Short Communications

Editorial Process: Submission:03/30/2025 Acceptance:03/16/2026 Published:04/07/2026

Surveilling Care, Protecting People: Legal Reforms for the Data-Driven Clinic

Minsoo Jung[1,2]*

Abstract

Abstract: This study explores the concept and potential applications of healthcare big data, focusing on the legal and institutional challenges arising from the imperative to protect personal information. Healthcare big data holds transformative promise across multiple domains, including precision medicine, public health policymaking, and drug development, contributing to enhanced population health and medical innovation. However, ongoing legal tensions between data utilization and privacy protection persist, largely due to the sensitive nature of medical data and the inherent risk of re-identification. **Method:** This paper conducts a comparative legal analysis to examine the consistency of legal frameworks governing healthcare big data. It analyzes U.S. laws specifically, the Health Insurance Portability and Accountability Act (HIPAA) and the 21st Century Cures Act alongside corresponding Korean statutes, including the Medical Service Act and the Bioethics and Safety Act. **Results:** The comparative analysis identifies ongoing legal tensions between data utilization and privacy protection. These tensions arise from ambiguous definitions of pseudonymized and anonymized data, the limited flexibility of consent mechanisms for data subjects, and inconsistent de-identification standards. Additionally, although technology-based security safeguards are advancing, legal frameworks have not evolved at the same pace, resulting in regulatory gaps that hinder effective governance of healthcare big data. **Conclusion:** This paper proposes several key reforms: refining the legal definitions of pseudonymized and anonymized data, introducing more flexible consent mechanisms for data subjects, enhancing de-identification standards, and strengthening technology-based security safeguards. The paper emphasizes the urgent need to reconcile conflicts and inconsistencies among these laws. For healthcare big data to serve as a trust-based public resource, a regulatory environment that ensures both robust privacy protections and the responsible use of data must be established.

Keywords: Healthcare big data- Personal information protection- De-identification- Legal system improvement

Asian Pac J Cancer Prev, **27 (4)**, 1149-1153

Legal and Ethical Considerations in the Conceptualization and Use of Healthcare Big Data

In the era of the Fourth Industrial Revolution, healthcare data is increasingly recognized as a core asset in technology-driven health management. Among its various forms, healthcare big data represents a high-value resource with broad applicability across domains such as national health promotion, disease prediction, precision medicine, pharmaceutical innovation, and public health policy. However, because this data inherently contains highly sensitive personal information, its utilization inevitably raises legal and ethical tensions regarding privacy protection. Healthcare big data refers to large-scale health-related datasets collected from diverse sources, including public institutions. It is characterized by the traditional "3Vs" of big data: volume, velocity, and variety. These datasets extend beyond clinical records to include genomic information, biometric data from wearable devices, and lifestyle-related indicators. Healthcare big data is distinguished by three key attributes. First, it is centered on sensitive personal information including health status, diagnoses, treatments, and biometric identifiers with an ever-present risk of re-identification. Second, it contains expert-based evaluative content, such as clinical judgments, diagnoses, and prescriptions, making it more complex than quantitative datasets. Third, it comprises longitudinal, patient-level time-series data, making it particularly well-suited for precision medicine. The potential social value of healthcare big data includes: (1) enabling personalized treatment through the integration of genomic and clinical information; (2) supporting prevention-focused public health policies via the linkage of screening and clinical data; (3) facilitating early detection of adverse drug reactions through analysis of medication and treatment outcomes; and (4) advancing new drug development and medical research by enabling

*[1]Department of Health Science, Dongduk Women's University, Seoul, South Korea. [2]Center for Community-Based Research, Dana-Farber Cancer Institute, Boston, MA. *For Correspondence: mins.jung@gmail.com*

large-scale clinical trial design and efficacy assessments.

