BMC Medical Ethics

RESEARCH Open Access

Exploring barriers and ethical challenges to medical data sharing: perspectives from Chinese researchers

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

Abstract

Background The impetus for policies promoting medical data sharing in China has gained significant traction. Nonetheless, the present legal and ethical framework governing the research use of medical data in China, is characterized by a more restrictive rather than permissive approach. The proportion of Chinese medical data being leveraged for scientific research still has room for improvement at present, indicating a significant untapped potential for advancing medical knowledge and improving healthcare outcomes. Building upon this research, we aim to delve deeper into the challenges researchers encounter in the sharing of medical data through focus group interviews.

Methods We conducted two focus group interviews study with researchers representing diverse disciplines to explore their perspectives on 21 June 2021 and 28 July 2021. A total of seventeen researchers willingly participated in this study, representing various professional backgrounds. Similar codes were merged. Research team discussions were also utilized to select interviewees' statements that were regarded as typical or representative.

Results The respondents demonstrated a strong understanding that medical data should not be disseminated arbitrarily, recognizing the importance of sharing data in compliance with laws. Through the interview, we found that although respondents stressed the importance of careful consideration regarding if and when this information can be responsibly released, none of the respondents raised the issue of necessitating consent from data subjects for the research use of medical data. This observation sharply contrasts with the stringent separate consent provisions for secondary data use outlined in the PIPL.

Conclusions The findings from the focus group studies shed light on researchers' barriers and ethical challenges towards medical data sharing for scientific research, highlighting their deep concern for data security and cautious approach to sharing. The key objectives aimed at facilitating and enabling the reuse of medical data encompass enhancing interoperability, harmonizing data standards, improving data quality, safeguarding privacy, ensuring informed consent, incentivizing patients, and establishing explicit regulations pertaining to data access and utilization.

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

*Correspondence:
Yali Cong
ethics@bjmu.edu.cn
[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

Background

Medical data, compiled from diverse sources including research objectives, patients' clinical information, and government-released public data, boasts a rich history of fulfilling pivotal secondary functions that have significantly contributed to a wide array of societal benefits. These include, but are not limited to, population health monitoring, enhancements in healthcare quality, and groundbreaking advancements in biomedical research [1, 2]. In the contemporary interconnected environment, the flow of medical data transcends national boundaries and research teams, encompassing information from clinical and population research. For example, the interconnectedness of medical data empowers researchers to establish links between diverse datasets and access the necessary expertise to unravel the intricate molecular foundations and complexities of disease etiology [3].

The impetus for policies promoting medical data sharing in China has gained significant traction, driven by the national strategy to unlock the potential of big data for technological innovation, industrial growth, and societal benefits [4]. According to the *Opinions on Further Improving the Medical and Health Service System*, issued in March 2023, the government has explicitly aligned enhanced medical data sharing with its overarching vision of promoting equity, accessibility, and quality in the delivery of healthcare services [5]. Nonetheless, the current legal and ethical framework in China, which oversees the utilization of medical data collected for non-research purposes in research endeavors, is distinguished by an approach that is more restrictive in nature than permissive [3]. The enactment of Data Security Law (DSL), and Personal Information Protection Law (PIPL), demonstrates the government's commitment to reinforcing the protection of personal information. In our previous article, we identified that specific provisions within these laws failed to sufficiently address the distinctive challenges associated with scientific research in the era of big data [6].

The proportion of Chinese medical data being leveraged for scientific research still has room for improvement at present, indicating a significant untapped potential for advancing medical knowledge and improving healthcare outcomes. The findings from the 2021–2022 China Hospital Informatization Survey, conducted by the Information Professional Committee of the Chinese Hospital Association, provide valuable insights into the state of medical data reutilization. Notably, the research utilization of medical data remains relatively underdeveloped, with a mere 14.97% of hospitals dedicating focus to this domain [7]. Additionally, research by Shi Jingjin and her team reveals a notable gap within specialized alliance data-sharing platforms. Despite the low utilization rate of system functionalities for research collaboration—only 11.76% of member institutions actively engage in these activities—there is a strong demand for such features. Specifically, 58.82% of member institutions have expressed a keen interest in developing research-oriented capabilities [8].

