qualitative research method of in-depth interview is helpful to obtain core stakeholders' views and opinions on the current status of data sharing and application, the implementation degree of policies and laws and regulations, existing problems, and key constraints from multiple perspectives and at a deep level, which are often difficult to obtain through the quantitative method.

Therefore, after clearly defining and classifying the stakeholders of the sharing and application of healthcare big data, this study invites key stakeholders in the sharing and application of healthcare big data for in-depth interviews, in order to find out the obstacles and problems in the sharing and application of healthcare big data, and make an effective plan to remove the obstacles in the sharing and application of healthcare big data.

3.4 Technical Route

Based on the above research contents and methods, the technical route of this research is shown in Figure 3.1.

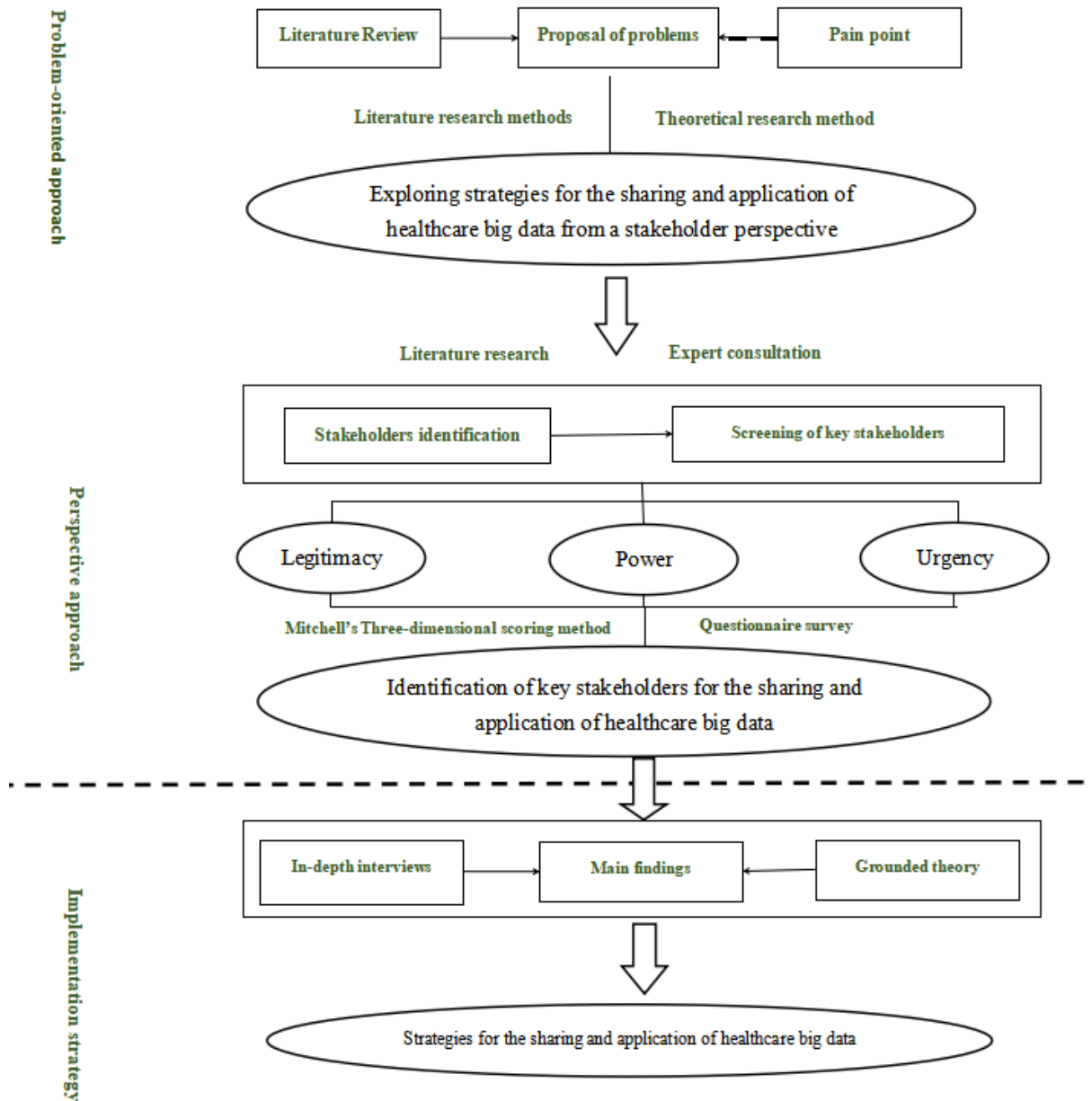


Figure 3.1 Technical route of this research

[This page is deliberately left blank.]

Chapter 4: Stakeholder Identification in the Sharing and Application of Healthcare Big Data

Building on the above-mentioned origins, development, classification, and application of Stakeholder Theory in the field of data sharing, the introduction of Stakeholder Theory into the process of constructing application and sharing strategy of healthcare big data has a solid theoretical foundation. Stakeholders' influence on healthcare big data is characterized by inevitability, diversity, interactivity, and dynamism. Different types of stakeholders have varying interests, positions, values, and roles, which, through competition, cooperation, conflict, and confrontation, collectively promote or hinder the sharing and application of healthcare big data. In different historical and practical contexts, the value, position, role, and characteristics of the same type of stakeholders also dynamically change throughout the practical process. The formulation of effective strategies to drive the sharing and application of healthcare big data relies on the interactions and synergies among key stakeholders in the current era and medical context.

To construct a healthcare big data sharing and application strategy based on a stakeholder perspective, it is necessary, firstly, to identify potential stakeholders who may participate in the sharing and application of healthcare big data in the current medical environment and specific scenarios of big data development and application. Analyzing potential stakeholders clarifies how different types of stakeholders play roles and hold positions in the process of sharing and applying healthcare big data. Among numerous stakeholders, accurately identifying and determining which stakeholders play crucial, and even decisive roles in promoting or hindering the sharing and application of healthcare big data is essential.

This chapter takes a medical university and its affiliated hospital as an example. Combining literature research and practical challenges, it comprehensively and systematically reviews the stakeholders involved in the sharing and application of healthcare big data. Through expert consultation involving 62 senior industry experts, a selection criterion of a 50% inclusion rate was used to identify the key stakeholders involved in data sharing and application. Building on this, a questionnaire was designed, and employing Mitchell's three-dimensional attribute classification method, a survey and statistical analysis, along with hypothesis testing, were conducted on 600 representative research subjects to determine the key stakeholders involved

in the sharing and application of healthcare big data.

4.1 Analysis of stakeholders engaging in healthcare big data sharing and application

In different healthcare environments, the process of sharing and applying healthcare big data involves conflicts and collaborations among multiple stakeholders. A comprehensive analysis of potential stakeholders in the healthcare and data sharing and application fields is essential. This analysis aims to clarify the positions, roles, and values of various stakeholders in the process of sharing and applying healthcare big data. Simultaneously, it seeks to elucidate the tasks and responsibilities that different stakeholders need to undertake during the sharing and application of healthcare big data, which holds significant importance for the formulation of strategies for healthcare big data sharing and application, as well as the establishment of collaborative mechanisms.

From the perspective of stakeholders, the researcher has explored various themes in the collection, storage, sharing, and application of healthcare data, including electronic health records, clinical trial data, biological samples, personal health data, and more. A review of previous literature on these topics has been conducted. The main findings are summarized in Table 4.1.

Table 4.1 Analysis of stakeholders engaging in healthcare big data application and sharing

Year	Proposers	Concepts	Stakeholders
2015	Mascalzoni, et al.	Bio-sample and data sharing	Patients, health volunteers, researchers, bio-banks, funders
2015	Husedzinovic, et al.	Bio-sample repository	Patients, public, researchers
2015	Na Xu	Resident electronic health record sharing	National Health and Family Planning Commission, Ministry of Finance, Primary Healthcare Institutions, medical staff, information product suppliers
2017	Packer, Roberts, & Parker	Medical research data sharing	Research participants, community representatives, researchers, Research Ethics Committees, data managers
2019	Miller, et al.	Clinical trial data sharing	Patients, industry, academia, regulatory agencies, and other stakeholders
2020	Wang Chunying	Open government data sharing	Government, corporations and data development teams, citizens, researchers, non-profit organizations related to data openness and application
2021	Dulhanty	Personal health data collection, sharing, and	Individuals, healthcare providers, public health authorities, government

Year	Proposers	Concepts use	Stakeholders
2021	Shi Jingjin, Yu Guangjun	Collaboration mechanism	Decision-makers, businesses collecting and using medical information Different types of medical institutions, centers for disease control and prevention, maternal and child health centers, medical security and social security and other administrative departments
2021	Liao Zirui	Sharing model	Hospitals, medical insurance, physical examination units, health industry, other departments
2022	Lan Lan, Li Rui, Bai Bo	Data sharing	Government departments, medical institutions, patients, medical IT enterprises, external scientific researchers

Based on stakeholders identified in the previous literature related to the themes of healthcare data sharing and application in both Chinese and English, and considering the current practical challenges and potential facilitating and hindering factors in the process of healthcare big data sharing and application, the researcher has initially determined 11 categories of stakeholders directly or indirectly involved in the generation, aggregation, storage, cleaning, sharing, and application stages of healthcare big data. These include government entities, health commissions, medical insurance departments, school administrators, hospital administrators, clinical department heads, information department leaders, medical staff, researchers, patients (residents), and third-party organizations.

4.2 Selection of key stakeholders in the sharing and application of healthcare big data

To identify the major stakeholders in the sharing and application of healthcare big data among the previously outlined potential stakeholders, the researcher employed a structured electronic consultation questionnaire. Combining the theoretical and purposive sampling methods, 62 experienced managers and expert scholars in the field of healthcare big data sharing and application were invited from the Guangdong Health Commission, Medical Insurance Bureau, Drug Supervision Bureau, medical universities, and affiliated hospitals. The consultation involved a 50% inclusion rate as the selection criterion (X. L. Xu, 2018) to ensure a representative sample.

The content of the expert consultation was divided into two parts: personal basic information and the selection of key stakeholders in healthcare big data sharing and application.

(1) Personal Basic Information of Interviewed Experts and Administrators: This section

covered gender, age, workplace, position, title, years of experience, and familiarity with stakeholders in healthcare big data sharing and application.

(2) Selection of Key Stakeholders: The experts were asked in the electronic questionnaire whether certain individuals belonged to the stakeholders in healthcare big data sharing and application. They were presented with a list of 11 potential stakeholders and asked to indicate “Yes” or “No” based on their insights.

The specific content of the expert consultation form can be found in Appendix A, titled “Expert Consultation Form for Identifying Key Stakeholders in Healthcare Big Data Sharing and Application in Medical Universities and Affiliated Hospitals.”

The expert consultation took place from June to August 2022. The interviewed experts included officials from health commissions, directors of public hospitals and relevant administrative departments, experienced medical professionals, and senior scholars from research institutes or universities.

Thanks to the keen interest and strong support from these experts, the effective response rate for the expert consultation was 100%. Based on the results from the 62 experts, scholars, and managers, with a 50% cutoff value, candidates with an inclusion rate of 50% or higher were considered as key stakeholders. The specific survey results are presented in Table 4.2.

Table 4.2 Expert consultation results on stakeholders in healthcare big data sharing and application

Stakeholders	Number of persons elected (person)	Inclusion ratio (%)
Government	59	95.2
Health Commission	60	96.8
Medical insurance department	58	93.5
School administrators	39	62.9
Hospital administrators	60	100.0
Directors of the clinical department	56	90.3
Leaders of the IT department	46	74.2
Medical staff	51	82.3
Scientific researchers	53	85.5
Patients (residents)	51	82.3
Third-party agencies	43	69.4

The results demonstrate unanimous agreement among all surveyed experts (100%) that hospital administrators are stakeholders in the application and sharing of healthcare big data. Additionally, over 90% of the surveyed experts acknowledged the roles of the Health Commission, government, medical insurance departments, and directors of the clinical department as stakeholders in healthcare big data sharing and application, with recognition rates reaching 96.8%, 95.2%, 93.5%, and 90.3%, respectively. Third-party agencies had a lower recognition rate at 69.4%. Drawing inspiration from the viewpoint proposed by Chinese scholar Q. Wang (2015), which considers a 50% inclusion rate as a reference standard, this study

identifies government entities, health commissions, medical insurance departments, school administrators, directors of clinical department, information department leaders, medical staff, researchers, patients (residents), and third-party organizations as the key stakeholders in the process of healthcare big data application and sharing. Hence, this research will focus on these 11 categories of stakeholders in subsequent studies to provide further elucidation and analysis, ultimately determining the critical stakeholders involved in healthcare big data application and sharing.

4.3 Mechanisms of key stakeholders in promoting healthcare big data sharing and application

Subsequently, the researcher conducted a detailed analysis of the mechanisms by which the 11 identified key stakeholders, based on expert consultation, play a role in promoting or hindering the process of healthcare big data sharing and application. This analysis aims to clarify the responsibilities and demands of each key stakeholder.

(1) Government: Government departments undertake the functions of establishing medical universities and hospitals, serving as the primary driving force for the development of healthcare and the formulators of healthcare policies. Relevant government departments aim to efficiently allocate various limited resources, including healthcare personnel, finances, and materials, to maximize benefits. By enhancing overall planning and resource investment, government departments actively engage different stakeholders in collaborative innovation and data sharing. They facilitate cross-departmental, cross-level, and cross-regional interoperability of data, narrowing the data divide, and improving mechanisms for data collection, exchange, and innovative sharing.

(2) Health Commission: The Health Commission utilizes healthcare big data to formulate national health policies, coordinate and advance the deepening of healthcare system reforms. It drafts local regulations and development plans based on the regional healthcare situation and the medical challenges faced by local residents. Leveraging information from institutions such as disease control centers, ports of entry, inspection and quarantine departments, and hospitals, the Health Commission promptly grasps the distribution characteristics and trend changes of various infectious and emerging diseases. This enables dynamic analysis and prediction of the risk of public health emergencies, enhancing the emergency response and comprehensive disposal capabilities for major public health emergencies.

(3) Medical Insurance Department: The medical insurance department establishes and

improves the management rules and regulations of medical insurance and general control work, answers questions about medical insurance policies in the process of patients' medical treatment, implements relevant provisions such as basic medical insurance and public health care policies, actively cooperates with medical insurance handling agencies at all levels and agricultural cooperative offices, provides data related to medical insurance and new agricultural cooperation in a timely manner, gives feedback on the opinions of the insured and patients participating in insurance and cooperation, and coordinates and handles the relationship with medical insurance handling agencies at all levels, agricultural cooperation offices, and patients participating in insurance and cooperation. The medical insurance department also makes use of big data to obtain timely information on patient visits and medical insurance, and obtain statistical data reflecting the overall picture of medical service utilization and population health level.

(4) School Administrators: In the process of healthcare big data sharing and application, the primary responsibilities of school administrators include data supervision and protection, investment in specialized hardware facilities, and facilitating the establishment of data governance systems and standards. Their interests lie in utilizing big data to assess the development of specialized construction in affiliated hospitals, coordinating school resources to support hospital specialization. They aim to promote the teaching quality, research capabilities, and innovation through data sharing and application and enhancing the academic reputation and overall competitiveness of the school and its affiliated hospitals.

(5) Hospital Administrators: Throughout the process of healthcare big data sharing and application, hospital administrators are primarily responsible for data resource management, ensuring data security, establishing data governance systems and standards, investing in specialized hardware facilities, and controlling data quality. Their interests involve fully utilizing clinical visit data from various hospital specialties to assess each specialty's development capabilities. They aim to understand the hospital's operational status, patient needs, and the allocation of medical resources promptly, providing reliable information for future management decisions and long-term development planning. Hospital administrators seek to optimize resource utilization and overall operational efficiency, enhance the quality of patient care and services, and improve patient satisfaction. Through data sharing, they establish collaborative relationships with other medical institutions, research organizations, and government departments, promoting resource sharing and knowledge exchange to achieve common medical goals.

(6) Directors of Clinical Department: In the process of healthcare big data sharing and application, the responsibilities of directors of clinical department include timely coordinating

and organizing the collection, recording, governance, and analysis of clinical data within their departments. They ensure the accuracy, completeness, and good visibility of data. Directors of clinical department establish relevant systems and regulations to ensure the rational management of data, guarantee data security, and reduce the risk of patient information leakage. Their demands primarily revolve around optimizing clinical decisions through the analysis of medical big data, improving clinical workflows and operational modes, enhancing clinical work efficiency, improving the quality of clinical services, and collaborating with other directors of clinical department and medical institutions to advance medical research and clinical practices.

(7) Information Department Leaders: In the process of healthcare big data sharing and application, the responsibilities of information department leaders include formulating data standards and data architecture to ensure the consistency, interoperability, and sustainability of data. They are responsible for the construction, maintenance, and updates of medical big data software and hardware equipment and network platforms, as well as overseeing processes such as data collection, storage, integration, cleaning, and data protection. As leaders of the information department, they have authority over the management of medical big data resources. Their interests primarily revolve around integrating and optimizing data resources to enhance their value and efficiency. They aim to drive innovation and development in medical information technology, improve institutional efficiency and reputation, and increase the competitiveness of the institution.

(8) Medical Staff: In the process of healthcare big data sharing and application, the main responsibilities of medical staff, including doctors and nurses, involve accurately recording patients' medical data, including medical histories, examination and test results, diagnostic outcomes, and treatment plans, ensuring the integrity and accuracy of the data. They adhere to data security and patient privacy protection policies related to medical big data, ensuring the non-infringement of patient privacy rights and the security of sensitive data. Their interests primarily lie in optimizing diagnosis and treatment plans, improving clinical practices, enhancing treatment outcomes, and increasing patient satisfaction through the analysis and application of medical big data. They seek to improve clinical workflows and operational modes, simplify the medical visit process, reduce waiting times, and enhance clinical efficiency and service quality.

