Editorial

Healthc Inform Res. 2017 January;23(1):1-3.
https://doi.org/10.4258/hir.2017.23.1.1
pISSN 2093-3681 • eISSN 2093-369X

HIR
Healthcare Informatics Research

Sharing Clinical Big Data While Protecting Confidentiality and Security: Observational Health Data Sciences and Informatics

Rae Woong Park, MD, PhD

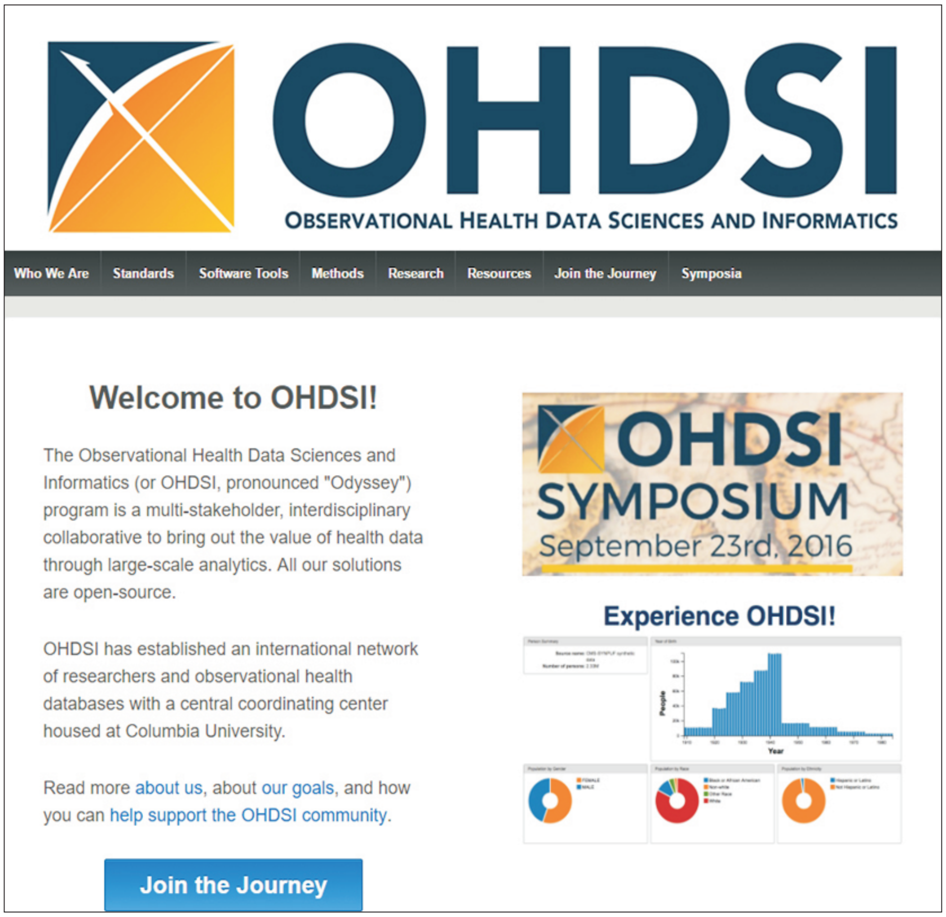
Observational Health Data Sciences and Informatics, New York, NY, USA; Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

Over the past decade, the amount of data we routinely create and collect has increased tremendously, and our ability to analyze and understand them has also greatly improved. The routine operation of a modern healthcare system enriches electronically stored data as a byproduct of clinical practice. Today, utilizing data is far less costly than ever before. Debates between studies on the same topic often arise from differences in participants, research design, or interpretation. If data can be collected and analyzed comprehensively and at scale, the fundamental need to work with small samples under strict assumptions can be eliminated [1]. Even a very large individual database, however, is not enough to meet the diverse needs of researchers.

In many countries, researchers and administrators have struggled to apply a standardized data model to harmonize and collect medical data from a variety of heterogeneous sources. However, various barriers exist, such as system heterogeneity, various data formats, changes in human protection rules around the world, trust building, contracting and coordination, and research governance policies.