Limited research has delved into the underlying reasons behind the low utility rate of medical data for scientific research. The sole study we could uncover, conducted by Zhang et al. in 2022, sheds light on this issue. Their findings reveal that despite a notable willingness, with 93.53% of respondents expressing interest in participating in data sharing, a majority of interviewed researchers reported unsuccessful experiences in this regard [10]. They concluded that this can be attributed to several key factors: challenges in obtaining administrative document support, a lack of access to database utilization rights with roles largely confined to data collection, and growing concerns among researchers about data subjects' privacy following data sharing. These concerns are exacerbated by the absence of clear legal and ethical regulations governing the research use of medical data [10]. The 2021-2022 China Hospital Informatization Survey has highlighted a critical challenge for tertiary hospitals regarding medical data sharing: the shortage of data scientists. Among the 684 hospitals surveyed, an overwhelming 514 hospitals, representing 75.15% of the total, identified this shortage as a primary obstacle. Additionally, more than 60% of the hospitals reported having fewer than 10 information technology staff members [8].

Building upon the above findings, we aim to delve deeper into the barriers and ethical challenges of sharing medical data for scientific research through focus group interviews. This inquiry seeks to address several critical questions:

How has the introduction of the PIPL and the DSL influenced medical data sharing practices?

Why is government-collected medical data not readily shared with researchers?

What factors contribute to researchers ' barriers to share data they collect with other researchers?

Are there any values that promote medical data sharing?

Methods

Focus groups

Given the limited data available, we used focus group interviews in this exploratory study, and attempted qualitative analysis of the results so that we could develop preliminary hypotheses regarding our research questions [9]. We conducted two focus group interviews study with researchers representing diverse disciplines to explore their perspectives on 21 June 2021 and 28 July 2021. The researchers we interviewed was comprised by physicians(P)and data scientists(S). The physicians

were from various medical specialties, including internal medicine, surgery, dentistry, and psychiatry. Meanwhile, the data scientists were from medical universities and research institutes, specializing in data analysis and processing. Rather than engaging in direct data collection, they acquired data from physicians or fellow researchers, subsequently subjecting it to analysis and processing (Additional file 1). The focus groups were conducted under the guidance of the corresponding author (YLC), who served as the moderator, with the first author (XJL) present to attend and observe the sessions. Respondents were encouraged to freely express their opinions, with the guarantee that their responses would be handled with utmost confidentiality and anonymity [10]. Special attention was given to ensure the inclusion of every participant, fostering a comprehensive and inclusive discussion environment [11, 12].

The study adheres to the rigorous reporting guidelines outlined in the Consolidated Criteria for Reporting Qualitative Studies (COREQ) [13]. By employing this robust methodology, we aimed to capture rich and in-depth insights into the ethical considerations and viewpoints of the participating researchers.

Respondents

Recruitment of interview respondents was conducted by investigators from the Institute of Science and Technology Policy and Management Science, Chinese Academy of Science. We have entered into an agreement with the Institute of Science and Technology Policy and Management Science, Chinese Academy of Sciences, as they possess access to the esteemed network of scientific experts in China's expert database system. This partnership enables us to connect with authoritative experts in the realm of medical data sharing for interviews.

Each focus group interview, conducted in the Chinese language, lasted approximately 2 h. To ensure a comprehensive and consistent approach, we collaborated with the broader research team associated with this study to develop a semi-structured interview guide[1]. The guide consisted of open-ended questions, with references provided in the Additional file 2.

To refine the interview guide, we conducted a pilot interview with two colleagues to assess the effectiveness of the questions in stimulating reflective responses [14]. Based on their feedback, we made minor adjustments, such as converting some questions into probing inquiries and modifying the order of the questions for improved flow and coherence.

[1] For more detailed information about the broader research team, please refer to the Acknowledgments section.

Analysis

The recorded interviews were transcribed verbatim by a professional transcription company. The transcripts were listened to in their entirety to verify the transcription. The authors read the transcripts several times, analyzed them line by line, and replaced individual statements with general concepts or themes, so that all the issues relevant to the expectations of the respondents were identified. An inductive approach was adopted. An initial list of bottom-up derived codes was composed and discussed between the first and corresponding author after a first analysis of the transcripts. Additional bottom-up codes were added during the continuation of the analysis, staying close to the content of the answers of respondents [15]. Similar codes were merged. Research team discussions were also utilized to select interviewees' statements that were regarded as typical or representative. We repeated these processes until we reached consensus regarding the final presentation of the results. The results were translated into English by the authors, two authors check separately ensuring that the content remains accurate and faithful to the original exchanges. Since the original interview was conducted in Chinese, we have made meticulous adjustments during the translation process to enhance clarity and readability, while meticulously ensuring that the content remains precise and loyal to the original exchanges.

Results

The implications of PIPL and the DSL

When discussing the impact of the PIPL and the DSL, the respondents recognized that strict adherence to these legal frameworks was essential for protecting privacy and upholding integrity, thereby maintaining high legal and ethical standards throughout the research process. The respondents demonstrated a strong understanding that medical data should not be disseminated arbitrarily, recognizing the importance of sharing data in compliance with laws.

