# Transforming the National Department of Veterans Affairs Data Warehouse to the OMOP Common Data Model

Fern FitzHenry[1,2], Jesse Brannen[1], Jason Denton[1,2], Jonathan R. Nebeker[3,4], Scott L. DuVall[3,4], Freneka Minter[1,2], Jeffrey Scehnet[3], Brian Sauer[3,4], Lucila Ohno-Machado[5], Michael E. Matheny[1,2]

[1]Tennessee Valley Healthcare System, Veterans Affairs Medical Center, Nashville, TN; [2]Vanderbilt University, Nashville, TN; [3]VA Salt Lake City Health Care System, Salt Lake City, UT; [4]University of Utah, Salt Lake City, UT; [5]Bioinformatics and Systems Biology, University of California, San Diego, CA;

Abstract: To describe the conversion of the national Department of Veterans Affairs (VA) healthcare network's corporate data warehouse to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) suitable for distributed observational research. Observational outcomes from electronic medical record systems are becoming more important in comparative effectiveness research, particularly as post marketing surveillance research.[1]

Background: The VA has the largest integrated healthcare system in the US with electronic health record coverage from the late 1990s. It is one of the few healthcare systems with lifetime coverage of covered veterans as well as greater continuity in patient coverage because eligibility is not related to the changes in health insurance seen in civilian health systems.

Methods: The common data model will include data from Veterans collected through the electronic health record and stored in the corporate data warehouse, clinical registries collected within the VA, and records provided to the VA from the Department of Defense, Medicare and USRDS extracts. The VA corporate data warehouse has undergone an initial transformation into OMOP CDM version 4 with large-scale data population beginning in 1999 (1998 for drugs). The OMOP CDM allows shared data analysis to have syntactic and semantic interoperability through standardized terminologies such as ICD, RxNorm, SNOMED-CT, CPT, HCPSC, and LOINC. Achilles Heel has been run to determine basic data validity and quality errors, and iterative fixes are underway.

Results: There are approximately 16,927 thousand unique patients in the dataset, with 11,368 thousand of those having at least one encounter (see Table 1). The ETL process for cost data have not yet been done. As part of a cooperative effort with the patient-centered Scalable National Network for Effectiveness Research (pSCANNER), we are also transforming the OMOP CDM to the PCORnet data model. The OMOP transformation of the VA data will be available for use within VINCI for any VA investigator in both formats. The Observational Health Data Sciences and Informatics (ODHSI) community is developing a suite of data characterization, data quality, cohort generation, comparative effectiveness analysis, and surveillance tools (www.ohdsi.org), and VINCI plans on integrating these tools into the VA research environment, and providing programming support to meet all federal software requirements, as well as participating in tool development.

Conclusions: The national data from the corporate data warehouse that is represented within the OMOP CDM was successfully transformed. As part of a cooperative effort with PCORI Clinical Data Research Networks and the pSCANNER CDRN, we are also transforming the OMOP CDM to the PCORnet data model. Ongoing work will be undertaken to iteratively improve data quality and maintain support of the CDMs for widespread use.

Impact: The VA OMOP and PCORNet CDMs will be available for the research initiatives of VA researchers. The Observational Health Data Sciences and Informatics (ODHSI) community is developing a suite of data characterization, data quality, cohort generation, comparative effectiveness analysis, and surveillance tools (www.ohdsi.org), and VINCI plans on participating in tool development and supporting use of these tools within the computing infrastructure to support observational cohort analytics.

Table 1: VA Domains and Record Counst

| Domains | Records (Thousands) |
|---|---|
| Persons | 16,927 |
| Deaths | 5,173 |
| Organizations | 3 |
| Locations | 20,539 |
| Care Site | 861 |
| Provider | 973 |
| Visits | 2,002,743 |
| Procedures | 2,817,943 |
| Drug exposures | 4,068,709 |
| Drug Era | 661,069 |
| Observations | 9,507,140 |
| Observation Period | 25,587 |
| Conditions | 2,470,374 |
| Condition Era | 1,038,159 |

References:

1. Olsen LA, McGinnis JM. Redesigning the Clinical Effectiveness Research Paradigm: Innovation and Practice-Based Approaches: Workshop Summary. In: Medicine Io, ed. Washington, DC: National Academies Press; 2010.