Using OHDSI Data Network for Capturing Real-World Evidence: Our Experience with a Multi-Country Study on an Obese and Overweight Cohort

Hamed Abedtash
Eli Lilly and Company, Indianapolis, IN, USA

Abstract

We assessed and explored the feasibility of conducting Network Research on OHDSI data network to capture evidences from real-world data (RWD). We tested the capability through a multi-country observational study to characterize obese and overweight cohort across six distributed electronic health records or claims databases. In this paper, we describe the process, lessons learned, and recommendations for a successful Network Research execution.

Introduction

Epidemiologists and real-world evidence (RWE) scientists have been using RWD such as electronic health records and administrative claims databases to capture evidence to support regulatory decisions and to assess the effectiveness and safety of current therapies. The OHDSI Network Research offers an alternative solution to the routine, time-consuming practice where analyzing each RWD source requires a dedicated team of scientists and analysts. By harmonizing data sources and statistical scripts into a common standard through the OHDSI standardized approach, it enables systematics research using the same protocol, statistical plan, and tools on all data nodes across the standardized federated data network.

The OHDSI data network is composed of over 82 data repositories (data nodes) of longitudinal electronic health records of more than 1 billion patients from 17 countries. All data within this network of distributed databases are in OMOP CDM, meaning that both the data structure and data representation (i.e., coding systems) are homogenous across all participating data sources. Adoption of OMOP CDM has enabled conducting "Network Research" by simultaneously executing the same analysis or protocol on a number of harmonized data sources independent of the source coding systems and database structure.

We aimed to assess the feasibility of conducting Network Research on OHDSI data network through a multi-country cohort characterization study on obese and overweight population using six distributed electronic health records or claims databases. In this paper, we share our experience on the process, the lessens we learned, and recommendations for a successful Network Research execution.

Methods

We partnered with a third-party organization (TPO) to execute this Network Research on five US, France, Germany, UK, and Belgium electronic medical records (EMR) and the national South Korean administrative claims databases participating in the OHDSI data network.

This was a cross-sectional, retrospective analysis of demographic and clinical characteristics of obese and overweight cohort. We selected patients 18 years or older with at least two events indicating obesity or overweight condition 6 months apart during the study period. These events included obesity diagnosis, bariatric surgery procedure, anti-obesity medication prescription, body-mass index (BMI) greater than 28 kg/m^2, or if there was any dietary counseling encounter. The first event date was defined as the index date. Continuous observation was required during the study period; continuous enrollment in medical benefit plan in case of claims database and continuous encounters in EMR data.

We analyzed age distribution at index date, gender distribution, and clinical characteristics including BMI, smoking status, systolic and diastolic blood pressure, weight, height, waist circumference, liver function lab tests, lipid profile, glucose profile, anti-obesity medications, concomitant drug use, hypertension occurrence, cardiovascular events, and metabolic syndrome.

We provided the TPO with a requirement document (protocol) describing the rationale, objectives, study design, data coverage tests, and statistical plan. We also evaluated the data coverage per each variable in each data source. We had access to the results throughout the project but not the patient-level data; therefore, Lilly scientist and TPO analyst were in constant communication throughout the study to develop analytical scripts, refine study design, and analyze the data. TPO distributed analytical scripts to data partners for data analysis execution; subsequently each data provider executed the scripts on their data residing on-site and returned the results for further analyses.

Lessons Learned

We successfully selected the cohort of obese and overweight patients from all participating data nodes, and assessed data coverage, and captured a comprehensive list of demographic and clinical characteristics of the cohort. It was also an exceedingly efficient method to analyze multi-database studies, whereas the average time to analyze each data node was 2.6 weeks.

The level of transparency during the analysis process was encouraging. We could ask varieties of questions from the vendor on existing data points, data coverage, and data transformation quality. Since the data was formatted in OMOP CDM, we could ask further quality-check questions to be performed on the data to ensure calculated measures are valid.

Unmapped source codes may pose selection and measurement biases; therefore, we asked the analyst to curate the list of source codes that were not mapped to OMOP standard concepts. These unmapped codes could potentially adversely affect cohort attrition and final endpoint outcomes. To address this issue, we reviewed the list of concepts relevant to our study to ensure validity of mappings in each database.

Recommendations

Validate the quality of data transformation. Ensure relevant source codes are mapped to standard concepts in the OMOP repository to avoid any miscalculations and selection bias. Although OMOP CDM enables quick analysis of multiple distributed RWD repositories, there is a potential risk that not all OMOP-transformed databases have the same data transformation quality due to the complexity of embedded semantic network and OMOP-specific conventions. Thus, it is essential to ensure the quality of data transformation within each OMOP-based repository in every study.

Provide validated code lists before starting data analysis. In one instance, the data coverage of "fasting blood glucose" variable was reported very low (2%). Upon our review of the list of LOINC codes that TPO had used in the analysis, we noticed one important LOINC code (2345-7) was missing. The code was not included in the analysis because there was no mention of "fasting" in the code description, while the code is a legitimate laboratory test code for measuring fasting glucose level in blood because one of the requirements of the test is that "patients should fast for 12 hours". Upon including this code into the analysis, the data coverage jumped to 47% from 2% for the US database, and similarly data coverage improved in other data sources.

Conclusion

We successfully executed the study protocol on obese and overweight population in six countries in less than 4 months. We found the OHDSI Network Research a feasible and efficient approach to conduct multi-database observational studies, in particular on the databases that are not readily obtainable. We also recommended solutions that may improve the efficiency of Network Research and minimize the risk of bias in data analysis.