However, several critical considerations must be addressed to enable the responsible use of healthcare big data. First, the legal definition of personal information requires greater clarity. The current Personal Information Protection Act defines personal information too broadly, creating ambiguity in determining which datasets may be lawfully utilized. Clearer distinctions between pseudonymized, anonymized, and sensitive information are essential to enhance legal predictability and consistency. Second, the issue of data subject consent remains a significant challenge. Obtaining prior informed consent for large-scale healthcare datasets is often impractical. Therefore, a shift from an opt-in model to a conditional opt-out framework should be explored, particularly in contexts such as rare disease research or public health initiatives where public interest is paramount. Third, the principles of data minimization and purpose limitation should be interpreted with flexibility. Given the dynamic nature of big data analytics, secondary research purposes may emerge beyond the original scope of data collection, often leading to valuable medical insights. Rigid adherence to narrow legal interpretations may stifle innovation and public benefit. Fourth, anonymization poses inherent limitations. While it reduces re-identification risks, it also diminishes data utility. Moreover, complete prevention of re-identification is technically unfeasible. Therefore, dual-layered de-identification protocols such as expert determination and safe harbor standards, as outlined in the U.S. HIPAA framework should be considered (Health Insurance Portability and Accountability Act of 1996, 45 CFR Parts 160 and 164). Fifth, robust, technology-driven security systems are essential. Tools such as blockchain may enhance data integrity and transparency, while digital ID-linked access control and audit systems can help safeguard data subjects' rights in practice.

Healthcare big data has the potential to transform the medical paradigm through integrated applications across clinical care, biomedical research, and public health policy. However, realizing this potential depends on legal and institutional coordination that reconciles data use with privacy protection. The objective is not to dilute privacy principles, but to construct a "trust-based data ecosystem" grounded in public value and responsible governance. Achieving this vision requires a multi-pronged strategy: flexible legal interpretation, enhanced technological safeguards, robust rights protections, and cross-sectoral institutional collaboration.

This paper examines the conceptual framework and societal utility of healthcare big data, with a focus on privacy-related legal issues. It argues that, given the high utility of such data in precision medicine, policy development, and drug innovation, it is imperative to harmonize existing statutes namely, the Personal Information Protection Act, the Medical Service Act, and the Bioethics and Safety Act. Key recommendations include refining definitions of pseudonymized and anonymized information, operationalizing flexible consent mechanisms, strengthening de-identification protocols, and institutionalizing security measures. Together, these efforts aim to establish a balanced legal infrastructure that supports both privacy protection and data-driven innovation.

Legal Foundations in Korean Medical Law and Comparative Perspectives on Privacy Protection

With the rapid proliferation of digital healthcare and AI-driven medical technologies, the value of healthcare data has increased substantially. These technological advancements hold immense potential for delivering patient-centered precision care, designing evidence-based public health policies, and enabling predictive disease management and prevention. At the same time, the protection of personal health information has emerged as a core legal and ethical challenge, as improper use or disclosure of sensitive data can result in serious harms such as privacy violations, stigmatization, discrimination, and data misuse.

The U.S. experience demonstrates that privacy protection is not merely a matter of confidentiality but serves as the foundation for building trust in healthcare data systems. According to Kim (2021), U.S. laws particularly HIPAA, HITECH, and the 21st Century Cures Act constitute a multi-layered legal framework centered on patient rights [1]. These laws establish clear classifications of health information, mandate safeguards based on identifiability, prohibit information blocking, enforce interoperability, and institutionalize patients' rights to access, correct, delete, and withdraw consent regarding their personal data. Importantly, even de-identified data is subject to scrutiny, with limitations on use and disclosure based on scientifically grounded standards such as expert determination or the safe harbor method. Medical institutions are held legally accountable for ensuring these rights in practice.

In Korea, a similar framework has been developed through the Medical Service Act, the Personal Information Protection Act, and the Bioethics and Safety Act. Article 21 of the Medical Service Act requires the secure retention and confidentiality of patient records, prohibiting unauthorized disclosure by medical professionals. Article 28-2 of the Personal Information Protection Act permits the use of pseudonymized data without consent for public interest purposes such as statistical analysis and scientific research, provided that adequate safeguards are in place. Additionally, Article 16 of the Bioethics Act guarantees explicit consent and withdrawal rights for particularly sensitive data, such as genetic and life-related information.

Nonetheless, Korea's current privacy regime remains fragmented and does not fully ensure the data subject's practical control over personal information. For instance, patients are not always informed about when, or to whom, their data has been disclosed, and there is a lack of clear scientific criteria for assessing re-identification risk. Drawing on the U.S. model, Korea must move toward a legal framework that strengthens data subject rights and enhances public trust in privacy protections. Personal information protection should not be seen as an impediment to technological progress, but rather as a prerequisite for creating a trustworthy data ecosystem.

As in the United States, advancing healthcare data utilization in Korea requires reinforcing individual rights, establishing scientifically grounded de-identification standards, and increasing accountability among private enterprises and healthcare providers.

