Using medical big data for clinical research and legal considerations for the protection of personal information: the double-edged sword

Raeun Kim¹, Jiwon Shinn², Hun-Sung Kim²,³

¹Department of Convergence Public Administration, Hansung University Mirae Plus College, Seoul, Korea
²Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Korea
³Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea

The advent of medical big data has increased the scope of the clinical use of such data; however, these data have raised serious concerns regarding personal privacy protection, which hinders their usage. For instance, as the pseudonymization or anonymization of data increases, the quality of its clinical use decreases. Thus, a balanced approach is required to maximize clinical data use while protecting personal information as much as possible. However, Korea’s existing laws mandate several kinds of consent; soliciting some of these types of consent can be cumbersome. Moreover, while the collection of medical data by hospitals requires considerable time and money, its ownership is difficult to ascertain. To bridge the enormous gap between the protection of personal information and the use of clinical data, the European Union and countries such as Finland have already proposed various modes of guaranteeing the free movement of personal information that simultaneously strengthen people’s personal rights. Similarly, Korea has initiated the MyData Service, although it faces several limitations. Therefore, this study reviews Korea’s current healthcare big data system, the laws governing data sharing and usage, and compares them with similar laws enacted by the European Union and Finland. It then provides future direction for Korea’s personal information protection legislation. Ultimately, governments must expand and elaborate upon the scope and content of personal information protection laws to enable the development of healthcare and other industries without sacrificing either personal information protection or clinical use of medical data.

Keywords: Big data; Delivery of health care; Privacy

INTRODUCTION

Medical services based on digital technologies, such as big data, real-world data, and artificial intelligence, are becoming increasingly popular [1,2]. In the field of digital-based medicine, the emergence of personalized medical services employing big data has led most countries to create big-data platforms for healthcare, which can be used across the industry [3]. Amid increasing use of big data across domains, the European Union (EU) and countries such as Finland have already enacted laws to protect people’s personal information. The same is true for Korea; yet Korea’s policies for protecting personal information have obstructed data linkages between healthcare institutions or hospitals. Stud-
ies have shown that the legal basis for utilizing data in Korea is insufficient, which has led to a failure in meeting industry demands. In 2018, a big-data utilization platform for healthcare was launched in Korea for linking medical data from the National Health Insurance Service (NHIS), the Health Insurance Review and Assessment Service (HIRA), the Korea Disease Control and Prevention Agency (KDCA), and the National Cancer Center. In addition, “three data-related bills (the Personal Information Protection Act [PIPA], Act on Promotion of Information and Communications Network Utilization and Information Protection, and Credit Information Use and Protection Act)” were amended [4–8]. During this process, a special provision for processing pseudonymous information was made, for purposes of statistical writing, scientific research, preservation of public interest records, and establishing the legal basis for instituting a personal information manager (or an honest broker) to utilize pseudonymized personal information without requiring consent [4,7].

Recently, the Korean government established a close public-private cooperation system, with the larger goal of implementing the “My HealthWay System” [9]. While previously, only financial data were digitalized, the establishment of the “My HealthWay System” enables the integration of various types of citizen’s data, such as insurance information and credit scores, under the government’s My Data Service. These changes imply that Korea’s personal information protection system, which had been quite conservative in disclosing medical data, is approaching the assumed global standard. However, compared to other major countries, Korea’s current personal information protection laws continue to block the availability of information to a large extent. This paper reviews the current state of Korea’s healthcare big data and related laws, and compares it with the EU’s General Data Protection Regulation (GDPR) [10], and Finland’s Act on the Secondary Use of Health and Social Data [11], to suggest future direction for Korea’s personal information protection legislation.

