Opinion on the Implementation of the Act Partially Amending the Act on the Protection of Personal Information and Other Acts: Balancing the Protection and Use of Medical and Health Data Based on Scientific Evidence

July 10, 2026

Society for Health Data Science

The Society for Health Data Science aims to contribute to improving the health and welfare of the public through the advancement of scientific knowledge concerning the collection, management, analysis, and social implementation of health data.

In the use of medical and health data, it is important not only to appropriately protect the rights and interests of individuals but also to realize public benefits that are returned to society as a whole, including the prevention of disease, the development of new diagnostic and therapeutic methods, the improvement in the quality of health care, and the advancement of public health policy.

Medical and health data accumulated by health care institutions and other entities have the potential to contribute significantly to clinical research, the evaluation of health care quality, the safety assessment of pharmaceuticals and other products, the improvement of public health policy, and the development of new diagnostic and therapeutic technologies. Nevertheless, the appropriate use of such data has not progressed sufficiently. This is attributable not only to issues arising under the legal framework for the protection of personal information but also to multiple factors, including differences in understanding between those who hold and manage data and those who use them regarding purposes, necessity, risks, and responsibilities; insufficient systems for collaboration across organizations; inadequate data infrastructure, including data standardization and quality management; and the lack of clarity regarding available legal frameworks and procedures.

To promote the appropriate use of medical and health data throughout society, it is essential that health care institutions, government agencies, universities, research institutions, private-sector entities, and other relevant parties develop a common understanding of the purposes of use, the nature of the data, the required data quality, security management, the allocation of roles, and the attribution of responsibilities and that they collaborate with one another. Although the present amendment alone will not resolve all of these issues, the implementation of the Amendment Act should serve as an opportunity to clarify the lawful and trustworthy pathways for using medical and health data, the safeguards to be applied, and the responsibilities of the relevant parties while appropriately protecting the rights and interests of data subjects.

This opinion is submitted for the purpose of ensuring that the characteristics of medical and health data and relevant scientific knowledge are appropriately reflected in the Rules of the Personal Information Protection Commission, guidelines, and other implementation measures to be developed in light of the purpose of the recently enacted Amendment Act. Without favoring the position of any particular organization, industry, institutional framework, or technology, the Society for Health Data Science presents the following opinion on the implementation of the Amendment Act from the standpoint of balancing the protection of personal information with the appropriate use of data on the basis of scientific evidence and professional expertise.

1. Basic Direction of the Amendment Act

We consider that there is a certain degree of reasonableness in the direction taken by the Amendment Act, which seeks to reconsider the role of data subject consent and to enable the appropriate use of data for the "generation of statistics and related activities," in which the impact on the rights and interests of data subjects is limited.

In applying these provisions, however, it should be a prerequisite that the necessity of handling personal data in identifiable form during the generation of statistics and related activities be clearly demonstrated. In particular, the scope and duration of such handling should be limited to what is strictly necessary, and effective measures should be implemented to prevent statistical outputs from being repurposed for adverse assessments of, or decisions concerning, individuals. It is also important, to maintain trust in data use, to establish requirements such as limiting the use of data to the generation of statistics and related activities, public disclosure of such use, written agreements between data providers and data recipients, and restrictions on use for purposes other than the stated purpose and onward provision.

Medical and health data, on the other hand, include highly sensitive information, such as medical histories, and the associated risks vary considerably depending on factors such as the scale, granularity, rarity, and environment of use of the data. Fundamental obligations, prohibitions, and restrictions on rights that may have a significant impact on the rights and interests of data subjects should not be left solely to guidelines. Moreover, rather than relying exclusively on uniform statutory provisions, the Rules of the Personal Information Protection Commission and relevant guidelines should specify concrete safeguards appropriate to the characteristics of the data.

2. Promote Diverse Forms of Data Analysis and Model Development According to Their Purposes and Establish a Clear Framework According to the Nature of the Outputs and the Stage of Use

It is important for the Amendment Act to enable the appropriate use of medical and health data for data analysis and model development that falls within the scope of the generation of statistics and related activities, where the impact on the rights and interests of data subjects is limited. Such use is necessary for promoting the early detection of disease, the development of diagnostic and therapeutic methods, disease prevention, and improvements in health care quality, public health, and other forms of medical innovation.

Methods used to analyze such data include biostatistics, epidemiology, operations research, mathematical modeling, simulation, machine learning, and artificial intelligence, and these methods should not be restricted in advance. Rather, appropriate methods should be selected and, where necessary, combined, considering the purpose of the analysis, the nature of the data, scientific validity, and feasibility of implementation.

