

Opinion  
Medical Informatics



# Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records

Hun-Sung Kim ,<sup>1,2</sup> Suehyun Lee ,<sup>3</sup> and Ju Han Kim <sup>3</sup>

<sup>1</sup>Department of Medical Informatics, The Catholic University of Korea College of Medicine, Seoul, Korea

<sup>2</sup>Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, Seoul, Korea

<sup>3</sup>Division of Biomedical Informatics, Systems Biomedical Informatics Research Center, Seoul National University College of Medicine, Seoul, Korea



Received: Apr 17, 2018

Accepted: May 24, 2018

**Address for Correspondence:**

**Ju Han Kim, MD, PhD**

Division of Biomedical Informatics, Systems Biomedical Informatics Research Center, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea.  
E-mail: juhan@snu.ac.kr

© 2018 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ORCID iDs**

Hun-Sung Kim   
<https://orcid.org/0000-0002-7002-7300>

Suehyun Lee   
<https://orcid.org/0000-0003-0651-6481>

Ju Han Kim   
<https://orcid.org/0000-0003-1522-9038>

**Funding**

The research was supported by the funding of Korean Society of Hypertension (2018).

**Disclosure**

The authors have no potential conflicts of interest to disclose.

## ABSTRACT

Real-world evidence (RWE) and randomized control trial (RCT) data are considered mutually complementary. However, compared with RCT, the outcomes of RWE continue to be assigned lower credibility. It must be emphasized that RWE research is a real-world practice that does not need to be executed as RCT research for it to be reliable. The advantages and disadvantages of RWE must be discerned clearly, and then the proper protocol can be planned from the beginning of the research to secure as many samples as possible. Attention must be paid to privacy protection. Moreover, bias can be reduced meaningfully by reducing the number of dropouts through detailed and meticulous data quality management. RCT research, characterized as having the highest reliability, and RWE research, which reflects the actual clinical aspects, can have a mutually supplementary relationship. Indeed, once this is proven, the two could comprise the most powerful evidence-based research method in medicine.

**Keywords:** Real-world Evidence; Randomized Control Trial; Real-world Data

## INTRODUCTION

Real-world data (RWD) refer to data collected from diversified areas of daily life that are outside the scope of highly controlled randomized control trials (RCTs).<sup>1</sup> Examples of RWD are the large quantities of data collected broadly through the use of a diverse range of medical devices by various users, including self-checked glucometer or sphygmomanometer. These real, practical data cannot be obtained through contemporary or traditional clinical trials. A wide range of data already accumulated, including the electronic medical records (EMRs) used in hospitals,<sup>2-4</sup> Korea National Health and Nutrition Examination Survey (KNHANES) data,<sup>5</sup> Health Insurance Review and Assessment service data<sup>6</sup> or Korea Adverse Event Reporting System (KAERS) database<sup>7</sup> are also included in RWD.

EMR, in allowing for the accumulation of a large amount of clinical data, are increasingly used for medical research.<sup>8,9</sup> Worldwide, EMR-based clinical research has been reported to increase the quality of research and satisfaction of clinicians.<sup>10-13</sup> Depending on study design and objectives, EMR research may utilize a variety of data: medication prescription,

**Author Contributions**

Conceptualization: Kim HS, Kim JH. Data curation: Kim HS. Formal analysis: Kim HS. Investigation: Kim HS, Lee S, Kim JH. Writing - original draft: Kim HS. Writing - review & editing: Kim JH.

treatment decision making, disease management, and clinical research data. For these reasons, EMR data are recognized as definitive data with the highest reliability among RWD. Analysis of RWD yields real-world evidence (RWE), and many scholars are focusing their interest on RWE research using RWD accumulated broadly, along with the issue of big data. Efforts have been made toward expanding the range of relevant clinical research.

Fundamentally, it is desirable for the new research format, RWE, and the traditional research format, RCT, to be mutually complementary. However, despite the comprehensive reports on the importance of RWE, there is a tendency among scholars of mistrusting the findings of RWE compared with RCT. For many researchers, overcoming a diverse range of biases that can occur in EMR-based RWE research requires trial and error. Therefore, we intend to present opinions on precautions that can be used effectively in executing research using RWE from the clinical rather than technical perspectives.

