




Commentary

A brief introduction to research based on real-world evidence: Considering the Korean National Health Insurance Service database

Eun Kyoung Ahn *

Korean Medicine Data Division, Korea Institute of Oriental Medicine, Daejeon, South Korea



ARTICLE INFO

Article history:

Received 23 July 2021

Revised 1 September 2021

Accepted 29 September 2021

Available online 15 October 2021

Keywords:

Real-world data

Real-world evidence

National health insurance service database

1. Introduction

Real-world data (RWD) on patient health status or the delivery of health care are routinely collected from a variety of sources, including electronic medical records (EMRs), claims and billing databases, product and disease registries, and devices used in the home by patients, including mobile devices. Real-world evidence (RWE) regarding the clinical utility and potential benefits and risks of a medical product can be derived via analysis of RWD.¹ Examples of real-world health and medicine data include National Health Insurance claims data, hospital EMRs, cancer registry data, and reports on adverse drug reactions.

Among them, data on national health insurance claims are provided for research purposes by the Health Insurance Review and Assessment Service and the National Health Insurance Corporation. In Korea, social insurance was made compulsory in 1977 and expanded to all citizens in 1989. In addition to western medicine, Korean medical services were added to the National Health Insurance in 1987. Korea has a universal health insurance system covered by the National Health Insurance. The insurance program covers nearly 100% of the Korean population of approximately 50 million. The database contains individual beneficiary information and healthcare information, such as diagnosis, procedures, prescriptions, and type of institution or department. Claims data for research are shared by two national institutes, i.e., the Korean Na-

tional Health Insurance Service (NHIS) and Health Insurance Review and Assessment Service (HIRA).

These administrative claims datasets have high completeness for variables required for reimbursement and low completeness for health screening records. The strengths of RWD studies include cost-effectiveness due to the use of already existing data, efficient long-term follow up, and generalizability to the entire population (guaranteed external validity). A major strength of RWD studies is the analysis of insurance or medical record databases comprising data from thousands of patients with specific conditions. The availability of such data saves time by avoiding the need to enroll participants in clinical trials or observational studies. RWD is particularly advantageous for research on diseases with low prevalence. In addition, RWD can provide insight into non-licensed indications and rare side effects of licensed drugs not revealed by clinical trials².

RWD studies also have limitations. Insurance claims data are collected for the purpose of payment/reimbursement, which limits the validity of definitions of terms such as disease, condition, and comorbidity.³ Additionally, data is shared between agencies, limited laboratory results are available, and details of adverse events and use of non-reimbursed medications cannot be detected. While RWD provide information on the radiology/laboratory tests ordered by physicians, the results may be confounded by factors that are not captured by insurance claims. Incomplete data are also an issue, and may be seen in the context of sensitive diagnoses, drug dosages and information on pharmaceutical companies. As a matter of principle, institutions cannot disclose all of their data. Even EMRs, which include test results, can be very difficult to obtain from certain agencies due to policy issues, among other reasons.

* Corresponding author at: Korean Medicine Data Division, Korea Institute of Oriental Medicine, 1672, Yuseong-daero, 34051 Daejeon, South Korea.