(9) Researchers: Although hospitals are service-oriented institutions providing medical services and not pure research organizations, the provision of high-quality medical services requires support from high-quality research outcomes. The selection of scientific research directions should prioritize a focus on clinical services. As the possibility of cross-database and

cross-platform data interoperability becomes feasible, the analysis and exploration of massive data combining clinical, population, genetic, and mobile healthcare databases will help researchers obtain more accurate information. This effectively harnesses the role of big data in precision disease prevention and treatment.

(10) Patients (Residents): In the context of big data health care, the establishment of electronic medical records and electronic resident health archives databases will facilitate the creation of a new model for big data health care resources and “Internet + healthcare” services. In this backdrop, patients can access their personal electronic health care information at any time for self-health management. However, while big data health care sharing and application bring convenience to residents, it also implies an unpredictable increase in the risk of information leakage related to health and diagnosis, as well as genomic information based on personal biological samples. Balancing convenience and risk is a crucial factor that requires careful consideration.

(11) Third-Party Agencies: This category primarily includes developers and builders of platforms for the collection, extraction, cleaning, analysis and visualization of big data in health care, as well as agencies providing biological sample testing for medical research projects. They serve as assistants in data production and have a primary goal of obtaining direct economic benefits. However, in the process of data construction and biological sample testing, they inevitably acquire access permissions to the data, potentially processing and utilizing it to gain additional economic benefits and commercial value.

4.4 Identification of key stakeholders in the sharing and application of healthcare big data

4.4.1 Approach to identification

The various major stakeholders in the process of sharing and applying healthcare big data exhibit heterogeneity, with significant differences in their demands, rights, status, and roles in participating in this process. This heterogeneity among key stakeholders leads to varying levels of impact and interest in healthcare big data sharing and application. Some major stakeholders play a crucial role, while others have a relatively minor impact. In this study, those stakeholders pivotal to the healthcare big data sharing and application process are referred to as key stakeholders.

Establishing collaboration mechanisms and coordination strategies based on the 11

categories of major stakeholders in healthcare big data sharing and application not only complicates the difficulty of building such mechanisms and strategies but also makes the entire system complex and potentially chaotic. Moreover, it increases the operational costs of collaboration mechanisms, potentially hindering the overall coordination system.

The identification of key stakeholders offers a new perspective for formulating strategies in healthcare big data sharing and application. Starting from key stakeholders allows for better coordination of the interests of stakeholder groups with decisive roles in the data sharing and application process, reducing conflicts and facilitating the development of effective strategies to drive healthcare big data sharing and application.

To identify and determine key stakeholders in the process of healthcare big data sharing and application, this study combines a review of high-quality literature and the practical experience gained by the researchers in over a decade of management work. The study proposes that key stakeholders are those individuals or entities that contribute significant human, financial, and material capital in various forms during the process of healthcare big data generation, aggregation, storage, sharing, and application. These stakeholders directly participate in one or more of these processes, assuming relatively high risks and responsibilities. Their contributions and actions significantly impact the achievement of goals in healthcare big data sharing and application.

The researchers argue that relying solely on theoretical analysis or general statistical descriptions to identify key stakeholders in healthcare big data sharing and application is not appropriate. It may not accurately comprehend the value and role positioning of different types of key stakeholders. Identifying and determining key stakeholders requires rigorous screening and effective statistical hypothesis testing.

Therefore, the researchers, based on theoretical studies, systematically review potential stakeholders in healthcare big data sharing and application through literature research and practical challenges. They use expert consultations to determine major stakeholders with a 50% inclusion rate as the standard and analyze their mechanisms in the process. Finally, through a questionnaire and the three-dimensional scoring method proposed by Mitchell - evaluating stakeholders based on power, urgency, and legitimacy - key stakeholders in healthcare big data sharing and application are identified through statistical hypothesis testing.

4.4.2 Identification method

The “Multidimensional Segmentation” was once the primary analytical tool for classifying stakeholders, mainly based on differences in characteristics among different stakeholder groups.

While this method achieved some success in refining the features of stakeholders, it lacked quantitative differentiation and statistical testing for differences in the impact of stakeholders. In 1997, Mitchell et al. (1997), after thoroughly exploring the mechanisms of stakeholder impact on the survival and development of enterprises and the interactions between classified stakeholders in the process of enterprise survival and development, innovatively proposed three crucial attributes to differentiate the impact of stakeholders on enterprises, namely legitimacy, power, and urgency. Legitimacy refers to the entitlement granted to a certain stakeholder group or individual under legal regulations, ethics, social norms, or other specific conditions regarding organizational construction, survival, development, and profit distribution of the enterprise. Power indicates whether a certain type of stakeholder group possesses the status and capability to influence the strategic goals and operational decisions of the enterprise. Urgency is based on time sensitivity and criticality, and used to measure whether a stakeholder group can immediately attract the attention of enterprise managers and elicit a proactive response to their demands. If a stakeholder group only possesses one of these three attributes, it is termed a potential stakeholder, with a relatively minor role and impact. If it possesses any two, it is referred to as an expected stakeholder. Stakeholder groups possessing all three important attributes are termed identified stakeholders. Identified stakeholders should be the primary focus, as their role and influence are crucial, and their interests are often significant. Additionally, Mitchell pointed out that the impact of stakeholders on enterprises is not static. With the development of enterprises and organizations, stakeholders may undergo attribute changes under different conditions, losing their original attributes or acquiring new ones.

The impact of stakeholders on the healthcare big data sharing and application is dynamic. Stakeholders of the same type may exhibit different characteristics, roles, and impacts in different real-world contexts. Building upon theoretical research and expert consultations, the researcher employed Mitchell's dynamic categorization method. Using a questionnaire, they invited experts, scholars, and practitioners in the medical field to rate key stakeholders involved in healthcare big data sharing and application. The assessment was conducted based on three dimensions: legitimacy, power and urgency, utilizing a Likert scale with a 5-point scoring system to score the major stakeholders involved in the healthcare big data sharing and application. Statistical methods were then applied to statistically verify the scores of different attributes to quantitatively identify the key stakeholders involved in healthcare big data sharing and application. This study lays the groundwork for subsequent in-depth interviews with key stakeholders and the development of strategies for healthcare big data sharing and application, with a focus on these critical contributors.

4.4.3 Questionnaire

In an effort to effectively promote the sharing and application of healthcare big data and unleash its intrinsic value, the research team designed a questionnaire to identify the attributes of stakeholders in the strategy for healthcare big data sharing and application. The team sought input from leaders and experts in the field of medical health, including leaders from the Health Commission, government departments, medical universities, management personnel, research professionals, and directors of clinical departments in affiliated hospitals. The questionnaire's content underwent professional validation based on the feedback received, leading to a reordering and refinement of the survey's content and a more targeted and effective approach.

After incorporating feedback and suggestions from leaders and experts, the research team finalized the questionnaire for identifying the attributes of stakeholders in healthcare big data sharing and application. Adhering to the principles of anonymity and medical ethics, the introduction of the questionnaire clearly stated the research purpose and the research context in which the collected information would be used. All the respondents were ensured anonymity, allowing them to express their genuine thoughts freely to the greatest extent. The questionnaire was divided into two main parts based on the different attributes of the content.

(1) Personal Basic Information of the Respondents: This involves gender, age, nature of work unit, job position, professional title, and years of work experience.

(2) Mitchell's Three-Dimensional Attribute Scoring Table: This table lists the 11 major stakeholders identified in the early stage of participation in healthcare big data sharing and application in medical universities and their affiliated hospitals. The stakeholders include government departments, health commissions, medical insurance departments, school administrators, hospital managers, clinical department directors, information leaders, medical staff, researchers, patients (residents), and third-party organizations. Respondents are required to score each of the major stakeholders on the dimensions of legitimacy, power and urgency, using a Likert scale ranging from 1 to 5. A score of 1 indicates the perception of the weakest influence in that dimension, while a score of 5 represents the strongest influence. Higher scores in each dimension for a stakeholder imply higher legitimacy, stronger power, and more pronounced urgency in their participation in healthcare big data sharing and application. For detailed information, refer to Appendix B, the Questionnaire for Identifying Mitchell's Three-Dimensional Attributes of Major Stakeholders in Healthcare Big Data Sharing and Application.

The survey was conducted from September to November 2023. To facilitate information collection and reduce the potential errors that may arise from manual data entry, the researchers

imported the survey content into So Jump, a questionnaire platform, creating an electronic questionnaire.

Prior to the formal survey, a pre-test evaluation was conducted to mitigate the risk of low reliability in the questionnaire. Additionally, this pre-test provided a basis for comparison and reference for later data analysis in the formal survey. Using the roster of currently employed employees provided by the personnel management departments of medical universities and their 13 affiliated hospitals, 50 individuals were randomly selected as pre-survey participants. After collecting data, the researchers utilized SPSS 22.0 statistical software to assess the reliability of the 50 pre-survey samples, determining the internal consistency coefficient of the questionnaire, namely the Cronbach's α coefficient. This coefficient's value ranges from 0 to 1, with the closer the value is to 1 indicating higher reliability. Nunnally (1978) suggested that Cronbach's α should not be less than 0.7; otherwise, it is considered too low. According to the SPSS statistical analysis results, the Cronbach's α value for our designed questionnaire was 0.954 (greater than 0.7), indicating high reliability of the questionnaire. The KMO value for the sample data was 0.77 (greater than 0.5), indicating the suitability of the questionnaire for factor analysis. The final survey questionnaire was then confirmed, and the research was extensively carried out.

Subsequently, combining random and purposive sampling methods based on the personnel lists of medical universities and the 13 affiliated hospitals and publicly available information from government departments, we selected 600 participants from government departments, schools, research institutions, hospitals, and other units. The participants' professional titles covered primary, intermediate, associate, and full professor levels, while their positions included researchers, medical professionals, managers, department leaders, and leaders at the deputy director level and above. The range of work experience varied from less than 5 years to over 30 years.

A total of 600 questionnaires were distributed, and 573 responses were collected. All questionnaire responses were complete, with no inconsistencies or invalid questionnaire with uniform scores across all dimensions for all stakeholders. Therefore, all 573 collected questionnaires were considered valid, resulting in an effective response rate of 95.5%.

4.4.4 Statistical analysis

4.4.4.1 Statistical description

The data collected from the survey on So Jump was exported and imported into R 4.0.4 software

(R Development Core Team, 2021) for statistical analysis. Basic characteristics of the participants were described using frequencies (n) and percentages (%). The major stakeholders, selected through expert consultation, were described in terms of their scores on the dimensions of legality, power and urgency using mean and standard deviation. Mean values were initially used to assess the influence of different types of major stakeholders on each dimension.

4.4.4.2 Hypothesis testing

Merely comparing the mean values of key stakeholders on a single attribute dimension does not accurately determine their differences and rankings in that dimension. Mean differences in survey samples, may resulting from systematic errors, do not represent the truth and are the statistically significant differences. Following the research design, we further employed the “paired-sample T-test” method to assess whether the differences in mean scores between pairs of the 11 major stakeholders in healthcare big data sharing and application on each dimension were statistically significant compared to “0.” The purpose of this test is to eliminate the possibility of mistakenly assuming ranking differences due to sampling errors when the overall mean of the sampled population happens to fall within the same confidence interval. If the difference in mean scores between two types of stakeholders on a specific dimension is statistically significant compared to “0,” we can then use the magnitude of the mean to determine their ranking order in that dimension.

4.4.5 Survey results

4.4.5.1 Basic characteristics of the participants

The survey encompassed participants engaging in various job types within the medical universities and its affiliated hospitals, including unit leaders, department directors, doctors, nursing staff, medical technicians, administrative personnel, and researchers. The selection proportions generally aligned with the composition of different employee types. Additionally, leaders and staff from government and the Health Commission were included. The basic characteristics of the study participants are presented in Table 4.3.

Table 4.3 Basic characteristics of e-survey participants

Items	Category	Sample	Percentage (%)
Gender	Male	342	59.69
	Female	231	40.31
Age	20-29	63	10.99
	30-39	188	32.81
	40-49	235	41.01
	50-59	83	14.49

Items	Category	Sample	Percentage (%)
Work unit	60 and above	4	0.70
	Government departments	18	3.14
	Universities	187	32.64
	Scientific research institutions	28	4.89
	Hospitals	340	59.34
Years of working	Less than 5 years	110	19.20
	5-10 years	81	14.14
	11-15 years	116	20.24
	16-20 years	110	19.20
	21-25 years	62	10.82
	26-30 years	47	8.20
	30 years and above	47	8.20
Title	Senior	124	21.64
	Deputy senior	190	33.16
	Intermediate	153	26.70
	Junior	42	7.33
	No	64	11.17
Position	Departmental leaders	5	0.87
	Section-level leaders	61	10.65
	Division-level leaders	67	11.69
	Administrative staff	80	13.69
	Medical staff	185	32.29
	scientific researchers	175	30.54

Based on the basic characteristics of survey participants, we observed that male participants constituted 59.69% of the total, while females comprised 40.31%. Regarding the age dimension, the majority fell into the 30-39 and 40-49 age brackets, totaling 423 individuals and accounting for 73.82% of all participants. The primary source of participants was hospitals (59.34%), followed by universities (32.62%), aligning with the demographic composition of the institution. Participants exhibited a relatively balanced distribution across various work experience intervals. In terms of professional titles, which also exhibits relatively balanced distribution, individuals with associate, intermediate, and senior titles collectively represented 81.5% of the total, with 42 holding junior titles (7.33%) and 64 having no specified title (11.77%). Examining positions, medical and research staff constituted the majority, each representing over 30% of the participants. Administrative staff (13.69%), department leaders (11.69%), and division-level leaders (10.65%) followed, each accounting for 10-15% of the total research subjects.

4.4.5.2 Descriptive analysis

We utilized mean and standard deviation to describe the scores given by 573 qualified study participants in the dimensions of legitimacy, power and urgency to the major stakeholders identified earlier.

(1) Legitimacy dimension scores for the 11 major stakeholders

The mean and standard deviation of legitimacy dimension scores for the 11 major

stakeholders in healthcare data application and sharing are presented in Table 4.4.

Table 4.4 Legitimacy dimension scores for 11 major stakeholders in healthcare data

Stakeholders	Mean	Sd
Government	3.58	1.91
Health Commission	3.83	1.74
Medical insurance department	3.59	1.87
School administrators	1.82	2.13
Hospital administrators	3.22	1.98
Directors of the clinical department	2.42	2.17
Leaders of the IT department	1.91	2.14
Medical staff	2.83	2.12
Scientific researchers	2.81	2.08
Patients (residents)	2.72	2.16
Third-party agencies	1.62	1.95

Solely relying on the magnitude of means doesn't allow for a straightforward assessment of the relative power of various stakeholder groups in healthcare data application and sharing. To address this, we imported the data into the R software and employed paired-sample T-tests to statistically examine whether the mean differences in the legitimacy dimension between the 11 major stakeholders were significantly different from "0". This approach aims to eliminate the possibility of misinterpreting ranking differences due to sampling errors, assuming the overall mean values of the sampled population fall within the same confidence interval. Based on the descriptive statistical analysis results of legitimacy dimension scores for the 11 major stakeholders presented in Table 4.5, where the mean values for government, health commission, medical insurance, and hospital managers are all greater than 3, it is noteworthy that the health commission has the highest legitimacy dimension score.

(2) Descriptive analysis of power dimension for 11 major stakeholders

The mean values and standard deviations for the power dimension scores of the 11 major stakeholders in healthcare data application and sharing are presented in Table 4.5.

Table 4.5 Descriptive statistical analysis results for power dimension scores of 11 major stakeholders in healthcare data

Stakeholders	Mean	Sd
Government	3.68	1.90
Health Commission	3.86	1.74
Medical insurance department	3.60	1.87
School administrators	1.81	2.11
Hospital administrators	3.24	1.99
Directors of the clinical department	2.45	2.18
Leaders of the IT department	1.90	2.14
Medical staff	2.83	2.12
Scientific researchers	2.89	2.10
Patients (residents)	2.70	2.11
Third-party agencies	1.69	1.95

Based on the results, government, the Health Commission, medical insurance, and hospital

managers could be significant stakeholders in healthcare data application and sharing, with high average scores. Directors of the clinical department, researchers, medical staff, and patients (residents) have average scores ranging between 2 and 3, while school managers, IT department leaders, and third-party agencies have mean values below 2. However, relying solely on mean values does not accurately indicate the ranking order of stakeholders in the power dimension of health data application and sharing, as this result lacks statistical significance.

Therefore, we need to use the paired-sample T-test method to statistically examine whether the mean differences in the power dimension between the major stakeholders exhibit significant differences from “0.” This aims to eliminate the possibility of misinterpreting ranking differences due to sampling errors, assuming the overall mean values of the sampled population fall within the same confidence interval.

(3) Descriptive analysis of urgency dimension for 11 major stakeholders

The mean values and standard deviations for the urgency dimension scores of the 11 major stakeholders in healthcare data application and sharing are presented in Table 4.6.

Table 4.6 Descriptive statistical analysis results for urgency dimension scores of 11 major stakeholders in healthcare data

Stakeholders	Mean	Sd
Government	3.44	1.85
Health Commission	3.71	1.71
Medical insurance department	3.45	1.85
School administrators	1.72	2.04
Hospital administrators	3.14	1.96
Directors of the clinical department	2.36	2.14
Leaders of the IT department	1.83	2.07
Medical staff	2.73	2.07
Scientific researchers	2.68	2.00
Patients (residents)	2.53	2.05
Third-party agencies	1.57	1.86

Based on the aforementioned results, it is observed that the government, the Health Commission, medical insurance, and hospital managers could be crucial stakeholders in healthcare data application and sharing, as these groups exhibit significantly higher average scores. Directors of the clinical department, researchers, medical staff, and residents (patients) have average scores ranging between 2 and 3, indicating moderate perceived urgency. Conversely, school managers, IT department leaders, and third-party agencies have mean values below 2, suggesting lower perceived urgency. However, it is essential to note that relying solely on the average values does not precisely depict the hierarchical order of stakeholders in the urgency dimension of healthcare data application and sharing. This outcome lacks statistical significance.

