MEDINFO 2019: Health and Wellbeing e-Networks for All L. Ohno-Machado and B. Séroussi (Eds.) © 2019 International Medical Informatics Association (IMIA) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHT1190561

# Characterizing VA Users with the OMOP Common Data Model

## Benjamin Viernes, Kristine E Lynch, Brett South, Gregorio Coronado, Scott L DuVall

VA Informatics and Computing Infrastructure, Department of Veterans Affairs, Salt Lake City, UT, USA, Department of Internal Medicine, Division of Epidemiology, University of Utah, Salt Lake City, UT, USA

## Abstract

In 2015, the VA Informatics and Computing Infrastructure, a resource center of the Department of Veterans Affairs, began to transform parts of its Corporate Data Warehouse (CDW) into the Observational Medical Outcomes Partnership) Common Data Model for use by its research and operations communities. Using the hierarchical relationships within the clinical vocabularies in OMOP we found differences in visits, disease prevalence, and medications prescribed between male and female veterans seen between VA fiscal years 2000-17.

## Keywords:

Electronic health ecords; systematized nomenclature of medicine; veterans

## Introduction

The development and implementation of common data models (CDMs) address some of the logistical challenges of performing research on data generated from disparate healthcare systems by using standardized terminology to express clinical information unambiguously.[1] Even within a single healthcare system, a CDM can be advantageous as the standard representation of clinical concepts improves data quality, allows for rapid and reproducible execution of data management and analysis, and reduces barriers to entry for new researchers to complicated data systems such as electronic health records (EHR). The Veterans' Affairs (VA) is the largest integrated healthcare system in the United States and its EHR is especially complex, with data being sourced from what is analogous to >130 distinct medical systems. Theoretically, the same clinical concept (e.g., blood glucose lab test) may be represented more than 130 different ways, to represent each distinct VA medical facility.

In 2015, the VA Informatics and Computing Infrastructure (VINCI) begain transforming parts of the VA's Corporate Data Warehouse (CDW) into the Observational Medical Outcomes Partnership (OMOP) CDM (hereinafter referred to as VA OMOP). VA OMOP version 5.2 contains a large portion of the CDW and allows for less complex, high-level views of many parts of the VA healthcare system. While OMOP is available to VA researchers, to date no one has provided a comprehensive clinical characterization of the VA population using OMOP. Thus, we sought to complete a characterization of veterans who utilize the VA for healthcare services using VA OMOP.

## Methods

We compared male and female veterans with respect to visits, conditions, and medications. We utilized the Visit\_Occurrence, Drug\_Exposure, Condition\_Occurrence, Death, and Person fact tables of the OMOP CDM for this analysis. Comparatively, these 5 OMOP tables comprised data transformed from >14 patient level fact tables and a large number of meta-data tables from the CDW data model.

#### **Inclusion Criteria**

Using the Visit\_Occurrence table, we required patients to have an outpatient visit between fiscal years (FY) 2000 to 2017. Although VA enrollment data was not yet available in VA OMOP, we required patients to have been enrolled in the VA after 10-01-1999 (FY 2000) to ensure that we were only analyzing patients whose VA medical history began during our study period. Patients had to have a valid birthdate (i.e., be 18-114 years of age on the first day of the FY of their first visit), have a defined gender (male or female), and a date of death, if applicable, within or after the FY of their first visit.

## Person and Visit Characteristics

We assessed basic patient characteristics using the Person table, a one row per patient table that contains information on demographic factors. We described the VA population using data representing characteristics of race, ethnicity, and deaths. Using the Visit Occurrence table we calculated the number of outpatient visits, counts of new patients, and counts of returning patients, each overall and by year. The Visit Occurrence table specified inpatient, outpatient, and emergency room visit types, and for this analysis, outpatient and emergency room visits were combined as outpatient visits. To understand the difference between genders, we compared mean visits per year by gender, percent of all new patients by year and gender, and percent of all patients who are returning patients by year and gender. Finally, we adjusted the mean visits per year per patient by creating rules for patient activity. Assuming that patient activity required constant system use, and based on distributions of visits, a FY was counted as active for a patient if a visit occurred during the year and there was no 1 year no gap preceding a year of no visits.

#### **Condition Occurrence**

We extracted the most common conditions by Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) classifications, OMOP's standard vocabulary for conditions. Within the SNOMED hierarchy, many levels were available to potentially evaluate the clinical concepts that described records in the OMOP Condition Occurrence table. After identifying the most common overall SNOMED concepts in the Condition Occurrence table, it was determined that the SNOMED 'Disease' concept (i.e., SNOMED code 64572001) and its related descendants was an optimal method for evaluating the burden of different diseases in the VA. This approach created disease classifications with minimal deviation from the standard mappings of International Classification of Disease diagnostic codes, the VA's condition vocabulary, to SNOMED. There were 89 immediate descendants, of the SNOMED disease hierarchy, of which only 74 had been observed in VA from FY 2000-2017. These 74 descendants helped to organize the 13,922 descendant concepts that described diagnoses in the VA OMOP.

