PRiME-R

Case Studies of Health Data Standardization - Japan Use Cases

Masafumi Okada M.D. Ph.D.
PRiME Research Institute for Medical RWD Inc.

Aug 7, 2023 APEC Forum

Contents

PRiME-R

- Self introduction
- Overview of Health data standards in Japan
- Real world data interoperability
- EMR data interoperability
- Clinical trial data interoperability
- Summary

Self introduction

PRiME-R

- Roles
  - Currently working as an Epidemiologist and the Director of Data Science at PRiME Research Institute for Medical RWD Inc.
  - Lecturer for a real-world data course at the University of Tokyo.
  - Designed a standard EMR template for the collection of cancer treatment data.
- Research activity
  - Collaborator with OHDSI in Japan.
  - Developed a validation tool for CDISC datasets using R language.
  - Primarily interested in real-world data analysis, utilizing data obtained from EMRs.

Overview of Health Data Standards in Japan

PRiME-R

Administrative Claims Data

- Effectively standardized using a straightforward domestic code
- Although the data format is also well standardized, it originated from paper-based claims and requires certain techniques for tabulation

Electronic Medical Records (EMR)

- Diagnoses, prescriptions, and injections are coded using a de-facto standard code in most hospitals
- Other codes are not standardized at all
- Over 50% of larger hospitals (>400 beds) have standardized data storage. However, they do not adhere to FHIR, but to a domestic standard known as SS-MIX2

Clinical Trials

- Data sets collected from clinical trials are converted to CDISC tables for submission to the regulatory agency (PMDA)
- The original data (data from case report forms) is collected in various formats and manually converted to the standard format by data managers

Real world data interoperability

PRiME-R

- Almost all administrative claims are processed online. This means every prescription, measurement, and procedure is recorded using a domestic standard code, along with the date they were performed.
- Diagnoses are also recorded using this domestic standard code alongside the claims, however, this is considered to be inaccurate for medical research purposes, because it is recorded only to justify contents of claims
- As a result, administrative claims data in Japan can be integrated into a single database. This whole Japan database is provided by the government, known as the National Database (NDB), and can be utilized by academia or various industries.
- However, all codes are domestically used and no concept hierarchy or structure is provided. Consequently, we must individually select each code carefully for database study.
- A new claims scheme, the Diagnosis Procedure Combination (DPC), has been applied to approximately 35% of all hospitals.

DPC

PRiME-R

- The DPC scheme of administrative claims is applied to approximately 35% of all hospitals equipped with beds.
- In addition to the claims record, hospitals participating the DPC scheme are required to report diagnoses in a very detailed, yet domestic, classification along with the claims.
- Various patient background data is also reported in a domestic standard format, known as “DPC Form #1”, which includes classifications of heart failure, existence of hospital-acquired pneumonia, emergency room visits, and in-hospital deaths.
- Although data from the DPC is more detailed and standardized than a typical administrative claims dataset, it does not include the results of any measurements or laboratory tests.

EMR data interoperability

PRiME-R

- For diagnosis, a domestic de-facto standard disease code, accessible at byomei.org, is widely used in Japan, along with ICD-10.
- For medication, another widely utilized domestic de-facto standard code is the YJ code.
- For measurements, laboratory tests and procedures, there is NO standard code. Every hospital uses their own unique code.
  - Some procedures are also documented in administrative claims, which allows us to collect information in a domestic standard code.
- In addition to the issue of code incompatibility, the structure of databases is not standardized.
  - However, there exists a domestic data structure standard for clinical data archive : SS-MIX2.
- Consequently, merging EMR data across different hospitals remains a significant challenge in Japan. A combination of diagnosis data from EMR and drugs and procedure data from administrative claims could offer a better data source. However, due to the incompatibility of codes and formats, collecting measurement or test results from Japanese hospitals is the most challenging task.

SS-MIX2

PRiME-R

- 59.4% of large hospitals (>400 beds), accounting for 22.1% of all hospitals, have implemented SS-MIX2, a Japanese domestic standard for EMR archive data storage.
- SS-MIX2 is an archive of HL7 (v2) messages transferred within hospital information systems, stored in file system folders labeled by patient ID and date.
- It is primarily intended to serve as a backup for the hospital's database for disaster recovery.
- As a result, SS-MIX2 could function as a simple API for accessing the data in EMR via HL7v2 messages.
- In implementing FHIR endpoints, I anticipate that the existing SS-MIX2 could be utilized as a common data source independent of each hospital's individual database.
- However, given that the data within SS-MIX2 consists solely of HL7v2 messages, there is a substantial variety in the codes used and the types of messages used to store information.

Clinical trial data interoperability

PRiME-R

- As the Japanese regulatory agency (PMDA) has been accepting clinical trial data in CDISC standard format since 2016, pharmaceutical industries are now equipped to maintain their clinical trial data sets in this standard format.
- While standardization of the source data (recorded in the case report form) is not mandated, analysis-ready datasets are typically available in the CDISC SDTM and ADaM format for trials conducted by industries.
- For trials conducted by academia, the majority of data is not standardized, especially for protocols that are not intended for data submission to PMDA.
- Therefore, while the integration of data from multiple clinical trials could technically be straightforward due to their high level standardization, such data typically includes private assets of pharmaceutical companies and are not shared.

Summary

PRiME-R

Health Data Standardization in Japan: Challenges and Opportunities

- **Administrative Claims Data:** Uniform domestic codes have been successfully applied, but the data is limited in precision for research due to its granularity and the lack of hierarchical structure in coding.
- **Electronic Medical Records (EMR):** While most hospitals use a de-facto standard code for diagnoses, prescriptions, and injections, lack of standardization for other codes and differing database structures present major obstacles for data integration.
- **Clinical Trials:** Conversion to CDISC tables provides high level of standardization, but source data formats can vary widely. Most trial data remains proprietary and unshared.
- **Real World Data & DPC:** Despite challenges with domestic codes and lack of measurement results, administrative claims data and the DPC scheme offer potential for valuable real-world data collection and integration.
- **SS-MIX2:** Though it presents its own set of challenges, the domestic data structure standard could provide a path toward easier EMR data access and integration with the FHIR.

Overall, while Japan has made strides in health data standardization, significant challenges remain, particularly around the use of domestic codes, lack of standardization in EMR. Future efforts should focus on addressing these issues by providing mapping between global standard and domestic code to make Japanese health data ready to integrate with global standard.