To address this limitation, we employed the “paired-sample T-test” to conduct statistical tests, assessing whether the mean differences between pairs of major stakeholders in the urgency dimension are statistically significant. This approach helps eliminate the possibility of misinterpreting ranking differences due to sampling errors when the overall population mean falls within the same confidence interval.

4.4.5.3 Statistical tests

(1) The paired sample T-test results of the score difference of legitimacy dimension

The paired-sample T-tests were conducted to examine the score differences in the legitimacy dimension among the 11 major stakeholders involved in healthcare data application and sharing. The results are presented in Table 4.7.

The results allow for the assessment of mean differences in the legitimacy dimension scores among the 11 major stakeholders involved in healthcare data application and sharing, enabling the establishment of a preliminary ranking. From Table 4.7, it can be observed that, for certain pairs such as government and medical insurance department, school managers and IT department leaders, medical staff and researchers, as well as researchers and patients (residents), there are differences in the mean scores in the legitimacy dimension. However, it is important to note that these differences do not hold statistical significance. In other words, relying solely on the calculated mean values from the sample results cannot be used to determine significant differences in the legitimacy dimension scores between these pairs of stakeholders. On the contrary, significant statistical differences exist in the mean scores among other pairs of stakeholders.

Table 4.7 Paired-sample T-test results for legitimacy dimension score differences

Item	1	2	3	4	5	6	7	8	9	10
1. Government										
2. Health Commission	0.246**									
3. Medical insurance department	0.010	-0.236**								
4. School administrators	-1.757**	-2.003**	-1.768**							
5. Hospital administrators	-0.360**	-0.606**	-0.370**	1.398**						
6. Directors of the clinical department	-1.161**	-1.407**	-1.171**	0.597**	-0.801**					
7. Leaders of the IT department	-1.672**	-1.918**	-1.682**	0.086	-1.312**	-0.511**				
8. Medical staff	-0.752**	-0.998**	-0.763**	1.005**	-0.393**	0.408**	0.920**			
9. Scientific researchers	-0.771**	-1.017**	-0.782**	0.986**	-0.412**	0.389**	0.901**	0.019		
10. Patients (residents)	-0.850**	-1.096**	-0.860**	0.908**	0.490**	0.311*	0.822**	-0.098	-0.078	
11. Third-party agencies	-1.955**	-2.201**	-1.965**	-0.197*	-1.595**	-0.794**	-0.283**	-1.202**	-1.183**	-1.105**

Note: ***, **, and * indicate P -values < 0.01 , < 0.05 , and < 0.10 in the statistical tests, respectively.

(2) Paired-sample T-test results for differences in power dimension scores

The Paired-sample T-test results for differences in power dimension scores among the 11 major stakeholders involved in healthcare data application and sharing are presented in Table 4.8.

Government verses insurance department, school managers verse IT department leaders, medical staff verses researchers, medical staff verses patients (residents), and researchers verses patients (residents) showed differences in mean scores in the power dimension. However, these differences were not statistically significant. In other words, relying solely on the calculated mean values from the sample results is insufficient to determine the differences in the power dimension between the aforementioned pairs of stakeholders. Significant statistical differences in mean scores exist among the other pairs of stakeholders.

(3) Paired-sample T-Test results for differences in urgency dimension scores

The Paired-sample T-test results for differences in urgency dimension scores among the 11 major stakeholders in healthcare data application and sharing are presented in Table 4.9.

Government verses the insurance department, school managers verse IT department heads, medical staff verses researchers, researchers verse patients(residents) showed differences in mean scores in the legitimacy dimension. However, these differences were not statistically significant. In other words, solely relying on the calculated mean values from the sample results is insufficient to determine differences in legitimacy dimension between the aforementioned pairs of stakeholders. Significant statistical differences in mean values were observed among other pairs of stakeholders, allowing for an assessment of their urgency dimension.

Table 4.8 Paired-sample T-test results for differences in power dimension scores

Item	1	2	3	4	5	6	7	8	9	10
Government Health Commission	0.179*									
Medical insurance department	-0.080	-0.260**								
School administrators	-1.866**	-2.045**	-1.785**							
Hospital administrators	-0.436**	-0.616**	-0.356**	1.429**						
Directors of the clinical department	-1.232**	-1.412**	-1.152**	0.634**	-0.796**					
Leaders of the IT department	-1.777**	-1.956**	-1.696**	0.089	-1.340**	-0.545**				
Medical staff	-0.853**	-1.033**	-0.773**	1.012**	-0.417**	0.379**	0.923**			
Scientific researchers	-0.871**	-1.017**	-0.791**	0.994**	-0.435**	0.361**	0.906**	0.068		
Patients (residents)	-0.972**	-1.152**	-0.892**	0.894**	-0.536**	0.260**	0.804**	0.119	0.187	
Third-party agencies	-2.000**	-2.180**	-1.920**	-0.134*	-1.564**	-0.768**	-0.223*	-1.146**	-1.215**	-1.028**

Note: ***, **, and * indicate P -values < 0.01 , < 0.05 , and < 0.10 in the statistical tests, respectively.

Table 4.9 Paired-sample T-test results for differences in urgency dimension scores

Item	1	2	3	4	5	6	7	8	9	10
Government Health Commission	0.261*									
Medical insurance department	0.009	-0.253**								
School administrators	-1.720**	-1.982**	-1.729**							
Hospital administrators	-0.304**	-0.565**	-0.312**	1.417**						
Directors of the clinical department	-1.085**	-1.347**	-1.094**	0.635**	-0.782**					
Leaders of the IT department	-1.620**	-1.881**	-1.628**	0.101	-1.316**	-0.534**				
Medical staff	-0.707**	-0.969**	-0.716**	1.014**	-0.403**	0.379**	0.913**			
Scientific researchers	-0.637**	-0.899**	-0.646**	1.084**	-0.333**	0.449**	0.983**	0.056		
Patients (residents)	-0.908**	-1.169**	-0.916**	0.813**	-0.604**	0.178**	0.712**	-0.201*	-0.145	
Third-party agencies	-1.871**	-2.132**	-1.880**	-0.150	-1.567**	-0.785**	-0.251*	-1.164**	-1.108**	-0.963**

Note: ***, **, and * indicate P -values < 0.01 , < 0.05 , and < 0.10 in the statistical tests, respectively.

4.4.5.4 Classification results of stakeholders

Tables 4.4 to 4.6 display the mean scores of the 11 major stakeholders involved in healthcare big data application and sharing across the dimensions of legitimacy, power and urgency. During the survey, the respondents were asked to rate the major stakeholders in the 11 categories of healthcare big data sharing and application using a 5-point Likert scale. To analyze the data, we referred to the classification method proposed by Chinese scholars H. H. Chen and Jia (2004) when applying Stakeholder Theory in empirical research. The classification method involves dividing the average scores of stakeholders on a particular dimension into three ranges: greater than 1.0 but less than 2.0, greater than 2.0 but less than 3.5, and greater than 3.5. Subsequently, we categorized the 11 major stakeholders in healthcare big data application and sharing into the corresponding intervals based on their average scores in each dimension. This process resulted in the final classification presented in Table 4.10.

Table 4.10 Classification results of 11 major stakeholders in healthcare big data

Item	[1.0, 2.0)	[2.0, 3.5)	[3.5, 5.0) *
Legitimacy	Leaders of the IT department, school administrators, third-party agencies	Medical staff, scientific researchers, patients, directors of the clinical department	Health Commission, government, medical insurance department, hospital administrators
Power	Leaders of the IT department, school administrators, third-party agencies	Scientific researchers, medical staff, patients, directors of the clinical department	Health Commission, government, medical insurance department, hospital administrators
Urgency	Leaders of the IT department, school administrators, third-party agencies	Medical staff, scientific researchers, patients, directors of the clinical department	Health Commission, government, medical insurance department, hospital administrators

Note: * indicated stakeholders falling entirely within the third group are defined as core stakeholders.

Based on the three-dimensional classification results of the 11 stakeholders in Table 4.10, it is evident that the health commission, government, medical insurance department, and hospital management fall into the highest category in terms of legitimacy, power and urgency values for healthcare big data application and sharing, with mean scores falling within the “3.5-5.0” range. Therefore, in the subsequent research of this paper, the National Health Commission, government, medical insurance department, and hospital management are considered key stakeholders in the application and sharing of healthcare big data. This strategic focus allows for a better understanding of the key stakeholder groups involved in healthcare big data application and sharing, clarifying the relationship positioning of the health commission, government, medical insurance department, and hospital management. This, in turn, provides a reference basis for the formulation of strategies related to healthcare big data sharing and application.

Chapter 5: Interview Analysis of Healthcare Big Data Sharing and Application

In the current era of informatization, healthcare big data, as a crucial strategic resource, should be shared and utilized, aligning with the prevailing trend. It is an essential pillar supporting the deepening reform of the medical and health system. Against this backdrop, effectively balancing and meeting the interests of key stakeholders in healthcare big data application and sharing is imperative. This further drives the efficient sharing and utilization of healthcare big data, fully harnessing the intrinsic value of data to promote progress and development in the medical and health industry.

In the realm of social sciences, in-depth interviews are commonly employed as a qualitative research method. The interview process typically involves using open-ended questions, engaging in face-to-face, detailed communication and interaction with respondents. Through probing and guidance, this method is able to gain a profound understanding of respondents' viewpoints, perspectives, and motivations regarding specific issues. Consequently, in-depth interviews can comprehensively and accurately reveal respondents' genuine thoughts, thought processes, and behavioral motivations. This approach aids in better understanding the essence, internal mechanisms, and relevant influencing factors of the researched problem (Brinkmann & Kvale, 2018; Gerson & Damaske, 2020; Guest et al., 2006; Knott et al., 2022). In light of this, after clearly defining and categorizing stakeholders in healthcare big data sharing and application, the researchers invited key stakeholders from the medical and health departments such as the National Health Commission, the Health Insurance Bureau, the Drug Administration, and managers from Grade A tertiary hospitals to engage in in-depth interviews. They also interviewed administrators from medical universities closely associated with the circulation and application of healthcare big data. The objective was to uncover obstacles and challenges in healthcare big data sharing and application and develop effective solutions to overcome these barriers.

5.1 Selection of interview subjects

In line with the research objectives, in-depth interviews were conducted with key stakeholders

in healthcare big data sharing and application from the Health Commission of Guangdong Province, the Health Insurance Department (referred to as government departments), a certain medical school, and its 13 directly affiliated hospitals. The 13 affiliated hospitals of the medical school are distributed across four cities with significant differences in economic development levels: Guangzhou (Baiyun District, Haizhu District, Yuexiu District, Tianhe District, Conghua District), Shenzhen (Bao'an District, Pingshan District, Luohu District), Foshan (Shunde District, Nanhai District), and Dongguan (Wanjiang District). This distribution ensures a certain degree of representativeness among interviewees. The study invited a total of 7 managers from government departments, including 5 males with an average age of 47.4 years. Two of them held intermediate or higher professional titles (40%), and three had work experience of 15 years or more (60%). There were also 2 females with an average age of 33.5 years, both holding intermediate or higher professional titles (100%), and one of them had work experience of 15 years or more (50%). Furthermore, 8 managers from a certain medical school and its affiliated hospitals were invited for interviews. This group included 6 males with an average age of 48.5 years, all holding intermediate or higher professional titles (100%), and 5 of them having work experience of 15 years or more (83%). There were also 2 females with an average age of 52.5 years, both holding intermediate or higher professional titles (100%), and both having work experience of 15 years or more (100%). In summary, a total of 15 key stakeholders in healthcare big data sharing and application participated in the interviews. To protect the privacy of the interviewees, interview subjects are identified using interview numbers rather than their names. The basic information of the interview subjects is detailed in Table 5.1.

Table 5.1 Basic characteristics of interview subjects

No.	Nature of the unit	Gender	Age	Position	Professional title	Years of working	Familiarity ^a
A1	Government	Female	43	County- level leader and administrative personnel	Intermediate	15-20 (included) years	Relatively familiar
A2	Government	Female	33	Administrative personnel	No	5-10 (included) years	Generally familiar
A3	Government	Female	41	County- level leader and administrative personnel	No	20-25 (included) years	Relatively familiar
A4	Government	Male	28	Administrative personnel	Intermediate	Less than 5 years	Relatively familiar
A5	Government	Female	39	Medical personnel and government	Deputy senior and	5-10 (included) years	Very familiar

No.	Nature of the unit	Gender	Age	Position	Professional title	Years of working	Familiarity ^a
				administrators	deputy director		
A6	Government	Male	39	Administrative personnel	Intermediate	15-20 (included) years	Generally familiar
A7	Government	Female	42	Department leader	No	15-20 (included) years	Very familiar
B1	Colleges and universities	Female	40	Department leader	Senior	10-15 (included) years	Very familiar
B2	Hospitals	Male	51	Department leader and scientific research personnel	Senior	25-30 (included) years	Very familiar
B3	Hospitals	Male	54	Department leader and medical personnel	Senior	More than 30 years	Very familiar
B4	Hospitals	Female	46	Department leader and medical personnel	Senior	20-25 (included) years	Relatively familiar
B5	Hospitals	Female	51	Department leader and administrative personnel	Deputy senior	25-30 (included) years	Relatively familiar
B6	Hospitals	Female	51	County-level leader and administrative personnel	Senior	25-30 (included) years	Very familiar
B7	Hospitals	Female	52	Administrative personnel	Senior	25-30 (included) years	Relatively familiar
B8	Hospitals	Female	51	Department leader	Deputy senior	25-30 (included) years	Relatively familiar

Note: ^aFamiliarity indicates familiarity with the stakeholders of the sharing and application of healthcare big data.

5.2 In-Depth interview research process

To conduct in-depth interviews with key stakeholders in healthcare big data, the researcher went through four stages. The first stage, the preparation stage, took place from September 11, 2022, to November 20, 2022. During this stage, the researcher designed the interview outline, initially formulated the interview plan, and made repeated revisions to the outline and plan based on feedback from tutors, experts, professors, and the trial interviews. The researcher also underwent training in interview techniques. The second stage was the implementation stage of

in-depth interviews, occurring from December 11, 2022, to May 30, 2023. The primary tasks included preparing interview materials and tools, contacting and scheduling interviews with participants, and conducting the interviews. The third stage involved organizing interview records and data entry, covering the period from June 1, 2023, to July 10, 2023. Using textual records and audio files from the interviews, the researcher meticulously compiled and organized the interactions with each interviewee. The fourth stage encompassed the analysis of interview results, extraction and condensation of key information, in-depth reflection, and summarizing major findings. This stage occurred from July 15, 2023, to August 30, 2023.

5.2.1 Preparation for in-depth interviews

(1) Development of interview outline and plan

Initially, the researcher focused on the core proposition of “clarifying obstacles and challenges in healthcare big data sharing and application, and devising effective solutions to eliminate barriers.” Based on the literature review and the results of the previous study on identifying key stakeholders in healthcare big data sharing and application, along with the practical challenges faced in the process by a medical university and its affiliated hospitals, the researcher identified two major categories of key stakeholders for interviews: 1) Managers from government departments such as the Health Commission, Medical Insurance Bureau, and Drug Administration, and 2) Managers from higher medical institutions or hospitals. The researcher designed an initial interview outline and formulated an interview plan. Subsequently, the researcher sought input from the advisory group, relevant experts, scholars, and leaders of relevant departments. Two managers from each category - government departments, universities, and hospitals - were invited for pilot interviews. Based on feedback from experts, leaders, and the six interviewees, the interview outline and plan were revised and refined. The detailed interview outline can be found in Appendix C: Interview Outline for Healthcare Big Data Sharing and Application at a Medical University and Its Affiliated Hospitals (Government Managers), and Appendix D: Interview Outline for Healthcare Big Data Sharing and Application at a Medical University and Its Affiliated Hospitals (University or Hospital Managers).

(2) Learning and training in interview techniques

Conducting in-depth interviews poses challenges such as interviewees having an excessive sense of self-protection, low cooperation, potential strong subjective biases, and the need for more time and resources. Therefore, interviewers need to possess excellent communication skills and analytical abilities to ensure the quality and effectiveness of the interviews. The

researcher, along with two graduate students specializing in health management, systematically underwent learning and training in interview techniques. Before the formal interviews, several mock interviews were conducted to clarify how to communicate with interviewees in case of semantic ambiguity or understanding deviations, ensuring misunderstandings were addressed. Additionally, when probing and exploring interviewees' opinions on specific questions, the training emphasized maintaining a balance, respecting interviewees while extracting rich information to the fullest extent possible (Gerson & Damaske, 2020).

5.2.2 Implementation of in-depth interviews

(1) Preparing interview data

The interviews involved in-depth communication and interaction with interviewees in a relatively warm and relaxed environment, guided by the predefined outline and plan. The aim was to inspire interviewees to engage in divergent thinking and express their opinions and perspectives based on their own experiences. Simultaneously, recording and reviewing the information gathered during the interviews were crucial to ensure reliability and avoid overlooking important details. Therefore, preparing interview data includes printing the interview outline, plan and record sheets, as well as procuring interview tools such as recording devices, pens and batteries, and preparing the interview venue.

(2) Selecting and scheduling interviewees

The researcher employed a combination of purposive and heterogeneous sampling methods to select 15 managers from government departments, hospitals, and medical universities for the interviews. Contacting interviewees via mobile phones or WeChat, the researcher provided clear explanations and details about the interview's purpose, content, expected duration, and assured the interviewees of privacy protection. Specific times and locations for the interviews were then scheduled with the interviewees.

