Common data models (CDMs) are used to organize clinical data from complex, disparate source systems into a standard format. CDMs can provide the opportunity to run standardized queries on multiple healthcare information systems. Approach can be used to increase size of studies with typically small cohorts, such as rare disease and post-market surveillance research. Previously, no CDM available at YNHH, which prevented some opportunities for collaboration.

Objectives

- Implement an extract, transform, and load (ETL) of the gold standard data from the YNHH clinical data warehouse (Epic Clarity) into a CDM (PCORnet)
- Incorporate CDM into a data science platform
- Validate the ETL and use of the tool with a post-market surveillance use case at multiple institutions
- Assess efficacy of CDM to identify potential adverse events in post-market surveillance use case

Methods

- Data is extracted from the Epic Clarity data warehouse with custom SQL queries
- Custom Python script applies mappings to the CDM attributes and validates referential integrity
- Records that could not be transformed to the CDM are flagged with descriptive error logging

Results

Figure 1: PCORnet CDM v3.1 Schema

Figure 2: Project and Validation Workflow

Figure 3: YNHH Epic Clarity to PCORnet Architecture

Conclusions

- CDMs can be used to represent commonly accessed data for outcomes-based research
- Big data tools can be used to parallelize ETL processes for CDM implementation
- Data science platforms may allow researchers to easily link and extend CDMs without altering the approved CDM architecture

Future Work

- Complete validation testing of the YNHH CDM implementation
- Complete post-market surveillance validation with partner institution (Mayo Clinic) and the FDA based on an immuno-oncology use case
- Assess the performance of the PCORnet CDM in traditional SQL and within the YNHH data science platform based on real-world use case
- Identify opportunities to extend the CDM with other big data sources within the data science platform (i.e., genomics, real-time monitoring) to support outcomes research

References


This work was supported by broad agency announcement award from the US Food & Drug Administration (HHSF223201710167) and by a Center of Excellence in Regulatory Science and Innovation (CERSI) grant to Yale University and Mayo Clinic from the US Food & Drug Administration (U01FD005938). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the HHS or FDA.

Acknowledgements