**CONFLICT BETWEEN PROTECTION OF PRIVACY AND PUBLIC INTEREST**

**Sensitive information**

Sensitive information refers to information regarding an individual’s ideology, beliefs, labor-union affiliation, political party membership, political opinions, health status, and other kinds of personal information that may significantly infringe on their privacy [12,13]. According to Article 15(1) of the PIPA [4], an information holder can process another person’s general personal information after obtaining their consent; however, sensitive information closely concerns an individual’s most private aspects. Therefore, for processing sensitive information, Article 23(1) of the same Act [4] mandates seeking consent from the subject of the information, in addition to seeking their consent for processing personal information (i.e., double consent is required). Ultimately, the person concerned must agree to provide their personal information. Considering that simply knowing which disease an individual is afflicted with could reveal their physical or mental inadequacies, medical history is part of sensitive information, as it is closely tied to one’s core personality. Other related aspects that require protection include information about who, when, where, what kind of treatment, and the amount of money spent. For example, in the case of hospitals specializing in gynecology, obstetrics, urology, psychiatry, etc., one’s possible ailment can be predicted simply by the medical institution’s name; combining this with information about medical-care benefits, number of days of hospitalization and visits, and self-payments, one could even infer the individual’s specific physical and mental limitations.

**Protection of personal information over public interest**

The aforementioned context is associated with many legal cases in which judges have set precedents to deal with the contradictions between public interest and protection of personal information [4,14,15]. In the scenario presented in Table 1 [14], the NHIS provided sensitive information of wanted criminals to the chief-of-police that indicated the claimant’s type of disease and health status. Considering that sensitive information constitutes the core of an individual’s personality and privacy, the judges concluded that the claimant’s right to self-determination of personal information had been violated, as the act of providing information was very serious. Therefore, the Constitutional Court ruled that the actions of the NHIS and the chief-of-police violated the individual’s right to self-determination regarding information.
Most Korean citizens are enrolled in the NHIS, which contains a vast amount of health information included in the National Health Insurance Corporation [16,17]. The NHIS’s information about medical-care benefits includes not only individual details but also sensitive information; together, they constitute comprehensive and integrated information about a person’s health. The subject of information must provide consent for processing such sensitive information, notwithstanding the provisions of Article 18(2), seventh subparagraph of the PIPA [4].

FACTORS IMPEDING THE USE OF PERSONAL INFORMATION

Restrictions to processing sensitive information: pseudonymized and anonymized information

In principle, a safe pseudonymization method enables the use of sensitive information as pseudonymized information [18]. However, unless existing technology provides a perfectly safe pseudonymization method, an individual’s personal information cannot be processed without obtaining their consent. Notably, most of the discussions regarding the processing of personal and sensitive information occurred in the past. The crux of the problem related to processing sensitive information is that the availability of data decreases with the increasing legal requirements for pseudonymization [19]. This is clearly highlighted in the provisions made through Article 3(7) of the PIPA [4]. If personal information can be either anonymously or pseudonymously processed by a personal information manager, it is ideal to stick to the former, as the latter presents several difficulties.

Unrealistic consent requirements of certain laws

Existing laws mandate seeking consent several times, which is another factor hindering the use of personal information. For example, according to Article 15(1) of the PIPA [4], a personal information manager can collect personal information only when the subject of information has provided consent. In addition, according to Article 17(1) of the same
Act [4], the collected personal information can be provided to a third party that solely aims to collect this information (and not use it), only after the individual’s consent has been obtained. Again, according to Article 18(1) of the same Act [4], consent must be obtained from the subject of information when the information is used beyond the scope of the original purpose and provided to a third party. The conditions for consent are also quite strict. According to Article 22 of the same Act [4], personal information managers must clearly recognize certain key points, such as whether personal information has been collected, the purpose of its use, the items to be collected and used, and the retention period. They are responsible for informing the subjects and seeking their consent. The fact that it is a legal regulatory provision to protect data subjects’ self-determination rights, makes it reasonable at first glance, but further examination clearly shows that seeking “consent” to protect data subjects’ rights and interests blocks the possible use of this data to a large extent.