(3) Conducting and recording interview

This stage represents the primary phase of information collection and acquisition. The interview content focused on six main aspects: the foundation and current status of healthcare data sharing, interviewees' attitudes toward healthcare data sharing, benefits and risks of healthcare data sharing, problems and suggestions regarding healthcare data sharing, legal safeguards for healthcare data sharing, and security and privacy protection in healthcare data sharing. Following the predefined interview plan and outline, the researcher conducted face-to-face interviews with the 15 healthcare data stakeholders. Each interview lasted approximately 32-59 minutes, spanning a total of four months. Prior to the formal interviews, the researcher

provided detailed explanations about the purpose, content and expected duration of the interviews. They reaffirmed adherence to the principles of anonymity and minimal harm, securing informed consent from the interviewees. Specific details about the interviews are outlined in Appendix E.

The entire interview process involved one graduate student specializing in health management, serving as a dedicated recorder. This recorder diligently and promptly documented essential contextual details and non-linguistic information. This approach ensured the subsequent organization and extraction of information based on the researcher's and recorder's varying interpretations of the interview process. It also allowed the interview personnel to concentrate on actively listening to the interviewees' narratives and adequately prepare for posing follow-up questions.

5.2.3 Data compilation and entry

In this phase, the recorded audio materials and physical documents from the interview process are organized into standardized written materials. The total duration of the interviews was 634 minutes. This process aims to faithfully reconstruct the entire authentic interview process and preserve the obtained valuable information.

5.2.4 Data analysis and results synthesis

The fourth stage involves the analysis of interview results, extraction of key information, condensation of findings, and in-depth contemplation. This study employs the template organization method in qualitative research, categorizing interviewees into groups A and B. Group A represents government officials, with interview materials coded as A1-A7, while Group B represents university or hospital administrators, with interview materials coded as B1-B8.

In this study, NVivo 12 for Windows was employed as the tool for organizing and summarizing data, the raw data collected from in-depth interviews underwent processes such as coding, word frequency queries, sentiment analysis, and hierarchical visualizations. These techniques assist researchers in managing, discovering and identifying patterns within the data, facilitating clearer thinking and enhancing research efficiency. The software analysis, complemented by the application of professional knowledge and skills, was further enriched by referencing relevant literature in the research field. This comprehensive approach allowed the researcher to delve into the obtained information, engage in profound reflections, and

summarize the key findings throughout the research process.

5.3 Key findings from interviews

5.3.1 Foundation and current state of healthcare big data sharing

The foundation of healthcare big data encompasses its composition, the information systems supporting data sharing, and the financial resources dedicated to this endeavor. Examining the current state of healthcare big data sharing involves an investigation into the sharing practices within the affiliated hospitals of the studied medical universities.

5.3.1.1 Diverse forms of healthcare big data composition

In terms of the data composition of healthcare big data, according to the respondents' answers, clinical diagnosis and treatment data, medical data, cost information, and patient personal information were mentioned more, followed by data from medical institutions as well as hospital operation and management. According to the source and attribution of the data, the data can be divided into patient-related data and hospital-related data. The patient-related data includes the patient's basic information such as gender, age, ID card, home address, contact phone number and medical insurance information, as well as the patient's health file data such as the patient's disease history, family genetic diseases, and others, in addition to the patient's diagnosis and treatment data such as the patient's biochemical indicators, imaging examinations and doctor's diagnosis and treatment data. The hospital-related data includes hospital operation and management data, such as hospital income and expenditure, equipment and facilities, hospital personnel, and others, as well as hospital specialty diagnosis and treatment data, such as physician diagnosis data, medication information and treatment costs. Figure 5.1 shows the data classification hierarchy.

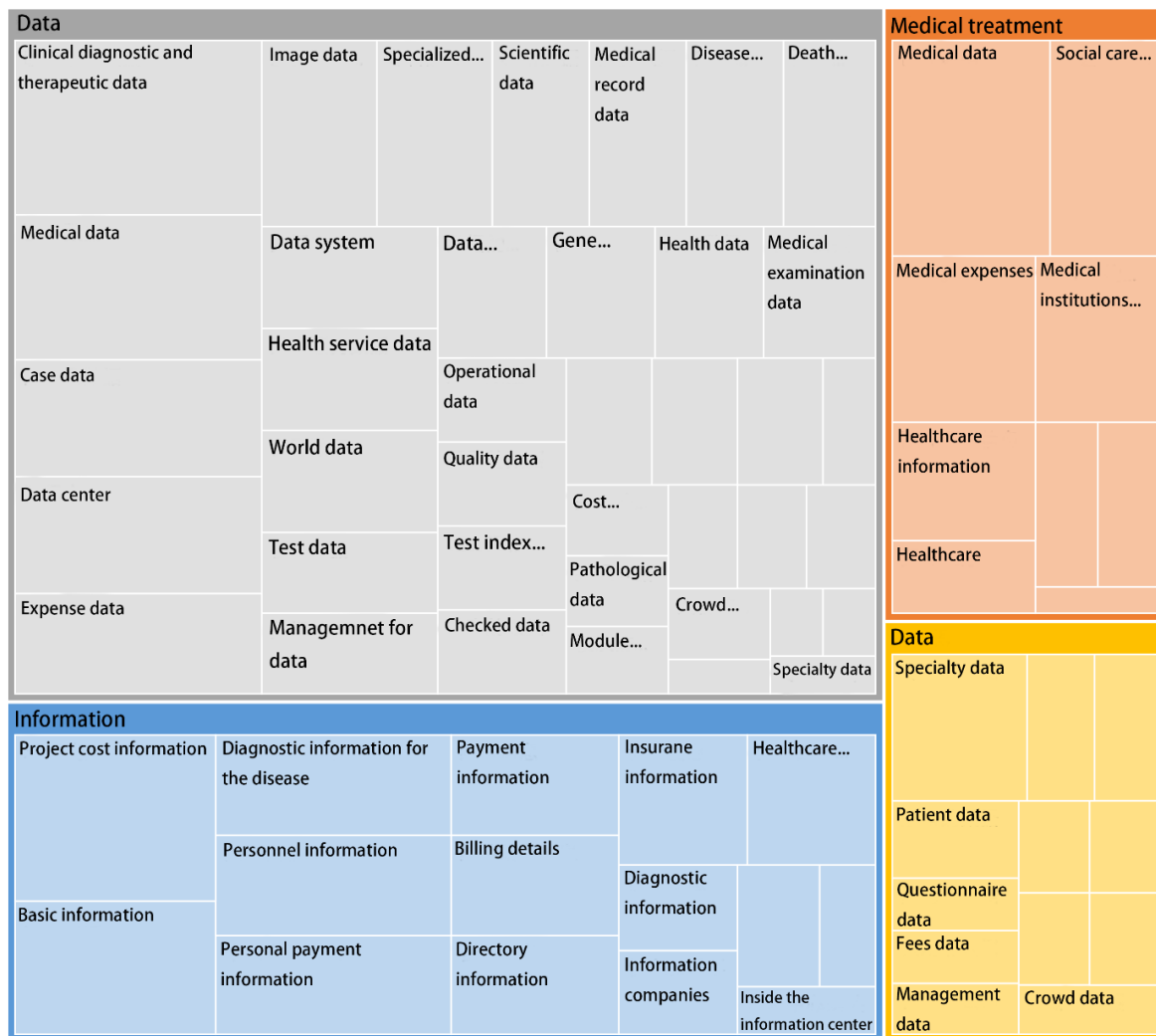


Figure 5.1 Composition of healthcare big data

5.3.1.2 Ineffective coordination between the system and the platform hindering the data sharing among hospitals

In the context of supporting healthcare big data sharing information systems, data sharing can be categorized into intra-hospital and inter-hospital data sharing. Most respondents indicated that the current situation of intra-hospital data sharing is good, while the situation for inter-hospital data sharing needs improvement. High-frequency keywords extracted from interview data include hospital, data, sharing, absence, system, platform, etc. This implies that there is currently no dedicated system or platform specifically designed for healthcare data sharing to ensure seamless inter-hospital collaboration. Hospitals tend to operate independently when procuring and establishing data collection and storage systems. There is a lack of uniformity between systems used by different affiliated hospitals and even among departments within the same affiliated hospital, leading to obstacles in data sharing and application. Refer to Figure 5.2 (a) for detailed keywords.

Table 5.2 Interview of the status quo of healthcare big data sharing

Status quo of sharing	Transcripts of interviews
No sharing	B1: There is no such sharing, and the so-called automatic exchange cannot be done. B2: The sharing of this test is now agreed, that is, the results are mutually recognized, but you actually want the patient to bring it himself, and you can't read or retrieve him remotely. Now it seems that there is none, and maybe it seems that there is none temporarily. I am not very clear about whether the mutual recognition of results among the Internet hospitals has been realized. B3: There is no automatic exchange among Southern Hospital and the various branches of Southern Hospital. B4: The data is actually for each hospital. In fact, the data information among various hospitals is not shared, including among the Southern Hospital system. B5: It is difficult to share our online and offline services if smart management does not reach a certain level. B8: Almost zero. There is basically no automatic data exchange and sharing among affiliated hospitals, and the only data sharing is carried out by relying on the mutual recognition platform of inspection results in Guangzhou. A1: From the perspective of clinical research, a disease-specific database is established to bring together data from various hospitals for specific diseases. A3: Smooth. A4: Good.
Limited sharing	A5: Linkage and sharing can be achieved within the county; interconnection can be achieved within large hospitals; interconnection between medical institutions and commercial institutions cannot be formed; and the public health system and the medical system cannot be interconnected. A6: The quality of relevant data still needs to be improved. A7: Sharing is limited; data is incomplete; and user experience needs to be improved.

5.3.2 Healthcare big data sharing benefits and risks

If the effective sharing of healthcare big data is realized, the diagnosis and treatment team can quickly read the patient's past history, allergy history, drug history and related test records when the patient is treated, without the need for repeated filing and repeated examination. It not only improves the efficiency of the entire diagnosis and treatment process, but also effectively avoids the potential risks caused by missing or misreporting information such as allergy history, as well as the additional costs and waste of medical resources caused by repeated tests.

The effective sharing of healthcare big data enables medical researchers to effectively use representative large-scale clinical samples and medical data sets from multiple centers for scientific research, so as to solve key scientific problems in clinical work and promote the improvement of medical level and academic influence of research units.

In addition, the healthcare big data sharing can promote the improvement of diagnosis and treatment. See Table 5.3 for details.

Table 5.3 Potential benefits of sharing and application of healthcare big data

Benefits	Transcripts of interviews
Convenience for patients	A2: It can reduce the burden on the masses, such as mutual recognition of inspection results and duplicate documentation. B1: Patients who see a doctor at this hospital do not have to undergo a repeat examination at another hospital. In addition, after the data is shared, just like when I receive a patient, I can see his historical medical records clearly, so I can understand this person in all aspects. B2: You can expand the amount of data, which may actually be convenient for patients to seek medical treatment.
Conducive to the development of scientific research in hospitals	A1: In terms of scientific research, the patient data of a single hospital in clinical research is not enough, and the data of applied research can be expanded through healthcare data sharing, which makes the research more valuable. B1: From the hospital's point of view, if my data is shared, then it is very valuable for me to do scientific research. B2: There must be some benefits to disease management, and it becomes polycentric. The management of some diseases is useful for some scientific research. B4: For example, in scientific research, my intuitive idea is that sharing data may actually be more convenient for patients to seek medical treatment. I see a doctor at Zhujiang Hospital, and I can directly retrieve its results here. A3: Through the analysis of big data, we can further find the relevant factors of disease development, and make breakthroughs in digital therapy, health management throughout the life cycle, and primary general medicine diagnosis and treatment.
Conducive to the progress of disease research	A6: Big data analysis can provide assistance for disease research, clinical and management decisions, and medical resource management. A7: Improving the level of diagnosis and treatment: Through the analysis of a large amount of medical data, commonalities and differences between different patients can be found, so as to improve the accuracy and effect of diagnosis and treatment. Healthcare big data can provide a more detailed and comprehensive data foundation for medical research, thereby helping doctors and researchers better understand the mechanism and treatment of diseases.

The potential risks associated with healthcare big data sharing and application can be categorized into patient privacy leakage risk, network and device security issues, and leakage of core competitive data.

The large-scale sharing of medical data may give rise to issues concerning data security and privacy. Due to the lack of strict standards and robust technical measures in areas such as data anonymization and encryption, coupled with widespread reports of data breaches in various industries, there is a potential risk in healthcare records. These records often contain sensitive information that patients may be reluctant to disclose voluntarily. Inability to appropriately and effectively handle and safeguard healthcare big data could lead to the risk of data leakage or misuse. This, in turn, may erode patients' trust in healthcare institutions and health departments. A majority of respondents (7 out of 13) expressed concerns about the risk of patient privacy breaches during the process of sharing and utilizing healthcare big data. This

concern is particularly heightened when personal medical data is disclosed to third parties such as insurance companies and pharmaceuticals or when it is maliciously exploited, causing significant distress and losses for the patients.

As medical testing technologies advance, the routine and widespread inclusion of patient-identifying information in various datasets, such as genomics and transcriptomics, becomes more prevalent. The leakage of such information, especially with the progress of scientific technologies like genomic mapping, poses increasingly significant and potential risks to patients. The disclosure of multi-omic information containing racial health characteristics also carries the possibility of harming public interests and jeopardizing national security.

The concerns of respondents also extend to network security and device security issues in the data sharing process. One of the worries expressed is related to network security risks during data exchange and sharing, where the bridging of internal and external network barriers could create vulnerabilities. As stated by a hospital manager (B8), “Network security risks: Data exchange and sharing break through internal and external network barriers, resulting in network security risks.” Currently, the implementation of healthcare big data sharing often requires the introduction of data extraction interfaces in various hospitals. However, the introduction of these data interfaces has raised concerns among hospital managers about the controllability of the future scope of data sharing. One manager (B2) expressed concern about the intrusion of systems, stating, “It just breaks into your system and it causes a whole host of problems. Because if you want to share data, you have to make an interface, and if you make an interface, it will break into your system. Then maybe it gets what it wants from you, and in the end, it can do whatever damage it wants to do to you. There is this risk.”

Furthermore, the varying levels of diagnostic and treatment capabilities among different medical institutions contribute to the uneven landscape of healthcare services. The quality of diagnosis and treatment is a core competitive factor for hospitals. Healthcare big data typically encompass therapeutic methods and diagnostic information, which may lead to concerns from institutions with currently high comprehensive diagnostic and treatment capabilities. These institutions may fear that sharing such data could result in the loss of competitive advantage and may therefore adopt a cautious or negative attitude towards data sharing. For specific details, refer to Table 5.4.

Table 5.4 Potential risks of sharing and application of healthcare big data

Types of risks	Transcripts of interviews
Disclosure of patient privacy	A1: Disclosure of patient privacy to third-party institutions such as insurance, medicine and other industries, which may cause malicious use of patient privacy. A2: Privacy disclosure may cause damage to citizens' interests such as nuisance. A4: Healthcare big data involves the privacy of individuals' biology and genes, and massive medical data is related to national security, but it lacks standards and technical means in desensitization and encryption. With the exponential growth of data and the increase in data correlation interactions, anonymized data privacy protection becomes difficult. A7: Healthcare big data contains a large amount of personal privacy information, such as medical records, diagnoses, treatment plans, and others. If these data are attacked or misused by criminals, it will cause great damage to personal privacy. B1: I don't want my medical record to be seen by others, and I don't want everyone to know what disease I have. Didn't it happen abroad? Even if I hide the name, gender, and age, you can still guess who this person is. B3: For example, if my patient information or something others are taken away and seen by others at any time, this is okay, but if it is publicly published, it must be consented. B4: It is because it involves the privacy of the patient. We have recently arrested a lot of doctors in the past two months, because they gave their accounts to others, which leads to a lot of problems on the disclosure of patient privacy.
System security issues	A7: The processing and storage of healthcare big data requires the use of advanced technology and equipment, and if there is a failure or security breach, it will pose a threat to data security and use. B8: Network security risks: Data exchange and sharing break through internal and external network barriers, resulting in network security risks.
Loss of competitive advantage	B4: If some trade secrets, intellectual property secrets, or the means and methods of the diagnosis and treatment are seen by others, there will be this potential risk, that is, this cannot be completely shared, but only limited sharing. B5: In fact, potential risks to the hospital may exist in some advantageous disciplines, because its own valuable resources may be taken away by others if data sharing is conducted.
Emergence of social problems	A7: The sharing and use of healthcare big data also involves some social issues, such as resource allocation, fairness, and intellectual property rights, which requires the establishment of sound systems and norms.

5.3.3 Attitudes of interviewees towards healthcare big data sharing

Two-thirds of the interviewed individuals express support for healthcare big data sharing, believing that it aligns with the overall interests of hospitals and patients. They see data sharing as beneficial for advancing medical research, education, and quality improvement. However, 20% of the respondents express concerns about the risks associated with data sharing, such as leakage of hospital and patient data and infringement on patient privacy. The interviewees' attitudes towards data sharing are summarized in Table 5.5.

Table 5.5 Attitudes of healthcare big data sharing respondents

Attitude	Interview Record
Positive	A1: Data is an asset, and the value of data assets should be fully explored to empower high-quality development of healthcare. A3: Promote the development of the healthcare industry, especially medical services. A4: The sharing and application of healthcare big data should be actively promoted. B1: I personally believe that sharing is very necessary. B2: Managing some diseases is useful for some scientific research. B3: Sharing is a trend and also beneficial for social development. B4: The benefits of sharing must be better than those of not sharing. B5: Contribution is currently a widely recognized concept and definitely the direction for the future. B7: Supporting wider sharing can effectively save social resources. B8: I support sharing. However, it is necessary to achieve healthcare big data sharing and application while ensuring personal privacy and information security.
Neutral	A1: Open attitude. A2: To be prudent. Medical big data sharing application is a general trend, but its application should comply with the protection of citizens' personal privacy rights and their health needs, and should not be out of control. A6: It is important to promote the circulation of data elements and develop and apply data effectively while relevant policy documents and personal information protection laws and regulations also clearly emphasize the requirements for data security and personal information processing.
Combined	A7: Healthcare big data refers to a large amount of medical data collected and integrated from multiple sources such as hospitals, institutions, and research institutions. These data are analyzed and explored to improve diagnostic accuracy, treatment effectiveness, and support public health decision-making. However, due to privacy and confidentiality issues related to medical data, data sharing and usage have always faced challenges.