In other countries, infrastructure for accessing such data, secure analytical environments, regulatory sandboxes, and systems for validation in real-world settings are being developed. If Japan uniformly restricts data use and model development on the grounds of risk, this may not only reduce Japan's international competitiveness in research and development and medical innovation but also compel Japan to use technologies and models developed overseas that have not been adequately validated with respect to the characteristics of the Japanese population or the realities of health care in Japan.

Therefore, for data analysis and model development falling within the scope of the generation of statistics and related activities, where the impact on the rights and interests of data subjects is limited, it is necessary to promptly establish an environment in which permissible uses and required management measures are clearly defined, so that researchers, health care institutions, and private-sector entities can engage in research and development with a reasonable degree of predictability.

The Act on the Protection of Personal Information should not be designed as a framework that obstructs appropriate data analysis or model development. Rather, it should serve as a foundation that enables beneficial data use and medical innovation while maintaining public trust.

On the other hand, when statistical outputs or developed models are applied to specific individuals and used for clinical care, assessment, prediction, intervention, advertising, solicitation, or other decisions or actions directed at those individuals, the impact on their rights and interests may differ from that arising during the generation of statistics and related

activities. In particular, where artificial intelligence or other methods are used to infer health conditions, disease risks, psychological states, or other characteristics that have not been explicitly provided by the data subject, appropriate measures should be considered, including transparency, accountability, human oversight, logging of data use, and ongoing performance and impact assessments. Such measures should consider the entity using the information, the purpose of use, the information being inferred, the method of intervention, and the benefits and disadvantages that may arise for the data subject and for society.

3. Clearly Distinguish the Generation of Statistics and Related Activities from Uses That Affect Individuals

It should be clearly explained to the public and to entities handling data that what may fall within the scope of the generation of statistics and related activities under the Amendment Act is not data analysis or model development in general using artificial intelligence or other analytical methods but statistical use in which the resulting outputs are not used for decisions or actions directed at specific individuals.

In applying the Amendment Act, whether an activity falls within the scope of the generation of statistics and related activities should not be determined solely on the basis of the name of the method employed, such as artificial intelligence, machine learning, or statistical analysis. When the outputs generated are used for clinical care, assessment, prediction, intervention, insurance benefits, or other decision-making concerning specific individuals, such uses should be distinguished from statistical uses that do not directly affect individuals. The appropriate form of data subject involvement, transparency, fairness, and safety should then be considered separately.

Data analysis, model training, performance evaluation, further training, deployment in real-world settings, and application to specific individuals may each have different purposes and different effects on the rights and interests of data subjects. Therefore, whether an activity falls within the scope of the generation of statistics and related activities should not be determined collectively for the entire sequence of activities involved in developing an artificial intelligence system or model. Instead, the determination should be made in accordance with the handling of data, the outputs generated, and the manner in which those outputs are used at each stage.

Where a developed model or the right to use such a model is provided to a third party, consideration should be given not only to the purpose for which the model was developed but also to the uses that may reasonably be anticipated after its provision. Where the model may be used for clinical care, assessment, prediction, screening or selection, or other decision-

making concerning specific individuals, and such use cannot be effectively restricted through contracts, technical controls, audits, or other measures, the activity should not uniformly be regarded as one having only a limited impact on individuals merely because statistical outputs were generated during the developmental stage.
Separately from the generation of statistics and related activities, where it is determined that there are reasonable grounds for not obtaining the data subject's consent on the basis of public health necessity, the public interest involved should not be recognized in an abstract manner. The determination should consider both the public benefit expected from the use of the data and the impact on the rights and interests of data subjects, considering the nature of the data, the necessity and urgency of the use, the availability of alternative means, possible disadvantages to data subjects, and the safeguards to be implemented.

4. Require Additional Measures for Medical and Health Data According to the Risk Level

When medical and health data are used, uniform measures should not be required for all entities. Rather, safeguards appropriate to the level of risk should be selected on the basis of a comprehensive evaluation of the following factors:

- The purpose of use and the manner in which the outputs generated will be used;
- The entity using the data, the users, the data recipients, and the scope of use;
- The scale and granularity of the sensitive personal information being handled;
- Whether the data include information that may facilitate the identification of individuals or the inference of sensitive attributes, such as information concerning rare diseases or genetic information;
- Whether the data include free-text entries, medical images, genomic information, detailed location data, or similar information;
- The potential for linkage with external data;
- The possibility that information concerning individuals may be inferred from statistical outputs or trained models;
- The data retention period; and
- The potential impact if the rights and interests of individuals are violated.