The Importance and Challenges of Personal Information Protection

As data-driven decision-making, precision medicine, and artificial intelligence continue to expand rapidly in the healthcare sector, the potential value of medical data is drawing increasing attention. These developments highlight the utility of data for advancing scientific research, improving care quality, and informing evidence-based policy. At the same time, the protection of personal health information is gaining prominence as a core condition for building the trust necessary to enable data utilization.

Kim (2021) emphasizes that the U.S. legal framework aims to prevent the ethical and social harms that may result from the misuse or unauthorized disclosure of sensitive health information [1]. Centered around the concept of protected health information (PHI), HIPAA, HITECH, and the 21st Century Cures Act collectively establish a complex regulatory architecture. This system guarantees data subject rights such as access, correction, and deletion prohibits information blocking, enforces stricter de-identification standards, and mediates the interplay between federal and state laws. Importantly, privacy protection in the U.S. is not merely defensive; it serves as a foundational enabler of data-driven healthcare innovation. High-quality data becomes accessible and usable only when patients trust that their information is being handled securely and ethically.

In Korea, patient privacy is protected under a dual framework consisting of the Personal Information Protection Act (as general law) and the Medical Service Act (as sector-specific law). According to Sung (2020), the Medical Service Act outlines several key protections [2]:

• Article 19 prohibits medical professionals from disclosing or improperly publishing information acquired in the course of treatment.

• Article 21-2 requires patient consent for the transmission of medical records and prohibits data leakage, alteration, or loss during electronic transfer.

• Article 23 bans unauthorized access, modification, or destruction of personal data stored in electronic medical records (EMRs).

• Articles 23-3 and 23-4 mandate notification to the Ministry of Health and Welfare and the implementation of response systems in the event of data breaches.

These provisions reflect the distinctive role of medical institutions in handling not only health data but also highly sensitive economic, social, and biometric information. However, the Medical Service Act has been criticized for its conceptual ambiguity. Terms such as "information," "personal information," and "medical information" are inconsistently used, and the scope of legal protections is often unclear. Additional challenges include the legal status of deceased patients' data, the permissible use of pseudonymized information, and the extent of healthcare providers' rights to supplement or interpret data issues that may lead to legal uncertainty [2].

Crucially, personal information protection should not be seen as a barrier to innovation, but rather as the legal infrastructure that enables responsible and sustainable data use. Medical institutions and regulatory authorities must go beyond statutory compliance and actively uphold core privacy principles, including data minimization, purpose specification, and security safeguards. As in the United States, Korea's privacy regime should evolve into a more coherent and transparent system one that robustly supports patients' right to informational self-determination while incorporating scientifically grounded standards for de-identification and access transparency.

Key Legal Issues in the Protection and Use of Healthcare Data in the United States

The United States has developed a complex legal framework to balance the protection and utilization of healthcare data. This system is designed to safeguard patient privacy while supporting secondary uses of data for research, policymaking, and technological innovation. Five key legal issues underpin this framework.

First, the definition and scope of Protected Health Information (PHI): Under HIPAA, PHI refers to individually identifiable health information created, collected, transmitted, or maintained by healthcare providers, insurers, or clearinghouses. It includes 18 identifiers such as name, date of birth, and medical record number and the line between PHI and non-PHI has become increasingly blurred due to digitization and the proliferation of data types [3].

Second, the risks of de-identification and re-identification: While HIPAA permits secondary use of data once it has been de-identified via either expert determination or the safe harbor method there remains a risk of re-identification when de-identified data is combined with other datasets. Addressing this risk remains a persistent legal challenge [4, 5].

Third, interoperability and the prohibition of information blocking: The 21st Century Cures Act promotes interoperability across health information systems and prohibits information blocking defined as unreasonably impeding data access or exchange. The 2020 ONC Final Rule further elaborates on these requirements, aiming to enhance patient access and enable seamless data sharing [6, 7].

Fourth, the use of broad consent for research purposes: The Revised Common Rule permits "broad consent" for future secondary use of identifiable biospecimens and data. If obtained at the time of initial consent, researchers are not required to seek re-consent for each subsequent study. However, questions remain about whether this approach adequately ensures participant understanding and the right to withdraw [8].

Fifth, conflicts between federal and state laws: HIPAA and HITECH serve as foundational federal laws, but

state-level legislation such as California's CCPA or New York's SHIELD Act may impose stricter protections. In cases of conflict, the law that offers greater rights to the patient takes precedence, adding to the legal compliance burden for healthcare organizations [9].