5.3.4 Problems of and recommendations for healthcare big data sharing

Healthcare big data sharing involves many complex issues such as the scope of sharing, forms of sharing, and the involved stakeholders. Keyword analysis indicates that data sharing faces challenges related to privacy, security, technology, standards, quality, barriers, formats, mechanisms, ethics, and more (see Figure 5.3 (a)).

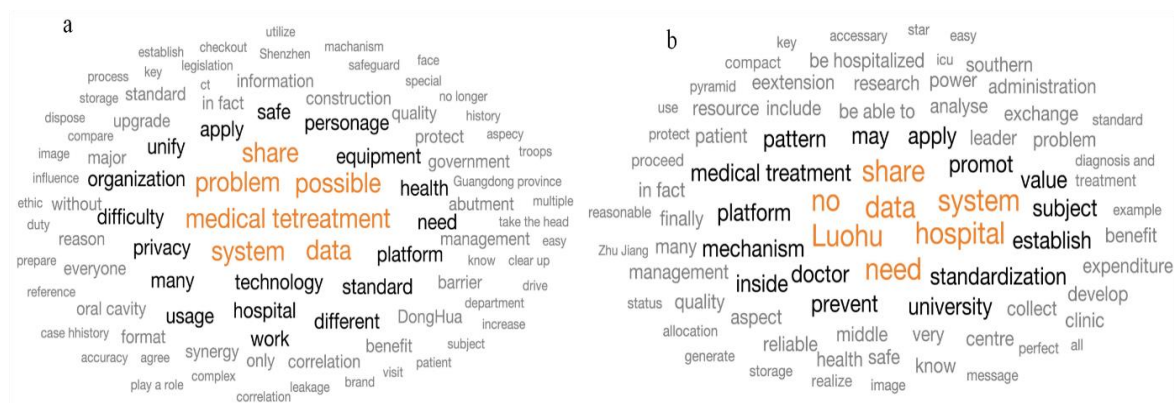


Figure 5.3 Keyword analysis of problems of and recommendations for healthcare big data sharing

Further analysis and classification reveal that the main issues in healthcare big data sharing include:

(1) Security and Privacy Issues: Healthcare big data sharing poses a risk of patient privacy data leakage. Opening up system interfaces for data sharing may expose hospital systems to external attacks, leading to system paralysis.

(2) Institutional Mechanism Issues: Lack of effective data sharing mechanisms at the upper levels of hospitals. Varying levels of information technology development in hospitals. Independence among different types of healthcare institutions and hospitals creates significant obstacles to medical data sharing.

(3) Standards, Regulations, and Ambiguous Responsibilities: Incomplete macro-level legal and regulatory support for data sharing. Lack of clarity in data standards, management, intellectual property rights, and responsibility allocation. Independence of information systems and data storage systems among hospitals creates significant difficulties in data sharing, with low operability, incompatible data systems, and formats being prominent issues. Problems of healthcare big data sharing are outlined in Figure 5.4.

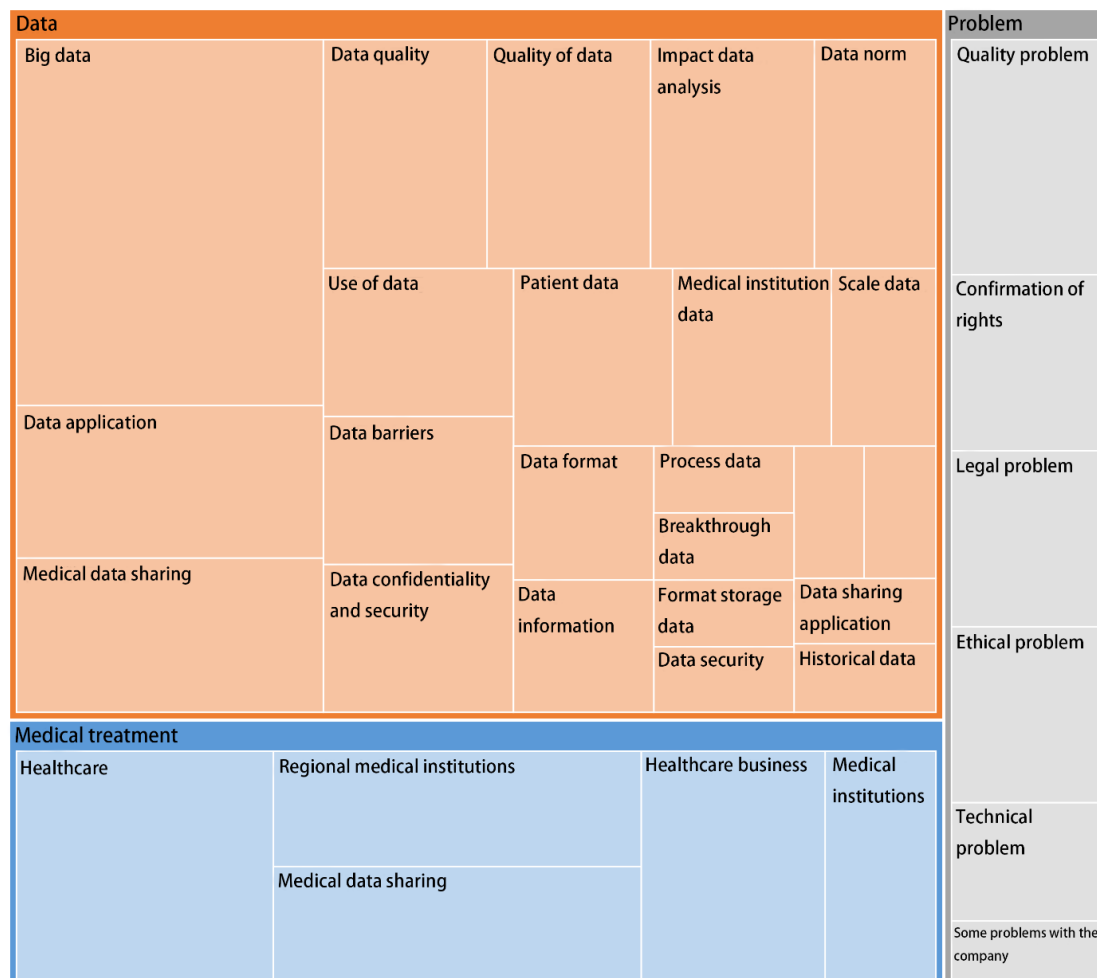


Figure 5.4 Problems of healthcare big data sharing

In addressing the challenges of healthcare big data sharing and application, interviewees earnestly proposed constructive measures to promote effective sharing. Key suggestions, as indicated by keywords in Figure 5.3 (b), include mechanisms, standards, leadership, management, funding, systems, cooperative healthcare organizations, and more. Specific details of the proposed sharing recommendations are outlined in Table 5.6.

Table 5.6 Recommendations for healthcare big data sharing

Suggestions	Keywords	Interview record
System and mechanism	Mechanism, system, administration, channels, models	A3: Be demand-oriented and improve the joint application mechanism. A4: Actively construct a system of healthcare big data resource elements. A6: Establish mechanisms, patterns and standards for data sharing and application, to make data sharing and application more standardized. A7: Establish a performance evaluation mechanism. Establishing a scientific and reasonable performance evaluation mechanism for hospitals can promote the enthusiasm and participation of medical staff, thereby pushing the promotion and implementation of data applications. Evaluation and rewards and punishments can be based on indicators such as clinical quality, medical efficiency, and patient satisfaction. To achieve data sharing, it is necessary to integrate and standardize data, establish a unified data model and metadata management mechanism. B3: So, it is possible to provide at least some policy guidance from the national level. So, everyone can contribute some money and ultimately construct a unified standard, right? B4: You need to strongly recommend the sharing, and then others need to extract value from it. I think there should be a drive from an administrative level, or it can be possible through a mechanism. For example, I add something, so I can also share it. A6: Industry standards need to be established to standardize medical data, which is more conducive to data sharing applications. A7: Data integration and standardization: different information systems and databases are used among different affiliated hospitals, with differences in data format, field names, and coding methods. To achieve data sharing, it is necessary to integrate and standardize data, establish a unified data model and metadata management mechanism.
		B2: Address standards while addressing sharing issues. B3: Try to unify the standards; from a professional perspective, I think it is very important to unify standards in all aspects, which is the unification of standardization. B4: But I think it may be difficult to promote with the
Unified norms and standards	Standards, norms, unification	

Suggestions	Keywords	Interview record
Investment	Funds, money, expenses	help of administration, but at least there would be a unified standard. That's a standard for data, right? It would be standardized and promoted. B5: Standardize diagnosis and treatment in various hospitals. This is the lack of standardization in diagnosis and treatment that I mentioned, and there will be some problems after data collection. A5: You need to design a plan. There should be various funds, such as the discipline construction funds. I once spent a certain amount of funds, and I mean an investment in this. B1: That actually requires a lot of money. B3: Just now, I mentioned that there should be money and a platform, right? So, it is possible to provide at least some policy guidance from the <u>national</u> level, so let everyone contribute some money and finally establish according to the same standard. B4: It must be a matter of money. It is the most important. A2: Taking work as an effective mean, and after the system is built, there should be corresponding means to promote the sharing of data among departments. A3: Firstly, establish an information platform, build a big data center, establish specialized departments, and first achieve the interconnection of medical related (HIS, LIS, imaging, human resources) systems.
System and platform	System, platform	A7: Establish a secure and reliable data sharing platform. In order to protect patient privacy and the commercial interests of medical institutions, the data sharing platform needs to be with features such as security and reliability, permission control, and integrity protection. During the construction, it is necessary to consider the sensitivity and confidentiality of data, and take corresponding technical measures to protect it. B2: The barrier of system. B7: To improve the system. A4: Improve data quality.
Quality	Quality and assessment	A7: Strengthening data quality management, the quality and accuracy of healthcare big data are key factors in promoting data application. In order to improve data quality, it is necessary to establish standardized data collection and processing processes, and introduce a quality inspection mechanism that combines automation and manual auditing. Strengthening data quality management, promoting data application scenarios, and establishing a scientific and reasonable performance evaluation mechanism. B8: Universities set up a data sharing center and conduct quality control and evaluation on the quantity and quality of data uploaded.

5.3.5 Legal guarantee for healthcare big data sharing

Overall, most respondents have a certain understanding of the relevant laws and regulations on the legal guarantee of healthcare big data sharing. Among them, respondents are familiar with *Data Security Law of the People's Republic of China* and *Personal Information Protection Law of the People's Republic of China* in the legal guarantee of healthcare big data sharing. The specific laws and regulations mentioned by the respondents are detailed in Table 5.7.

Table 5.7 Relevant legal regulations for healthcare big data sharing

Personnel No.	Interview record
A1	Data Security Law of the People's Republic of China
A6	Personal Information Protection Law of the People's Republic of China, Data Security Law of the People's Republic of China
A7	Personal Information Protection Law of the People's Republic of China, The Cyber Security Law of the People's Republic of China, Regulations on the administration of medical institutions of the People's Republic of China, Data Governance Act, Electronic Medication Administration Record Specification, Drug Administration Law of the People's Republic of China
B1	Data Security Law of the People's Republic of China, Personal Information Protection Law of the People's Republic of China
B6	Data Security Law of the People's Republic of China, National healthcare Big Data Standards, Safety and Service Management Measures (Trial), Guiding Opinions on the Protection of Information Security Levels in the Health Industry.
B8	The Cyber Security Law of the People's Republic of China, Data Security Law of the People's Republic of China, Personal Information Protection Law of the People's Republic of China, Notice of the State Council on Issuing the Action Outline for Promoting the Development of Big Data, Guidance of the General Office of the State Council on Promoting and Standardizing the Development of Healthcare Big Data Application, Opinions of the General Office of the State Council on Promoting the Development of "Internet + medical Health"

In response to the legal guarantee of healthcare big data sharing, respondents provided multi-dimensional answers from aspects such as patient privacy protection, intellectual property protection, data security guarantee, clear connotation principles, ethical norms, providing ideas for targeted measures to achieve healthcare big data sharing, and better promoting healthcare big data sharing from the perspective of legal guarantee. The specific content is detailed in Table 5.8.

Table 5.8 Legal guarantee for healthcare big data sharing

Guarantee category		Interview record
Patient protection	privacy	A7: Healthcare big data contains a large amount of sensitive information, such as personal identity, health status, and disease diagnosis. In order to protect the privacy rights of patients, relevant laws and regulations stipulate strict requirements for data collection, processing, storage, and transmission, such as obtaining informed consent and adhering to the principle of minimization when obtaining data. B1: Legally, the country is gradually improving, but the most important one is to protect patient privacy. B2: The issue of data security is the most basic requirement for management, so the protection I mentioned earlier refers to data protection, that is privacy protection, including how many antibiotics you used as mentioned earlier. For hospitals, this is also the privacy of the hospital. This is also not suitable for public disclosure. B3: Now, the main concern for patients is that there may be corresponding strategies in the Civil Code, which was originally called <i>Tort Law</i>, but now it is included in it, right? Of course, some related protection of intellectual property rights may also be secure for the hospital, doctors, or our relevant departments to share data, right? After all, there are still some people who really plagiarize, and they may not necessarily notify you. If they copy the figure, it will belong to him. They will definitely give lectures and publish papers, so this situation will exist.
Intellectual protection	property	A7: The sharing and application of healthcare big data require a safe and reliable data platform to prevent data disclosure, tampering, loss, and other issues. Relevant laws and regulations have standardized and guided the encryption, permission control, auditing, and backup of data.
Data security guarantee		B6. The principles and requirements for open data sharing should be clearly defined, the objects, forms, and boundaries of data sharing should be strictly defined, the conditions for data asset attributes should be clarified, and major consensus issues such as open ownership and privacy security of healthcare data sharing should be focused. The legal guarantee for the sharing of healthcare big data should be improved.
Clear principles, etc.	connotation	A3. Define the connotation of healthcare big data, determine management entities, clarify application scope, and improve regulatory and punishment mechanisms. A6. Clarify the connotation of healthcare big data, the conditions and principles of data sharing, the rights and responsibilities of all parties involved in data sharing and usage, as well as regulatory penalties.
Ethical norms		A7: The sharing and application of healthcare big data involves a series of ethical issues, such as fairness and transparency, informed consent, and rights protection. Relevant laws and regulations regulate the ethical behavior of data use through the establishment of ethics committees and the formulation of ethical standards.

5.3.6 Security and privacy protection of healthcare big data sharing

Respondents believe that the protection of privacy data in healthcare big data sharing relies on three key aspects: the enactment of national laws, data anonymization processes, and the implementation of hierarchical regulatory authorization. Detailed information is detailed in Table 5.9.

Table 5.9 Privacy data protection and protective measures in healthcare big data sharing and application

Terms	Category	Interview record
Privacy data protection	Laws introduced by the state	B1: You intuitively feel that the current protection is quite poor so there still exist such leaks or corresponding ones. For example, although the country has issued these laws, the legal concepts of some medical personnel are not so strong. Just as there would be disclosure if we use WeChat to transmit this data or something because the data is loaded to the cloud. A5: Public information desensitization, interconnecting within medical institutions using an ID card as the unique identification code. A7: By using data desensitization technology, sensitive personal identity information and medical records are transformed into unrecognized formats to ensure data privacy and security.
	Data desensitization processing	B2: Taking personal information as an example, the data can be shared if you cannot trace it back to this person. B3: For our own hospital, we can only conduct desensitization, which is to conceal one's name, contact information, and home address. B4: I think data cleaning is possible by hiding the names of these data.
	Graded regulatory authorization	A3: Set access permissions level by level. B2: Regulatory measures should be taken., For example, our data are accessible in the form of passwords, which means authorization at different levels. Through authorization at different levels, it grants someone permissions. A7: Adopt network security technologies such as firewalls and intrusion detection to monitor and defend data transmission channels and improve system security.
	Strengthen network security protection	B3: The systems in our hospital may not be connected to the internet. In terms of the interconnection of our internal system, once it is connected to the internet, there may be attacks, and the firewall is difficult to resist. This is usually processed through, such as adding a front-end processor or something. It actually is still a closed system, but if you want this data, I will extract the data system and put it in a place. If you read this part, you won't be able to get into it. It's called a front-end processor, which means I'll process the data to the front. This technology is mature.
Protective measures		B4: For example, none of the major systems are directly connected to the Internet. This is actually for security protection. Once the data needs to be shared, it is to create a data interface, which is a one-way export, and then manually export the data, and then share and exchange it, rather than automatically forming a switchboard.
	Legal restrictions	A4: Fully implement regulations and standards related to network security and data security. Implement the requirements of Cyber Security Law, Data Security Law, Personal Information Protection Law, Cryptography Law, Regulation on Protecting the Security of Critical Information Infrastructure, support standards and norms, and fulfill the network security, data security supervision and administrative law enforcement powers granted by law. On the basis of strictly implementing the network security level protection system and the basic security guarantee system for commercial password applications, focus on the security of key information infrastructure, implement the data exit security management system, strengthen the

Terms	Category	Interview record
		network and data security supervision related to medical equipment, and comprehensively implement the network security management requirements. Research and develop health information management methods and corresponding standards and norms, provide compliance guidelines for the rational use of data, and promptly correct violations.
		B5: Sign agreements. One is agreement on legal liability. For data flow in and out, I will sign some confidentiality and data security agreements with someone for a project, to protect through this commitment, goodwill, and law.
	Data closed-loop operation	A3: Relevant analysis is conducted in a closed-loop manner within the big data center, with only visual conclusions being given priority. B5: Try to keep the data from leaving our hospital as much as possible. This is a very effective protective measure, and so far, we don't have much data to offer.
	Improve data authorization	A2: Strengthen data management, standardize the entire process from data extraction to application, and ensure traceability. A3: Application system should be adopted for data usage. B8: Legal application methods such as authorization for the use of health data.
	Implement result sharing	B1: Tell you the results without the basic data and raw data, and then you will calculate in another hospital. This is actually a way of sharing, but it is indirect sharing, rather than direct sharing, which emphasizes result sharing.
	Inspect values standardization	B3: For the Luohu model, a unified standard is prepared for all hospitals in Luohu. For all the tests, for example, it is possible for Pingshan Medical Group in Pingshan to use unified standards, the same equipment, the same brand, and the same reagents, but it has not achieved standardization. Because it tests all values and it is faster and better to transmit, unlike the images which are relatively large. The test data is easy to be shared.
	Data backup and recovery	A7: Regularly backup and archive data, and establish an emergency recovery plan to ensure rapid recovery in the event of data damage or loss.