## RWE VERSUS RCTS

RCT conducts research on highly selective populations and is managed in tightly controlled settings.<sup>14</sup> RCTs are conducted under the premise that the research outcomes of the chosen sample population represent those of the entire population. This research format is deemed to have the highest reliability. The majority of clinical guidelines are given on the basis of the resultant values deduced from RCTs.<sup>15</sup> However, a disadvantage is that participants may fail to reflect the actual clinical site sufficiently to represent the entire population, given the clear set of inclusion or exclusion criteria. This drawback of exclusion has been deemed a bias.<sup>16,17</sup> That is, there would be difficulty in applying the resultant values of data analysis confirmed in a small number of patients selected in accordance with set criteria in RCTs to diversified actual situations that could occur in the real world. For this reason, the foundation on which to elevate definitive reliability through concurrent execution of RWE research, in addition to RCT, needs to be established; these two should ideally be developed in a mutually supplementary rather than competing relationship. For example, if the RCT concentrates on the efficacy of a drug, the RWE can focus on the epidemiology, effectiveness, safety, or costs of treatment related to the drug (Tables 1 and 2).<sup>18</sup> RWE research that can sustain follow-up observation after the completed RCT research holds important significance. Clinical trials and secondary syntheses of evidence-based information can be biased owing to undisclosed financial and nonfinancial conflicts of interests. Therefore, research should be undertaken with the perception that managing these conflicts of interests more thoroughly is crucial to maintaining the essence and credibility of research.<sup>19</sup>

**Table 1.** RCTs versus RWE<sup>18</sup>

Variables	RCTs	RWE
Purpose	Efficacy	Effectiveness
Setting	Experimental setting	Real-world setting
Follow up	Designed	In actual practice
Treatment	Fixed pattern	Variable pattern
Study group	Homogenous	Heterogeneous
Attending physician	Investigator	Many practitioners
Comparator	Placebo/selective alternative interventions	Many alternative interventions
Patient monitoring	Continuous, per protocol	Changeable

RCT = randomized clinical trial, RWE = real-world evidence.

**Table 2.** Advantages and limitations of real-world evidence compared with RCTs

Advantages
Less time and cost consumption compared with RCT
- Possible to shorten the duration of clinical research
- No period needed for patient recruitment/enrollment
Presentation of the directionality for RCT
Research that cannot be done with RCT is possible
- Safe research on high-risk groups
Detection of less frequent side effects
Rapid access/easier information and data retrieval
Prediction model or high-risk group selection
Set a foundation on AI
Limitation
For correct analysis, a massive amount of data should be collected.
Much time needed for DQM
Experienced experts needed for the analysis of the massive amount of data
Lack of privacy/confidentiality/lost data
Standardized research protocol needs to be established before research
- High possibility of bias
- Possible to interpret with bias or incorrectly based on research results

RCT = randomized control trial, DQM = data quality management.

## DIRECTIONALITY FOR EMR-BASED CLINICAL RESEARCH

Previous research on the use of EMR data have focused on their convenience and accessibility.<sup>20</sup> With the recent accumulation of extensive data in EMR systems, these data can be used in clinical studies. The use of EMR-based RWE addresses the disadvantages of both cohort studies (e.g., follow-up cost, long study periods, and maintenance of consistency during the study period) and RCTs (e.g., loss of participants at follow up, changes in treatment, long study periods, and expense). In EMR-based clinical research, the investigator has direct access to the EMR system and can quickly verify a hypothesis by sampling related variables from the system.

### Real clinical practice

Compared with RCT, the core aspect of RWE is its practical nature rather than a tightly controlled set group. EMR include clear diagnoses and prescriptions entered by physicians stored in a database that can be easily accessed and utilized. EMR research can broadly reflect actual practice as it is fundamentally research aimed at the analysis of vast quantities of data already accumulated. More specifically, EMR enables researchers to extract and analyze quickly the side effects,<sup>21</sup> progress, and prognosis of drugs<sup>22</sup> with low frequency. EMR allows for the quick and systematic collection of data on the effectiveness or side effects that can manifest when a specific drug has been broadly prescribed to unspecified masses. EMR research can practically approach a diverse range of safety issues not detected in three- and four-phase studies.<sup>23</sup> In addition to the evaluation of drugs, EMR research is able to evaluate the performance or clinical output of medical staff or hospital. With EMR data, any scholar can evaluate the actual achievement rate of goals of individual medical departments or hospitals.<sup>24</sup> For instance, whether prescriptions were made in accordance with or against the guidelines is being disclosed through such research activities. In other words, the one making the prescription is observed and data accumulated on what is being prescribed. Further, the diversification of approaches to data allows the evaluation of the prescription patterns of individuals. Consequently, this system can be expected to enable the most appropriate prescriptions for any given disease. Thus, the possibility of approaching system-based rather than experience-based medicine is extremely high.