> *I believe data security comes first, followed by sharing. Before sharing data, I would ensure that it complies with legal requirements, understand how the data will be used, determine who is responsible, and verify any legal obligations. Data sharing should never be done blindly. (S1, male)*

Certain respondents expressed the belief that privacy and security concerns could be effectively addressed as long as data usage and sharing adhered to legal provisions. For example, they proposed that if hospitals followed the PIPL's provision to de-identify patient information before sharing, the risk of medical data leakage would be minimal. Nonetheless, they underscored that achieving this

objective require legal safeguards as well as oversight by professional ethics committees.

Some respondents contended that the de-identification of medical data according to the law might diminish its scientific value and voiced concerns that, in their eagerness to share data, some researchers might disregard legal constraints.

> *The perceived low cost of legal infractions allows for the casual use of medical data[2] without fear of repercussions or penalties, such as fines. This results in instances where medical data is haphazardly employed for article writing, seemingly disconnected from any patient's welfare. Additionally, the prevailing belief is that patients are unlikely to pursue legal action against such practices due to the prohibitively high costs associated with litigation. (P3, female)*

None of the respondents raised the importance of re-consent from data subjects for the secondary use of data. This observation stands in stark contrast to the stringent separate consent provisions for secondary data use as stipulated in the PIPL. Notably, one respondent went so far as to express concerns that the patient's informed consent process might potentially pose future challenges for researchers:

> *Some patients will be more hypocritical. What have researchers done with my data, and what have researchers developed? Can I get feedback? The more patients know about what researchers do through the informed consent process, the more they will care about these things. The protection of rights of patients will have a lot of adverse effects on researchers. (P5, female)*

Some respondents expressed concerns regarding the lack of enforceability of legal provisions on informed consent:

> *Although patients had signed the informed consent form and confidentiality agreement, I know that thing has no binding effect on researchers, they can do whatever they want. (S3, male)*

Some respondents expressed concerns regarding the lack of enforceability of legal provisions on informed consent:

> *Although the academic community generally assumes that medical data is owned by the patient, there is no official regulation to enforce this. Implementing such a rule could introduce challenges: without explicit patient consent or a compelling public interest, data sharing would be limited, creating complications in how data is utilized and shared. (P2, female)*

Concerns about data sensitivity hampers government data sharing for scientific research

Several respondents have expressed concerns regarding the constraints imposed on the opening of medical data[3] housed within government databases. These restrictions have resulted in researchers being unable to access these valuable resources, thereby impeding the realization of their full potential value. They believed the non-public nature of government data, hampers their access to reliable domestic databases. Based on the responses from the participants, the refusal to share government data was anticipated to yield three negative consequences: first, it significantly limited researchers' capacity for medical development and innovation. Just as one participant said:

> *The concept of 'sensitivity' often serves as a self-imposed restraint, hindering progress and innovation. In reality, we have the capability to establish industry standards independently, yet we frequently find ourselves trailing behind the United States and Europe. This can be attributed to an outdated mindset and inadequate mechanisms, resulting in a squandering of valuable resources. (S1, male)*

Second, it improved the likelihood of researchers resorting to unethical methods of acquiring or trading hospital data.

> *By making data more accessible, scientific researchers will be less inclined to engage in unethical data purchases. In the absence of reliable domestic databases, purchasing data becomes the only viable option. (S4, female)*

Third, it impacted the efficacy of government's public functions, as one respondent articulated:

> *At the governmental level, the integration of medical data necessitates a profound understanding of the government's fundamental role in fostering the holistic advancement of society. Consequently, the utilization of medical data for scientific research emerges as a significant and valuable contribution, deserving of consideration within the framework of*

[2] The medical data here refers to data collected from clinical settings.

[3] The medical data mentioned here pertains to data gathered by governmental agencies for purposes that are not related to scientific research endeavors.

governmental evaluations and assessments. (P4, male)

Although certain government agencies have expressed their willingness to openly share medical data with researchers, respondents have underscored that concerns related to data sensitivity hinder these agencies from readily disclosing this information.

In order to gain a more comprehensive understanding of the concept of data sensitivity, participants were queried for their perspectives. A respondent elucidated:

The current landscape of medical data, especially within the domain of big data, raises substantial concerns regarding sensitivity, particularly with respect to genetic information. The dissemination and disclosure of genetic information present substantial challenges to security, demanding meticulous consideration of when and how this information should be shared (P1, male).

While medical data does not inherently constitute genetic information, it was believed that certain medical data may inadvertently disclose genetic specifics unique to the Chinese population. "This is due to the fact that, when interconnected, even scientific research data may present potential risks to national security" (S1, male). This realization has prompted respondents to adopt a cautious stance towards cross-border data sharing.

