Applying OMOP Tools and Methods to Nursing Home Data

Richard D. Boyce, PhD[1,2], Jeremy Jao[2], Steven M. Handler, MD, PhD[1,2]
[1]Center for Pharmaceutical Policy and Prescribing, University of Pittsburgh
[2]Department of Biomedical Informatics, University of Pittsburgh

Introduction:

Adverse drug events (ADEs), defined by the by the Institute of Medicine as injuries resulting from a medical intervention related to a drug, are a significant cause of harm to patients residing in the nursing home setting. Pharmacoepidemiology can identify and characterize drugs-event associations that occur in the nursing home from observational data. However, methods in the field depend on a deep understanding of the strengths and limitations of a given observational dataset. The objective of this study is to use OMOP tools and methods to characterize a de-identified dataset of nursing home drug dispensing, Minimum Dataset 3.0 (MDS), and fall incidence data in terms of its usefulness for pharmacoepidemiology.

Methods in Brief:

A de-identified dataset containing two years of drug dispensing and MDS 3.0 data from five nursing homes was translated and loaded into version 2 of the OMOP CDM [1,2]. In order to reduce noise of the many MDS 3.0 data categories, raw data was placed into a simpler pipe-delimited CSV file which was then loaded into an Oracle database. Java and Hibernate was used to translate the simplified MDS data into the CDM table structure and vocabulary.

OSIM2 was used to generate three simulated nursing home populations approximately 1, 2, and 4 times the size of the original population. In a separate process, we began to compare the accuracy of exposure periods generated from the drug dispensing and MDS data to nursing medication administration records.

Figure 1.The process for extracting, translating, and loading the nursing home data, and for running simulations.

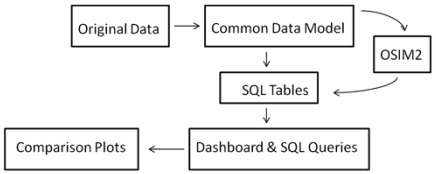

Figure 2. Female age frequencies in real data (n=~2,800) vs OSIM2 (n=12,000)

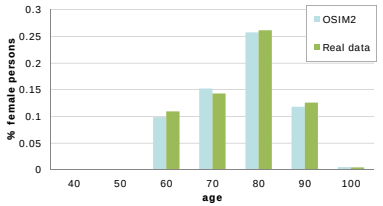

Figure 3. Male age frequencies in real data (n=~2,800) vs OSIM2 (n=12,000)

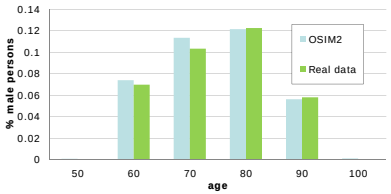

Figure 4. Distinct drugs in real data (n=~2,800) vs OSIM2 (n=12,000)

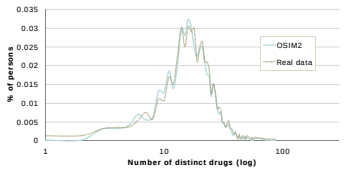

Figure 6. Drug co-occurrence prevalence in real data (n=~2,800) vs OSIM2 (n=12,000)

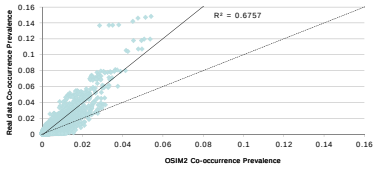

Figure 5. Distinct conditions in real data (n=~2,800) vs OSIM2 (n=12,000)

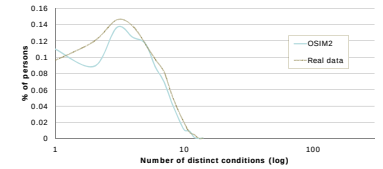

Figure 7. Condition co-occurrence prevalence in real data (n=~2,800) vs OSIM2 (n=12,000)

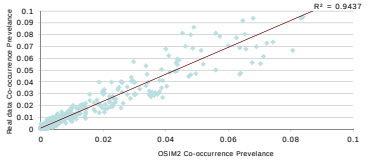

Figure 8. Issues with NH drug dispensing data

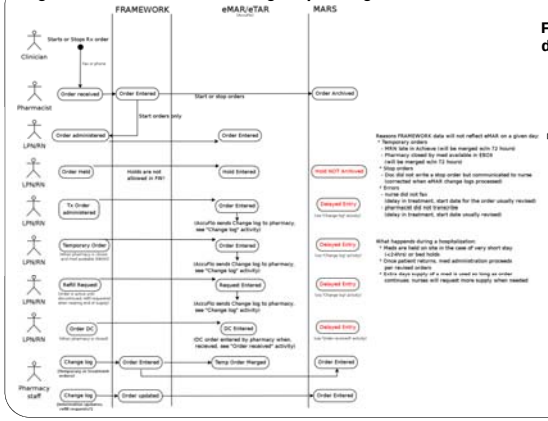

Figure 9. A pilot comparison of NH drug dispensing data to eMAR data

#Pt	Month/Year	%Start dates correct	%End dates correct
Pt1	Jan 2012	0.69	0.62
Pt2	April 2012	0.58	0.58
	Feb 2012	0.37	0.95
Pt3	Jan 2011	0.11	0.94
	Feb 2011	0.57	0.64
Pt4	Mar 2012	0.86	0.79
Pt5	Oct 2011	0.64	0.71
	Sept 2012	0.86	0.71
Pt6	Feb 2012	1.0	0.92
	April 2012	0.68	0.78
Pt7	Feb 2012	0.8	1.0

Results:

The final dataset contained drug exposure, demographics, and 59 health conditions for 2,859 individual nursing home residents. OSIM2 was ran successfully on the nursing home dataset to generate simulated populations of 3,000, 6,000, and 12,000 persons.

Figures 2 and 3 show a that the age and gender distribution between the original and simulated population of 12,000 was very similar. Likewise, the number of distinct drugs and conditions per person was very similar (Figures 4 and 5). There was a moderate correlation between the real and simulated data for drug co-occurrence prevalence (Figure 6), but a strong correlation between the datasets for condition co-occurrence (Figure 7). We could not compare age specific drug and condition prevalence between the datasets because of a "division by zero" error in OSIM2. This was likely due to the lack of patients in the real population that can be placed in certain age buckets (i.e., <18).

Figure 8 shows an activity diagram showing how data from drug prescribing, dispensing, and administration is captured from the study nursing homes. Figure 9 shows our early work comparing the drug exposure start and stop dates that we can derive from dispensing + MDS 3.0 data with what nurses record in medication administration records.

Conclusions and Future work:

We found the OMOP CDM and OSIM2 simulator is useful for generating simulated nursing home datasets, but the OSIM2 code will need revision to address division by zero errors. We plan to fix this issue and contribute the fix back to OMOP. We will also complete the validation of dispensing data. We then plan to alter the simulated datasets to include known drug-event associations and than test various pharmacoepidemiologic methods provided by the OMOP group. This will help to specify the kinds of novel drug-event associations that can be identified in the nursing home population.

Acknowledgement:

This work was funded in part by grant K12HS019461 from the Agency for Healthcare Research and Quality. Figure 9 was created by CoSBBI student David Inman.

References:

1. Murray R, Ryan P, Reisinger S: Design and Validation of a Data Simulation Model for Longitudinal Healthcare Data. UnitedBioSOurce Corporation, Harrisburg, PA
2. Ryan, P. B., D. Griffin, et al. (2009). "OMOP Common Data Model (CDM) Specifications." Retrieved 30 July 2013, from http://omop.fnih.org/CDMandTerminologies