Society of Medical Education & Research

2024, Volume 4, Page No: 337-347
ISSN: 3108-5059

Asian Journal of Ethics in Health and Medicine

Ethics, Law, and Practice in Medical Data Sharing: Empirical Evidence from Chinese Researchers

Xiaojie Li[1], Yali Cong[2]*

[1]Department of Situation and Policy, University of International Business and Economics, Beijing, China.
[2]Department of Medical Ethics and Law, Peking University Health Science Center, Beijing, China.

***E-mail** ✉ ethics@bjmu.edu.cn

Abstract

Policies encouraging the sharing of medical data have gained considerable momentum in China. However, the existing legal and ethical framework for using medical data in research remains predominantly restrictive rather than supportive. At present, the share of Chinese medical data utilized for scientific research still leaves substantial room for growth, revealing considerable untapped opportunities to advance medical knowledge and enhance healthcare results. Drawing on this foundation, the current study examines the obstacles researchers face in medical data sharing through focus group interviews. Two focus group interviews were conducted with researchers from a range of academic fields on 21 June 2021 and 28 July 2021. Altogether, 17 researchers from diverse professional backgrounds volunteered. Overlapping codes were combined, and group discussions within the research team helped identify representative or typical statements from the participants.

Participants clearly understood that medical data must not be shared without proper justification and emphasized the need to comply with relevant laws. The interviews revealed that although the researchers highlighted the importance of thoughtful evaluation before releasing such information, none of them mentioned the requirement of obtaining consent from data subjects specifically for research purposes. This finding stands in notable contrast to the strict rules on separate consent for secondary use of data stipulated in the PIPL.The focus group interviews reveal the obstacles and ethical challenges researchers face when sharing medical data for scientific purposes. They underscore the participants' strong emphasis on data security and their generally cautious stance toward data sharing. The primary goals for promoting the reuse of medical data include improving interoperability, standardizing data formats, enhancing data quality, protecting privacy, securing informed consent, motivating patient participation, and developing clear rules governing data access and use.

Keywords: Medical data, Data sensitivity, Data sharing, Benefits sharing, Data protection law, Data quality

Introduction

Medical data, gathered from various origins such as research projects, patients' clinical records, and publicly released government information, has long served important secondary roles that deliver broad societal advantages. These benefits range from population health surveillance and improvements in healthcare quality to major progress in biomedical research [1, 2]. In today's highly connected world, medical data flows across borders and beyond individual research groups, incorporating information derived from both clinical and population-based studies. For instance, linking datasets through interconnected medical data enables researchers to combine the resources and expertise needed to explore the complex molecular mechanisms underlying disease [3].

Policies promoting medical data sharing in China have garnered strong support, fueled by the national push to harness big data to drive technological breakthroughs, economic development, and broader societal gains [4]. According to the Opinions on Further Improving the Medical and Health Service System released in March

Access this article online https://smerpub.com/

Received: 28 May 2024; **Accepted:** 12 September 2024

How to cite this article: Li X, Cong Y. Ethics, Law, and Practice in Medical Data Sharing: Empirical Evidence from Chinese Researchers. Asian J Ethics Health Med. 2024;4:337-47. https://doi.org/10.51847/mYs0LBI8EN

2023, the authorities have directly linked greater medical data sharing to the broader goal of achieving fairer, more accessible, and higher-quality healthcare services [5]. Nevertheless, China's current legal and ethical system for employing medical data originally collected for non-research aims in scientific studies adopts a more restrictive stance than a permissive one [3]. The introduction of the data security law (DSL) and the personal information protection law (PIPL) reflects the government's determination to strengthen safeguards for personal information. In a prior publication, we noted that certain clauses in these laws do not adequately tackle the unique issues arising in scientific research within the big data context [6].

The amount of Chinese medical data currently used in scientific research still offers significant potential for expansion. This situation highlights a significant opportunity to advance medical knowledge and deliver better healthcare outcomes. Insights from the 2021–2022 China Hospital Informatization Survey, performed by the Information Professional Committee of the Chinese Hospital Association, shed useful light on how medical data is currently being reused. In particular, the use of medical data for research purposes continues to lag, as only 14.97% of hospitals actively concentrate on this area [7]. Furthermore, work by Shi Jingjin and her colleagues points to a clear deficiency in specialized alliance data-sharing platforms. Even though the functions supporting research collaboration are used by just 11.76% of member institutions, interest in these tools remains high. Notably, 58.82% of member institutions have shown a strong willingness to develop research-related features [8].

