Validation of a Common Data Model for Clinical Research and Post-Market Surveillance

H. Patrick Young PhD[1], Will M. Byron MBA[2], Charlie J. Torre Jr MS[2], Guoqian Jiang MD, PhD[3], Nilay D. Shah PhD[3], Harlan M. Krumholz MD, SM[1,4], Joseph S. Ross MD, PhD[1,4], and Wade L. Schulz MD, PhD[1,5]

[1]Yale New Haven Hospital, Center for Outcomes Research and Evaluation; 1 Church St #200, New Haven, CT 06510
[2]Yale New Haven Health, Information Technology Systems; 20 York St, New Haven, CT 06520
[3]Mayo Clinic, Department of Health Sciences Research; 200 First St SW, Rochester, MN 55905
[4]Yale University School of Medicine, Department of Internal Medicine, 330 Cedar St, Boardman 110, New Haven CT, 06520
[5]Yale University School of Medicine, Department of Laboratory Medicine; 55 Park Street PS345D, New Haven, CT 06511

Background

- ❖ 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
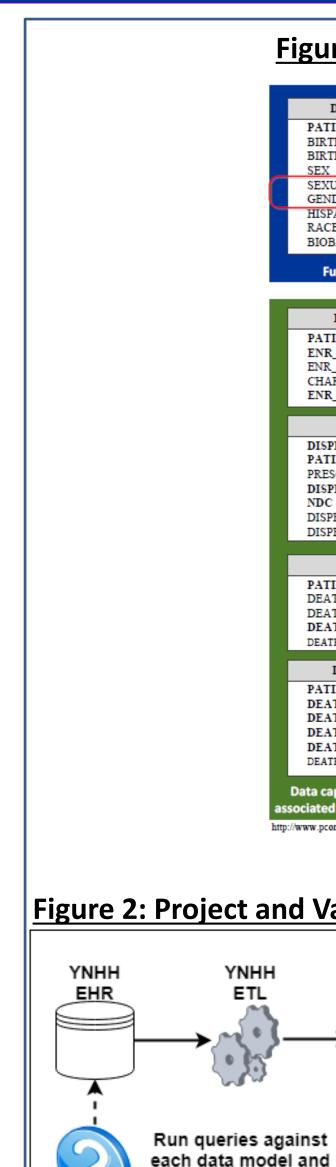- ❖ Previously, no CDM available at YNHH, which prevented some opportunities for collaboration
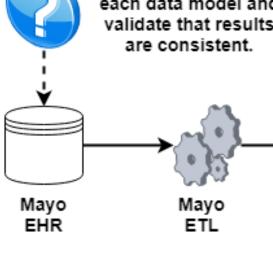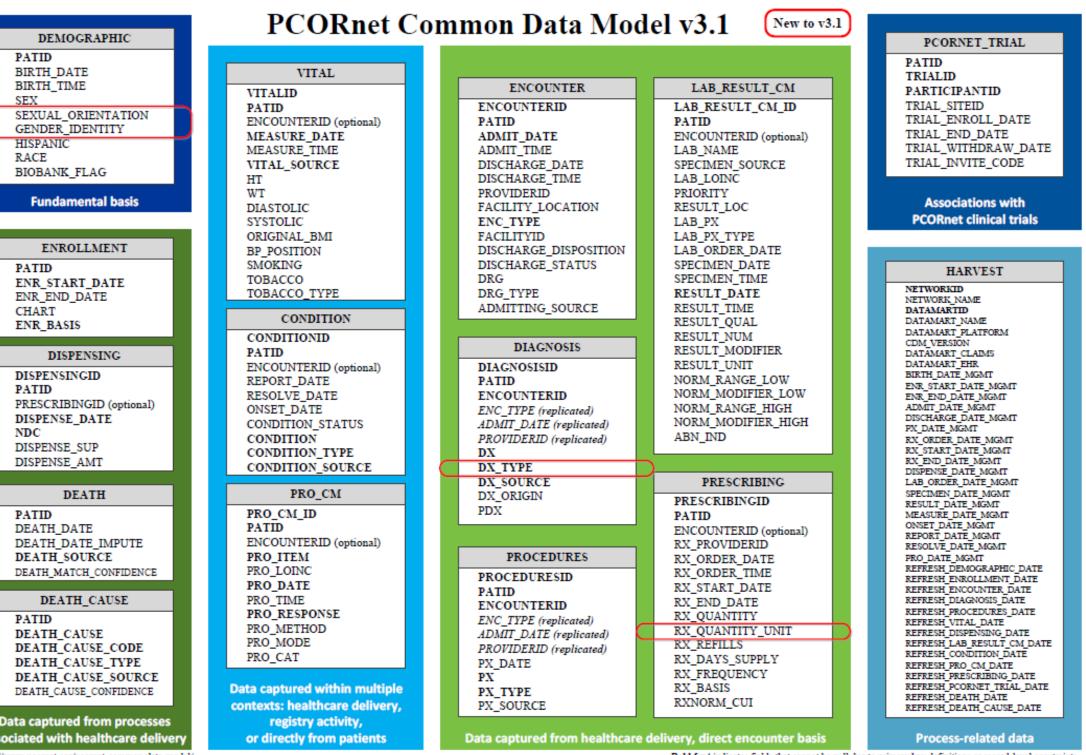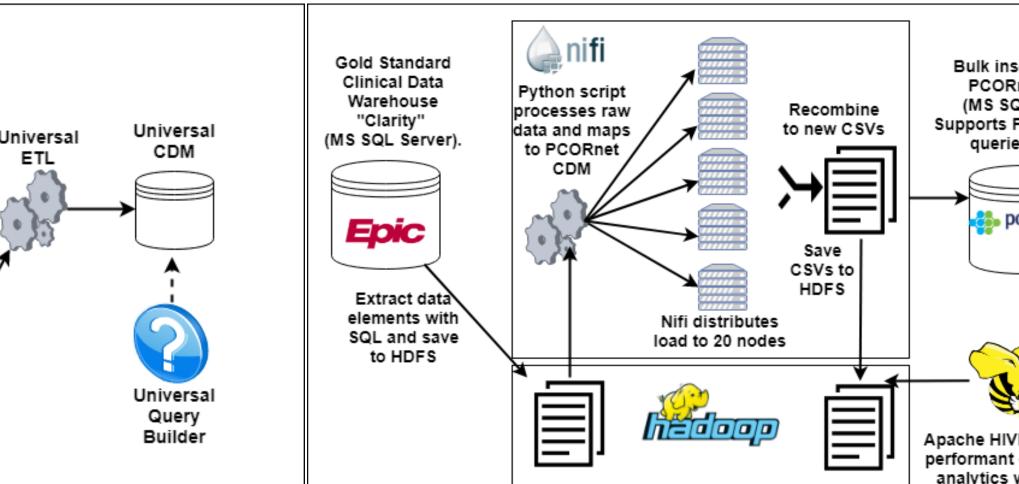
Objectives

- ❖ 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

1. Cook and Collins; Br J Surg; 2015; 102:e93-e101
2. Klann et al. JAMA 2016;909-915.

Acknowledgements