Some respondents expressed concerns about China's limited participation in cross-border sharing of health and medical data. They noted the predominant influence of European and American countries in formulating data sharing standards, leaving a gap in the exploration of suitable local standards. They argued that stringent confidentiality measures, while intended to protect privacy and security, have also limited growth and impeded innovation.

Addressing the issue of sensitivity concerns surrounding the government's medical data, some experts contend that it is imperative to engage a team of dedicated professionals who are equipped to discern between data that is highly sensitive and that which poses minimal risk for sharing. One respondent said:

I believe that a significant number of researchers are either unaware of or indifferent to this issue, and as a result, they lack the know-how to properly desensitize sensitive information. In some instances, they unwittingly share information such as patients' names, case numbers, and ID card numbers, simply because they do not possess the necessary technical skills to obscure or remove identifying information (P8, male).

However, they acknowledge that implementing such a desensitization mechanism required a sustained effort and could not be achieved overnight.

Researchers' aspiration is to make a vast array of medical data readily available for comprehensive research and analysis. However, the sensitivity of a considerable portion of medical data presents a significant challenge, leading to hesitancy in its release. In reality, it is possible that some of this data may not be as sensitive as was initially believed, yet the government has failed to give this matter the due attention it deserves. Consequently, a robust and well-defined mechanism for the responsible release of data has yet to be put in place, a fact that is highlighted by the pressing need for disease control data to be made openly accessible to the public. (S5,female)

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

When asked about their opinion on the major concerns related to sharing medical data collected by researchers themselves, the majority of respondents stressed the importance of benefit mechanisms. These may entail receiving financial rewards, gaining recognition through published articles, or being acknowledged for their valuable contributions to organizations." *This is the property of our hospital, and it is not acceptable to give it to you casually*", as a participant expressed, "*it is crucial to avoid situations where one party diligently contributes their data, only for the other party to reap the benefits or publish articles without providing any recognition or compensation in return*" (P10, male). Without a robust and sustainable benefit-sharing mechanism, it becomes increasingly difficult to foster long lasting and meaningful data sharing practices.

Take blockchain, for example—it's a popular topic, yet it doesn't truly solve the challenges faced in scientific research. Blockchain's primary strength lies in facilitating basic data sharing, such as granting access to medical data. However, when it comes to scientific research requiring vast amounts of data, blockchain falls short. This is exacerbated by the fact that stakeholders lack sufficient motivation to collaborate—why would they share their resources if there's no tangible incentive? While blockchain has the potential to allow medical data to be shared across hospitals, the cost and upkeep often outweigh the perceived benefits, leaving participants with little reason to engage. For technologies like blockchain to succeed in real-world applications—particularly

in healthcare, where even charitable efforts require sustainability—the cost-benefit equation must be addressed. This is where many current technology implementations fall short.(S4).

According to several respondents, the crux of an effective medical data sharing mechanism lies in the implementation of a benefit-sharing system. Based on their own experiences, individuals are often reluctant to share their data without a compelling incentive structure in place. "*The absence of a grassroots-driven interest significantly impedes the establishment of mutually beneficial sharing agreements*, consequently hindering the progress of data sharing endeavors at more elevated operational levels. (P6, female)." In the absence of such mechanisms, data processors perceive little disadvantage in withholding their data, while encountering potential challenges and drawbacks if they choose to share.

Formal institutional data sharing contracts leading by trusted experts

Many respondents advocated for a shift from individual-level data sharing to a more institution-based approach due to inherent compliance risks associated with sharing data among individual researchers. They proposed that medical data sharing was best facilitated between institutions, rather than relying solely on researchers. These respondents suggested that proactive measures, such as signing comprehensive cooperation agreements, could effectively address potential discrepancies and conflicts that may arise during the collaborative process.

According to one participant, academic institution specializing in data analysis should establish contractual arrangements with hospitals. This collaboration allowed for the exchange of expertise ("technology") and datasets ("data"), solidifying the partnership through formal data sharing contracts.

We share all our data openly and collaboratively with others, adhering to rigorous contractual agreements. Our agreements ensure complete transparency, with each party's responsibilities and rights being clearly outlined and mutually understood. (P7, male)

Respondents emphasized the importance of a reputable and trusted "big expert" within the alliance.

This esteemed individual assumes a leadership position, playing a pivotal role in guiding and orchestrating collaborative endeavors. They further contribute by formulating equitable rules that govern the distribution of benefits, ensuring universal acceptance and commitment from all alliance members. As a trusted expert, they serve as a recognized authority, offering invaluable guidance and fostering a culture of trust within the collaborative alliance. (P1, male).

