Security Risks and Governance Paths of Healthcare Data

Xing Jiang[1] Ke Huang[2] Qinglian Deng[3] Haicen Guo[4] Yuqing Zhang[5]

[1,2,3,4,5] *College of Humanities, Hubei University of Chinese Medicine, Wuhan, Hubei 430065, China*
[5]*Corresponding author. Email: zyq95@hbtcm.edu.cn*

ABSTRACT
Healthcare data covering the full life circle of individuals is a fundamental strategic resource and core asset of China. It is the fusion point of the two national strategies of "Healthy China" and "Digital China", and also the intersection point of the two national strategies of "Innovation Power" and "Digital China". The development of healthcare data should be standardized, orderly, safe and controllable. Currently, healthcare data in China faces security risks in areas such as security management, standard systems, lifecycle, supply chain, and external threats. This article proposes countermeasures from the aspects of establishing a healthcare data security service system, strengthening important data control through classification and grading, improving data security technology protection system, and strengthening health and medical data security supervision, in order to provide reference for building a legal ecosystem for the application of healthcare big data.

Keywords: *Healthcare data, Personal information, Security risks.*

1. INTRODUCTION

Data is both a resource and an asset, which is crucial for the future development of the healthcare industry. [1] Healthcare data is an important support for promoting high-quality development of the healthcare industry. In recent years, the application and development of healthcare data have risen to a national strategic level. Due to its carrying of massive data resources and the continuous emergence of various new businesses and applications, it also faces many security risks. Once healthcare data is tampered with, damaged, and leaked, it will inevitably pose a serious threat to the reputation of medical institutions, the privacy and health safety of both doctors and patients, and even affect social harmony and stability.

2. THE MEANING AND CHARACTERISTICS OF HEALTHCARE DATA

2.1 Meaning

Although healthcare data is widely used, its legal concept is unclear. The theoretical and practical understanding is not consistent, and the differences are reflected in the terminology and definition of healthcare data. In the theoretical field, there are terms such as "health data", "healthcare data", "medical data", "healthcare", "big data medical information", "personal medical information", "personal medical data", "medical data", and so on. At present, various levels and types of relevant legal norms, policy documents, and national standards also have different definitions of the concept of "healthcare data". The essence behind the controversy over the legislative name and connotation is that the development time of healthcare data is relatively short, it is an emerging physical object, and there are objective limitations such as insufficient legislative experience, immature technical and theoretical research, and insufficient legal practice experience.

The "Action Program for Promoting the Development of Big Data", released by the State Council in August 2015, is the first authoritative and systematic document in China to promote data development. The big data mentioned in the program refers to the massive data set generated and available in the process of modern informatization in China, the sum of resources and data in the contemporary information society, and the full data in the information age, including both

Internet data, government data and industry data [2]. The "Information Security Technology Healthcare Data Security Guidelines" (hereinafter referred to as the "Security Guidelines") do not have national mandatory force. As a recommended national standard, the "Security Guidelines" has a certain guiding role. However, there is an error in the "Security Guidelines" regarding the cyclical definition of healthcare data, and both healthcare data and personal healthcare data are defined as electronic data, excluding other forms of data. In combination with the above provisions and the development trend of big data and information technology, this paper defines healthcare data as all potentially valuable health related data collected by people in the whole cycle (birth, aging, disease, death), full coverage (departments, institutions, regions), all elements (traditional clinical, Internet medical and emerging genomics, pharmacomics, biomolecules, immune and other precision medicine), and multi industries (meteorology, environmental protection and other big health related industries).

2.2 Features

In addition to the 5V characteristics of big data (massive, velocity, diverse, validity, and high-value), healthcare data also has some unique features.

2.2.1 Publicity

Healthcare data is an important fundamental strategic resource in China, involving patient data, medical record information, medical insurance information, genetic inheritance, medical experiments, clinical data, in vitro diagnostic products and third-party testing data, medical institution operation data, scientific research data, and other comprehensive information. Healthcare data is not only closely related to everyone's life, but its mining and utilization also affect the prevention and control of diseases and the development of new drugs worldwide. Therefore, big data carries significant public interests and has a dimension of "publicity".

Unlike traditional production factors that either lose value or disappear as products after being utilized, healthcare data not only eliminates wear and tear depreciation, but also generates new factor products. With the passage of time and scene transitions, these data also exhibit a persistent and reproducible characteristic. Therefore, healthcare big data is non-competitive and non-exclusive, and has the attributes of a public good.