### Presentation of the directionality for RCT

A long period is necessary to conceptualize RCT for the effects or side effects of drugs that carry social issues. Indeed, there are cases in which clinical research is terminated owing to unexpected results of the RCT executed.<sup>25</sup> The organization that conducts clinical research must bear the enormous consequences if the results of RCT research turn out to be different from those anticipated. In such cases, a desirable RWE research model is one that presents the directionality of RCT by forecasting the research results under the concept of a pilot study by securing as extensive a range of data as possible in a short time period.<sup>26</sup> The advantages of EMR for pharmaceutical or medical device companies include precise market and customer analyses as well as access to the data of competitors. Further, information on the prescribing patterns of physicians would be useful for establishing marketing plans.

### Research not possible with RCT owing to research ethics

This point marks the most powerful advantage that can be done only with EMR. Research can be conducted for all clinical trials only after having acquired the approval of an Institutional Review Board (IRB). New topics of research outside the range of set guidelines cannot be approved by an IRB, and consequently, such research cannot be carried out. However, certain prescriptions that are outside the guidelines may exist within the actual EMR data.<sup>22</sup> These are mostly cases in which it was unavoidable to make the said prescription rather than intentionally infringing the guidelines with the determination that there are greater advantages than disadvantages in making the said prescription. As RWE research uses already accumulated data, there is no infringement on the rights and benefits or harm with respect to the physical states of patients. Therefore, research in itself is not impossible. Nonetheless, this aspect requires careful approach. Modification of existing guidelines can be proposed carefully, and additional large-scale RCT research can be proposed for this purpose. If clinical RWD are used well, various controversial issues can be easily studied, which may improve future diagnosis/treatment guideline modalities.

### Prediction model based on artificial intelligence

With deep learning becoming the talk of town in recent years, people are becoming increasingly interested in artificial intelligence that utilizes deep learning.<sup>27</sup> Deep learning-based artificial intelligence ultimately needs to the utilization of an extensive range of accumulated data. As EMR have been implemented widely in hospitals, we expect a drastic increase in the analysis of accumulated EMR data. To obtain clinically significant data from large hospitals, a clinical data warehouse can be an important tool, whereas EMR will potentially lead to various types of clinical research. However, the majority of actual EMR-based deep learning projects have failed to reach the level initially anticipated. At the heart of the failure is data collection. Various algorithms developed using unprocessed data have frequently failed to perform as anticipated. It is not possible to generalize and deduce outcomes with the results obtained from unprocessed and low-quality data. For these reasons, hospitals have recently commenced improving access to and systematic purification of data.

### Unknown information

Previously unknown information that clinical researchers fail to predict in the study planning phase may come to light, and EMR systems might lead to new knowledge and theories.<sup>21,22</sup>

## PRIVACY PROTECTION

Accumulation of large-scale patient data and clinical trials using such data raise the important issue of the protection of personal information. In the case of Korea, numerous limitations on the use of clinical big data have been enacted to address issues related to personal information leaks. Therefore, privacy protection must be mentioned clearly when preparing works on RWD. This practice must include the fact that the name and resident registration number will not be included in the accumulated data, and information directly related to privacy, such as address, telephone number, and e-mail, will be removed. Any combination that can lead to the identification of an individual, such as hospital registration number and date of birth, must be removed immediately prior to the stage of analysis. Further, all files on the extracted information must be encrypted and then stored in an encrypted computer, accessible only to the designated researcher. Additionally, contents of theses must not deal with the personal information of patients; researchers should emphasize that the research has no possibility of causing additional physical harm whatsoever to the subjects. Finally, it must be specified that the advance consent of the patient is not necessary given the anonymity of the subjects through encryption of data and that there is no effect on the rights and welfare of the subjects because past data, collected after having completed treatment, are used (informed consent not required). The statement “the research was conducted after having acquired the approval of IRB” must also be included.

## ISSUES TO BE CONSIDERED AT THE TIME OF RESEARCH ANALYSIS

Compared with RCT, EMR-based RWE study poses difficulties in controlling for biases that may affect the study outcomes. Research projects need to be commenced with a definitive design to minimize such biases. Data quality management (DQM) is the most important part of the extracted data in RWE. As extracted data could be deemed byproducts of the process of treating patients, rather than data elaborately organized and structured in accordance with the purposes of the research, the quality of the data can be low. There may even be many errors or erroneous inputting in the course of their collection. Through DQM, comparison with original data can be confirmed, including the mixed use of numbers and letters, erroneous descriptions, and serious errors. This phase may be time consuming, depending on the number or capacity of data. The most serious bias normally occurs at this stage.