For specialized disease data, appointing chief physicians or renowned professors as leaders is a more fitting choice. These esteemed individuals bring their profound expertise to the alliance, thereby guaranteeing a depth of specialized knowledge and fostering a sense of trust among its members. (P11, male)

By embracing institution-based collaborations and establishing robust leadership structures within alliances, respondents were of the conviction that compliance risks could be significantly reduced, expertise could be harnessed efficiently, and trust could be nurtured. These strategic approaches not only lay a solid groundwork for seamless data sharing but also pave the path towards achieving successful and impactful biomedical research collaborations.

Confronting the barrier of poor data quality

According to the respondents, even if data was shared, its usability was not guaranteed, and establishing a comprehensive set of standardized data across the country was impractical. Respondents emphasized that amassing a substantial volume of data alone is inadequate; the genuine value resided in the capacity to access and utilize the data effectively.

I have had the opportunity to participate in multicenter collaborative epidemiological research, where I observed that out of all the hospitals involved, only one site's data consistently met the established standards. This revelation highlighted the limitation in the availability of usable data despite the collaborative efforts, emphasizing the formidable challenge of achieving standardized data sharing at a national scale (S1, male).

When they were asked about their views on barriers of data utility to achieve the greatest overall well-being and positive outcomes, the reason they gave were as follows:

1. Elevated Sharing Costs and Fragmented Storage Infrastructures: The proliferation of disparate storage systems poses a significant barrier to seamless data sharing, ultimately driving up costs.
2. Scarcity of Data Scientists and Limited Data Literacy: The scientific value of shared data is significantly diminished by a shortage of skilled data scientists and a general lack of data analysis techniques among researchers. This deficiency necessitates extensive guidance for researchers to comprehend

fundamental concepts such as data sets, scientific data management, and advanced analysis techniques.

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 in order to mitigate these associated risks.

Recognizing the values of rewarding patients, yet facing challenges

All respondents unanimously voiced their conviction that researchers ought to ensure the welfare of data subjects, embodying the cherished principle of "What is garnered from the people should be utilized for their benefit". They said ensured a fair distribution of beneficence is quite important because on one hand, data sharing would bring privacy risk to patients, on the other, through data sharing, decisions based on big data became more scientifically grounded and reasonable, leading to tangible benefits for data subjects.

> *Patients find themselves in a position of increased transparency, as their health information becomes accessible to society at large, insurance companies, and hospitals alike. Despite this exposure, hospitals are well-positioned to offer enhanced services to patients. For instance, by conducting meticulous analyses of historical data pertaining to lung cancer patients, hospitals can improve screening processes during routine physical examinations. This refinement will undoubtedly benefit a pivotal segment of the population, such as petrochemical workers, who are often exposed to risk factors associated with the development of lung cancer. (P12, male)*

However, a noteworthy concern was raised by a respondent pertaining to the prevailing obstacles in facilitating the seamless sharing of medical data across various institutions. Specifically, she underscored the intricate challenges inherent in even the intrahospital sharing of medical data, encompassing the intricate hierarchy of hospitals, from first-level to second-level and up to third-level facilities.

> *Patients seeking medical consultation at tertiary hospitals often encounter a frustrating scenario where they are compelled to undergo redundant medical tests, despite having already undergone the same procedures at secondary hospitals. Facilitating seamless data sharing between various tiers of healthcare institutions holds the potential to significantly streamline the process, saving patients valuable time and alleviating financial burdens. (P6, female)*

The respondents advocated to develop a data sharing platform to foster patients' trust and facilitate the expansion of sharing initiatives. They believed that the sharing platform instilled a sense of confidence. Researchers could securely upload and utilize data, knowing that it will be accessed by a broader community of peers. This expanded accessibility enhanced collaboration and encourages the utilization of high-quality data, ultimately leading to improved research outcomes.

> *By implementing this platform, several pivotal benefits are envisioned. Firstly, patients attain enhanced clarity concerning which researchers possess access to their data and the precise objectives of their research endeavors. This heightened transparency equips patients with the knowledge necessary to make more informed choices about data sharing, empowering them to actively engage in and shape the research process. Additionally, the platform facilitates the dissemination of research findings, enabling patients to directly observe the societal advantages stemming from their data contributions. (P7, male)*

Discussion

Comprehending researchers' perspectives on the challenges surrounding medical data sharing for scientific research is paramount in the ever-evolving healthcare landscape [16]. The insights garnered from this study hold immense value for researchers in developing countries, as they provide a holistic understanding of strategies to cultivate an environment conducive to the ethical sharing of medical data for scientific endeavors. These findings resonate with the ethical principles outlined in various international guidelines and scholarly works [17].