2.2.2 High Professionalism

High professionalism is an important characteristic of healthcare data. Healthcare data has strong business relevance, high professionalism, multiple data variables, and is difficult to interpret, especially some clinical testing data, which contains a large number of disease-related attribute variables. Traditional data mining techniques lag behind the discovery of new knowledge hidden behind the data. High professionalism also creates barriers for deep application of big data, often requiring semantic processing. For example, when storing medical data, semantic tags should be added and documents should be split according to specific original data to avoid having to parse all documents without extracting a small amount of information. For another example, regarding the privacy protection of genetic data, even after de-identification, due to the strong personality of genes, even if a certain locus on the gene is removed, it can still be confirmed through other genes. It is necessary to cooperate with professionals in the fields of law, gene research, and data security technology to effectively protect genetic data and build a scientific protection mechanism.

2.2.3 Privacy

"Big data" refers to a collection of data with extensive application value that is formed through information processing, utilizing the circulation, dependence, dominance, and objectivity of data, presenting characteristics of scale, circulation, and diversity. Big data is a derivative of data, the result of information technology processing data pools, and it has an inclusive relationship with data. Big data includes both data and information. [3] Article 1032, Paragraph 34 of the Civil Code stipulates that privacy refers to the private space, activities, and information of a natural person's private life, which is peaceful and they do not want others to know. Most of the information in health and medical big data belongs to the category of privacy, especially personal attribute data, health status data, and medical application data, which require higher privacy protection requirements.

2.2.4 Temporality

Temporality is determined by human physiological and biological characteristics.

Diseases or health conditions are constantly changing and developing dynamically, and disease treatment often requires a process. Healthcare data is often accompanied by spatiotemporal data such as time, location, disease history, and environment. For example, waveform signals (ECG, EEG, etc.) and image signals (MRI, CT, etc.) from medical examinations belong to time functions and have temporality; At the same time, with the intensification of social aging, chronic diseases are gradually occupying a dominant position in the disease spectrum, and the proportion of time series data brought by chronic diseases in the entire regional medical and health big data is also increasing.

2.2.5 Diversity of Interests

In the process of collecting, storing, mining, predicting, and utilizing big data, there are different entities that have a significant impact on the generation of data value. These entities all have reasonable needs for data benefits and a legitimate basis for sharing benefits. Dutch scholar Zwitter distinguishes the stakeholders of big data into three types from the perspective of data generation and development, namely collectors, users, and producers [4]. The "Safety Guidelines" specify that the subjects of personal health and medical data include: controllers (organizations or individuals who determine the purpose, method, and scope of healthcare data processing); processors (representing relevant organizations or individuals who collect, transmit, store, use, process, or disclose healthcare data in their possession, or provide services related to the use, processing, or disclosure of health and medical data to controllers); users (for specific scenarios of data, they do not belong to the subject, nor do they belong to the relevant organizations or individuals who control and process it, but utilize healthcare data). The interests between these subjects mainly include personal interests, social interests, and public interests. Healthcare data inevitably involves personal privacy, including personal information, medical history, physical condition, etc. At the same time, the collection, storage, maintenance, and use of data also involve public interests and even national security.

3. SECURITY RISKS FACED BY HEALTHCARE DATA

3.1 The Urgent Necessity of Improving Safety Management and Standard System

The healthcare data security service system is not yet complete, and relevant standards are missing, mainly manifested in: (1) Chinese medical data is still isolated islands. There are standard barriers and policy barriers to the collection, exchange, sharing, and use of data across industries, institutions, and departments. There are no standards to rely on for healthcare data, and a truly comprehensive health and medical database has not yet been established. (2) Medical institutions lack guidance and reference in carrying out data classification and grading, lifecycle control, security monitoring and early warning and emergency response, data security operation management, data security supervision, etc., and have not formed an overall security protection system for healthcare data. (3) The protective measures for patient personal information still need to be improved. Personal information in the health care field is more private than that in other fields. Apart from basic personal information (name, ID number number, contact information, home address, etc.), it also involves treatment information, examination results, surgical information, etc. The disclosure of information will cause harm to patients' psychology, family, and social relations. Meanwhile, the spread of patient personal information will bring serious legal issues and social impacts to medical institutions. (4) The ownership of data is unclear, making it difficult to clarify data security responsibilities. Healthcare data is widely distributed in processes such as outpatient and emergency departments, hospitalization, physical examinations, insurance, and payment in hospitals, and is applied and transmitted through multiple channels such as local area networks, medical insurance systems, and government networks within medical institutions. The boundaries of rights and responsibilities are blurred, and the ways to protect rights are not clear.