E-mail address: ekahn@kiom.re.kr

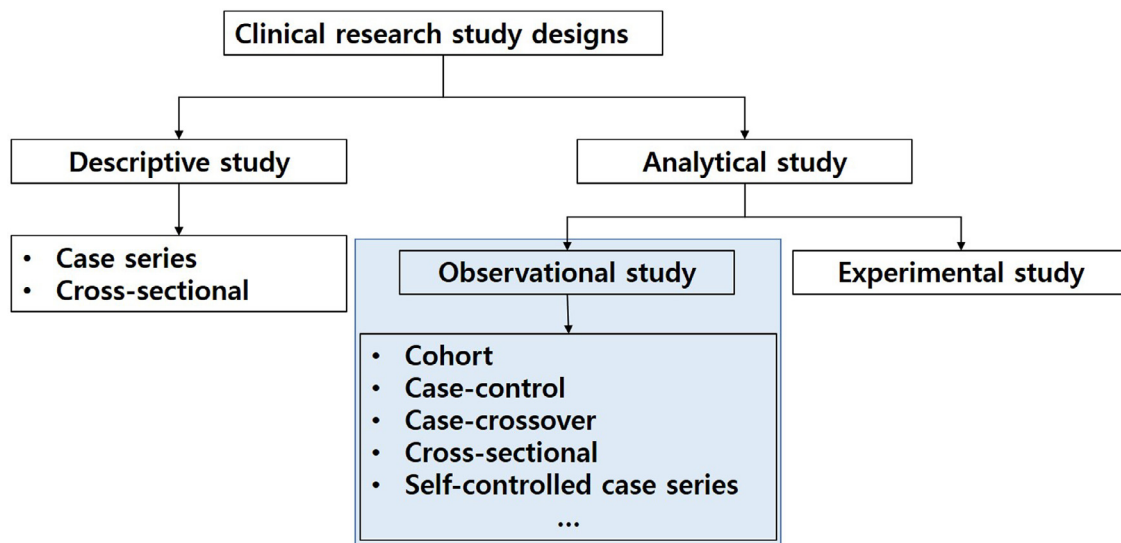


Fig. 1. The study design based on RWD.

A flowchart describing type of clinical research study designs and their categorization. Observational study is a sub-concept of analytical study, and cohort, case-control, case-crossover, cross-sectional, and self-controlled case series study can be performed as detailed research designs of it.

Therefore, researchers are not always able to conduct sophisticated randomized controlled trials using RWD. Nevertheless, RWD are useful for retrospective cohort studies. For example, after being released to the market, RWD can be analyzed to identify side effects of drugs, and the medical expenses associated with treating certain diseases.

In this article, study designs based on RWD, strategies and tools for ensuring high-quality data, and a procedure for conducting research based on RWE, i.e., using data from a health insurance corporation, are discussed.

2. Study design based on RWD

A researcher wanting to use RWD to understand a topic can perform a descriptive study as an initial step.⁴ RWD are feasible for use in retrospective, observational studies. Therefore, most studies with cohort, case-control, and cross-sectional designs can be conducted using such data (Fig. 1.). Additionally, RWD provide excellent material for healthcare research⁵. In a cohort study design, people who have a certain condition or receive a particular treatment are followed over time, and compared to another group of people who have not been exposed to the condition or treatment. The claims data can be used in a cohort study to emulate a randomized clinical trial, and a popular mechanism for adjusting for perturbation is a propensity score (PS).

Case-control studies select people with the outcome of interest, and do not follow them over time. Researchers select people with a particular outcome (the cases) and ascertain their exposures based on interviews or medical records. The odds of having an exposure with the outcome are compared to the odds of having an experience without the outcome. For case-control studies, the claims data can be used to compare the “cases,” i.e., people who experience the outcome of interest, and “controls,” i.e., people who did not experience the outcome of interest.

The case-crossover design is similar to the case-control design, except that controls and cases are the same individuals. Individuals are considered controls before the development of disease. Studies may compare two or more treatments or interventions by administering one treatment or intervention followed by switching to another. In the case of two treatments (e.g., A and B), half of the subjects are randomly allocated to receive treatment A followed by

B, and half to receive treatment B followed by A. This study design is criticized because of the possibility that the effects of the first treatment may continue into the period of the second treatment.

The cross-sectional study design involves observation of a defined population at a single point in time or time interval. Therefore, the exposure and outcome are simultaneously determined. The self-controlled cohort design compares the rate of outcomes during exposure to treatment to the rate of outcomes prior to the exposure. The self-controlled case series design compares the rate of outcomes during exposure to treatment (cohort A) to the rate of outcomes during the unexposed interval (cohort B), before, between, and after exposures. This method is vulnerable to confounding due to time-varying effects.

RWD can be analyzed during retrospective cohort studies to identify risk factors for a particular disease, or to establish a causal relationship between a particular behavior and outcome. There are some examples of ideas on research design; observational findings in health care: prevalence/incidence of target disease, large-scale evidence generation and evaluation, comparative effectiveness and safety of treatment or tests, risk factors of target disease, association between factor A and B, screening/pilot study of a pragmatic or randomized clinical trials, longitudinal outcomes of target disease or treatment, conducting/evaluation of a patient-level predicting model for target disease, and treatment patterns for chronic comorbid conditions in patient with target disease. The data must be processed appropriately in such studies.