Despite these constraints, the law does stipulate the possible uses of data, other than the case of Article 17(1) second subparagraph of the PIPA [4] mentioned above. The law also regulates possible use, and restrictions, other than those mentioned in Article 18 of the same Act [4]. This creates problems because, apart from the fact that collecting consent at each stage is quite cumbersome, the actual act of collecting information takes so long that the subject of the information cannot often be contacted for seeking additional consent for the subsequent stages. Because of these limitations, the demand to amend the relevant provisions for preventing the loss of valuable medical information resources is gaining strength. Indeed, according to Article 17(4) of the PIPA [4], if a personal information manager deems that there shall be no disadvantage to the data subject within the scope of the original purpose of collection, and necessary safety measures such as encryption have been taken, as prescribed by the Presidential Decree, personal information may be collected without the data subject’s consent. However, the reasonable circumstances in which personal medical information may be collected are unambiguous or not easy to define, because there is no unified view or guidelines about what is reasonably related. Whether something is reasonable must ultimately be determined by the court on a case-by-case basis. Consequently, personal information managers have no choice but to conform to requirements for seeking “consent,” to avoid legal responsibility. Thus, a lot of medical information is going in vain, without being utilized.

**COMPARISONS AND DIFFERENCES WITH FOREIGN CASES**

**GDPR of the EU**

With the EU’s formation as a single supranational entity, the economic and social functions of its member states were rapidly integrated [20]. In this process, people’s personal information was extensively exchanged, which led the EU to enact the GDPR as a unified guideline for processing personal information [21]. The GDPR inherits and develops the basic spirit of the previous Data Protection Directive (DPD), for protecting and standardizing personal information about the citizens of the member countries [21]. The GDPR establishes standard guidelines for the member states, which in turn have particular legal frameworks for protecting personal data [21,22].

The GDPR protects personal information while allowing for its free movement within the EU [21]. One of the GDPR’s characteristics, which acted as the basis of Korea’s “MyData Project” [23] is a substantial strengthening of the data subjects’ rights. The GDPR not only incorporates individuals’ rights to receive (that existed under the DPD), access, and correct information, but also strengthened other rights, such as the right to delete information (the so-called right to be forgotten) [24]. In addition, by newly establishing rights to restrict information processing, enable the portability of personal information, and object, it formed a structure to ensure that only information based on the data subject’s right to choose could circulate in the market [25]. According to the GDPR, private companies, including public institutions, are obliged to move personal data to other locations, and correct or delete existing information, if the data subject so desires (right to data portability, right to correct/delete data) [26].

**Finland’s Act on the Secondary Use of Health and Social Data**

In 2019, the Finnish Parliament passed the Act on the Secondary Use of Health and Social Data [11], which introduced strict data-security requirements and authorization proce-
The Act was based on the GDPR’s basic spirit, which stipulates that personal information can be secondarily used for preserving public records, and other scientific, historical, and statistical purposes. However, this Act goes a step further than the GDPR, by allowing the collected medical information to be used for development and innovation activities. Therefore, in Finland, private research institutes and companies can use public data for industrial purposes [27]. This Act consolidates fragmented national rules for the use of health and social security data. In addition, it was considered key for reviving the Finnish healthcare industry, as it fostered opportunities for research and innovation across the fields of health, welfare, disease prevention, novel treatment methods, and predictable personalized medicine.

However, in Finland, most of the human governance of the data permission organization is conducted by representatives of government departments, public institutions, and local governments, and only one private welfare and health service provider has been included as mentioned in section 8, first paragraph of the Act [11]. Nevertheless, the Act does allow the possibility of establishing an expert group in accordance with section 8, fourth paragraph, to receive assistance in preparing guidelines for anonymity, data protection, and security for business processing.

**Korea’s MyData Service**

With the amendment to Korea’s Credit and Information Use and Protection Act [6] in 2020, Article 33(2), the “personal credit information management business” was established to realize the right to request transmission of personal credit information, to introduce the “My Data Industry” in the financial field. However, MyData is not based on the PIPA, but the Credit Information Use and Protection Act, which applies only to credit information [28]. In this respect, the discussion about the right to self-determination of personal information is still in its infancy. Moreover, by assuming that the subject of information is the owner of data, the position of the institution or hospital remains ambiguous. Thus, it is difficult to expect that hospitals would voluntarily cooperate in responding to each data subject’s request, as it requires considerable time and money, in addition to depriving them of depriving them of the right to control their own data. After all, the current legislation does not enable a standardized and integrated way of providing information.