Very few investigations have examined why medical data is used so sparingly in scientific research [9, 10]. The only study we located on this topic was conducted by Zhang *et al*. Their results indicate that, despite high interest — with 93.53% of those surveyed willing to participate in data sharing — the majority of researchers interviewed had experienced failed attempts at data sharing [10]. The researchers concluded that several important obstacles were at play: difficulty obtaining official administrative backing, very limited rights to use databases (with most roles restricted to data collection), and rising concerns among researchers about the privacy of data subjects after information is shared. These difficulties are exacerbated by the lack of clear legal and ethical guidelines for conducting research with medical data [10]. The same 2021-2022 China Hospital Informatization Survey also highlighted a serious problem for tertiary hospitals in medical data sharing: the shortage of data scientists. Of the 684 hospitals included in the survey, 514—75.15%—named this shortage as one of their main hurdles. In addition, more than 60% of hospitals reported having fewer than 10 information technology personnel on staff [8].

Based on the results outlined above, this study examines, in greater detail through focus group interviews, the obstacles and ethical difficulties surrounding the sharing of medical data for scientific research. The inquiry aims to tackle several key questions:

How have the PIPL and the DSL changed the way medical data sharing is practiced?

Why is medical data collected by government bodies not easily accessible to researchers?

What are the main reasons that prevent researchers from sharing the data they gather with others in the field?

Do any particular values exist that actively encourage medical data sharing?

Materials and Methods

Focus groups

Because available information on the topic is scarce, this exploratory study relied on focus group interviews followed by qualitative analysis of the findings. The goal was to generate early hypotheses to address the research questions [9]. Two focus group interviews were held with researchers from varied disciplinary backgrounds on 21 June 2021 and 28 July 2021. The group included both physicians (P) and data scientists (S). Physicians came from a range of specialties, including internal medicine, surgery, dentistry, and psychiatry. Data scientists were drawn from medical universities and research institutes, where they focused on data analysis and processing. These data scientists did not collect data directly; instead, they received it from physicians or other researchers and then carried out analysis and processing (Additional file 1). The sessions were moderated by the corresponding author (YLC), with the first author (XJL) in attendance to observe. Participants were invited to speak candidly, assured that all contributions would remain strictly confidential and anonymous [10]. Care was taken to give every person a chance to contribute, creating an open and inclusive atmosphere for discussion [11, 12].

This study follows the detailed reporting standards described in the Consolidated Criteria for Reporting Qualitative Studies (COREQ) [13]. By adopting this

systematic method, the aim was to gather rich, detailed perspectives on the ethical issues and attitudes of the participating researchers.

Respondents

Interview participants were recruited by investigators from the Institute of Science and Technology Policy and Management Science at the Chinese Academy of Sciences. A formal agreement was made with the institute, which maintains access to China's respected expert database system that connects leading scientific specialists. This arrangement allowed the research team to reach well-qualified experts in medical data sharing for the interviews.

Each focus group session was conducted in Chinese and lasted roughly 2 hours. To maintain consistency and thoroughness, the full research team jointly developed a semi-structured interview guide. The guide featured open-ended questions, with further details supplied in Additional file 2.

Before final use, the interview guide was tested through a pilot session with two colleagues to check whether the questions successfully prompted thoughtful discussion [14]. Drawing on their suggestions, minor revisions were made — including changing some questions into follow-up probes and rearranging the order to improve overall flow and clarity.

Analysis

The audio recordings of the interviews were converted into full written transcripts by a professional transcription service. To confirm accuracy, the researchers listened to the entire recordings while reviewing the transcripts. The authors went through the transcripts multiple times, conducting a detailed line-by-line examination. Individual participant statements were then abstracted into broader concepts or themes to capture all aspects of participants' expectations. An inductive method was followed throughout. After the first round of review, an initial set of codes, developed directly from the data, was created and discussed between the first and corresponding authors. Further codes emerging from the data were incorporated as the analysis progressed, remaining faithful to the participants' actual responses [15]. Codes that overlapped were combined. Team discussions among the research group were also used to pick out participant statements considered typical or especially representative. These steps were repeated until the team reached full agreement on how the findings would be presented. The results were translated into English by the authors themselves. Two authors independently reviewed the translation to confirm that the meaning stayed accurate and true to the original conversations. Because the interviews had been conducted in Chinese, careful refinements were applied during translation to improve clarity and readability without compromising precision or fidelity to the source material.