As mentioned above, it is important to secure larger samples size to elevate fundamentally the reliability of RWE. Although larger quantities of data do not necessarily elevate accuracy and reliability, using the same would help minimize bias that occurs in the smallest number. Even with plenty of actually extracted data, a researcher may end up excluding a large number once the data are sorted in accordance with the purpose of a research. For example, there are many cases in which the basic baseline results, such as height/weight and blood pressure, end up omitted. Without baseline values, there can be no follow up, and changes in clinical laboratory data cannot be examined. There also are frequent cases in which the actual overall number of research subjects is reduced by more than half, such as when patients for whom follow up is not possible or patients who transferred to another hospital in the middle of the research are excluded. The exclusion of a large number of samples is in itself a significant bias, and it is the aspect of EMR-based clinical research pointed out most frequently by journal reviewers as a hard limitation.

Certainly, not all the data are available in EMR. For one, it is difficult to confirm the compliance of the subjects using EMR, including whether they took drugs as instructed in a drug-related research. Hence, assessment of the effectiveness or efficacy of drugs tends to be difficult in RWE research (This is generally carried out in RCTs). Checking whether subjects took the drugs as instructed by checking the number of pills prescribed and those remaining at the time of the next visit can be implemented as an alternative method for assessment of the compliance of subjects. Failure of the medical staff to make entries in the medical chart is also a problem. Proper research would be possible if the medical staff describe the state of patients accurately and in detail. Further, although patient conditions and treatments can be described, EMR cannot be used to elucidate scientifically causal relationships.

The domain of RWE in addition to EMR data mentioned in this work is extremely broad. It can be collected directly from the patient by means of a diverse range of medical devices and by utilizing a wide range of data. Accordingly, RWD are increasing in volume, whereas RWE is being gradually expanded, with continuous advancement of their respective analytic functions and protocols. In addition to the importance of the quantity of data, it is necessary to concentrate on the means of properly structuring and utilizing rapidly accumulated clinical data. RWE and RCT research are not in a mutually competitive relationship. Presently, RWE research cannot substitute RCT research. However, RCT research, bestowed the highest reliability, and RWE research, perceived to reflect the actual clinical aspect, can be proven to share a mutually supplementary relationship to become the most powerful resource in evidence-based medicine.

## REFERENCES

1. Garrison LP Jr, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health* 2007;10(5):326-35.  
[PUBMED](#) | [CROSSREF](#)
2. Kim HS, Lee H, Lee SH, Jeong YJ, Kim TM, Yang SJ, et al. Use of moderate-intensity statins for low-density lipoprotein cholesterol level above 190 mg/dL at baseline in Koreans. *Basic Clin Pharmacol Toxicol* 2017;121(4):272-8.  
[PUBMED](#) | [CROSSREF](#)
3. Kim HS, Lee H, Park B, Park S, Kim H, Lee SH, et al. Comparative analysis of the efficacy of low- and moderate-intensity statins in Korea. *Int J Clin Pharmacol Ther* 2016;54(11):864-71.  
[PUBMED](#) | [CROSSREF](#)
4. Jeong YJ, Kim H, Baik SJ, Kim TM, Yang SJ, Lee SH, et al. Analysis and comparison of the cost-effectiveness of statins according to the baseline low-density lipoprotein cholesterol level in Korea. *J Clin Pharm Ther* 2017;42(3):292-300.  
[PUBMED](#) | [CROSSREF](#)
5. Kweon S, Kim Y, Jang MJ, Kim Y, Kim K, Choi S, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). *Int J Epidemiol* 2014;43(1):69-77.  
[PUBMED](#) | [CROSSREF](#)
6. Kim L, Kim JA, Kim S. A guide for the utilization of health insurance review and assessment service national patient samples. *Epidemiol Health* 2014;36:e2014008.  
[PUBMED](#) | [CROSSREF](#)
7. Kim H, Kim N, Lee DH, Kim HS. Analysis of national pharmacovigilance data associated with statin use in Korea. *Basic Clin Pharmacol Toxicol* 2017;121(5):409-13.  
[PUBMED](#) | [CROSSREF](#)
8. Hanna K, Anderson SM, Maddox SD. *Think Research: Using Electronic Medical Records to Bridge Patient Care and Research*. Washington, D.C.: Center for Accelerating Medical Solutions; 2005.
9. Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. *Methods Inf Med* 2009;48(1):45-54.  
[PUBMED](#)