In terms of specific protective measures, respondents believe that achieving healthcare big data sharing requires strengthening network security protection, advancing legal regulations and restrictions, ensuring closed-loop data operation, improving data authorization systems and methods, implementing result sharing, standardizing examination numerical values, and enforcing data backup and recovery. Refer to Table 5.9 for detailed records.

[This page is deliberately left blank.]

Chapter 6: Strategies for Improving Healthcare Big Data Sharing and Application

In light of the previous analysis and key findings from in-depth interviews, the researcher proposes the following strategies to enhance the healthcare big data sharing and application.

6.1 Establishing a comprehensive system for healthcare data guided by policy

A robust healthcare data system serves as the cornerstone for the long-term development of a national health information platform. It is recommended to promptly introduce policies and implementation details related to the management of healthcare big data resources, issue supporting documents and formulate principles for the distribution of benefits from healthcare data sharing and application. At the policy level, it is to provide local authorities with a basis for data sharing. Encouraging diverse stakeholders to actively participate in the construction of healthcare data sharing and application through policy incentives will contribute to the overall development.

To enhance data collection, it is imperative to improve the top-level design of information system construction and establish a sustainable mechanism for data collection. Coordinating efforts across multiple departments, it is essential to achieve precise coverage of data resources throughout various stages of human life. Facilitating business collaboration, data gathering, and integrated sharing through multi-platform information systems will provide robust support for the development and application of data resources. Government-led initiatives should encourage diverse entities, including medical institutions, research institutes, universities, and medical device companies, to actively participate in healthcare data sharing and application. Guiding relevant stakeholders to engage in deep data sharing cooperation under government leadership will foster high-quality industrial development and create a favorable ecosystem for collaborative healthcare data.

In terms of data quality, it is proposed to establish a classification and grading standard for healthcare data. This standard should consider factors such as data classification, usage purposes, and other relevant elements. The objectives are to formulate transparent and objective evaluation criteria, incorporate third-party collaboration if necessary, and standardize the

assessment process. For local grassroots data concentration, a standardized technical and security standard system should be established. This system will clarify the technical requirements and assessment standards for local data governance. Through technical guidance, it will promote centralized construction at the data source, providing guidance and enhancing the quality of medical data.

Regarding data sharing, leveraging the existing national health information platform is essential. Continuous improvement of provincial-level coordinated regional health information platforms should be pursued, gradually achieving nationwide data connectivity across national, provincial, municipal, and county levels. Simultaneously, comprehensive management measures covering the entire process of data sharing and application should be gradually formulated and perfected. This will serve as a basis for local implementation of data sharing and application. Clear pathways and methods for sharing, unified data standards, and continuous monitoring of the implementation of data sharing policies should be established. This approach aims to avoid discrepancies where government officials may believe that data sharing has been achieved, while managers in universities and hospitals may feel otherwise.

In terms of usage, relevant departments of the national government should promptly promote standardized or guideline documents for the compliant use of data. It is crucial to clearly define the responsibilities and authority of the central management department and various other departments. The establishment of a comprehensive healthcare data quality supervision mechanism is essential to strengthen the overall supervision of medical services. Simultaneously, the construction of a data integrity system is crucial. Data generated by healthcare data services should be traceable throughout the entire process, leaving a comprehensive record. This ensures that data is searchable and traceable, meeting the regulatory requirements of the industry.

6.2 Establishing industry standards and unified platforms to promote data sharing

The diverse nature of healthcare big data requires the integration and standardization of data formats to greatly facilitate interconnected data sharing.

Firstly, data formats, interfaces and protocols should be unified to ensure seamless transmission and interpretation of data. This guarantees compatibility, exchange, and sharing of data among different systems.

Secondly, various departments within hospitals should be encouraged to adopt a unified

data management system or technical standards so as to simplify internal data sharing and integration.

Thirdly, a centralized platform or coordinating body is essential to coordinate data sharing among different hospitals and systems.

Meanwhile, Industry-wide planning and layout, guided by policies, can gradually establish and improve industry standard systems. Encouraging collaboration between medical institutions, forming industry alliances or professional organizations, and utilizing shared platforms or forums for sharing best practices and experiences collectively advance interconnected sharing of healthcare big data.

6.3 Ensuring funding and resource investment to support data platform construction

The construction of the healthcare data infrastructure requires substantial financial and resource investment.

Firstly, leveraging fiscal entities is crucial. This involves using special fiscal funds to support the construction of comprehensive health data management platforms at the provincial, municipal and district levels. Simultaneously, involving social capital in data platform construction can be achieved through industry development funds and industry guidance funds, promoting resource sharing, and capitalizing on complementary strengths. This approach aims to enhance the efficiency of fiscal fund utilization.

Secondly, entities like hospitals should assess and allocate internal resources based on the development process. Allocating a portion of funds to support the construction of healthcare big data sharing is essential. Viewing data sharing as a vital component of hospital informatization development, necessary financial support should be provided to data sharing projects without compromising other crucial initiatives.

Furthermore, strengthening external collaboration and joint construction is advisable. Collaborating with other medical institutions, research entities or technology companies to jointly apply for funds or share resources can facilitate resource complementarity and sharing. By collectively shouldering project costs, this approach helps alleviate the financial burden on individual institutions.

6.4 Nurturing specialized and interdisciplinary talents to support healthcare data sharing and applications

Talent plays a crucial role in realizing the sharing and application of healthcare big data. The collaborative efforts required for big data analysis, sharing and application demand individuals with expertise in medicine, data statistics analysis, computer applications and management skills. Therefore, while harnessing domain-specific knowledge to explore health big data, there is a need to cultivate interdisciplinary talents. From the construction of the big data platform itself to the development of talents in healthcare management and data intelligence analysis, it is essential to gradually establish a comprehensive talent development pipeline and a multi-level and multidimensional training system.

Collaborative initiatives involving healthcare management departments, research institutions, universities and their affiliated hospitals, and healthcare institutions can systematically train influential discipline leaders and industry pioneers, fostering the development of a competent talent pool. By promoting cognitive management transformation at the decision-making level and cultivating talents skilled in data analysis operations, a supportive environment is created for healthcare data sharing and application with a patient-centered value orientation and an evolving collaborative ecosystem.

6.5 Enhancing network and data security defense capabilities, strengthening data security

To bolster the protection of network and data security, several key measures should be implemented. Firstly, the establishment of a unified standard for medical data privacy protection is paramount. This includes stringent requirements for data anonymization, encryption technologies, and the widespread adoption of efficient permission management systems. The introduction of advanced anonymization techniques is essential to ensure the privacy and security of data during the sharing process.

Simultaneously, the implementation of a robust security assessment mechanism is crucial. This involves promptly adjusting technical measures to address the continuously evolving challenges in data security. Secondly, it is imperative to clearly define and regulate the compliance of data usage. This ensures that data is employed only within authorized boundaries, accompanied by the establishment of explicit review and monitoring mechanisms.

Additionally, providing patients with transparent explanations of data usage policies is essential. This empowers them to understand how their personal data will be utilized, thereby enhancing their trust in the use of their data.

Finally, offering training and guidance on data privacy protection for medical staff is crucial. Emphasizing the significance of patient privacy, educating them on the correct handling and usage of sensitive data is essential. The establishment of comprehensive internal regulations underscores individual responsibilities and obligations, ensuring that everyone is well-informed, respects and adheres to privacy protection policies.

6.6 Advancing healthcare big data applications for sustainable development

By integrating multidimensional aspects of healthcare data technologies, methods, and analytical decision-making, the unified application of diverse data from various sources, dimensions, types, and levels can significantly enhance personalized and refined healthcare services. Government departments, through in-depth exploration and analysis of healthcare data, can drive the establishment of public disease monitoring and epidemic analysis systems, providing a basis for decision-making related to healthcare resource allocation and public health resource scheduling. The continual improvement of data construction through the application of shared data ultimately supports the advancement of comprehensive reforms in public hospitals and enhances modern hospital management systems and services.

[This page is deliberately left blank.]

Chapter 7: Conclusion

7.1 Main Conclusions

In this study, we reviewed Stakeholder Theory, applying the Mitchell three-dimensional attribute classification method, empirical research, and in-depth interviews. Based on a crucial stakeholder perspective, the research scientifically, systematically and rigorously explored the key influencing factors and implementation strategies for healthcare big data sharing and application.

One main focus of the research revolved around identifying the key stakeholders in the process of healthcare big data sharing and application, and which stakeholders play the most crucial roles.

Additionally, the study delved into understanding what are the interest demands of key stakeholders in participating in the process of healthcare big data sharing and application, and also what are the decisive factors that promote or hinder healthcare big data sharing and application, from the perspective of key stakeholders.

Finally, the study aimed to identify “strategies that can be constructed to achieve effective healthcare big data sharing and application. The main conclusions include:

(1) In the current landscape of information technology and medical practice, researchers conducted consultations with 62 experienced managers and researchers in the field of the sharing and application of healthcare big data, who hold positions in various institutions such as the government, the Health Commission, Medical Insurance Bureau, Medical Products Administration, medical schools, and affiliated hospitals in Guangdong Province. Their insights revealed that 11 primary stakeholders, including government entities, health commissions, insurance departments, school administrators, clinical department heads, IT leaders, medical personnel, researchers, patients (residents), and third-party organizations, are crucial stakeholders in the process of healthcare big data application and sharing. Subsequently, by integrating Mitchell’s three-dimensional attribute classification method with empirical research based on 573 valid survey responses, it was clarified that among these 11 primary stakeholders, the health commissions, government, insurance departments, and hospital administrators play pivotal roles in terms of legitimacy, power, and urgency. They are key stakeholders in the

application and sharing of healthcare big data in medical schools and their affiliated hospitals. The research emphasizes the importance of prioritizing and addressing the interests of these four groups during the process of the sharing and application of healthcare big data.

(2) Healthcare big data encompasses a wide range of information characterized by its vast volume and varied forms. While government officials have observed some level of sharing and application of healthcare big data, the extent of this activity is somewhat limited. Conversely, administrators in medical schools and affiliated hospitals perceive that the sharing and application of healthcare big data have not been effectively realized. Several factors contribute to the ineffective sharing and application of data: 1) Lack of top-level design: There is a need to establish a unified system and platform for data collection and sharing from a strategic level. 2) Disparities in information technology development: Different units have varying levels of development in information technology, leading to a lack of clear mechanisms for data sharing and application and equitable distribution of benefits. 3) Ambiguity in standards, regulations, and responsibilities: Unclear standards, laws, regulations, and responsibilities regarding data sharing hinder effective implementation. 4) Insufficient resources: The continuous investment of substantial resources, including finances, personnel, and equipment, is required for the sharing and application of healthcare big data. However, current support in terms of specialized funds, talents, and equipment is relatively limited. 5) Inadequate cybersecurity and data security measures: The incomplete development of cybersecurity and data security systems poses risks to patient privacy and hospital interests. 6) Discrepancies in data collection quality: Data collection quality varies among different units or even within the same unit's different departments, indicating a need for improvement in data cleansing and processing methods.

(3) To effectively implement strategies for the sharing and application of healthcare big data, it is essential to start with top-level design. This involves promptly issuing policies, implementation guidelines, and supporting documents related to the management of healthcare big data resources. Additionally, it is crucial to establish principles for the allocation of benefits in healthcare data sharing and application and to enhance the institutional system for data sharing and application from a policy perspective. Encouraging active participation from diverse stakeholders is key to building a robust framework for healthcare data sharing and application. Unified planning and layout are necessary to establish industry standards for data collection, aggregation, and cleansing, as well as standardized and normalized central integration platforms. Securing funding and resources through active efforts to garner support from financial entities, leveraging internal resource allocation, and strengthening external collaborations are essential. Moreover, establishing a multi-level and diversified talent training

system is imperative to cultivate high-level professional and interdisciplinary talents for the sharing and application of healthcare big data. Strict privacy protection standards and security mechanisms must be put in place, including providing training on data security and privacy protection for relevant personnel and reinforcing data network security barriers to ensure data security and personal privacy.

7.2 Research limitations and future research

This study identifies the key stakeholders influencing the sharing and application of healthcare big data. Based on the results of in-depth interviews, it develops strategies for the sharing and application of healthcare big data, enriching the connotations and implications of stakeholder theory.

The study promotes the concretization and orientation of responsibilities and rights for key stakeholders in the process of sharing and applying healthcare big data, establishing a symbiotic and mutually beneficial data sharing and application system. This is beneficial for advancing medical scientific research, improving clinical diagnosis and treatment capabilities, optimizing the allocation of public medical resources, and providing the population with higher quality, efficient, fair, accessible, convenient, and affordable medical services.

However, given this study's limitations due to the diverse types, extensive sources, and varied forms of healthcare big data, involving a wide range of stakeholders at different levels, future research on strategies for sharing and applying healthcare big data from the perspective of stakeholders needs to be further deepened and expanded.

(1) To address the current practical challenges in management, this study focused on the sharing and application of healthcare big data in a medical university and its affiliated hospitals in Guangdong Province. However, healthcare big data encompass all information related to an individual's entire life course, including health conditions, diseases, and diagnosis and treatment. Therefore, in future management practices and scientific research, it would be beneficial to further expand and extend the study horizontally by incorporating data from disease prevention and control centers, occupational disease prevention and control centers, inspection and quarantine departments, medical insurance departments, food and drug supervision and management departments, and data generated by third-party wearable devices.

(2) The researcher has proposed strategies to promote the sharing and application of healthcare big data in medical universities and their affiliated hospitals, based on the current state of information technology and medical practices. The roles, positions and influence of

stakeholders do not remain unchanged. Any change in the behavior strategy of a key stakeholder not only affects the process of sharing and applying healthcare big data but may also influence the behavior strategies of other key stakeholders. Therefore, the constructed strategies can be applied to subsequent management practices, evaluating their effectiveness in promoting healthcare big data sharing over time. Through continuous adjustments and improvements in practice, these strategies can enhance data application and management practices in the university and its affiliated hospitals. Additionally, this iterative process allows for the accumulation of experience, providing a standard paradigm for constructing strategies for healthcare big data sharing and application at different stages.

Bibliography

- Ansoff, H. I. (1965). *Corporate strategy: An analytical approach to business policy for growth and expansion*. McGraw-Hill.
- Antman, E. M., Benjamin, E. J., Harrington, R. A., Houser, S. R., Peterson, E. D., Bauman, M. A., Brown, N., Bufalino, V., Califf, R. M., Creager, M. A., Daugherty, A., Demets, D. L., Dennis, B. P., Ebadollahi, S., Jessup, M., Lauer, M. S., Lo, B., MacRae, C. A., McConnell, M. V., & Wayte, P. (2015). Acquisition, analysis, and sharing of data in 2015 and beyond: A survey of the landscape: A conference report from the American heart association data summit 2015. *Journal of the American Heart Association*, 4(11), e2810.
- Armstrong, S. (2017). Data, data everywhere: The challenges of personalised medicine. *British Medical Journal*, 359, j4546.
- Bahga, A., & Madiseti, V. K. (2013). A cloud-based approach for interoperable electronic health records (EHRs). *"IEEE" Journal of Biomedical and Health Informatics*, 17(5), 894-906.
- Berg, J. (2017). Data in public health. *Science*, 355(6326), 669.
- Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., Claffey, M. P., Sheehan, J., & Patrick, K. (2016). Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association*, 23(e1), e42-e48.
- Blair, J. D., Slaton, C. R., & Savage, G. T. (1990). Hospital-physician joint ventures: A strategic approach for both dimensions of success. *Hospital & Health Services Administration*, 35(1), 3-26.
- Blair, J. D., & Whitehead, C. J. (1988). Too many on the seesaw: Stakeholder diagnosis and management for hospitals. *Hospital & Health Services Administration*, 33(2), 153-166.
- Brugha, R., & Varvasovszky, Z. (2000). Stakeholder analysis: A review. *Health Policy and Planning*, 15(3), 239-246.
- Bull, S., Roberts, N., & Parker, M. (2015). Views of ethical best practices in sharing individual-level data from medical and public health research: A systematic scoping review. *Journal of Empirical Research On Human Research Ethics*, 10(3), 225-238.
- Bull, S., & Bhagwandin, N. (2020). The ethics of data sharing and biobanking in health research. *Wellcome Open Research*, 5, 270.
- Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X., & Greene, C. S. (2020). Responsible, practical genomic data sharing that accelerates research. *Nature Reviews Genetics*, 21(10), 615-629.
- Carr, D., & Littler, K. (2015). Sharing research data to improve public health: A funder perspective. *Journal of Empirical Research On Human Research Ethics*, 10(3), 314-316.
- Carroll, A. B. (1996). *Business and society: Ethical and stakeholder management* (3rd ed.). Southwestern College Publishing.
- Chan, A. T., Drew, D. A., Nguyen, L. H., Joshi, A. D., Ma, W., Guo, C. G., Lo, C. H., Mehta, R. S., Kwon, S., Sikavi, D. R., Magicheva-Gupta, M. V., Fatehi, Z. S., Flynn, J. J., Leonardo, B. M., Albert, C. M., Andreotti, G., Beane-Freeman, L. E., Balasubramanian, B. A., Brownstein, J. S., & Spector, T. (2020). The coronavirus pandemic epidemiology (cope) consortium: A call to action. *Cancer Epidemiology Biomarkers & Prevention*, 29(7), 1283-1289.
- Charkham, J. (1992). Corporate governance: Lessons from abroad. *European Business Journal*, 2(4), 8-16.