Personal information can only be transferred from one medical institution to another. A truly national system of maintaining health records can be implemented only when the relevant laws and regulations are revised.

**IMPLEMENTING THE MYDATA SERVICE IN THE MEDICAL FIELD**

The Korean Ministry of Health and Welfare has launched a pilot project in 2022. It has recruited around 400 patients from Seoul St. Mary’s Hospital (Seoul, Korea) and the Pusan National University Hospital (Busan, Korea), as participants for the MyData Service [23]. The participants were required to use a mobile application for requesting, receiving, and viewing their medical records. A quick glance would show how their data was being used.

**Significance of the current MyData Service**

The launch of the MyData pilot project in the medical field can be interpreted as acceptance of the basic idea of the GDPR, the EU’s personal information protection law. Korea had regarded subjects of information as passively “consenting” while providing information. By contrast, the MyData Service in the medical field assumes the subject of information as an active agent, and grants them the right to self-determination. It allows individuals to decide the amount and the type of information they wish to provide and goes beyond simply guaranteeing the right to decide whether to provide information or not.

Depending on the specifics of the case, the subject of information may withdraw the information provided, or request it to be provided to a third party. This enables them to actively decide on all aspects related to their personal information [9]. Individuals can choose how to move their personal information. This problem-solving method harmonizes the two incompatible goals of “protection of privacy” and “free distribution of data,” which were considered sensitive for personal information protection. This change is quite forward-looking, in that Korea’s information protection policy is converging with the assumed global standard.

**Limitations of the current MyData Service**

Scholars have pointed out that related laws and regulations
are still unable to keep pace with technological development. For example, to establish an exclusively medical MyData Service, private institutions must be able to utilize the data collected by medical institutions. In this case, they would naturally need to have the authority to process the personally identifiable information, including the data subject’s resident registration number [27]. However, so far, the current legislation does not specify institutions’ legal authority to use medical data. In addition, existing laws (PIPA, Enforcement Decree of Bioethics and Safety Act, Medical Service Act, National Health Insurance Act, etc.) for regulating personal information are ultimately intended to suppress its use. Moreover, they do not provide a unified legal guideline, because they were either enacted or revised according to the needs of the time. Finally, the ambiguous definition of “consent” must be improved. As per existing legislations, only a simplistic service for receiving and utilizing medical information for administrative convenience can be implemented. Without revising the relevant laws and regulations, it shall be impossible to create services that can utilize information from medical and other public institutions, as well as personal health information.

NECESSITY OF CHANGING THE CONCEPT OF PUBLIC INTEREST AND GOOD FOR PERSONAL INFORMATION PROTECTION: THE PRINCIPLE OF PROPORTIONALITY AS A STANDARD FOR PERSONAL INFORMATION PROTECTION

The principle of proportionality originally originated in French and German laws, but over time, it become increasingly important in many legal systems, including the EU laws [29]. The principle of proportionality is derived from the general principle of public law, which weighs measures in favor of public interest against the resulting damages to private interests and fundamental rights [30]. Therefore, administrative measures must be appropriate and necessary for enforcing public interest while balancing any public interest intervention with private interests [21]. Therefore, the purpose and the effect of the intervention must be proportionate. For example, measures that are favorable to public interest and strongly interfere with private freedoms should be abandoned. Proportionality requires that where several possible measures are available, measures that cause milder interference to private freedoms should be preferred. The GDPR emphasizes that personal data processing must be designed to be in the public interest [10]. An example is that the right to personal data protection is not an absolute right and must be balanced with other fundamental rights based on the principle of proportionality.

The same principle implies that certain categories of data processing may be necessary for protecting public interest in the field of public health without requiring the consent of the data subject [10]. However, the EU’s GDPR only strengthens the basic rights of the data subjects in some situations [31]. Interestingly, Finland’s secondary health-use law [11] is the only one that applies the principle of proportionality. Thus, countries should strive to balance the protection of personal data and privacy with economic interests.