Customizing data protection laws for scientific research and data subject rights

While imposing laws and regulations that govern data sharing is undoubtedly a vital step, it is insufficient in itself. It is of paramount importance that researchers are not only informed about these legal frameworks but also actively engaged in their enforcement. Upon conducting the interview, we discerned a notable disparity in

the respondents' perspectives. While they unanimously underscored the cruciality of meticulous deliberation regarding the responsible disclosure of information, the issue of securing explicit consent from data subjects for the secondary utilization of their data remained unaddressed. This observation stands in stark contrast to the stringent provisions outlined in the PIPL, which mandate separate consent for secondary use.

Several respondents expressed concerns about the practical enforceability of legal frameworks governing informed consent, highlighting a potential gap between legislation and its implementation. Concerns pertaining to potential data misuse, the risks associated with full transparency, and the specter of data breaches, particularly in the context of long-term storage, have been documented to influence patients' cautious approach and reluctance to grant consent for data reuse in research endeavors [18–21]. The physicians we interviewed, who conduct research and collect patient data in clinical settings, carry the dual responsibility of protecting patient privacy and advancing biomedical research. According to current law, informed consent from patients is required for the use of samples and medical data in scientific research. However, challenges remain. For instance, many samples stored in hospitals were collected years ago, before comprehensive informed consent protocols were established. Although these samples and data could be valuable for research, they are not usable under the PIPL.

Nevertheless, research consistently indicates that the majority of Chinese citizens favor the secondary use of medical data, contingent upon its contribution to the broader societal welfare [22–24].Given this backdrop, we posit that a more flexible consent model, which avoids the rigidity of a "one-size-fits-all" approach, could garner acceptance among the Chinese population, as they aspire to contribute to the common good. However, it is imperative that researchers adhere strictly to legal guidelines to ensure that data is handled responsibly and without abuse.

Another crucial aspect that significantly contributes to researchers' barriers in sharing medical data is the ambiguity surrounding legal regulations pertaining to data ownership. China's State Council recently released the "Opinions on Building a Data Infrastructure System to Better Leverage Data Elements" [25], a document that purposefully avoids delving into the matter of personal data ownership. During its formulation, the document adhered to the principle of "downplaying ownership, emphasizing usage rights, and concentrating on the circulation of data usage rights." [26] It refrained from explicitly designating data ownership to the data source. However, it's important to note that data ownership serves a dual purpose: safeguarding personal data against infringement and enabling data subjects to delegate data management to processors that align with their objectives [27].

Clarity in sensitivity: the need for government agency guidelines

The transformation underway in facilitating the reuse of open government data by third parties for public benefit, both in China and globally, is marked by a pivotal shift in policy landscapes. This necessitates a thorough examination of mechanisms to ensure data sensitivity standards and judicious decisions regarding the scope of data openness.

The research use of medical data is intricately tied to local frameworks and pathways, presenting researchers with a highly variable landscape. Despite the vast and diverse trove of medical data held by Chinese government entities like the Health Commission and the Natural Science Foundation of China, these resources remain underutilized and their sharing hindered by sensitivity concerns. A notable gap exists in the form of standardized guidelines or best practices for the provision of open government data, leaving researchers to navigate this terrain with limited direction. Addressing these challenges is crucial to harnessing the full potential of open government data for the benefit of society.

Ensuring motivated and equitable benefit distribution in medical data sharing

Ensuring motivated and equitable distribution of benefits in medical data sharing provides an incentive for researchers to actively engage in data sharing activities. Respondents suggest that addressing this issue involves platform building, embracing institution-based collaborations and instituting leadership structures within alliances. Our findings align with Federer et al.'s survey results, which suggest that systematic methods of data sharing can promote broader access to and reuse of research data [28].

All respondents unanimously expressed the viewpoint that researchers have an obligation to ensure that data subjects derive tangible benefits from their contributions. They brought to light the prevailing challenges in facilitating the sharing of medical data among institutions within the hospital hierarchy, which unfortunately tends to impose additional financial burdens on patients during their treatment journey. In light of these concerns, the respondents passionately advocated for the establishment of a unified data-sharing platform. Such a platform would not only streamline the treatment process for patients, making it more efficient and less stressful, but also foster scientific research endeavors, ultimately leading to rewards for the very individuals whose data

has been shared, thereby recognizing and valuing their contributions.

Medical data sharing demands adequate technical support

The recognition that research environments in numerous developing regions significantly diverge from those in high-income areas, particularly in terms of resource availability, research support, and infrastructural capabilities, poses a complex challenge for the integration of data. Although most researchers have the motivation to share data, respondents believe that they lack of professionals in data analysis and use, resulting in insufficient utilization of data and poor data quality [29–33]. These complexities stem from the inherent variations in treatments, outcomes, study designs, analytical methods, and the diverse approaches employed for collecting, processing, and interpreting data in the field of medicine [34].