3.2 Existing Security Risks in the Data Lifecycle

The circulation of healthcare data in various medical scenarios runs through the entire process of the data lifecycle, and each link faces potential

hazards. From the perspective of data exposure risks alone, according to IDC Market Research, by 2020, 42% of global electronic health data will be in an unprotected state. For example, data collection security risks caused by scattered information and diverse sources, data transmission security risks caused by frequent data flow, data storage security risks caused by massive data and complex scenarios, and data usage, openness, and exchange security risks caused by business development needs and user privacy. (See "Table 1").

Table 1. Potential risks of data lifecycle

Link	Potential risk
Data acquisition	Illegal (unauthorized) collection, sensor failure, data source distortion, malicious code injection, etc.
Data transmission	Data theft, data leakage, data tampering, illegal transmission, integrity damage, etc
Data storage	Unauthorized access, data destruction and tampering, unauthorized storage, storage medium damage, etc.
Data processing	Data not desensitized or declassified, unauthorized processing, data tampering, data forgery, data abuse, etc.
Data usage	Data abuse, unauthorized use, data leakage, abnormal retrieval, data theft, etc.
Data openness	Data crawling, illegal opening, data not desensitized, unauthorized access, etc.
Data exchange	Abuse of shared interfaces, unauthorized access, leakage of sensitive data, unauthorized exchange, etc.
Data destruction	Overdue retention, incomplete destruction, non-standard destruction process, missing destruction verification, etc.

3.3 Increased External Security Threats

The high value and privacy of healthcare data are the focus of attention for hackers or Advanced Persistent Threat (APT) organizations. After the outbreak of the COVID-19 pandemic in 2020, the healthcare industry surpassed government, finance, defense, energy, telecommunications, and other fields for the first time in history, becoming the primary target of global APT (network attacks and invasive behavior launched by hackers targeting customers for the purpose of stealing core information) activities. The "2021 HIMSS Healthcare Industry Cybersecurity Survey" released by the American Society for Healthcare Information and Management Systems (HIMSS) shows that nearly half of the network attacks suffered by the healthcare industry in 2021 were high-risk and ultra high risk events. 23.7% of global APT activity events are related to the healthcare industry. For the first time, China has surpassed countries and regions such as the United States, South Korea, and the Middle East, becoming the primary regional target of global APT activities [5].

3.3.1 Data Leakage Threat

Taking patient related data as an example, medical institutions store personal identity information, bank card information, medical data, diagnostic results, medication information, etc. These highly sensitive data contain great value. Once leaked, it can cause consequences such as medical identity fraud, phone fraud, property theft, etc., and seriously affect national security, social order, and public interests. The "2020 Digital Healthcare: Network Security Risk Research Report during Epidemic Prevention and Control" released by the China Academy of Information and Communications Technology shows that nearly 30% of the surveyed medical units have a risk of data asset leakage, and 7080 units use low version component services with public vulnerabilities, accounting for 44.39%. In April 2020, the source code of the experimental data of AI testing COVID-19 technology of Chinese medical company was stolen and sold by hackers.

3.3.2 Ransomware Attacks

The healthcare sector is a high-risk area for ransomware attacks. In July 2019, the Springhill Medical Center in Alabama, USA was attacked by a ransomware virus, which prevented related

medical equipment from monitoring the baby's condition properly, delayed patient medical treatment and treatment time, and resulted in the death of a newborn baby. In November 2021, German medical software giant Medatixx was subjected to ransomware attacks, affecting the internal IT systems of multiple medical institutions and forcing operational systems to collapse. In May 2021, Sophos released the "2021 Health Ransomware Status" report. According to a survey of 328 heads of information technology departments in medium-sized organizations in 30 countries/regions worldwide, approximately 34% of medical institutions were attacked by ransomware in 2020.