Until now, the state was seen as a protector of personal information that prevented data leaks in public interest. However, governments must explore the possibility of an environment wherein corporate autonomy and creativity are respected, and data subjects are provided enhanced options. Greater availability of information would lead to the emergence of new industries, improving the growth potential of existing companies, ultimately creating quality jobs. In the current age of low fertility and an aging population, encouraging data use and corresponding healthcare innovations could be an effective solution for saving the human kind from diseases and aging. To this end, the concept of public interest and good, as we know it, must be redefined and the public discourse should focus on the issue of what protects public interest.

CONCLUSIONS: NECESSITY OF CREATING A GOVERNANCE STRUCTURE FOR HEALTHCARE DATA

In the era of digital transformation, the demand for innovation, problem solving, and value creation using medical data is increasing [1]. The basic rights of data subjects are still core values that cannot be waived off. Therefore, despite the changing times, thinking about ways to protect the rights of data subjects and establishing a safe transaction environment will remain the biggest task for legislators. While avoiding unnecessary or overly restrictive regulations, legislators must strive to protect the core values of the community.

The structured generation of knowledge through the
formation of fair processes also requires a well-formed decision-making body. In a highly developed information society, decision-makers need a fair and rational governance processes, irrespective of whether it is an official or private organization [32]. Based on the examples of the EU and Finland, establishing data governance that facilitates the smooth sharing and utilization of healthcare data can be an effective way to facilitate the reuse and sharing of data. High-level administration is increasingly dependent on the participation of private regulators in carrying out the state’s takes. Therefore, a consensual arrangement between the state and private actors must be established in many areas [32]. The concept of "governance" is as old as human civilization. It involves the processes of decision-making and implementation. Governance analysis focuses on formal and informal actors involved in decision-making, and formal and informal structures for reaching and executing decisions [33]. To arrive at a more balanced legislation, businesses, experts, civic groups, consumers, governments, and public institutions should engage in dialogue and invite diverse opinions, and the impact of data reuse on data subjects should be thoroughly analyzed. This is because the legitimacy, reliability, and acceptability of decisions will improve only when the knowledge, experience, and wisdom created by collective intelligence are actively respected and utilized.

**ARTICLE INFORMATION**

**Conflicts of interest**
The authors have no conflicts of interest to declare.

**Funding**
This study was supported by the Medical AI Education and Overseas Expansion Support Research Program through the National IT Industry Promotion Agency (NIPA) funded by the Korean Ministry of Science, ICT and Future Planning.

**Author contributions**
Conceptualization: RK and HSK; Data curation: all authors; Formal analysis: all authors; Funding acquisition: HSK; Investigation: HSK; Methodology: HSK; Project administration: HSK; Resources: HSK; Software: HSK; Supervision: HSK; Validation: HSK; Visualization: RK; Writing—original draft: RK; Writing—review & editing: HSK. All authors read and approved the final manuscript.

**ORCID**
Raeun Kim, https://orcid.org/0000-0003-1245-7289
Jiwon Shinn, https://orcid.org/0000-0002-9131-6128
Hun-Sung Kim, https://orcid.org/0000-0002-7002-7300

**REFERENCES**

9. Korean Ministry of Health and Welfare. Press release: the launch of the My HealthWay (a digital highway) [Internet]. Korean Min-


16. National Health Insurance Sharing Service (NHIS). Data provision guide [Internet]. NHIS of Korea; c2019 [cited 2023 May 17]. Available from: https://nhiss.nihs.or.kr/bd/ab/bdabat001cv.do?jsessionid=s21o35g1UT7BIU5ahIl3iFEUFH10e4zlLiezGMJqja00jmpbG3jv7SmEFN9Xmp1.primrose22_servlet_engine10


31. Molnar-Gabor F, Sellner J, Pagil S, Slokenberga S; Tzortzatou-Nanopoulou G, Nystrom K. Harmonization after the GDPR?: divergences in the rules for genetic and health data sharing in four member states and ways to overcome them by EU measures: insights from Germany, Greece, Latvia and Swe-