Recognizing these differences suggests that the need for more nuanced solutions. This should include establishing a unified, cost-effective storage and sharing platform that integrates disparate systems, reducing both costs and barriers to seamless data exchange. Additionally, investing in the training and development of data scientists and enhancing data literacy among researchers is crucial to maximize the scientific value of shared data. Furthermore, implementing robust quality control measures and fostering a culture of transparency and accountability can alleviate institutional concerns, ensuring that data shared is of the highest quality and mitigating the risk of unintended revelations or legal consequences [35, 36].

Limitations

During the focus group interviews, the term "medical data" was interpreted differently by individuals based on their respective research backgrounds. For those requiring special clarification, we have provided explanation in footnote below the quoted interview texts. A limitation of our research is that it exclusively focused on researchers and did not include the perspectives of patients. While this approach was intentional to gain insights into the challenges researchers encounter in the sharing of medical data for scientific research, it does restrict our understanding of the patient viewpoint. To address this limitation, we are currently conducting a follow-up study to explore patient attitudes and perceptions regarding medical data sharing for research purposes in the Chinese context. While information acquired from focus group interviews can give rise to hypotheses about a specific target population, it is crucial to subject these hypotheses to quantitative verification through surveys to ensure their precision. The strength of incorporating focus group interviews into our study lies in their ability to provide profound insights into attitudes and beliefs, facilitated by the interaction among respondents, which fosters in-depth discussions on contentious subjects [37].

Conclusion

The findings from the focus group studies provide insights into strategies for incentivizing researchers in developing countries to participate in the sharing of medical data for scientific research. Regulatory frameworks must navigate the complex landscape encompassing privacy preservation, open science, and national security. Encouraging data sharing necessitates reciprocal actions, incentives, technical support, clarity in data sensitivity, and effective mechanisms for sharing benefits. By addressing these challenges, ethical data sharing practices can be promoted, ultimately benefiting patients, research, and society as a whole. The discussion should move beyond the binary question of whether to share data or not and focus on the optimal ways to share data that minimize potential harms and uphold patients' reasonable expectations. To foster data sharing, researchers require reciprocal actions and incentives from their peers and society as a whole, highlighting the necessity of recognizing and rewarding data contributors.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12910-024-01135-8.

Supplementary Material 1: Focus group guide.(DOCX ,37KB)

Supplementary Material 2: Characteristics of participants. (DOCX,16KB)

Supplementary Material 3: Reflexivity statement

Acknowledgements

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.

Author contributions

Y.L.C and X.J.L. conceptualized the study. X.J.L. drafted the initial version of this manuscript. Y.L.C subsequently made substantial revisions to the manuscript. X.J.L reviewed the manuscript for intellectual contents. All authors read and approved the final manuscript.

Funding

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

Data availability

The data analysed during the current study are not publicly available due to them containing information that could compromise participants' privacy and consent, but are available from the corresponding author on reasonable request.

Declarations

Ethical approval and consent to participate

This investigation was reviewed and approved by the Health Science Center of Peking University IRB (No IRB00001052-19013). To safeguard the privacy of the respondents, we proactively requested them to mask any identifying characteristics that could potentially lead to recognition. Moreover, all respondents provided informed consent, having received comprehensive written and oral information about the project. This information included details about 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, as well as the option to withdraw from the study at any point 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 posed a risk of identifying individuals was handled with utmost caution.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 October 2023 / Accepted: 7 November 2024
Published online: 15 November 2024