Results and Discussion

The implications of PIPL and the DSL

When discussing the effects of the PIPL and the DSL, participants agreed that strict compliance with these laws was vital to safeguard privacy and preserve integrity, thereby upholding strong legal and ethical standards throughout the research process. They clearly understood that medical data must not be shared without good reason and stressed the need to follow legal requirements when sharing.

I believe data security should come before sharing. Before releasing any data, I would ensure it meets all legal standards, clarify how the data will be used, identify who is responsible, and confirm any relevant legal duties. Data sharing should never happen without careful thought (S1, male).

Several participants felt that privacy and security issues could be adequately managed, provided that data use and sharing stayed within legal boundaries. For instance, they suggested that if hospitals properly removed identifying information from patient records before sharing — as required under the PIPL — the risk of medical data leakage would be very small. At the same time, they emphasized that achieving this goal would still require robust legal protections and oversight by professional ethics committees.

A few respondents noted that removing identifiers from medical data in accordance with legal rules could reduce its usefulness for scientific research. They worried that some researchers, in their rush to share data, might ignore legal limits.

The perceived low risk of punishment for breaking the law makes it easy for people to use medical data casually without worrying about consequences or penalties such as fines. As a result, medical data is sometimes used loosely for writing papers, apparently with little regard for patient interests. Moreover, it is widely thought that

patients are unlikely to sue because the costs of taking legal action are extremely high (P3, female).

None of the participants mentioned the need to obtain fresh consent from data subjects when using medical data for secondary research. This finding contrasts sharply with the strict requirement for separate consent for secondary use of data set out in the PIPL. Interestingly, one participant even raised worries that the informed consent process itself could create future difficulties for researchers:

Some patients tend to be overly demanding. They might ask what researchers have done with my data and what results have been achieved. Can I receive any feedback? The more patients learn about researchers' activities through the consent process, the more attention they will pay to these matters. Stronger protection of patient rights could create many disadvantages for researchers (P5, female).

Some participants raised issues about how difficult it is to actually enforce the legal rules on informed consent:

Even though patients had signed the informed consent form and confidentiality agreement, I know that these documents have almost no real power over researchers. They can essentially do whatever they like (S3, male).

Some participants also expressed concerns about the lack of enforceability of legal provisions on informed consent:

Although the academic community tends to assume that medical data belongs to the patient, there is no formal regulation that actually enforces this idea. Implementing such a rule could pose problems: without clear patient consent or a strong public-interest justification, data sharing would be severely restricted, leading to serious difficulties in how data can be used and exchanged (P2, female).

Concerns about data sensitivity hamper government data sharing for scientific research

Several participants expressed concerns about the tight restrictions on releasing medical data stored in government databases. These limitations prevent researchers from accessing important resources and from realizing the full potential of the data. Participants felt that the non-public status of government-held data seriously restricts their ability to use reliable domestic databases. According to the interviewees, the decision not to share government data would lead to three main negative outcomes. First, it greatly reduces researchers' ability to drive medical progress and innovation. As one participant noted:

The idea of 'sensitivity' often acts as a self-imposed barrier that slows down progress and innovation. In practice, we are fully capable of setting our own industry standards, yet we repeatedly fall behind the United States and Europe. This situation stems from outdated thinking and weak systems, which ultimately result in the wasteful underuse of valuable resources (S1, male).

Second, it increases the likelihood that researchers will resort to unethical means to obtain or trade hospital data. If data were made more accessible, scientists would have less reason to buy data through questionable channels. Without reliable domestic databases, purchasing data becomes the only practical option (S4, female).

Third, it weakens the government's ability to perform its public duties effectively, as one respondent explained:

At the government level, combining medical data requires a deep appreciation of the state's core responsibility to support overall societal development. Therefore, using medical data for scientific research should be seen as a meaningful and worthwhile contribution that deserves proper recognition in official government reviews and evaluations (P4, male).

Although some government departments have shown readiness to share medical data with researchers, participants stressed that worries over data sensitivity continue to prevent these agencies from releasing the information easily.

To develop a clearer picture of what data sensitivity means, participants were asked to share their views. One respondent explained:

In today's environment, especially with big data, medical information raises serious sensitivity issues, particularly when it involves genetic data. Releasing or disclosing genetic information poses significant security challenges and requires careful consideration of the timing and manner of sharing (P1, male).

Although medical data is not necessarily the same as genetic information, some participants believed that certain types of medical data could, unintentionally, reveal genetic characteristics specific to the Chinese population. "This happens because, when different datasets are linked together, even ordinary scientific research data may create potential risks to national security" (S1, male). This awareness has led participants to take a very cautious approach toward sharing data across borders.