We created a frequency table of the SNOMED disease child (e.g., mental disorder), the disease descendant (e.g. depressive disorder), and assessed differences between genders. The most prevalent diseases were then organized by overall patient prevalence during the study period, the most prevalent by each gender, and the comparative rankings for the opposite gender for highly prevalent diseases among each.

## Drug exposure

RxNorm is the standard vocabulary to represent drug exposures. We used the RxNorm hierarchy that has been preprocessed and built into the OMOP Concept\_Ancestor table to describe outpatient prescriptions at the level of ingredient. The ingredient can be derived from any RxNorm clinical drug form through the hierarchy of the RxNorm vocabulary. Of 7,612 unique RxNorm ingredient values present in the OMOP Concept table, 1,688 RxNorm ingredients had been prescribed in the VA with 1.1 million clinical drug descendants.For the most common prescribed ingredients, we assessed the patient frequency of prescription ingredients and gender differences.

## Results

After applying our inclusion criteria, our final analytic sample contained 8,401,080 distinct Person\_IDs with at least one outpatient visit from FY 2000-17. The majority (93%) were male, white (65%), and not Hispanic/Latino (77%). For the male population, 66.2% identified as white, and 10.9% identified as black. For the female population, 54.4% identified as white and 24.8% identified as black. Females were statistically significantly younger at age of first visit (39.75) compared to men (56.28). During the 18-year study period, 27% of males died compared to 7% of women.

The most common conditions as defined by SNOMED disease

Table 1. Ten Most Common SNOMED Diseases and Corresponding Rank and Percentage by Gender, VA 2000-17

Disease concept P	ercent	Male	Female
1. Hypertension	53.3	1 (55.1)	4 (28.3)
2. Hyperlipidemia	47.9	2 (49.3)	3 (28.4)
3. Presbyopia	25.3	3 (25.7)	8 (19.6)
4. Sensorineural hearing loss, ilateral	22.5	4 (23.8)	62 (6.5)
5. Obesity	22.6	5 (22.2)	5 (27.3)
6. Type 2 diabetes	20.9	6 (21.8)	42 (8.8)
7. Gastroesophageal reflux disease	21.4	7 (21.7)	9 (18.5)
8.Tobacco dependence syndrome	20.9	8 (21.1)	11 (17.4)
<ol><li>Depressive disorder</li></ol>	21.9	9 (21.1)	1 (32.9)
10. Nuclear senile cataract	17.2	10 (17.9)	43 (8.6)

descendants were hypertension and hyperlipidemia with over 50% of all veterans having a recorded diagnosis (see Table 1 for comparison of conditions between genders). The order of most prevalent conditions in women (highest to lowest) was depressive disorder (32.9%), anxiety disorder (28.7%), hyperlipidemia and hypertension. Of the top ten prescribed medications, five were for pain management, two for blood pressure, one for cholesterol, one was an antibiotic, and one for gastroesophageal reflux disease (see Table 2). Female veterans had lower prevalence of many of the most commonly prescribed medications for chronic diseases. This makes sense intuitively due to the fact that on average VA women are younger than men and have lower prevalence of hypertension and diabetes. The top three drugs for women were for pain. Aside from aspirin, men had lower rates of prescription for all pain medications in the top ten.

Table 2. Ten Most Common RxNorm Drug Ingredents and Corresponding Rank and Percentage by Gender, VA 2000-17

Ingredient	Percent	Male	Female
1.Acetaminophen	33.9	1 (33.5)	1 (39.0)
2.Simvastatin	30.3	2 (31.5)	33 (13.9)
3.Lisinopril	29.8	3 (31.0)	35 (13.4)
4.Omeprazole	26.2	4 (26.3)	4 (24.7)
5.Hydrocodone	22.6	6 (22.3)	3 (25.4)
6.Hydrochlorothiazide	22.3	5 (22.8)	25 (14.8)
7.Ibuprofen	18.4	7 (17.5)	2 (28.8)
8.Amoxicillin	17.5	9 (17.1)	7 (22.4)
9.Aspirin	16.8	8 (17.4)	62 (8.8)
10.Naproxen	16.5	11 (16.0)	6 (23.3)

Outpatient visits increased over time from FY 2000-17. This occurred for both genders, but for women, the number of visits had a higher rate of increase than men. In 2000, patients had an average of 4.84 visits per year with little difference between men and women. By FY 2017, that average had increased to 13.31 visits with women averaging almost three more visits per year than men. Although the majority of new patients each year were male, the relative proportion of new women veterans increased from 5% in 2000 to 9.4% in 2017.

## Conclusions

A characterization analysis of VA users is important to best align research questions and methodological approaches to the VA EMR data, and aligns with the educational mission of VINCI. Additionally, by leveraging the hierarchical features of the OMOP CDM, we were able to understand gender differences in ways that would have been more difficult with VA CDW.

#### Acknowledgements

This work was supported using resources and facilities at the (VINCI), VA HSR RES 13-457.

### References

[1] Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, DeFalco FJ, Londhe A, Zhu V, and Ryan PB, Feasibility and utility of applications of the common data model to multiple, disparate observational health databases, *J Am Med Inform Assoc.* 22 (2015), 553-564.

## Address for correspondence

Scott L DuVall,VINCI, Director Scott.Duvall@utah.edu