Where personal data in identifiable form are handled for the generation of statistics and related activities, the necessity of such handling should be clearly demonstrated in light of the purpose of use and the analytical process. The scope of handling should be limited to the data elements, data subjects, duration, and users necessary to achieve the purpose.

When the use of identifiers is unnecessary, or once the need to use them has ceased, the scope of data handled in identifiable form, the users who may access those data, and the duration of such handling should be limited to what is strictly necessary, in accordance with the principle of data minimization. Measures may include pseudonymization, separate management of identifiers, reduction of data elements, aggregation, and other methods. Moreover, a particular processing method should not be uniformly mandated. It should remain possible to select measures appropriate to the purpose and level of risk, including handling necessary for data linkage, longitudinal follow-up, and quality management.

In the medical and health fields, it may be necessary to handle identifying information for a certain period to generate accurate statistics or conduct epidemiological follow-up. Examples include eliminating duplicate registrations in the national cancer registry, matching registry data with death information, and ascertaining long-term outcomes. Uniform anonymization before the commencement of data use may make longitudinal follow-up, data linkage, deduplication, and quality management difficult, thereby impairing the accuracy of statistics and the public benefits derived from them.

Therefore, whether personal data in identifiable form may be used should not be determined solely on the basis of whether data subject consent has been obtained or whether a particular processing method has been adopted. Rather, the determination should be based on the overall governance framework, including the necessity of the use, the availability of alternative means, the scope and duration of handling, the entity using the data, restrictions on use for purposes other than the stated purpose and on onward provision, security management, and the deletion of data or reduction of identifiability after use.

Where appropriate, a combination of measures should be required, including secure analytical environments, user authentication, access controls, audit logging, output disclosure review, prohibitions on reidentification, periodic audits, and incident response procedures. In considering such measures, it is important to clearly identify the expected benefits of data use for the health and welfare of the public, appropriately assess risks to the rights and interests of individuals, and implement effective safeguards that consider both the significance of the use and the associated risks.

With respect to statistical outputs, trained models, and similar products, the risk that personal information contained in the data used for analysis or training may be inferred should be assessed separately from the benefits and disadvantages arising from applying those outputs or models to individuals. Such benefits may include clinical or public health benefits, whereas disadvantages may include adverse assessment, screening or selection, or discriminatory treatment.

Even when individuals are not directly identified, risk assessments should also consider the

possibility that statistical outputs may result in adverse assessments, prejudice, or discrimination against particular patient populations, geographic communities, or groups defined by particular characteristics.

Even when publicly available sensitive personal information is obtained for the generation of statistics and related activities, the mere fact that the information has been made public should not be regarded as demonstrating that the data subject reasonably anticipates all forms of collection, analysis, and secondary use. Appropriate safeguards should be applied, including an assessment of the necessity of collection, data minimization, and transparency, considering the context in which the information was made public, the sensitivity of the information collected, the scale of collection, the potential for linkage with other information, the purpose of use, and the impact on the data subject.

5. Design Safeguards Systematically and According to Their Purposes

The protection of medical and health data may involve a variety of measures, including ensuring the appropriateness of the purpose of use and scope of handling, managing access to data, transforming information, controlling the environment in which data are used, conducting audits, imposing contractual restrictions, and using privacy-enhancing technologies.

These measures address different risks and differ in their effects and conditions of application. No single measure can eliminate all risks.

Laws and regulations, the Rules of the Personal Information Protection Commission, and relevant guidelines should organize and clarify, on a scientific basis, the purpose and role of each measure. They should require an appropriate combination of legal, organizational, personnel, and technical measures according to the nature of the data, the purpose of use, the entity using the data, the environment of use, and the potential impact on individuals.

It is also important not only to formally determine whether a particular measure has been adopted but also to continuously assess the extent to which the anticipated risks have actually been reduced and what residual risks remain. Safeguards should be reviewed in response to changes in technology and in the manner in which data are used.

6. Ensure Transparency and Ongoing Review

To obtain public trust, it is necessary not only to formally disclose the names of the entities using the data and their purposes of use but also to explain clearly to the public what data are being used, by whom, for what purposes, and under what safeguards.