Several respondents expressed concern about China's relatively low involvement in international sharing of health and medical data. They observed that European

and American countries largely dominate the creation of data-sharing standards, leaving insufficient effort to develop standards suited to China's own context. They argued that overly strict confidentiality rules, although designed to protect privacy and security, also restrict development and slow down innovation.

When discussing how to handle sensitivity issues related to government medical data, some experts argued that it is essential to assemble a specialized team of professionals skilled at distinguishing between highly sensitive data and information that carries little risk if shared. One participant stated:

I think many researchers are either unaware of this problem or/ do not care about it. As a result, they do not know how to properly de-sensitize sensitive information. In some cases, they accidentally share details such as patients' names, case numbers, and ID card numbers because they lack the technical ability to mask or remove identifying details. (P8, male)

However, the participants recognized that building an effective desensitization system would demand continuous long-term effort and could not be accomplished quickly.

Researchers hope to see a wide range of medical data made easily available for in-depth research and analysis. Yet the sensitivity of a large portion of this data creates significant hesitation about releasing it. In practice, some of the data may not actually be as sensitive as it is perceived to be, but the government has not given this issue the serious attention it requires. As a result, a clear and reliable system for responsibly releasing data has still not been established. This gap is especially evident in the urgent need to make disease control data publicly accessible (S5, female).

The sharing of medical data collected by researchers must guarantee a fair distribution of benefits

When participants were asked what worried them most about sharing medical data that they had gathered themselves, the clear majority highlighted the need for proper benefit-sharing arrangements. These arrangements might involve monetary rewards, co-authorship on publications, or formal recognition of their contributions to institutions or projects. "This data is the property of our hospital — we cannot just hand it over freely," one participant remarked. "It is unfair when one side puts in serious effort to collect the data while the other side gains all the advantages, publishes papers, or receives credit without offering any form of return or compensation" (P10, male). In the absence of a robust and enduring benefit-sharing system, establishing sustained, meaningful data-sharing practices is extremely challenging.

Consider blockchain, for instance. Although it receives considerable attention, it does not genuinely resolve the difficulties encountered in scientific research. Its main advantage is enabling basic data access, such as allowing others to view medical records. However, when large-scale studies demand enormous quantities of data, blockchain proves inadequate. This limitation is worsened because the parties involved often lack real motivation to work together — there is little incentive to contribute resources without clear rewards. Even though blockchain could, in theory, support data exchange between hospitals, the financial costs and ongoing maintenance usually exceed the expected gains, leaving little reason for participants to get involved. For such technologies to succeed in actual healthcare environments — where even voluntary projects must remain sustainable — the overall cost-benefit balance must be carefully resolved. This is exactly where many existing technological solutions fall short (S4).

Several participants argued that the foundation of any successful medical data-sharing system is a well-designed benefit-sharing framework. From their own experience, people are generally unwilling to release their data unless there is a convincing incentive. "Without strong bottom-up motivation, it is very hard to create truly mutual sharing arrangements, which in turn slows down larger-scale data-sharing initiatives" (P6, female). When such mechanisms are missing, data holders see almost no downside to keeping their data private, while facing various risks and disadvantages if they decide to share.

Formal institutional data sharing contracts led by trusted experts

A large number of participants recommended moving away from researcher-to-researcher data sharing toward institution-level arrangements. They pointed to compliance risks arising when individuals share data on their own. Instead, they believed medical data sharing should primarily occur between organizations rather than rely on individual researchers. These participants suggested that formal steps, such as signing detailed cooperation agreements, could help prevent misunderstandings and conflicts during joint projects.

One participant explained that institutions specializing in data analysis should enter into formal contracts with hospitals. This setup would enable the exchange of technical know-how ("technology") and actual datasets ("data"), with the relationship secured through formal data-sharing contracts.

We openly share all our data with partners under strict contractual terms. These agreements ensure full transparency, with each side's duties and rights clearly defined and mutually agreed upon (P7, male).

Participants also stressed the value of having a respected and trustworthy "big expert" to lead the alliance.

This person takes on a central leadership role, coordinating joint activities and helping to create fair rules for sharing benefits. These rules are designed so that every member of the alliance accepts and commits to them. As a recognized authority, the expert provides valuable direction and helps cultivate a strong atmosphere of trust within the group (P1, male).

For data focused on particular diseases, it was considered more suitable to appoint chief physicians or prominent professors as leaders. These individuals bring deep, specialized knowledge to the alliance, strengthening expertise and building greater confidence among all members (P11, male).