3.4 Existed Data Security Threats in Healthcare Equipment

Healthcare equipment that plays an extremely important role in disease prevention, diagnosis, and treatment, and safety issues cannot be ignored. According to the "2022 Medical Internet of Things Device Security Status Report", 53% of connected health medical devices have known vulnerabilities, and 33% of bedside medical devices have significant security risks. Intelligent infusion pumps are networked devices that deliver drugs and liquids to patients. After investigating data from over 200000 infusion pumps on hospital and other healthcare organization networks, researchers found that 75% of infusion pump devices have known security vulnerabilities, which greatly increase the risk of attackers invading. There are also shortcomings in the field of high-end medical equipment in China, and the industrial and supply chains are subject to certain constraints. According to public information, imported CT equipment, magnetic resonance imaging (MRI) diagnostic equipment, surgical robots, extracorporeal membrane oxygenators (ECMOs), and other medical devices occupy the main market of tertiary hospitals in China. Although some medical devices have achieved localization, their core components, raw materials, manufacturing equipment, and testing equipment still rely on imports, making it difficult to achieve localization. These devices store a large amount of detection data. If there are vulnerabilities and preset backdoors, they can be exploited by attackers, causing data leakage and directly endangering the life and health of patients.

3.5 Data Security Threats in the Healthcare Supply Chain

The healthcare supply chain involves various aspects such as enterprises, personnel, technology, management, products, services, etc., with a wide scope, multiple links, long duration, complex situations, and relatively hidden security threats. The reasons for this, the first is that medical institutions lack management in software supply chain security and rarely carry out software supply chain security reviews, source code testing, and independent controllability evaluations. It is difficult to predict open source software threat events caused by vulnerabilities such as Apache Log4j2. The second is that medical institutions outsource some modules or functions when carrying out network security and information construction, and third-party manufacturers are responsible for module research and development, testing, or operation support services. If relevant data is transmitted overseas without reporting, approval and review by relevant institutions, it will bring huge risks to China's public safety, biosafety, and national security. At the same time, attackers may also use third-party services as attack points, detour into the network environment or platform systems of medical institutions, and steal or destroy data. Finally, outsourcing personnel find it difficult to effectively control and data theft incidents occur frequently.

4. HEALTHCARE DATA SECURITY GOVERNANCE PATHS

4.1 Promoting the Construction of Data Security Service System

The completeness and quality of the healthcare data security service system directly determine the level of data security. There is a must to stand at the height of national security and use relevant laws, regulations, and policies as the starting point to promote the construction of a health and medical data security service system and achieve a full life cycle control mechanism. Specifically, it includes: the first is to establish a health and medical big data platform. The platform integrates electronic health records, electronic medical records, medical imaging and examination results, daily health sign data, medical institution information, etc. [6] Medical institutions do not need to open data interfaces, and can collect and manage medical data "in real-time and in full from" all medical institutions in the region through government data

platforms. The data format should adopt unified standards to further support data applications. The second is to develop encryption rules and hierarchical authorization rules, control the legitimate storage and circulation of data through the health and medical big data platform, and ensure that data is "usable but not visible, visible but not desirable". The third is to build a comprehensive cross departmental and cross hierarchical security organizational structure, clarify the responsibilities and obligations related to data security of each department, and divide data ownership. The forth is to focus on promoting the implementation rules of key systems such as healthcare data classification and grading, data export management, and data security assessment, so that the norms can be effectively implemented and more practical.

4.2 Classification and Grading to Strengthen Data Control

As responsible units, medical and health institutions of all levels and types, as well as related enterprises and institutions, are not only required to comply with legal provisions on healthcare data, but also to fulfill corresponding obligations in accordance with laws and regulations on personal information protection, privacy protection, and network security management. Different healthcare data have different sensitivities, therefore there are different requirements for openness. The responsible unit needs to establish a data classification system to standardize the scope, degree, and method of openness of healthcare data.

According to the "Security Guidelines", data in the open scenario of healthcare data can be classified and managed according to different elements such as the purpose of opening, data attributes, importance and risk level of the data, recipient type, etc. For example, for healthcare data involving trade secrets and personal privacy, it is classified as non-public data; Healthcare data with high requirements for data security and processing capabilities, strong timeliness, or the need for continuous access shall be classified as conditionally open data; For those listed as non-open data or conditionally open data, relevant laws, regulations, and other basis should be listed in the corresponding list.

