DND Dementia and Neurocognitive Disorder

Editorial

Check for updates

Dementia Research Using Healthcare Big Data

Hun-Sung Kim [ID],[1,2] **Dai-Jin Kim** [ID] [1,3]

[1]Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Korea
[2]Department of Endocrinology and Metabolism, College of Medicine, The Catholic University of Korea, Seoul, Korea
[3]Department of Psychiatry, College of Medicine, The Catholic University of Korea, Seoul, Korea

OPEN ACCESS

Correspondence to
Hun-Sung Kim
Department of Medical Informatics, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 06591, Korea.
E-mail: 01cadiz@hanmail.net

Dai-Jin Kim
Department of Medical Informatics, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 06591, Korea.
E-mail: kdj922@catholic.ac.kr

ORCID iDs
Hun-Sung Kim [ID]
https://orcid.org/0000-0002-7002-7300
Dai-Jin Kim [ID]
https://orcid.org/0000-0001-9408-5639

The prevalence rate of dementia is steadily increasing.[1,2] Diverse government-level efforts are being initiated in Korea as well as globally, to reduce the dementia-related financial burden on the state. In the case of Korea, the state's role in the prevention of dementia has been emphasized in the past, and planning for the Korean National Responsibility for Dementia Care (NRDC) initiative is under way.[3] NRDC includes offline roles, such as increasing dementia relief centers and Alzheimer's disease relief nursing hospitals, as well as diverse fields of Research and Development, such as prevention and management of dementia at all stages of life. However, such roles ultimately require a vast amount of resources, and it is essential to procure and verify several databases before these roles are implemented. In other words, to vitalize NRDC, certain verification based on sound grounds is mandatory. The use of accumulated medical big data is becoming a possible solution to this problem.

Big data, renamed 'real-world data' (RWD), is a major talking point in the medical world.[4,5] The Food and Drug Administration defines RWD as 'data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources'.[6] RWD is all medical data not used for clinical research. Four types of RWD are becoming major issues in Korea: 1) Electronic medical records (EMR) data; 2) Claim data, such as Korea National Health and Nutrition Examination Survey, Health Insurance Review and Assessment Service data; 3) Patient-generated health data; and 4) Genomic data. Findings from research conducted using RWD is referred to as real-world evidence (RWE).[4,5]

EMR data comprises all information about actual patients, including their medical history, medical activities, diagnoses and prescriptions, treatment results, and so on.[4] EMR data has the major advantage of allowing the chronological observation of a person's health.[7,8] Claim data, another aspect of RWD, boasts a massive sample size of 1 million cases in Korea.[9,10] Generally speaking, RWD-based research uses the diverse and vast datasets that have already been accumulated by various structures in the healthcare system, and leverages the undeniable advantage of low-cost, immediate availability of large-scale clinical data. Data is accumulating even now; we need no longer worry about the body of data available, but rather about structuring and using such clinical data efficiently, usefully and in ways applicable to the clinical context. Another major strength is that this data is not strictly monitored for research purposes but rather reflects the actual treatment scene, which can generate markedly interesting results.[4] However, RWD is collected for the filing of claims, not for the purpose of research. This presents an innate limitation to the nature and extent of research

Generated by xmlinkpress

that can be conducted using this type of data.[5] At this stage, thus, there are more concerns regarding its use than expectations, and diverse efforts are being initiated in the medical field to overcome such limitations, including quality control and operant definitions of data.

Dementia differs markedly from other diseases regarding major medical big data, requiring a different and varied range of methods, such as health-screening diagnosis and treatment, and clear-cut protocol for approaching data that considers the diversity of symptoms and payment for medication, among others, does not exist.[11-13] Thus, a comprehensive assessment considering diverse information is mandatory, while diagnoses from numerical values that lack precision must be avoided. Realistically, it is difficult to use big data only regarding dementia. The active participation of government entities and hospitals, academies, data providers such as the National Health Insurance Corporation and Health Insurance Review & Assessment Service, and diverse data scientists who can analyze the data with medical teams diagnosing dementia, are crucial. A network that seamlessly incorporates all these aspects must be established.[5,14]

A large body of data does not necessarily equate access to a large body of information.[4,5] The task of extracting information from data is not difficult, and medical sense must support this process to apply the information to the practice of medicine. Through such a process, information is transformed into knowledge that can be practically and clinically used. Medical personnel in the clinical field are currently limited in obtaining information from data, and are not at the stage of obtaining knowledge. This is due to the lack of medical principles reflected in the data, and the subsequent plethora of incorrect information results in studies obtaining inconsistent findings. Medical personnel are most familiar with medical data, and it is essential for them to actively engage in the entire process of gaining information from data, and knowledge from such information.