- Chen, H. H., & Jia, S. H. (2004). 企业利益相关者三维分类的实证分析 [Empirical analysis of three-dimensional classification of enterprise stakeholders]. *Economic Research*, (4), 80-90.
- Chen, Y. Y., & Wang, Y. Y. (2020). 科研数据开放共享的利益相关者互动关系 [Research on stakeholder interaction in the open sharing of scientific research data]. *Library Tribune*, 40(5), 55-63.
- Chen, Z. M., Chen, J. S., Collins, R., Guo, Y., Peto, R., Wu, F., & Li, L. M. (2011). China Kadoorie Biobank of 0.5 million people: Survey methods, baseline characteristics and long-term follow-up. *International Journal of Epidemiology*, 40(6), 1652-1666.
- Clarkson, M. B. E. (1994, June 3-5). *A risk based model of stakeholder theory*. The Second Toronto Conference on Stakeholder Theory, Toronto, ON, Canada.
- Clarkson, M. B. E. (1995). A stakeholder framework for analyzing and evaluating corporate social performance. *The Academy of Management Review*, 20(1), 92-117.
- Cornell, B., & Shapiro, A. (1987). Corporate stakeholders and corporate finance. *Financial Management*, 16(1), 5-14.
- Costa, R., & Menichini, T. (2013). A multidimensional approach for CSR assessment: The importance of the stakeholder perception. *Expert Systems with Applications*, 40(1), 150-161.
- Costeloe, K., Turner, M. A., Padula, M. A., Shah, P. S., Modi, N., Soll, R., Haumont, D., Kusuda, S., Göpel, W., Chang, Y. S., Smith, P. B., Lui, K., Davis, J. M., & Hudson, L. D. (2018). Sharing data to accelerate medicine development and improve neonatal care: Data standards and harmonized definitions. *The Journal of Pediatrics*, 203, 437-441.
- Creswell, J. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Sage Publications, Lnc.
- Damiani, A., Onder, G., & Valentini, V. (2018). Large databases (big data) and evidence-based medicine. *European Journal of Internal Medicine*, 53, 1-2.
- Denny, S. G., Silaigwana, B., Wassenaar, D., Bull, S., & Parker, M. (2015). Developing ethical practices for public health research data sharing in South Africa: The views and experiences from a diverse sample of research stakeholders. *Journal of Empirical Research On Human Research Ethics*, 10(3), 290-301.
- Dexheimer, J. W., Greiner, M. V., Beal, S. J., Johnson, D., Kachelmeyer, A., & Vaughn, L. M. (2019). Sharing personal health record data elements in protective custody: Youth and stakeholder perspectives. *Journal of the American Medical Informatics Association*, 26(8-9), 714-721.
- Dulhanty, A. (2021). Present value of future health data: Ethics of data collection and use. *Bulletin of the World Health Organization*, 99(2), 162-163.
- Dymond, S., Nix, T. W., Rotarius, T. M., & Savage, G. T. (1995). Why do key integrated delivery stakeholders really matter? Assessing control, coalitions, resources and power. *Medical Group Management Journal*, 42(6), 26-38.
- Elliott, J. H., Grimshaw, J., Altman, R., Bero, L., Goodman, S. N., Henry, D., Macleod, M., Tovey, D., Tugwell, P., White, H., & Sim, I. (2015). Informatics: Make sense of health data. *Nature*, 527(7576), 31-32.
- Elsayed, A. M., & Saleh, E. I. (2018). Research data management and sharing among researchers in Arab universities: An exploratory study. *International Federation of Library Associations and Institutions Journal*, 44(4), 281-299.
- Fan, Y. F., Wang, Y. J., Chen, J. M., & Xia, Z. W. (2016). 基于共享的区域医疗大数据发展 [Development of regional medical big data based on information sharing]. *Chinese Journal of Medical Library and Information Science*, 25(12), 7-10.
- Foster, M. W., & Sharp, R. R. (2007). Share and share alike: Deciding how to distribute the scientific and social benefits of genomic data. *Nature Reviews Genetics*, 8(8), 633-639.

- Frederick, W. C. (1988). Creatures, corporations, communities, chaos, complexity: A naturological view of the corporate social role. *Business & Society*, 37(4), 358-389.
- Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. MA Pitman.
- Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B., & de Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge University Press.
- Freeman, R. E., & Evan, W. M. (1990). Corporate governance: A stakeholder interpretation. *Journal of Behavioral Economics*, 19(4), 337-359.
- Freeman, R. E., & Reed, D. L. (1983). Stockholders and stakeholders: A new perspective on corporate governance. *California Management Review*, 25(3), 88-106.
- Gao, J., Zhao, X., Hu, S., Huang, Z., Hu, M., Jin, S., Lu, B., Sun, K., Wang, Z., Fu, J., Weersma, R. K., He, X., & Zhou, H. (2022). Gut microbial DL-endopeptidase alleviates Crohn's disease via the NOD2 pathway. *Cell Host & Microbe*, 30(10), 1435-1449.
- Geissbuhler, A., Safran, C., Buchan, I., Bellazzi, R., Labkoff, S., Eilenberg, K., Leese, A., Richardson, C., Mantas, J., Murray, P., & De Moor, G. (2013). Trustworthy reuse of health data: A transnational perspective. *International Journal of Medical Informatics*, 82(1), 1-9.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. Jossey-Bass.
- Guo, H. W., Yin, W. Q., Zhao, Y. K., Hu, J. W., & Hu, S. L. (2014). 新型农村合作医疗制度利益相关者分析与评价 [Analysis and evaluation of stakeholders of the new rural cooperative medical system]. *Chinese Health Service Management*, 31(6), 451-453.
- Halamka, J. D. (2006). Harmonizing healthcare data standards. *Journal of Healthcare Information Management*, 20(4), 11-13.
- Han, W., Gang, N., Lu, H., Liu, Q. Y., & Wang, J. S. (2020). 基于健康医疗大数据的智能辅助诊断系统研究 [Research on intelligent assisted diagnosis system based on big data of health medical]. *Information Medical*, 17(10), 159-161.
- Hazra, R., Tenney, S., Shlionskaya, A., Samavedam, R., Baxter, K., Ileakis, J., Weck, J., Willinger, M., Grave, G., Tsilou, K., & Songco, D. (2018). DASH, the data and specimen hub of the National Institute of Child Health and Human Development. *Scientific Data*, 5, 180046.
- Hill, C. W. L., & Jones, T. M. (1992). Stakeholder-agency theory. *Journal of Management Studies*, 29(2), 131-154.
- Husedzinovic, A., Ose, D., Schickhardt, C., Frohling, S., & Winkler, E. C. (2015). Stakeholders' perspectives on biobank-based genomic research: Systematic review of the literature. *European Journal of Human Genetics*, 23(12), 1607-1614.
- Javanparast, S., Coveney, J., & Saikia, U. (2009). Exploring health stakeholders' perceptions on moving towards comprehensive primary health care to address childhood malnutrition in Iran: A qualitative study. *Bmc Health Services Research*, 9, 36.
- Jawahar, I. M., & McLaughlin, G. L. (2001). Toward a descriptive stakeholder theory: An organizational life cycle approach. *The Academy of Management Review*, 26(3), 397-414.
- Jia, S. H., & Chen, H. H. (2002). 利益相关者的界定方法述评 [A review of the stakeholder approach]. *Foreign Economy and Management*, 24(5), 122-176.
- Jin, Y., Lu, S. K., & Li, S. H. (2013). 我国医疗联合体的利益相关者分析 [Analysis of Healthcare Alliance based on Stakeholders Theory in China]. *Chinese Hospital Management*, 33(10), 3-4.
- Jones, T. M. (1980). Corporate social responsibility revisited, redefined. *California Management Review*, 22(2), 59-67.
- Kalkman, S., Mostert, M., Gerlinger, C., van Delden, J., & van Thiel, G. (2019). Responsible data sharing in international health research: A systematic review of principles and norms. *Bmc Medical Ethics*, 20(1), 21.

- Kazanjian, A. (1998). Understanding women's health through data development and data linkage: Implications for research and policy. *Canadian Medical Association Journal*, 159(4), 342-345.
- Kenny, G. (2013). The stakeholder or the firm? Balancing the strategic framework. *Journal of Business Strategy*, 34(3), 33-40.
- Kim, K. K., Browe, D. K., Logan, H. C., Holm, R., Hack, L., & Ohno-Machado, L. (2014). Data governance requirements for distributed clinical research networks: Triangulating perspectives of diverse stakeholders. *Journal of the American Medical Informatics Association*, 21(4), 714-719.
- Knut, H. M., & Svein, J. (2001). From user-groups to stakeholders? The public interest in fisheries management. *Marine Policy*, 25(4), 281-292.
- Koutkias, V. (2019). From data silos to standardized, linked, and fair data for pharmacovigilance: Current advances and challenges with observational healthcare data. *Drug Safety*, 42(5), 583-586.
- Laird, Y., Manner, J., Baldwin, L., Hunter, R., McAteer, J., Rodgers, S., Williamson, C., & Jepson, R. (2020). Stakeholders' experiences of the public health research process: Time to change the system? *Health Research Policy and Systems*, 18(1), 83.
- Lei, H. G., Chen, X. C., Yang, G. B., Wang, N., & Liu, Y. X. (2015). 双向转诊中利益相关者的利益诉求实证研究—以东莞市为例 [Empirical analysis on profit-relaters demand in dual referral—Dongguan city as an example]. *Health Service Management in China*, 32(10), 737-742, 764.
- Lemke, A. A., & Harris-Wai, J. N. (2015). Stakeholder engagement in policy development: challenges and opportunities for human genomics. *Genetics in Medicine*, 17(12), 949-957.
- Leyens, L., Reumann, M., Malats, N., & Brand, A. (2017). Use of big data for drug development and for public and personal health and care. *Genetic Epidemiology*, 41(1), 51-60.
- Li, H. Q., Yin, C. Q., & Fan, Y. J. (2019). 中国健康医疗大数据国家战略发展研究 [Research on the national strategy development of China's health and medical big data]. *Library*, (11), 30-37.
- Li, X. H. (2001). 面向可持续发展的利益相关者管理 [Management of the interested parties geared to the needs of sustainable development]. *Modern Finance and Economics*, (1), 66-70.
- Li, Z. H., Zhong, W. F., Liu, S., Kraus, V. B., Zhang, Y. J., Gao, X., Lv, Y. B., Shen, D., Zhang, X. R., Zhang, P. D., Huang, Q. M., Chen, Q., Wu, X. B., Shi, X. M., Wang, D., & Mao, C. (2020). Associations of habitual fish oil supplementation with cardiovascular outcomes and all cause mortality: Evidence from a large population based cohort study. *British Medical Journal*, 368, m456.
- Li, Z., Chen, Y., & Ke, H. (2023). Investigating the Causal Relationship Between Gut Microbiota and Crohn's Disease: A Mendelian Randomization Study. *Gastroenterology*, 166(2), 354-355.
- Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boulton, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., Sheard, S., Smith, S. M., Sudlow, C., Matthews, P. M., & Allen, N. E. (2020). The UK Biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. *Nature Communications*, 11(1), 2624.
- Liu, D., Huang, Y., Huang, C., Yang, S., Wei, X., Zhang, P., Guo, D., Lin, J., Xu, B., Li, C., He, H., He, J., Liu, S., Shi, L., Xue, Y., & Zhang, H. (2022). Calorie restriction with or without time-restricted eating in weight loss. *The New England Journal of Medicine*, 386(16), 1495-1504.

- Love, D. E., Paita, L. M., & Custer, W. S. (2001). Data sharing and dissemination strategies for fostering competition in health care. *Health Services Research*, 36(1), 277-290.
- Luo, Y., & Jiang, J. M. (2011). 利益相关者理论及其分析方法在卫生领域的应用进展 [Application of stakeholder theory and its analytical method in health]. *Chinese Health Service Management*, 28(2), 84-85.
- Luper-foy, S. (1988). The Moral Obligations of Multinational Corporations. In N. Bowie (Ed.), *Problems of international justice* (pp. 97-113). Westview Press.
- Lusher, S. J., McGuire, R., van Schaik, R. C., Nicholson, C. D., & de Vlieg, J. (2014). Data-driven medicinal chemistry in the era of big data. *Drug Discovery Today*, 19(7), 859-868.
- Ma, S. S., Yu, G. J., & Cui, W. B. (2018). 区域卫生信息化环境下健康医疗大数据共享应用思考与建议 [Reflection on the sharing application of health big data in regional health informatization]. *China Digital Medicine*, 13(4), 11-13.
- Mackey, T. K., Kuo, T. T., Gummadi, B., Clauson, K. A., Church, G., Grishin, D., Obbad, K., Barkovich, R., & Palombini, M. (2019). 'Fit-for-purpose?' -challenges and opportunities for applications of blockchain technology in the future of healthcare. *Bmc Medicine*, 17(1), 68.
- Magness, V. (2008). Who are the stakeholders now? An empirical examination of the Mitchell, Agle, and Wood theory of stakeholder salience. *Journal of Business Ethics*, 83, 177-192.
- Malykh, V. L., & Rudetskiy, S. V. (2018). Approaches to medical decision-making based on big clinical data. *Journal of Healthcare Engineering*, (2018), 3917659.
- Martin, E. G., Helbig, N., & Shah, N. R. (2014). Liberating data to transform health care: New York's open data experience. *The Journal of American Medical Association*, 311(24), 2481-2482.
- Mascalzoni, D., Dove, E. S., Rubinstein, Y., Dawkins, H. J., Kole, A., McCormack, P., Woods, S., Riess, O., Schaefer, F., Lochmuller, H., Knoppers, B. M., & Hansson, M. (2015). International charter of principles for sharing bio-specimens and data. *European Journal of Human Genetics*, 23(6), 721-728.
- Mazor, K. M., Richards, A., Gallagher, M., Arterburn, D. E., Raebel, M. A., Nowell, W. B., Curtis, J. R., Paolino, A. R., & Toh, S. (2017). Stakeholders' views on data sharing in multicenter studies. *Journal of Comparative Effectiveness Research*, 6(6), 537-547.
- McKeown, A., Mourby, M., Harrison, P., Walker, S., Sheehan, M., & Singh, I. (2021). Ethical issues in consent for the reuse of data in health data platforms. *Science and Engineering Ethics*, 27(1), 9.
- Miller, J., Ross, J. S., Wilenzick, M., & Mello, M. M. (2019). Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: Cross sectional descriptive study and pilot of a tool to improve company practices. *British Medical Journal*, 366, 14217.
- Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *The Academy of Management Review*, 22(4), 853-886.
- Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, 22(2), 303-341.
- Na, X. (2016). *Study of electronic health records sharing impact factors bases on stakeholders analysis* [Master's thesis]. Chinese Academy of Medical Science Peking Union Medical College.
- Na, X., Guo, M. J., Xie, L. Q., & Hu, H. P. (2015). 国外居民电子健康档案共享服务体系建设及启示 [Construction of service system for sharing electronic health records in foreign countries and its enlighten-ments]. *Chinese Journal of Medical Library and Information*, 24(10), 18-21.
- Naci, H., Cooper, J., & Mossialos, E. (2015). Timely publication and sharing of trial data: Opportunities and challenges for comparative effectiveness research in cardiovascular

- disease. *European Heart Journal. Quality of Care & Clinical Outcomes*, 1(2), 58-65.
- Nahar, J., Imam, T., Tickle, K. S., & Garcia-Alonso, D. (2013). Issues of data governance associated with data mining in medical research: Experiences from an empirical study. *Studies in Health Technology and Informatics*, 193, 332-361.
- Nandraj, S., Khot, A., Menon, S., & Brugha, R. (2001). A stakeholder approach towards hospital accreditation in India. *Health Policy and Planning*, 16 Suppl 2, 70-79.
- Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.
- Obenshain, M. K. (2004). Application of data mining techniques to healthcare data. *Infection Control and Hospital Epidemiology*, 25(8), 690-695.
- Packer, M. (2018). Data sharing in medical research. *British Medical Journal*, 360, k510.
- Parasidis, E., Pike, E., & McGraw, D. (2019). A belmont report for health data. *The New England Journal of Medicine*, 380(16), 1493-1495.
- Patrick, K. (2016). Harnessing big data for health. *Canadian Medical Association Journal*, 188(8), 555.
- Pisani, E., Aaby, P., Breugelmans, J. G., Carr, D., Groves, T., Helinski, M., Kamuya, D., Kern, S., Littler, K., Marsh, V., Mboup, S., Merson, L., Sankoh, O., Serafini, M., Schneider, M., Schoenenberger, V., & Guerin, P. J. (2016). Beyond open data: Realising the health benefits of sharing data. *British Medical Journal*, 355, i5295.
- Rachel, F., & Baskerville, M. (2004). Dangerous, dominant, dependent, or definitive: Stakeholder identification when the profession faces major transgressions. *Accounting and the Public Interest*, 4(1), 24-42.
- Room, S. (2004). Data protection, informed consent, and research: Data protection act does not bar medical research. *British Medical Journal*, 328(7453), 1437.
- Rowhani-Farid, A., Allen, M., & Barnett, A. G. (2017). What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review*, 2, 4.
- Savage, G. T., Nix, T. W., Whitehead, C., & Blair, J. D. (1991). Strategies for assessing and managing organizational stakeholders. *Academy of Management Executive*, 5(2), 61-75.
- Sheng, X. P., & Wang, Y. (2019). 利益相关者在科学数据开放共享中的责任与作用—基于国际组织科学数据开放共享政策的分析 [The responsibilities and roles of stakeholders in the open sharing of scientific data: An analysis based on policies of open sharing of scientific data in international organizations]. *Library and Information Service*, 63(17), 31-39.
- Song, Y., Jia, W. P., Han, K., Liu, M., & He, Y. (2021). 健康医疗大数据的应用及其挑战 [Application and challenges of big data in health care]. *Chinese Journal of Prevention and Control of Chronic Non-Communicable Diseases*, 29(3), 220-223.
- Starik, M. (1995). The Should trees have managerial standing? Toward stakeholder status for non-human nature. *Journal of Business Ethics*, 14(5), 207-217.
- Sullivan, F., McKinstry, B., & Vasishta, S. (2019). The "All of Us" research program. *The New England Journal of Medicine*, 381(19), 1883-1884.
- Szirbik, N. B., Pelletier, C., & Chausalet, T. (2006). Six methodological steps to build medical data warehouses for research. *International Journal of Medical Informatics*, 75(9), 683-691.
- Topping, S., & Fottler, M. D. (1990). Improved stakeholder management: The key revitalizing the HMO movement? *Medical Care Review*, 47(3), 365-393.
- Varvasovszky, Z., & Brugha, R. (2000). How to do (or not to do) : A stakeholder analysis. *Health Policy and Planning*, 15(3), 338-345.
- Walker, D. H. T., & Hampson, K. D. (2003). *Implications of human capital issues. Procurement strategies: A relationship-based approach*. Blackwell Publishing.
- Wan, J. H., Dai, Z. W., & Chen, J. (1998). 利益相关者管理 [Stakeholder management].