4.3 Improving the Data Security Technology Protection System

For medical institutions, large-scale data leaks are mostly caused by invading information systems, and reliable technical means are the key to ensuring the integrity, confidentiality, security, and controllability of healthcare data. The specific measures are as follows: the first is to collect data in accordance with the principle of minimum necessity and obtain authorization from patients to strictly control the source of data collection. The second is that at critical data storage nodes, gateways and firewalls should be established, and security cloud technology should be fully utilized to conduct real-time evaluation of visitor behavior. Once dangerous operations or data are discovered, immediate intervention and control should be implemented. If necessary, two firewalls can be set up, internal and external. The former is responsible for monitoring and filtering data on the internal network, as well as effectively identifying user identities, while the latter is used to prevent external intrusion behavior and effectively prevent the theft of healthcare data during transmission and use. The third is to ensure the security of healthcare data transmission by using technologies such as data encryption, personal information deidentification, and transmission verification. The forth is to ensure that data usage does not exceed permissions and reduce the risk of information leakage by fully informed authorization, utilizing blockchain technology to record the usage process of data in detail, and utilizing privacy computing and other technologies. The fifth is to provide anti-crawling technology and digital watermarking technology to prohibit the crawling of networked data during open sharing of healthcare data, and add watermarks on the network to avoid data leakage caused by screenshots or photos.

4.4 Strengthening the Supervision of Healthcare Data Security

The competent authorities in the field of healthcare should conduct comprehensive security supervision of big data ownership, personal privacy protection, data confidentiality, integrity, and availability in accordance with national network security level protection policies and industry security protection policies and technical standards, to ensure that big data is "manageable, controllable, and trustworthy". [7] The specific measures are as follows: The first is to strictly abide by relevant laws, regulations and standards on data security,

comprehensively use various means such as supervision and inspection, pilot supervision, evaluation and testing, guide medical institutions to timely discover loopholes and identify hidden dangers, and avoid serious consequences. The second is to build industry data security situational awareness capabilities, strengthen monitoring and management of data security, and achieve the sharing, reporting, and emergency response of threat information. The third is to conduct important data security risk assessments to enhance the ability of medical institutions to prevent security risks throughout the full life cycle of important data. The forth is to conduct data security supply chain security inspections to identify risks and hidden dangers from multiple dimensions including enterprise, product, service, and personnel.

5. CONCLUSION

Healthcare data, as an important fundamental strategic resource of the country, is related to national strategic security, national biosafety, people's life safety, and personal privacy security of citizens. It is necessary to have a "bottom line" thinking, establish and improve the regulatory mechanism for healthcare data, strengthen the legal and efficient use of health and medical data, continuously improve the level of healthcare big data services for the benefit of the people, in order to meet the multi-level and diversified health needs of the people, and serve the strategic goal of "Healthy China"

ACKNOWLEDGMENTS

Fund Project: General Project of Sichuan Health Rule of Law Research Center, Key Research Base of Philosophy and Social Sciences in Sichuan Province, "Research on Privacy Protection in the Context of Medical Artificial Intelligence" (YF21-Y41).

REFERENCES

[1] Ye Qing, Liu Xun, Zhou Xiaomei, etc., Research on the Problems and Countermeasures for the Development of Health Care Big Data [J]. Chinese Hospital Management, 2022, 42(01): 83-85.

[2] Shan Zhiguang, Interpretation of the Action Plan for Promoting the Development of Big Data, published on the National Information Center website http://www.sic.gov.cn/News/609/9713.htm (Visited December 18, 2022)

[3] Li Aijun, Data: Its Rights Attribute and Legal Nature [J]. Oriental Law, 2018, No.63(03): 64-74.

[4] ZWITTER A. Big Data Ethics [J]. Big Data & Society, 2014. 1(2): 1-6.

[5] Ji Peng, Xu Kaiyi, The Generation of Political Security Risk in Cyberspace and Its Resilience Governance [J]. Journal of Socialist Theory Guide, 2022, No.448(03): 78-83.

[6] Qin Panpan, Xie Liqin, Chen Quan, etc., Research on the Implementation Path of Hierarchical Diagnosis and Treatment Based on Health Care Big Data [J]. Chinese Hospital Management, 2021, 41(06): 75-78.

[7] Qin Xiaodong, Architecture Design and Operation Practice of Government Big Data Platform [J]. Digital Technology & Application, 2021, 39(08): 130-134+137. DOI:10.19695/j.cnki.cn12-1369.2021.08.43.