In conclusion, the U.S. legal landscape governing health data continues to evolve in response to tensions between privacy and innovation. As technological capabilities grow and public health demands intensify, legal systems must adapt to ensure the protection of individual rights while enabling the responsible and ethical use of data.

Legal Conflicts in Personal Information Governance in the Health and Life Sciences

The health and life sciences rely heavily on sensitive personal information such as medical records, genetic data, and human-derived materials for both research and clinical practice. Given the nature of this data, protecting personal information is not only a legal mandate but also an ethical imperative. However, as Jeong (2015) argues, Korea's current legal framework reveals significant conflicts between personal information protection principles and the public interest in biomedical data use [10].

First, tensions exist between open science mandates and privacy restrictions. The Framework Act on Science and Technology (Articles 11 and 26) and the Regulations on the Management of National R&D Projects require that government-funded research results be registered, disclosed, and disseminated. Similarly, the Act on the Acquisition, Management, and Utilization of Life Science Research Resources mandates the deposit and shared use of collected resources. Yet, Article 18 of the Personal Information Protection Act prohibits secondary use or third-party provision of personal data, with only narrow exceptions. As a result, researchers working on projects of public interest often face structural disincentives to access or share personal information.

Second, inconsistencies between the Bioethics Act and the Personal Information Protection Act create double regulation. Article 43(2) of the Bioethics and Safety Act prohibits the provision of human-derived materials containing identifiable information without explicit donor consent. In contrast, Article 18 of the Personal Information Protection Act allows the use of de-identified data for research purposes without consent. This creates a conflicting regulatory scenario in which, despite compliance with anonymization standards, researchers may still be required to obtain IRB approval or additional consent undermining research efficiency.

Third, practical ambiguities hinder implementation in the field. This legal discord often translates into confusion in research practice. For instance, projects such as the Clinical Omics Data Archive (CODA) are designed to integrate genomic and clinical data, yet researchers using personal identifiers like resident registration numbers face regulatory uncertainty under the Personal Information Protection Act. While submission of a research plan under the Bioethics Act is expected to suffice, IRB approval is frequently still required, creating administrative delays and barriers.

Fourth, legislative inconsistency undermines research autonomy. Jeong (2015) frames the issue not as a matter of legal interpretation but as a systemic lack of legislative coherence [10]. In many cases, multiple laws impose contradictory requirements for the same activity, leaving practitioners particularly those in public institutions or without legal expertise uncertain about compliance standards. This not only hampers research productivity but also discourages innovation in sensitive fields.

Addressing these challenges requires structural legal reform. First, legal terminology and interpretive standards must be harmonized across relevant statutes, including the Personal Information Protection Act, the Bioethics Act, and the Framework Act on Science and Technology. Second, a unified legal framework or comprehensive guideline should be established to govern research-related data use. Third, IRB exemption criteria must be clearly defined and enforced so that review is not mandated for truly anonymized data. Finally, legal education and administrative support systems must be strengthened to ensure that researchers have practical guidance and institutional backing.

Although the use of personal data is essential for biomedical advancement, Korea's fragmented and conflicting legal structure currently impedes the autonomy and efficiency of research environments. This is not simply a matter of interpretation but a fundamental issue of legal consistency and systemic design. Moving forward, legislators and policy leaders must work toward a coherent legal framework that integrates ethical safeguards with the practical needs of research.

The Future of Healthcare Data: Toward a Trust-Based Data Ecosystem

Healthcare data is rapidly emerging as a core asset for comprehensively understanding individual and population health. Drawn from a wide array of sources including clinical records, genetic data, lifestyle behaviors, and environmental exposures healthcare data serves as a foundation for innovation across multiple domains. These include improving the quality of care, enhancing public health outcomes, reducing healthcare expenditures, developing AI-driven medical technologies, and informing policy design. Practical applications such as precision medicine, predictive public health, chronic disease and elderly care management, and platform-based health services are gaining traction. Data generated by wearable devices and lifelogging technologies is already playing a central role in delivering preventive and personalized healthcare services. The integration of healthcare big data with artificial intelligence is also contributing to greater diagnostic accuracy, prescription standardization, and patient-centered care.