By adopting institution-level partnerships and establishing strong leadership structures within alliances, participants believed that compliance risks could be substantially reduced, expertise could be used more effectively, and trust could be strengthened. These approaches were viewed as creating a solid base for smooth data sharing and supporting productive, high-impact biomedical research collaborations.

Confronting the barrier of poor data quality

Participants noted that even when data is shared, its practical usefulness cannot be taken for granted, and that developing a single set of data standards for the whole country is unrealistic. They emphasized that simply gathering large amounts of data is not enough; the true value lies in effectively accessing and using the data.

I have been involved in multi-center epidemiological studies with many participating hospitals. In one project, only a single hospital's data consistently reached the required quality level. This showed that usable data remains scarce even in great collaborative efforts and highlighted how difficult it is to achieve standardized nationwide data sharing (S1, male).

When asked about the main obstacles that prevent shared data from producing the greatest overall benefits and positive results, the participants gave the following reasons:

1. Elevated sharing costs and fragmented storage infrastructures: The existence of many different, incompatible storage systems creates major difficulties for smooth data exchange and significantly increases overall costs.
2. Scarcity of data scientists and limited data literacy: The scientific usefulness of shared data is greatly reduced due to a shortage of trained data scientists and the widespread lack of data analysis skills among researchers. This gap means researchers often need substantial guidance to understand basic concepts such as datasets, scientific data management, and advanced analytical methods.
3. Given the presence of data with suboptimal quality, there exists a tangible risk that its dissemination may inadvertently reveal sensitive information, including instances of misdiagnosis or even medical errors. This revelation could subsequently trigger heightened regulatory scrutiny from superior authorities, posing a significant challenge for institutions. Consequently, the fear of stricter monitoring, enhanced accountability, and potential legal repercussions has instilled hesitancy among these entities. As a result, many are reluctant to share their data to mitigate these risks.

Recognizing the value of rewarding patients while acknowledging the difficulties involved

All participants unanimously agreed that researchers have a duty to protect the well-being of data subjects, upholding the important principle that "What is taken from the people should be used for their benefit". They emphasized that achieving a fair balance of benefits is essential. On the one hand, sharing data exposes patients to privacy risks; on the other, it enables decisions grounded in big data to be more scientifically sound and rational, ultimately delivering real advantages to those whose data is used.

Patients now experience greater openness because their health records may be available to the wider public, insurance providers, and hospitals. Even with this increased visibility, hospitals are well-positioned to deliver better care. For example, by examining detailed records of past lung cancer cases, hospitals could refine screening methods used in routine health check-ups. Such improvements would particularly help key groups in the population, including petrochemical workers who

frequently face risk factors linked to lung cancer [P12, male].

Nevertheless, one respondent highlighted significant barriers to the smooth exchange of medical data between organizations. She specifically pointed out the complex difficulties that exist even within a single hospital when sharing data, due to the layered hospital system ranging from first- to third-level institutions.

People visiting tertiary hospitals for treatment often face the annoying situation of having to repeat medical examinations they already completed at secondary hospitals. Enabling efficient data exchange across different levels of healthcare facilities could greatly simplify this process, reducing both the time and money patients spend [P6, female].

The participants strongly recommended creating a dedicated data-sharing platform to build greater patient confidence and promote wider sharing efforts. They felt that such a platform would create a feeling of security. Researchers would be able to safely upload and use data, confident that it would be available to a larger group of fellow researchers. This broader access would improve teamwork and encourage the use of reliable, high-quality data, thereby leading to stronger research results.

Establishing this platform is expected to bring several important advantages. First, patients would gain clearer information about exactly which researchers can access their data and what specific goals those researchers aim to achieve. This increased openness would allow patients to make better-informed decisions about data sharing and to participate more actively in shaping research activities. In addition, the platform would make it easier to share research outcomes, so patients could see for themselves the public benefits that result from contributing their data [P7, male].

Understanding researchers' views on the obstacles to sharing medical data for scientific purposes is essential in today's rapidly changing healthcare environment [16]. The findings from this study offer valuable guidance for researchers in developing countries by providing a comprehensive framework for creating conditions that support the ethical sharing of medical data for research. These results align well with the ethical standards described in multiple international guidelines and academic publications [17].

Customizing data protection laws for scientific research and data subject rights

Although establishing laws and rules to control data sharing is clearly a necessary measure, it is not enough on its own. It is equally critical that researchers are not only aware of these legal requirements but also actively involved in implementing them. During the interviews, we observed a clear difference in how respondents viewed certain issues. While everyone stressed the need for careful thought when deciding how to disclose information responsibly, the question of obtaining specific consent from data subjects for any secondary use of their data was not discussed. This stands in sharp contrast to the strict rules in the PIPL, which require separate consent for secondary use.