With respect to information that must be publicly disclosed concerning the generation of statistics and related activities, we request the establishment of a mechanism under which, in addition to disclosure by each entity using the data, the Personal Information Protection Commission or another appropriate body aggregates information in a standardized format. Such a mechanism should enable members of the public to easily search for and confirm the entity using the data, the purpose of use, the data concerned, the source from which the data were obtained or the recipient to which they were provided, and the safeguards applied.

Even after the relevant framework has been implemented, there should be a mechanism for continuously reviewing actual uses, incidents, audit results, technological advances, and developments in international legal and regulatory frameworks and for periodically revising the Rules of the Personal Information Protection Commission and relevant guidelines.

Such reviews should involve diverse stakeholders, including experts in law, ethics, medicine, information science, and statistics, as well as patients and members of the public. Discussions should be based on scientific evidence, and the review process should be broadly disclosed to society.

Where appropriate, it is also important to establish contact points and procedures through which members of the public may make inquiries concerning the use of data and raise concerns or objections. Where procedures for the cessation of use or similar measures are established, their scope and effect should be clearly defined, considering their relationship with statistical information that has already been generated, the integrity and completeness of research, and public health necessity.

The fact that data subject consent is not obtained does not mean that transparency or opportunities for data subjects to make inquiries, request corrections, lodge complaints, request appropriate handling, or otherwise become involved, should be uniformly denied. Effective forms of data subject involvement should be established while maintaining consistency with the nature of statistical information and the integrity and completeness of research.

7. Establish a Technology-Neutral Framework that is Capable of Accommodating Future Developments

The legal framework and relevant guidelines should not be based on particular products, technologies, or centralized data infrastructures. Rather, they should be capable of continuously accommodating technological advances.

In Japan and other countries, diverse forms of data use are being developed and implemented, including distributed analytics that do not require data to be collected in a single location,

federated learning, secure computation, trusted research environments such as data enclaves, and data spaces.

To ensure that research and health care in Japan are not left behind by international developments, the framework should be technology-neutral and capable of accommodating future developments. It should enable centralized, distributed, and other architectures to be appropriately selected according to the purpose of use and level of risk and combined where necessary.

8. Ensure Multiple Lawful Pathways for Data Use

Medical and health data are used for diverse purposes, including clinical care, clinical research, epidemiological research, quality management, safety assessment of pharmaceuticals and other products, the generation of statistics, and artificial intelligence development.

Because medical and health data have characteristics that cannot be adequately addressed by general laws alone, rules concerning the protection of rights, governance, and data use that are appropriate for the characteristics of medical and health data should be developed systematically while ensuring consistency with existing frameworks.

In doing so, data use should not be consolidated into a single institutional framework. Rather, mechanisms with different legal bases, conditions for application, and governance arrangements should be used in a complementary manner. These include the Act on the Protection of Personal Information, the Ethical Guidelines for Life Science and Medical Research Involving Human Subjects, entrustment, joint use, pseudonymized personal information, anonymized personal information, provisions concerning academic research, and the Next-Generation Medical Infrastructure Act.

Universities, health care institutions, research institutions, private-sector entities, and other organizations should be able to select appropriate institutional frameworks and safeguards according to the purpose of use, the nature of the data, the impact on individuals, the manner in which data are managed and distributed, and the organizational system for allocating responsibility.

With respect to the distinctions among entrustment, joint use, provision to third parties, and other arrangements, the roles and responsibilities of each party should be clarified not merely on the basis of the contractual name assigned to the arrangement but also on the basis of who substantively determines the purpose of use and the manner in which the data are handled.

When mechanisms are considered for ensuring the eligibility and trustworthiness of entities using data, a single method, such as certification, should not be uniformly required. Instead, a flexible institutional design should be adopted that enables the substantive management

capabilities of an entity to be evaluated according to the risks associated with the data handled and the manner of use. Such an evaluation should be able to take into account existing governance mechanisms, including institutional management, ethics review, information security management, and audits.

Conclusion

The protection and use of medical and health data are not necessarily conflicting objectives. Through appropriate governance and scientifically grounded risk management, it is possible to protect the dignity, rights, and interests of individuals while returning the knowledge obtained from data to the health and welfare of the public.

In light of the purpose of the Amendment Act, the Society for Health Data Science requests that risk-based safeguards appropriate to the characteristics of medical and health data, technology-neutral institutional design, multiple lawful pathways for data use, and transparent governance be given concrete form through an appropriate allocation of roles among laws and regulations, the Rules of the Personal Information Protection Commission, and relevant guidelines.