The research potential of healthcare big data is particularly significant in oncology, as the global burden of cancer continues to rise. Cancer, being a complex and multifactorial disease, demands the comprehensive analysis of longitudinal clinical data, genetic profiles,

treatment responses, and lifestyle factors. Healthcare big data enables large-scale, multidimensional analysis that supports research breakthroughs in early diagnosis, treatment response prediction, and survivorship care through precision medicine approaches. To realize these benefits, a carefully balanced legal framework must be in place one that enables secure linkage and use of cancer-related data while upholding privacy protections. Such infrastructure is critical to improving survival outcomes and the quality of life for cancer patients.

Despite this promise, the full potential of healthcare data remains underutilized due to regulatory tension with privacy protection. Korea's Personal Information Protection Act classifies health data as sensitive information and imposes strict limitations on its collection, processing, and linkage. Even after the introduction of the pseudonymized data framework, concerns about data merging and re-identification risks continue to restrict access and use. To address these barriers and unlock the future value of healthcare data, several policy responses are necessary.

First, data standardization and interoperability must be ensured through robust technical and institutional frameworks that support secure data exchange across organizations. Second, clear legal guidelines must be established regarding the use of pseudonymized data, particularly in defining consent exemptions for scientific research. Third, a trust framework must be developed to reinforce accountability among healthcare providers and data custodians while guaranteeing the rights of data subjects. Fourth, public–private data sharing systems must be strengthened to mitigate data monopolies and promote broader public benefit.

In conclusion, healthcare data is the fuel and infrastructure for next-generation health innovation. Yet without a legal and ethical system that enables responsible use grounded in public interest and safety this potential cannot be realized. Like crude oil, data is of little value unless refined; it is society's task to convert raw data into a public asset through the refining mechanism of trust.

Conclusion

Healthcare big data represents a critical asset poised to drive innovation across a wide range of domains, including precision medicine, public health, and biomedical research. However, its effective utilization is not without legal and ethical challenges particularly those concerning personal information protection. These challenges extend beyond technical considerations and demand systemic legal coherence and the construction of a trust-based governance framework. This paper has examined the structural limitations of Korea's legal landscape by comparing relevant statutes with those of the United States, highlighting the dual regulations and legal inconsistencies that currently impede the responsible and public-minded use of healthcare data. It argues for reforming the legal framework in a way that simultaneously safeguards individual rights and facilitates the ethical use of data. Key recommendations include clarifying the legal definitions of pseudonymized and anonymized data, introducing greater flexibility in consent mechanisms, establishing scientifically grounded de-identification standards to mitigate re-identification risk, and adopting technology-based security infrastructures. These measures are presented as both urgent and feasible steps toward regulatory improvement. Looking ahead, national healthcare data policy must aim to construct a sustainable and inclusive data ecosystem grounded in public trust. This requires a balanced perspective one that harmonizes the imperatives of data security with the social and scientific value of data utilization.

Author Contribution Statement

MJ wrote and revised the manuscript.

Acknowledgements

None.

Funding

This study was supported by the Dongduk Women's University grant (2025-06660).

Competing interests

I declare that I have no conflict of interest.

References

1. Kim JS. Legal issues in protecting and utilitizing medical data in the United states-focused on hipaa/hitech, 21st century cures act, common law, guidance. The Korean Society of Law and Medicine. 2021;22(4):117-57.
2. Sung SY. A Study on the Protection of Personal Information in the Medical Service Act. The Korean Society of Law and Medicine. 2020;21(2):75-103.
3. Zubrzycki CM. Privacy from Doctors. Yale Law & Policy Review, 2021; 39(2): 526–580.
4. Clayton EW, Evans BJ, Hazel JW, Rothstein MA. The law of genetic privacy: Applications, implications, and limitations. J Law Biosci. 2019;6(1):1-36. https://doi.org/10.1093/jlb/lsz007.
5. Enriquez-Sarano L. Data-rich and knowledge-poor. Columbia Law Review. 2020 Dec 1;120(8):2319-58.
6. Rowe EA. Sharing Data. Iowa Law Review, 2018; 104: 287–340.
7. Office of the National Coordinator for Health Information Technology. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program Proposed Rule, 2020.
8. Konnoth C, Scheffler G. Can electronic health records be saved? Am J Law Med. 2020;46(1):7-19. https://doi.org/10.1177/0098858820919552.
9. Edwards BN. The 21st Century Cures Act: A Patient's Miracle or Demise?. J. Nat'l Ass'n Admin. L. Judiciary. 2021;40:79.
10. Jeong CR. Contradictions in the application of biomedical and health law in South Korea. Korean Journal of Medical Ethics. 2015;18(4):407-23.