Many respondents raised doubts about the practicality of enforcing legal rules on informed consent, pointing to a possible divide between the law as written and how it operates in practice. Issues such as the risk of data misuse, the dangers of complete openness, and the possibility of data breaches — especially when data is kept for long periods — have been shown to make patients hesitant and unwilling to agree to their data being reused for research [18-21]. The doctors we spoke with, who both conduct research and gather patient data in clinical settings, have the dual duty of safeguarding patient privacy while advancing biomedical science. Under existing regulations, patient-informed consent is mandatory for the use of samples and medical data in scientific studies. However, difficulties persist. For instance, numerous samples kept in hospitals were gathered many years ago, long before detailed informed consent procedures were in place. Although these older samples and data could prove highly useful for research, they cannot be used under the current PIPL rules.

Despite this, studies repeatedly show that most Chinese citizens support the secondary use of medical data, provided it serves the greater public interest [22-24]. In light of this, we suggest that a more adaptable consent approach — one that does not follow a rigid "one-size-fits-all" model — would likely be well received by the Chinese public, who generally wish to contribute to societal benefits. At the same time, researchers must carefully comply with all legal requirements to ensure that data is managed ethically and without misuse.

A further major factor that hinders researchers from sharing medical data is the lack of clarity in legal rules regarding data ownership. China's State Council recently issued the "Opinions on Building a Data Infrastructure System to Better Leverage Data Elements" [25], a document that deliberately avoids addressing the

question of personal data ownership. When this document was prepared, it followed the guiding idea of "downplaying ownership, emphasizing usage rights, and focusing on the circulation of data usage rights." [26]. It stopped short of clearly assigning ownership of data to the original data provider. Nevertheless, data ownership plays two important roles: it protects personal data from misuse. It allows data subjects to entrust the handling of their data to processors whose goals align with their own [27].

Clarity in sensitivity: the need for government agency guidelines

The ongoing shift toward enabling third parties to reuse open government data for public good, both in China and around the world, is accompanied by significant changes in policy direction. This development calls for a careful review of systems designed to maintain appropriate data sensitivity standards and to make well-informed choices about the extent of data openness.

The application of medical data in research is closely linked to local regulations and procedures, creating a highly inconsistent environment for researchers. Although Chinese government bodies such as the Health Commission and the Natural Science Foundation of China possess extensive and varied collections of medical data, these resources remain underused, largely because of sensitivity issues that prevent their sharing. A clear shortfall persists in the absence of standardized guidelines or recommended practices for releasing open government data, leaving researchers with insufficient guidance. Overcoming these obstacles is essential to unlock the complete value of open government data for societal benefit.

Ensuring motivated and equitable benefit distribution in medical data sharing

Creating a motivated, fair system for distributing benefits from medical data sharing is a strong incentive for researchers to participate in data-sharing efforts. Participants indicated that tackling this matter requires building dedicated platforms, promoting institution-level partnerships, and establishing clear leadership roles within collaborative alliances. Our results are consistent with the survey findings of Federer *et al*. [28], which indicate that structured approaches to data sharing can increase access to and reuse of research data.

All respondents firmly stated that researchers bear a responsibility to guarantee that data subjects receive concrete benefits from their involvement. They drew attention to the existing difficulties in exchanging medical data between institutions within the hospital system, which often results in extra financial pressure on patients throughout their care. In response to these issues, the participants strongly supported the creation of a single, integrated data-sharing platform. Such a platform would simplify the treatment journey for patients, making it smoother and less burdensome, while also advancing scientific research and ultimately delivering rewards to the very people whose data is shared, thereby acknowledging and appreciating their input.

Medical data sharing demands adequate technical support

The fact that research settings in many developing regions differ markedly from those in wealthier countries — especially in terms of available resources, research support, and infrastructure — creates substantial challenges for effective data integration. Although the majority of researchers are willing to share data, participants noted that they often lack skilled professionals in data analysis and application. This shortage leads to underuse of data and compromises data quality [29-33]. These issues arise from fundamental differences in treatments, patient outcomes, study designs, analytical techniques, and the varied methods used to collect, process, and analyze data in medicine [34].