10. Waller PC, Coulson RA, Wood SM. Regulatory pharmacovigilance in the United Kingdom: current principles and practice. *Pharmacoepidemiol Drug Saf* 1996;5(6):363-75.  
[PUBMED](#) | [CROSSREF](#)
11. Hauben M, Patadia V, Gerrits C, Walsh L, Reich L. Data mining in pharmacovigilance: the need for a balanced perspective. *Drug Saf* 2005;28(10):835-42.  
[PUBMED](#) | [CROSSREF](#)
12. Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011;20(1):1-11.  
[PUBMED](#) | [CROSSREF](#)
13. Waller PC, Evans SJ. A model for the future conduct of pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2003;12(1):17-29.  
[PUBMED](#) | [CROSSREF](#)
14. Lee JY, Park BJ. A method to evaluate scientific evidence in clinical preventive medicine. *J Korean Med Assoc* 2011;54(10):1006-12.  
[CROSSREF](#)
15. Roche N, Reddel HK, Agusti A, Bateman ED, Krishnan JA, Martin RJ, et al. Integrating real-life studies in the global therapeutic research framework. *Lancet Respir Med* 2013;1(10):e29-30.  
[PUBMED](#) | [CROSSREF](#)
16. May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation* 1981;64(4):669-73.  
[PUBMED](#) | [CROSSREF](#)
17. Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Fam Med* 2007;39(2):132-7.  
[PUBMED](#)
18. Wierzbicka N, Jahnz-Różyk K. The evolving landscape for real world evidence in Poland: physicians' perspective. *JHPOR* 2015;1:15-33.
19. Gasparyan AY, Ayzazyan L, Akazhanov NA, Kitas GD. Conflicts of interest in biomedical publications: considerations for authors, peer reviewers, and editors. *Croat Med J* 2013;54(6):600-8.  
[PUBMED](#) | [CROSSREF](#)
20. El-Kareh R, Gandhi TK, Poon EG, Newmark LP, Ungar J, Lipsitz S, et al. Trends in primary care clinician perceptions of a new electronic health record. *J Gen Intern Med* 2009;24(4):464-8.  
[PUBMED](#) | [CROSSREF](#)
21. Kim HS, Kim H, Jeong YJ, Lee H, Yim HW, Kim JI, et al. Comparative analysis of the suspected heparin-induced thrombocytopenia level in Korea. *Basic Clin Pharmacol Toxicol* 2017;121(4):360-7.  
[PUBMED](#) | [CROSSREF](#)
22. Kim HS, Lee SH, Kim H, Lee SH, Cho JH, Lee H, et al. Statin-related aminotransferase elevation according to baseline aminotransferases level in real practice in Korea. *J Clin Pharm Ther* 2016;41(3):266-72.  
[PUBMED](#) | [CROSSREF](#)
23. Kim HS, Kim H, Jeong YJ, Yang SJ, Baik SJ, Lee H, et al. Comparative analysis of the efficacy of omega-3 fatty acids for hypertriglyceridaemia management in Korea. *J Clin Pharm Ther* 2016;41(5):508-14.  
[PUBMED](#) | [CROSSREF](#)
24. Kim HS, Kim H, Lee H, Park B, Park S, Lee SH, et al. Analysis and comparison of statin prescription patterns and outcomes according to clinical department. *J Clin Pharm Ther* 2016;41(1):70-7.  
[PUBMED](#) | [CROSSREF](#)
25. James WP, Caterson ID, Coutinho W, Finer N, Van Gaal LF, Maggioni AP, et al. Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects. *N Engl J Med* 2010;363(10):905-17.  
[PUBMED](#) | [CROSSREF](#)
26. Kim TM, Kim H, Jeong YJ, Baik SJ, Yang SJ, Lee SH, et al. The differences in the incidence of diabetes mellitus and prediabetes according to the type of HMG-CoA reductase inhibitors prescribed in Korean patients. *Pharmacoepidemiol Drug Saf* 2017;26(10):1156-63.  
[PUBMED](#) | [CROSSREF](#)
27. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol* 2018;15(3 Pt B):504-8.  
[PUBMED](#) | [CROSSREF](#)