3. Tools

Generally, research involves data preprocessing, analysis, and visualization using tools optimized for different functions, which may or may not be familiar to the researcher. Pre-processing of large-scale EMR data can be performed with commercial programs such as MS-SQL and Oracle database management systems, or using free programs such as My-SQL and PostgreSQL. Other tools can be used depending on the hospital information system. After refining the data set, the researcher can analyze it using programs like R, Python, and SAS. For the claims data provided by NHIS or HIRA, the SAS Enterprise Guide was previously used as standard, but is gradually being replaced by analysis programs such as R, Python, and SPSS. Nevertheless, researchers should be familiar with the ba-

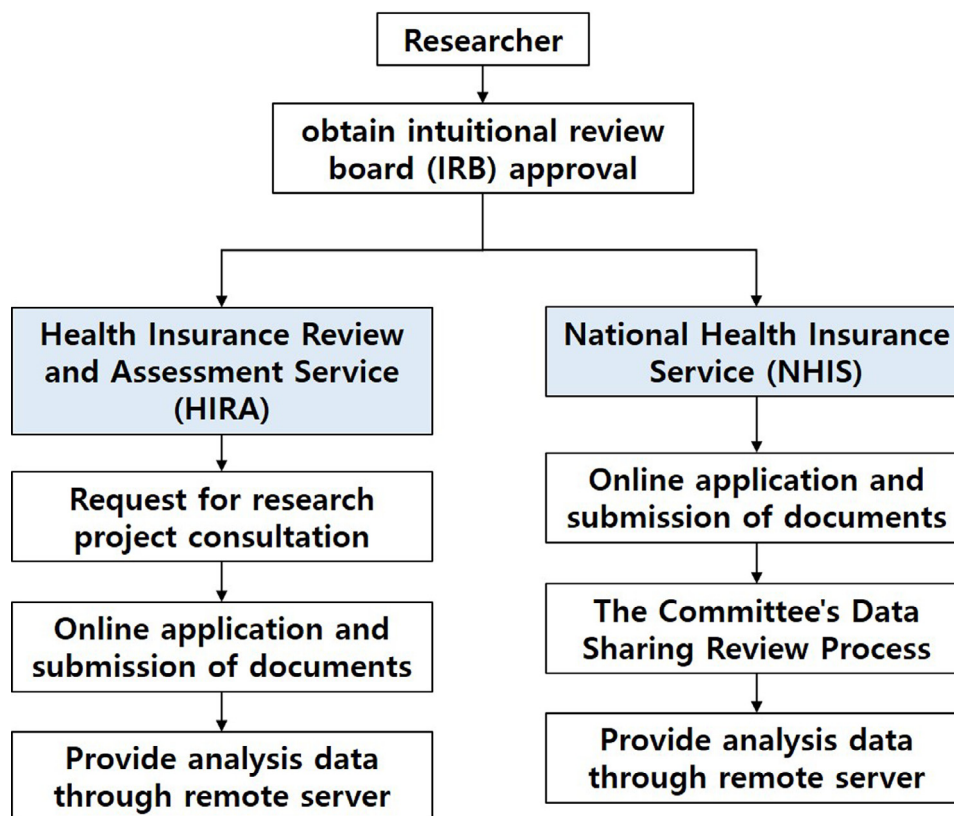


Fig. 2. Data sharing process from two type of Korean health insurance claim database

A schematic flowchart describing data sharing process of two type of Korean health insurance claim database. 1st step is that the researcher obtains intuitional review board (IRB) approval. Participate in the data sharing review process in accordance with the procedures of each organizations. The procedures for the two institutions are almost identical. However, in HIRA, consultation with the person in charge before the data request process is added.

sic data preprocessing functions of SAS Enterprise before performing further statistical analyses using more familiar programs.

4. Brief procedure for conducting research based on RWE

There are two main sources of health insurance claims data for research purposes: the HIRA Big Data Hub and the NHIS Sharing Service (Fig. 2.). To use the data from either of these sources, the researcher must obtain intuitional review board (IRB) approval and participate in a data-sharing review process.