Acknowledging these variations highlights the demand for more tailored approaches. Solutions should involve developing a single, affordable storage and sharing platform that connects different systems, thereby lowering costs and removing obstacles to smooth data exchange. In addition, investing in the training and development of data scientists and in improving data literacy among researchers is vital to fully realizing the scientific potential of shared data. Moreover, applying strict quality assurance procedures and cultivating an environment of openness and responsibility can reduce concerns at the institutional level, ensure that shared data meets the highest standards, and reduce the risk of accidental disclosures or legal problems [35, 36].

Limitations

In the focus group interviews, the phrase "medical data" was interpreted differently by participants with different research backgrounds. Where extra clarification was needed, explanations were added in a footnote beneath

the quoted interview excerpts. One limitation of this study is that it focused solely on researchers and did not capture patients' views. Although this focus was deliberate to explore the specific difficulties researchers face when sharing medical data for scientific purposes, it necessarily limits insight into the patient perspective. To help address this gap, we are now conducting a follow-up study examining patients' attitudes and opinions toward medical data sharing for research in the Chinese setting. While focus group interviews can generate useful hypotheses about a particular group, it is important to test these hypotheses using quantitative methods, such as surveys, to confirm their accuracy. The value of including focus group interviews in our research lies in their capacity to deliver a rich, detailed understanding of attitudes and beliefs through the dynamic exchange among participants, which encourages deeper exploration of complex or controversial topics [37, 38].

Conclusion

The results of the focus group studies offer valuable guidance on how to motivate researchers in developing countries to share medical data for scientific research. Regulatory systems must carefully balance the competing demands of privacy protection, open science, and national security. Promoting data sharing requires mutual efforts, appropriate incentives, sufficient technical support, clear definitions of data sensitivity, and effective systems for distributing benefits. By tackling these challenges, ethical data-sharing practices can be encouraged, ultimately delivering advantages to patients, scientific progress, and society at large. Discussions should move past the simple question of whether data should be shared and instead focus on the best methods for sharing that minimize potential risks and respect patients' legitimate expectations. To encourage greater data sharing, researchers need reciprocal support and incentives from their colleagues and society as a whole, underscoring the importance of recognizing and rewarding those who contribute their data.

Acknowledgments: We express our profound gratitude to the esteemed members of our extended research team, namely Zhang Haihong, Zhang Luxia, Fengyao, Zhao Liyan, and Liu Ruishuang, for their invaluable contributions and unwavering support in meticulously crafting the semi-structured interview guide. Our heartfelt thanks also go out to the diligent investigators from the Institute of Science and Technology Policy and Management Science, affiliated with the prestigious Chinese Academy of Sciences, for their diligent efforts in recruiting the interview respondents. Furthermore, we are deeply indebted to the Brocher Foundation for granting Xiaojie Li a unique and focused opportunity to pen this paper.

Conflict of Interest: None

Financial Support: This study was funded by the National Institute of Health Data Science at Peking University, the National Social Science Fund of China(20BZX128,23BZ095), and Beijing Social Science Fund(306/Z2301031).

Ethics Statement: This investigation was reviewed and approved by the Health Science Center of Peking University IRB (No IRB00001052-19013). To safeguard respondents' privacy, we proactively requested that they mask any identifying characteristics that could lead to recognition. Moreover, all respondents provided informed consent after receiving comprehensive written and oral information about the project. This information included the objectives and topics of the focus group interviews, the confidentiality measures in place to protect their data, the storage and use of the data for scientific papers and presentations, and the option to withdraw from the study at any time without providing a specific reason. Throughout the transcription process, careful consideration was given to protecting the identities of individuals, institutions, and locations involved in the study. Names were changed, and any information that could identify individuals was handled with the utmost caution.

References

1. National Committee on Vital and Health Statistics. Enhanced protections for uses of health data: a stewardship framework for secondary uses of electronically collected and transmitted health data. Report to the Secretary of the US Department of Health and Human Services; 2007.
2. Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information; Nass SJ, Levit LA, Gostin LO, editors. Beyond the HIPAA privacy rule: enhancing privacy, improving