Recently, distributed research networks (DRNs), such as Observational Health Data and Informatics (OHDSI, pronounced "Odyssey"), the National Patient Centered Clinical Research Network (PICORNET), or Sentinel Initiatives have gained popularity among clinical data partners worldwide. DRN uses the same data structure, called the Common Data Model (CDM), to run the same analysis program for participating organizations and then combine the results summarized over the network to provide results across the network. Data is stored in a CDM that achieves both syntactic and semantic interoperability, enabling consistent queries to be applied to databases around the world [2].

OHDSI (http://www.ohdsi.org) (Figure 1) is an organization of international collaborators with more than 120 researchers committed to making this clinical big data analysis a reality. OHDSI has transformed more than 680 million patient data from 56 databases in 12 countries into the Observational Medical Outcomes Partnership (OMOP) CDM format. It carried out a monumental study on chronic diseases prescription patterns using 250 million patient data items from four countries in the network [3]. In Korea, 4.6 million patient data from two university hospitals have been converted to CDM v5. In addition, seven Korean university hospitals and one Korean government agency joined OHDSI to convert their data into CDM. By the end of 2017, over 20 million patient data are expected to be available in CDM v5 format in Korea. The advantage of Electronic Medical Record (EMR) data for observational study data is that there is a detailed timestamp for each activity as well as laboratory test results and outcomes [4]. However, events that occur outside the hospital cannot be captured in the hospital EMR data. Conversely, insurance claim data does not include laboratory test results or detailed timestamps, but there are



Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University.

Read more about us, about our goals, and how you can help support the OHDSI community.

Figure 1. The Observational Health Data Sciences and Informatics (OHDSI) at http://www.ohdsi.org/.

virtually all diagnoses, prescriptions, and procedures that have occurred in the country. Fortunately, one million randomly drawn Korean patient data from the past 9 years of claim data have been converted to OMOP CDM v5, and the size of the conversion is likely to expand to the entire Korean population within a few years.

By providing all of their work products as open source, OHDSI has lowered the technical barriers required for participation in a DRN. With more and more data partners participating in the OHDSI network, it is expected that everyone, including students, researchers, and developers from hospitals, governments, pharmaceutical companies, and related industries seeking medical evidence, knowledge, or even artificial intelligence algorithms, will benefit from the network with unprecedentedly low cost and effort in a timely manner.

However, differences in data structures and coding system between or within countries are still major barriers to being a data partner in a DRN. For Korea, we had to map local codes for diagnoses, drugs, procedures, and laboratory tests into the OMOP standard vocabulary. As there are hundreds of thousands of local codes, mapping them all into OMOP vocabulary requires huge amounts of effort, time, and financial expense. One coordination center for code mapping for each country is strongly recommended because the code mapping process requires huge and continuous efforts together with nationwide consensus.

DRNs will enable researchers to access a network of billions of patients to generate evidence about all aspects of healthcare. This collaborative analysis of clinical big data across the world will lead to medical innovation along with the fourth industrial revolution [5]. In conclusion, clinical big data analysis using DRNs will serve as a key means to improve health by empowering a community to collaboratively generate evidence that can be used to promote better health decisions and better care. Patients and clinicians and all other decision-makers around the world are expected to be able to use DRN tools and evidence every day.

References

1. Mayer-Schonberger V. Big data for cardiology: novel discovery? Eur Heart J 2016;37(12):996-1001.
2. Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J,

McClellan M. The new Sentinel Network: improving the evidence of medical-product safety. N Engl J Med 2009; 361(7):645-7.

3. Hripcsak G, Ryan PB, Duke JD, Shah NH, Park RW, Huser V, et al. Characterizing treatment pathways at scale using the OHDSI network. Proc Natl Acad Sci U S A 2016;113(27):7329-36.
4. Go AS, Magid DJ, Wells B, Sung SH, Cassidy-Bushrow AE, Greenlee RT, et al. The Cardiovascular Research Network: a new paradigm for cardiovascular quality and outcomes research. Circ Cardiovasc Qual Outcomes 2008;1(2):138-47.
5. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013;309(13):1351-2.