Researchers should consider the particular characteristics of the HIRA and NHIS databases when selecting a data source. The HIRA database includes unit and cost price data that and can be accessed via a remote server. These data are useful when cost analysis is difficult. The NHIS database contains data on date of death and the results of medical examinations. Since claims databases do not include medical test results, obtaining blood test and lifestyle-related questionnaire data is important. In summary, the researcher should select the most appropriate data source for the particular research topic.

5. RWD in integrative medicine

Current RWD research has mostly focused on analyzing the current level of reporting and using world-wide RWD-based signal detection for adverse drug events. Additionally, effectiveness data may be generated using claims data. However, a limitation of studies of integrative medicine using RWD is that only limited prescription herbal preparations and treatments are covered by insurance. Therefore, RWD were applied late to integrative medicine research,

but their use has increased gradually since 2010 RWD-based research of claims in the field of integrative medicine has mainly focused on the use of medical services^{6, 7}, reimbursement costs⁸, effects of acupuncture⁹, comparisons with medical care¹⁰, and the status of oriental medicine prescription¹¹. However, as previously described, RWD-based research of claims should use various sophisticated research designs, which may help to generate evidence that reduces costs and time.

6. Conclusion

When using large datasets comprising data of unknown origin or quality for RWD studies, a lack of access to analytical tools for non-experts, or a lack of researchers familiar with the most appropriate statistical methodologies, can lead to inaccurate or unreliable conclusions¹². Researchers conducting studies based on RWE must preprocess claims data, which requires an understanding of the origin of those data. Furthermore, a research design appropriate for the research question must be used, along with optimal analysis methods and data analysis tools. Lastly, collaboration with experts in other fields may be necessary.

Author contribution

This is the sole author's work.

Conflict of interest

The author has no conflict of interest to declare.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (2016R1C1B1008425) and the Ministry of Science and Technology, ICT & Future Planning (2019R1C1C1006973).

Ethical statement

Not applicable.

Data availability

Not applicable.

References

1. Framework for FDA's Real-World Evidence Program. US Food and Drug Administration website., <<https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>>; Accessed March 05.
2. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med*. 2016;375(23):2293–2297.
3. Koto R, Nakajima A, Horiuchi H, Yamanaka H. Real-world treatment of gout and asymptomatic hyperuricemia: A cross-sectional study of Japanese health insurance claims data. *Mod Rheumatol*. 2021;31(1):261–269.
4. Parab S, Bhalerao S. Study designs. *Int J Ayurveda Res*. 2010;1(2):128–131.
5. Ranganathan P, Aggarwal R. Study designs: Part 1 - an overview and classification. *Perspect Clin Res*. 2018;9(4):184–186.
6. Lim B. Korean medicine coverage in the National Health Insurance in Korea: present situation and critical issues. *Integr Med Res*. 2013;2(3):81–88.
7. Huang CW, Hwang IH, Lee YS, Hwang SJ, Ko SG, Chen FP, et al. Utilization patterns of traditional medicine in Taiwan and South Korea by using national health insurance data in 2011. *PLoS One*. 2018;13(12).
8. Huang SK, Lai CS, Chang YS, Ho YL. Utilization Pattern and Drug Use of Traditional Chinese Medicine, Western Medicine, and Integrated Chinese-Western Medicine Treatments for Allergic Rhinitis Under the National Health Insurance Program in Taiwan. *J Altern Complement Med*. 2016;22(10):832–840.
9. Koh W, Kang K, Lee YJ, Kim MR, Shin JS, Lee J, et al. Impact of acupuncture treatment on the lumbar surgery rate for low back pain in Korea: A nationwide matched retrospective cohort study. *PLoS One*. 2018;13(6).
10. Jung B, Bae S, Kim S. Use of Western Medicine and Traditional Korean Medicine for Joint Disorders: a Retrospective Comparative Analysis Based on Korean Nationwide Insurance Data. *Evid Based Complement Alternat Med*. 2017;2017.
11. Kim H, Choi JY, Hong M, Suh HS. Traditional medicine for the treatment of common cold in Korean adults: a nationwide population-based study. *Integr Med Res*. 2021;10(1).
12. Gliklich RE, Leavy MB. Assessing Real-World Data Quality: the Application of Patient Registry Quality Criteria to Real-World Data and Real-World Evidence. *Ther Innov Regul Sci*. 2020;54(2):303–307.