Whether medical big data can be trusted unconditionally is debatable. For example, there have been cases wherein randomized controlled trials, a type of research that is highly trusted, and RWE, research that uses existing medical big data, generated opposing findings. In such cases, which of the 2 should be trusted? The data used in big data research must first be assessed. In other words, look for flaws in the actual data (this is also why data quality management is crucial).[5] Additionally, efforts must be initiated to identify flaws in the research process, including the principal agent, method, and evaluation; an operant definition of data is necessary. Too, the study results of data scientists who research dementia with superficial information, and those of medical personnel disinterested in data, may run counter to each other, and to restate an earlier point, continuous cooperation between the 2 groups is mandatory. This need is amplified in the case of dementia as it is markedly difficult to obtain a clear diagnosis using only data, compared to other diseases. Diagnoses must distinguish between ambiguous stages of dementia and failure of memory due to elderly age, and data interpretation is thus crucial. For results that are reliable, cooperation is the key.

One of the major issues related to big data management is the protection of personal information and security.[14-16] This issue is highly controversial. For dementia patients, it may be impossible to acquire their informed consent. Due to increasing cognitive impairment over time,[17] consent provided at the beginning of the research must be checked for validity throughout the process. Also, other nuances should also be considered, including the status of "elderly" of those older than age 70, categorized as vulnerable subjects.[18] Although the debate is ongoing, big data allows one freedom in this field because previously accumulated data is being used.

We are at a crossroads with diverse challenges to overcome, but RWD and big data usage in the medical field are generating much attention and interest. While still developing, the medical field's role in the use of RWD is expected to become more crucial in the future. The diverse limitations inherent to the field can be minimized through an integrative approach to data generation, analysis, and interpretation, based on cooperation between data analysts, dementia specialists, and the Korean Dementia Association. Ultimately, the focus should be on the realities of medical big data and identifying paths to supplementing it, rather than wavering under the plethora of superficial information.

REFERENCES

1. Alzheimer's Association. 2017 Alzheimer's disease facts and figures. Alzheimers Dement 2017;13:325-373.
   CROSSREF

2. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013;9:63-75.e2.
   PUBMED | CROSSREF

3. Kim SH. Future policy directions for planning of national responsibility for dementia care. J Korean Med Assoc 2017;60:622-626.
   CROSSREF

4. Kim HS, Lee S, Kim JH. Real-world evidence versus randomized controlled trial: clinical research based on electronic medical records. J Korean Med Sci 2018;33:e213.
   PUBMED | CROSSREF

5. Kim HS, Kim JH. Proceed with caution when using real world data and real world evidence. J Korean Med Sci 2019;34:e28.
   PUBMED | CROSSREF

6. U.S. Food and Drug Administration. Real-world evidence [Internet]. Silver Spring (MD): U.S. Food and Drug Administration; 2019 [cited 2019 Jul 15]. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.

7. Jones EB, Furukawa MF. Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010-12. Health Aff (Millwood) 2014;33:1254-1261.
   PUBMED | CROSSREF

8. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010;363:501-504.
   PUBMED | CROSSREF

9. Ryu DR. Introduction to the medical research using national health insurance claims database. Ewha Med J 2017;40:66-70.
   CROSSREF

10. Cho GJ. Clinical research using medical big data. Anesth Pain Med 2017;12:9-14.
    CROSSREF

11. Boyd D, Crawford K. Critical questions for big data - Provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc 2012;15:662-679.
    CROSSREF

12. Park HS. A review on the applications of medical big data and ergonomic implications. J Ergon Soc Korea 2018;37:143-154.

13. Park SH, Lee JH. National dementia research and development project. J Korean Med Assoc 2018;61:304-308.
    CROSSREF

14. Shin SY, Lyu Y, Shin Y, Choi HJ, Park J, Kim WS, et al. Lessons learned from development of de-identification system for biomedical research in a Korean tertiary hospital. Healthc Inform Res 2013;19:102-109.
    PUBMED | CROSSREF

15. Ienca M, Vayena E, Blasimme A. Big data and dementia: charting the route ahead for research, ethics, and policy. Front Med (Lausanne) 2018;5:13.
    PUBMED | CROSSREF

16. Vayena E, Mastroianni A, Kahn J. Caught in the web: informed consent for online health research. Sci Transl Med 2013;5:173fs6.
    PUBMED | CROSSREF

17. Yanhong O, Chandra M, Venkatesh D. Mild cognitive impairment in adult: A neuropsychological review. Ann Indian Acad Neurol 2013;16:310-318.
    PUBMED | **CROSSREF**

18. Park YS. Why do you need subject protection? Korean J Med 2009;77:S1062-S1064.