MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231267

An International Standard RWD Database Designed – Taiwan Experience

Hsiu-An LEE[a], Po-Sheng YANG[b,c,1] and Chien-Yeh HSU[d]
[a] *National Institute of Cancer Research, National Health Research Institutes, Tainan, Taiwan*
[b] *Department of Medicine, MacKay Medical College, New Taipei City, Taiwan*
[c] *Department of General Surgery, MacKay Memorial Hospital, Taipei, Taiwan*
[d] *Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan*

Abstract. The use of Real-World Data (RWD) in medical data analysis is nowadays required. RWD may come from electronic health records (EHRs), insurance claims, medical products, Internet of Things sensors, health screenings, etc. The goal of RWD is that the data used for analysis should not be affected by environmental variables, experimental control, research context settings, etc. RWD can effectively reduce the cost and improve the accuracy of medical research. The clinical data and patient self-report are integrated, cleaned and pre-processed, and the data format is unified and standardized by international standard formats to provide a structured database for clinical research by this study. An international standard real-world research database was established through the breast cancer database of a medical center in Taipei.

Keywords. Real-world data, precision medicine, international standard, breast cancer research, mCODE

1. Introduction

The use of Real-World Data (RWD) in medical data analysis is nowadays required. Actual data are observational data, not data collected in experimental settings such as randomized controlled trials (RCTs), but RWD may come from electronic health records (EHRs), insurance claims (health insurance data), medical products, IoT sensors, health screenings, etc. The goal of RWD is that the data used for analysis should not be affected by environmental variables, experimental control, research context settings, etc. The analysis of RWD is needed to assess patient outcomes and to ensure that patients receive the appropriate treatment for them. The analysis of Real-World Evidence(RWE) requires that RWD data must be of adequate quality, and data quality is defined as consisting of three components [1]. In order to comply with RWE, clinical data databases should be standardized to have adequate data quality. In this study, we designed an RWD compliant medical clinical data research database based on the breast cancer research database of a medical center in Taipei. Through this study, clinical data and patient self-report are integrated, cleaned and pre-processed, and the data format is unified and standardized by international standard formats to provide a structured database for clinical research.

[1] Corresponding Author: Chien-Yeh Hsu, email: cyhsu@ntunhs.edu.tw.

2. Methods

This study consists of three parts, 1. FHIR international standard and mCode for data content, 2. Data conversion quality control process, and 3. de-identification of private data. Through a complete data structure planning and reference of international standards, to design a clinical research database for breast cancer in Taiwan. This study introduces a three-stage operation mechanism including pre-processing operation, data conversion operation and data quality assurance for heterogeneous data pre-processing intermediary system to integrate heterogeneous data in each database.

3. Results

The minimal common oncology data elements of breast cancer were designed based on Taiwan health insurance data structure. The FHIR standard was used to data storage and exchange. Six subgroups were designed for research dataset, which including Patient, Disease, Assessment, Genomics, Treatments, Outcomes, and Organization. Total 11 kinds FHIR resources (Patient, Condition, Observation, Diagnostic Report, Specimen, Medication Request, Medication Administration, Procedure, Body Structure, Identifier, and Organization) were used in this study and 41 resource definitions based on 11 resources. After data conversion by ETL, data scattered throughout the hospital can be retrieved as a single patient. The retrieved data has the advantage of being standardized, structured, and formalized. The overall data can also be quickly and completely integrated across hospitals to meet RWD requirements. The study designed a front-end viewer for the mCODE FHIR data, which allows for quick internal viewing of the data within the hospital.

4. Conclusions

In this study, mCODE FHIR was used as a case study of a hospital breast cancer research database for data format adaptation, automated conversion, and data review. This study confirmed the feasibility of data use and standard adoption, and the next stage should be to establish a data validation and conversion integrity verification process to facilitate the conversion of huge amount of data.

Acknowledgements

This project has received funding from the National Science and Technology Council, Taiwan, under the project No. 111-2321-B-195-001.

References

[1] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, Estiri H, Goerg C, Holve E, Johnson SG, Liaw ST. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC). 2016 Sep;4(1):1244, doi:10.13063/2327-9214.1244.