- Haitian Publishing House.
- Wang, H. H., Wu, X., & Wang, H. (2017). 区域健康医疗数据共享及隐私保护策略研究 [Research on regional health and medical data sharing and privacy protection strategy]. *Technology Innovation and Application*, (31), 181-182.
- Wang, Q. (2015). *Research on implementation mechanism of corporate social responsibility based on stakeholder theory* [Doctoral dissertation]. Harbin Institute of Technology.
- Wheeler, D., & Maria, S. (1998). Including the stakeholders: The business case. *Long Range Planning*, 31(2), 201-210.
- Wu, J. H., Wang, Y., Tao, L., & Peng, J. M. (2019). Stakeholders in the healthcare service ecosystem. *Procedia Cirp*, 83, 375-379.
- Wu, K., & Ren, Y. F. (2017). 医疗机构与IT企业合作共享医疗健康数据的策略研究 [Research of medical data sharing strategies under the collaboration between medical institutions and information technology enterprises]. *Chinese Journal of Health Informatics and Management*, 14(5), 649-651.
- Wu, L. R., & Zeng, J. (2019). 健康医疗大数据在商业健康保险中的应用探讨 [Discussion on the application of health and medical big data in commercial health insurance]. *Business Study*, 1(34), 69-70.
- Wu, L., & He, H. M. (2005). 基于企业生命周期的利益相关者分类及其实证研究 [Empirical study of the classification of stakeholders based on enterprise lifecycle]. *Journal of Sichuan University (Philosophy and Social Sciences Edition)*, (6), 34-38.
- Xu, T., & Yu, G. (2020). 健康医疗大数据共享的应用场景及价值探析 [Exploration and analysis of application scenarios and value of healthcare big data sharing]. *China Digital Medicine*, 15(7), 1-3.
- Xu, X. L. (2018). *Division and cooperation mechanism of medical institutions and its implementation strategy under the background of grading treatment-based on the stakeholder's perspective* [Doctoral dissertation]. JiangSu University.
- Xu, X. L., Zhou, L. L., & Wei, J. J. (2017). 医疗卫生服务体系整合研究的回顾与展望 [Review and Prospect of Integrated Medical and Health Service System]. *Chinese Health Economy*, 36(7), 17-21.
- Yang, L., & Yi, K. J. (2003). 从股东赢利到利益相关者共赢——重塑上市公司价值观 [From shareholder's profit to stakeholder's win & win-remould the value of city corporation]. *Financial Theory and Practice*, (1), 67-69.
- Yang, R. L., & Zhou, Y. A. (2000). *企业的利益相关者理论及其应用* [The stakeholder theory of enterprises and its application]. Economic Science Press.
- Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage Publications, Lnc.
- Yu, G. P., Bao, X. Y., Huang, X. T., Liu, W., Xu, B. B., Yu, N., & Zhang, J. (2014). 医疗健康大数据的种类、性质及有关问题 [Medical and health big data: Types, characteristic and relevant issues]. *Journal of Medical Informatics*, 35(6), 9-12.
- Zhang, J. N., Li, Y. Y., Gu, Y. J., Zhu, Y. L., & He, Q. F. (2020). 健康医疗数据共享基本原则探讨 [Discussion on the basic principles for health and medical data sharing]. *Engineering*, 22(4), 93-100.
- Zhang, L. X., Wang, H. B., Li, Q. Z., Zhao, M. H., & Zhan, Q. M. (2018). Big data and medical research in China. *British Medical Journal*, 360, j5910.
- Zhang, P. D., Zhang, X. R., Zhang, A., Li, Z. H., Liu, D., Zhang, Y. J., & Mao, C. (2022). Associations of genetic risk and smoking with incident COPD. *The European Respiratory Journal*, 59(2), 2101320.
- Zhang, S. (2021). 企业社会责任对消费者忠诚度的影响研究 [Research on the influence of corporate social responsibility on consumer loyalty]. *Journal of Henan University (Social*

- Sciences Edition*), 61(06), 24-32.
- Zheng, X., Sun, S., Mukkamala, R. R., Vatrupu, R., & Ordieres-Mer é J. (2019). Accelerating health data sharing: A solution based on the internet of things and distributed ledger technologies. *Journal of Medical Internet Research*, 21(6), e13583.
- Zhou, J., Shi, Z., & Shen, Y. W. (2021). 转化医学实验平台建设 with 共享机制探讨 [Construction and sharing mechanism of translational medicine experimental platform]. *Chinese Hospital Management*, 41(4), 79-82.
- Zinkhan, G. M., & Balazs, A. L. (2004). A stakeholder-integrated approach to health care management. *Journal of Business Research*, 57(9), 984-989.

Webliography

- General Office of Guangdong Provincial People's Government. (2017). *Implementation plan for the construction of the Pearl River Delta national comprehensive pilot zone for big data*. General Office of Guangdong Provincial People's Government. Retrieved November 20, 2020, from http://www.gd.gov.cn/gkmlpt/content/0/145/post_145958.html#7
- General Office of Guangdong Provincial People's Government. (2020). *Look at the Guangdong Provincial government work report in 2020*. General Office of Guangdong Provincial People's Government. Retrieved February 16, 2021, from http://www.gd.gov.cn/gdywdt/xwdt/content/post_2864189.html
- General Office of Guangdong Provincial People's Government. (2021). *Look at the Guangdong Provincial government work report in 2021*. General Office of Guangdong Provincial People's Government. Retrieved February 16, 2021, from http://www.gd.gov.cn/gdywdt/xwdt/content/post_3183270.html
- General Office of the State Council of the People's Republic of China. (2016). *Guiding opinions of the General Office of the State Council on promoting and regulating the application and development of health and medical big data*. The People's Republic of China Website. Retrieved October 02, 2020, from http://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm
- Ma, X. R. (2021). *The work report of Guangdong Provincial government in 2021*. Nanfang Daily. Retrieved January 28, 2021, from http://news.southcn.com/gd/content/2021-01/28/content_192037869.htm

[This page is deliberately left blank.]

Annex A: Expert Consultation Form for Identifying Key Stakeholders in the Sharing and Application of Healthcare Big Data in Medical Colleges and Their Affiliated Hospitals

Dear experts and leaders,

Hello!

In October 2020, with the approval of the National Health Commission, the National Medical and Health Big Data Research Institute was established based on Southern Medical University. In order to further advance institutional reforms and accelerate the construction of the National Medical and Health Big Data Research Institute, a survey is now being conducted to effectively promote the sharing and application of healthcare big data, fully unleash the intrinsic value of data.

We sincerely invite you to participate in this questionnaire survey. The questionnaire is divided into two parts: the first part is a basic information survey, and the second part is an expert opinion consultation form. Thank you for your strong support in helping us with this survey. We kindly request you to provide feedback on the questionnaire results to our office, and we sincerely appreciate your assistance!

Development Planning Department

June 7, 2022

Part 1: Personal Basic Information

1. Gender: ①Male ②Female
2. Date of Birth: ①20-29 years old ②30-39 years old ③40-49 years old ④50-59 years old ⑤60 years old and above
3. Workplace: ①Government Department ②University ③Research Institution ④Hospital
4. Title: ①Senior ②Associate Senior
5. Position: ①Department-level Leader ②Section-level Leader ③Department Head ④Administrative Staff ⑤Medical Staff ⑥Research Staff
6. Years of Work Experience: ①Less than 5 years ②5-10 years ③11-15 years ④16-

20 years ⑤21-25 years ⑥26-30 years ⑦30 years and above

7. Familiarity with the Field of Stakeholders in the Sharing and Application of Healthcare Big Data: ①Very familiar ②Quite familiar ③Moderately familiar ④Not very familiar ⑤Not familiar at all

Part 2: Expert Consultation Form

Table A.1 Expert Consultation Form

Stakeholders in Healthcare Big Data Sharing and Application	Yes (✓)	No (×)
Government		
Health Commission		
Health Insurance Department		
School Administrators		
Hospital Administrators		
Clinical Department Head		
Information Department Leader		
Medical Staff		
Research Staff		
Patients (Residents)		
Third-party Institutions		

Please indicate whether the provided candidates belong to the stakeholders in the sharing and application of healthcare big data in medical colleges and their affiliated hospitals. If you believe they do, mark “✓” in the second column. If you believe they do not, mark “×” in the second column.

Annex B: Questionnaire for Identifying the Three-dimensional Attributes of Mitchell Stakeholders in the Sharing and Application of Healthcare Big Data

Dear experts and leaders,
Hello!

In October 2020, with the approval of the National Health Commission, the National Medical and Health Big Data Research Institute was established based on Southern Medical University. In order to further advance institutional reforms and accelerate the construction of the National Medical and Health Big Data Research Institute, a survey is now being conducted to effectively promote the sharing and application of healthcare big data, fully unleash the intrinsic value of data.

We sincerely invite you to participate in this questionnaire survey. The questionnaire is divided into two parts: the first part is a basic information survey, and the second part is the three-dimensional attributes of Mitchell stakeholders in the sharing and application of healthcare big data. Thank you for your strong support in helping us with this survey. We kindly request you to provide feedback on the questionnaire results to our office, and we sincerely appreciate your assistance!

Development Planning Department

September 6, 2022

Part 1: Personal Basic Information

1. Gender: ①Male ②Female
2. Date of Birth: ①20-29 years old ②30-39 years old ③40-49 years old ④50-59 years old ⑤60 years old and above
3. Workplace: ①Government Department ②University ③Research Institution ④Hospital
4. Title: ①Senior ②Associate Senior
5. Position: ①Department-level Leader ②Section-level Leader ③Department Head ④Administrative Staff ⑤Medical Staff ⑥Research Staff

6. Years of Work Experience: ①Less than 5 years ②5-10 years ③11-15 years ④16-20 years ⑤21-25 years ⑥26-30 years ⑦30 years and above
7. Familiarity with the Field of Stakeholders in the Sharing and Application of Healthcare Big Data: ①Very familiar ②Quite familiar ③Moderately familiar ④Not very familiar ⑤Not familiar at all

Part 2: Mitchell Three-dimensional Attributes

The following entities are the main stakeholders in the sharing and application of healthcare big data in medical colleges and their affiliated hospitals. Please rate them based on your perspective in terms of Legitimacy, Importance, and Urgency. The scoring is on a 5-point scale, where 1 indicates the weakest and 5 indicates the strongest (Table B.1).

Table B.1 Mitchell Three-dimensional Attributes Questionnaire

Stakeholders in Healthcare Big Data Sharing and Application	Legitimacy	Power	Urgency
Government			
Health Commission			
Health Insurance Department			
School Administrators			
Hospital Administrators			
Clinical Department Head			
Information Department Leader			
Medical Staff			
Research Staff			
Patients (Residents)			
Third-party Institutions			

Legitimacy: Indicates whether the actions of an individual or group are acceptable, correct, or appropriate within a certain societal norm, value, belief, or rule system. It also refers to the social, moral, legal, or other specific types of rights conferred on an individual or group for participating in organizational construction and rights for claiming profit distribution.

Power (Importance): Refers to whether a certain interest group possesses or can obtain a status, ability, and means of influencing organizational decisions in terms of mandatory, utilitarian, or normative effects. It explains the reliability of a certain stakeholder's existence in the organization.

Urgency: Highlights the urgency of a certain stakeholder for the development of the organization, indicating the crucial role of the stakeholder for the organization. At the same time, the stakeholder demands an immediate response and attention to its interest requests.

Annex C: Interview Outline for Healthcare Big Data Sharing and Application in a Medical College and Its Affiliated Hospitals (Government Department Manager)

Part 1: Personal Basic Information

1. Gender: ①Male ②Female
2. Date of Birth: ①20-29 years old ②30-39 years old ③40-49 years old ④50-59 years old ⑤60 years old and above
3. Workplace: ①Government Department ②University ③Research Institution ④Hospital
4. Title: ①Senior ②Associate Senior
5. Position: ①Department-level Leader ②Section-level Leader ③Department Head ④Administrative Staff ⑤Medical Staff ⑥Research Staff
6. Years of Work Experience: ①Less than 5 years ②5-10 years ③11-15 years ④16-20 years ⑤21-25 years ⑥26-30 years ⑦30 years and above
7. Familiarity with the Field of Stakeholders in the Sharing and Application of Healthcare Big Data: ①Very familiar ②Quite familiar ③Moderately familiar ④Not very familiar ⑤Not familiar at all

Part 2: Interview Content

1. Does your department manage healthcare big data? What are the main components of the data?
2. How do you perceive and approach the sharing and application of healthcare big data?
3. In your opinion, what potential benefits can the sharing and application of healthcare big data bring to the development of the healthcare industry?
4. What potential risks do you foresee in the construction of healthcare big data sharing and application?
5. What are the main challenges faced in the process of building healthcare big data sharing and application?
6. Does your department have information systems that support the sharing and

application of healthcare big data?

7. In your scope of work, how is the automatic exchange and sharing of data among hospitals?
8. What aspects do you think need to be addressed to achieve healthcare big data sharing and application within the university system's affiliated hospitals?
9. Is there any special funding currently allocated for the construction of healthcare big data sharing and application?
10. What legal regulations related to the legal protection of healthcare big data sharing are you aware of?
11. In your opinion, how is the legal protection of healthcare big data sharing primarily reflected?
12. How should privacy data in healthcare big data information be protected, in your view?
13. What measures should be taken for the security and protection of healthcare big data?
14. Do you have any additional opinions or suggestions regarding the sharing and application of healthcare big data?

Annex D: Interview Outline for Healthcare Big Data Sharing and Application in a Medical College and Its Affiliated Hospitals (University or Hospital Manager)

Part 1: Personal Basic Information

1. Gender: ①Male ②Female
2. Date of Birth: ①20-29 years old ②30-39 years old ③40-49 years old ④50-59 years old ⑤60 years old and above
3. Workplace: ①Government Department ②University ③Research Institution ④Hospital
4. Title: ①Senior ②Associate Senior
5. Position: ①Department-level Leader ②Section-level Leader ③Department Head ④Administrative Staff ⑤Medical Staff ⑥Research Staff
6. Years of Work Experience: ①Less than 5 years ②5-10 years ③11-15 years ④16-20 years ⑤21-25 years ⑥26-30 years ⑦30 years and above
7. Familiarity with the Field of Stakeholders in the Sharing and Application of Healthcare Big Data: ①Very familiar ②Quite familiar ③Moderately familiar ④Not very familiar ⑤Not familiar at all

Part 2: Interview Content

1. For your institution, what aspects do you believe constitute healthcare big data?
2. How do you perceive and approach the sharing and application of healthcare big data?
3. In your opinion, what potential benefits can the sharing and application of healthcare big data bring to the development of your hospital, university, or the healthcare industry in Guangdong Province?
4. What potential risks do you foresee in the construction of healthcare big data sharing and application for your hospital?
5. What are the main challenges faced in the process of building healthcare big data sharing and application?
6. Does your hospital have information systems that support the sharing and application

of healthcare big data?

7. In the Southern Medical University system, how is the automatic exchange and sharing of data among affiliated hospitals?
8. What aspects do you think need to be addressed to achieve healthcare big data sharing and application within the university system's affiliated hospitals?
9. Is there any special funding currently allocated for the construction of healthcare big data sharing and application?
10. What legal regulations related to the legal protection of healthcare big data sharing are you aware of?
11. In your opinion, how is the legal protection of healthcare big data sharing primarily reflected?
12. How should privacy data in healthcare big data information be protected, in your view?
13. What measures should be taken for the security and protection of healthcare big data?
14. Do you have any additional opinions or suggestions regarding the sharing and application of healthcare big data?

Annex E: Information Gathered During the Interviews on Healthcare Data Sharing

Table E.1 Information gathered during the interviews on healthcare data sharing

Inter view num ber	Interview time	Interview method	Interview length (min)
A1	2023.03.20	Face-to-face interview	36
A2	2023.04.08	Face-to-face interview	41
A3	2023.04.10	Face-to-face interview	36
A4	2023.04.15	Face-to-face interview	44
A5	2023.04.20	Face-to-face interview	33
A6	2023.04.21	Face-to-face interview	32
A7	2023.05.10	Face-to-face interview	51
B1	2023.02.15	Face-to-face interview	37
B2	2023.02.16	Face-to-face interview	54
B3	2023.02.17	Face-to-face interview	52
B4	2023.02.19	Face-to-face interview	59
B5	2023.03.14	Face-to-face interview	47
B6	2023.04.25	Face-to-face interview	42
B7	2023.05.11	Face-to-face interview	33
B8	2023.05.25	Face-to-face interview	37