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 ncvhs.hhs.gov/wp-content/uploads/2007/12/December-22-2007-Enhanced-Protections-for-Uses-of-Health-Data-A-Stewardship-Framework-for-Secondary-Uses-of-Electronically-Collected-and-Transmitted-Health-Data.pdf
2. Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule; Nass SJ, Levit LA, Gostin LO, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington (DC): National Academies Press (US). 2009;4.
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 of the General Office of the State Council on Promoting and Standardizing the Development of Health and Medical Big Data. 2016. https://www.gov.cn/gongbao/content/2016/content_5088769.htm
5. State Council. Opinions on Further Improving the Medical and Health Service System. 2023.https://www.gov.cn/zhengce/2023-03/23/content_5748063.htm
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), February 2023https://www.chima.org.cn/Html/News/Articles/8688.html
8. Jingjin S, Xueming Y, Rui Y, et al. Research on the current status of Health and Medical Data Sharing of Maternity and Child Healthcare specialist alliances in the Yangtze River Delta Region [J]. China Digit Med. 2023;18(02):93–8.
9. Cresswell J. 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 for interviews and focus groups. Int J Qual Heal 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, et al. COREQ (consolidated criteria for reporting qualitative studies). Guidelines for reporting health research: a user's manual. Hoboken: Wiley; 2014.
14. Majid MA, Othman M, Mohamad SF, et al. Piloting for interviews in qualitative research: operationalization and lessons learnt. Int J Acad Res Bus Social Sci. 2017;7(4):1073–80.
15. Burnard P, Gill P, Stewart K, Treasure E, Chadwick B. Analysing and presenting qualitative data. Brit Dent J. 2008;204(8):429–32.
16. Karasneh R, Al-Mistarehi AH, Al-Azzam S, et al. Physicians' knowledge, perceptions, and attitudes related to patient confidentiality and data sharing. Int J Gen Med. 2021;14:721–31.
17. Kalkman S, Mostert M, Gerlinger C, et al. Responsible data sharing in international health research: a systematic review of principles and norms. BMC Med Ethics. 2019;20(1):21.
18. Godard B, Schmidtke J, Cassiman JJ, et al. Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective[J]. Eur J Hum Genet. 2003;11(2):S88–122.
19. Delacroix S, Lawrence ND. Bottom-up data trusts: disturbing the 'one size fits all'approach to data governance[J]. Int data Priv law. 2019;9(4):236–52.
20. Skovgaard LL, Wadmann S, Hoeyer K. A review of attitudes towards the reuse of health data among people in the European Union: the primacy of purpose and the common good. Health Policy. 2019;123(6):564–71.
21. Mannheimer S, Pienta A, Kirilova D, et al. Qualitative data sharing: data repositories and academic libraries as key partners in addressing challenges. Am Behav Sci. 2019;63(5):643–64.
22. Gao Z, Huang Y, Yao F, et al. Public awareness and attitudes toward biobank and sample donation: a regional Chinese survey. Front Public Health. 2022;10:1025775.
23. Ma Y, Dai HL, Wang LM, et al. Consent for use of clinical leftover biosample: a survey among Chinese patients and the general public. PLoS ONE. 2012;7(4):e36050.
24. Gao H, Jiang J, Feng B, et al. Parental attitudes and willingness to donate children's biospecimens for congenital heart disease research: a cross-sectional study in Shanghai, China. BMJ open. 2018;8(10):e022290.
25. China's State Council. Opinions on Building a Data Infrastructure System to Better Leverage Data Elements, 2023 https://www.gov.cn/gongbao/content/2023/content_5736707.htm
26. National Development and Reform Commission. Accelerate the construction of a data infrastructure system with Chinese characteristics to promote the sharing of the dividends of digital economic development by all people. Macroeconomic Manage. 2023(03):13–5.
27. Shen Weixing L. Empowerment logic and institutional development of personal data ownership. Law Rev. 2023;41(05):114–28.
28. Federer LM, Lu YL, Joubert DJ, et al. Biomedical data sharing and reuse: attitudes and practices of clinical and scientific research staff[J]. PLoS ONE. 2015;10(6):e0129506.
29. Zhou Weneng L, Yun W. Analysis of domestic and foreign scientific data management and sharing policies and implications for the National Natural Science Foundation. Chin Sci Foundation. 2023;37(01):150–60.
30. Xu Yinghui. Research on the integration and application of medical data in the context of hospital informatization. Netw Secur Informatization. 2023;5(09):86–9.
31. Runkang S, Huoming W, Min C. etc. Research on data exchange system between regional medical data center and national medical center. Chinese Hospital Management.2023; 43(08): 56–59 + 63.
32. Li Juping. The big health era: a perfect path from data medical openness to joint governance. Med Law. 2023;15(04):36–43.
33. Zhang Zedan W, Bin C, Ting, et al. Problems and suggestions for storage and sharing of clinical data of traditional Chinese medicine in my country. Chin J Traditional Chin Med Libr Inform. 2023;47(04):1–5.
34. Lee CH, Yoon H-J. Medical big data: promise and challenges. Kidney Res Clin Pract. 2017;36(1):3.
35. Semler SC, Wissing F, Heyder R. German medical informatics initiative. Methods Inf Med. 2018;57(S 01):e50–6.
36. Haarbrandt B, Schreiweis B, Rey S, et al. HiGHmed–an open platform approach to enhance care and research across institutional boundaries. Methods Inf Med. 2018;57(S 01):e66–81.

37. Nyumba O, Wilson T, Derrick K. The use of focus group discussion methodology: insights from two decades of application in conservation. Methods Ecol Evol. 2018;9(1):20–32.
38. Zhang Jingyu N, Yuxian Z, Yubo, et al. Research on the current status of scientific research sharing of clinical real-world data. Chin Hosp. 2022;26(11):49–52.

Publisher's note