health through research. Washington (DC): National Academies Press; 2009.
3. Dove ES, Phillips M. Privacy law, data sharing policies, and medical data: a comparative perspective. In: Gkoulalas-Divanis A, Loukides G, editors. Medical data privacy handbook. Cham: Springer; 2015.
4. State Council. Guiding opinions on promoting and standardizing the development of health and medical big data. 2016.
5. State Council. Opinions on further improving the medical and health service system. 2023.
6. Li X, Cong Y, Liu R. Research under China's personal information law. Science. 2022;378(6621):713–5.
7. China Hospital Association Information Committee. China hospital informatization survey (2021–2022). 2023.
8. Shi J, Yuan R, Yan X, Wang M, Qiu J, Ji X, et al. Research on the current status of health and medical data sharing of maternity and child healthcare alliances in the Yangtze River Delta region. China Digit Med. 2023;18(2):93–8.
9. Creswell JW. Research design: qualitative, quantitative, and mixed methods approaches. 4th ed. New York: Sage; 2013.
10. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist. Int J Qual Health Care. 2007;19(6):349–57.
11. Kvale S, Brinkmann S. Interviews: learning the craft of qualitative research interviewing. Los Angeles: Sage; 2009.
12. Krueger RA, Casey MA. Focus groups: a practical guide for applied research. Sage; 2014.
13. Booth A, Hannes K, Harden A, Noyes J, Harris J, Tong A. COREQ (consolidated criteria for reporting qualitative studies). In: Guidelines for reporting health research. Hoboken: Wiley; 2014.
14. Majid MAA, Othman M, Mohamad SF, Lim SAH, Yusof A. Piloting for interviews in qualitative research. Int J Acad Res Bus Soc Sci. 2017;7(4):1073–80.
15. Burnard P, Gill P, Stewart K, Treasure E, Chadwick B. Analysing and presenting qualitative data. Br Dent J. 2008;204(8):429–32.
16. Karasneh R, Al-Mistarehi AH, Al-Azzam S, Abuhammad S, Muflih SM, Hawamdeh S, et al. Physicians' knowledge and attitudes toward data sharing. Int J Gen Med. 2021;14:721–31.
17. Kalkman S, Mostert M, Gerlinger C, van Delden JJM, van Thiel GJMW. Responsible data sharing in international health research. BMC Med Ethics. 2019;20(1):21.
18. Godard B, Schmidtke J, Cassiman JJ, Aymé S. Data storage and DNA banking for biomedical research. Eur J Hum Genet. 2003;11(Suppl 2):S88–122.
19. Delacroix S, Lawrence ND. Bottom-up data trusts and data governance. Int Data Priv Law. 2019;9(4):236–52.
20. Skovgaard LL, Wadmann S, Hoeyer K. Attitudes toward reuse of health data in EU. Health Policy. 2019;123(6):564–71.
21. Mannheimer S, Pienta A, Kirilova D, Elman C, Wutich A. Qualitative data sharing and repositories. Am Behav Sci. 2019;63(5):643–64.
22. Gao Z, Huang Y, Yao F, Zhou Z. Public attitudes toward biobank donation. Front Public Health. 2022;10:1025775.
23. Ma Y, Dai HL, Wang LM, Zhu LJ, Zou HB, Kong XM. Consent for use of clinical biosamples. PLoS One. 2012;7(4):e36050.
24. Gao H, Jiang J, Feng B, Guo A, Hong H, Liu S. Parental willingness for biospecimen donation. BMJ Open. 2018;8(10):e022290.
25. State Council. Opinions on building a data infrastructure system. 2023.
26. National Development and Reform Commission. Data infrastructure system development. Macroecon Manage. 2023;(3):13–5.
27. Shen W. Empowerment logic of personal data ownership. Law Rev. 2023;41(5):114–28.
28. Federer LM, Lu YL, Joubert DJ, Welsh J, Brandys B. Biomedical data sharing and reuse. PLoS One. 2015;10(6):e0129506.
29. Zhou W, Yun W. Scientific data sharing policies analysis. Chin Sci Found. 2023;37(1):150–60.
30. Xu Y. Integration and application of medical data. Netw Secur Inform. 2023;5(9):86–9.
31. Sun R, Wang H, Chen M, Li J, Zhang Y, Liu Q. Regional medical data exchange systems. Chin Hosp Manag. 2023;43(8):56–9.
32. Li J. Big health era and data governance. Med Law. 2023;15(4):36–43.
33. Zhang Z, Wang B, Chen T, Li S, Liu H, Yang X, et al. Storage and sharing of clinical data in traditional

Chinese medicine. Chin J Tradit Chin Med Libr Inform. 2023;47(4):1–5.
34. Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract. 2017;36(1):3–11.
35. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018;57(S1):e50–6.
36. Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O. HiGHmed platform for care and research. Methods Inf Med. 2018;57(S1):e66–81.
37. Nyumba TO, Wilson K, Derrick CJ, Mukherjee N. Focus group methodology insights. Methods Ecol Evol. 2018;9(1):20–32.
38. Zhang J, Niu Y, Zhang Y, Zhang Z, Li X, Wang H. Real-world clinical data sharing. Chin Hosp. 2022;26(11):49–52.