Generating realistic simulated administrative health data for population-based drug safety and effectiveness research and training

CURRENT STATUS: POSTED

Olawale Ayilara
University of Manitoba

Robert W Platt
McGill University

Matthew Dahl
Manitoba Centre for Health Policy

Janie Coulombe
McGill University

Pablo Gonzalez Ginestet
Karolinska Institutet

Dan Chateau
University of Manitoba Faculty of Health Sciences

Lisa M Lix

✉ lisa.lix@umanitoba.ca *Corresponding Author*
ORCID: <https://orcid.org/0000-0001-8685-3212>

DOI:

10.21203/rs.2.18325/v1

SUBJECT AREAS

Health Economics & Outcomes Research

KEYWORDS

administrative health records, computer simulation, prescription drug records, 69 validation

Abstract

Background: Administrative health records (AHRs), which are generated primarily for management and billing purposes, are now widely used in drug safety and comparative effectiveness studies. The development of analytic methods for multi-site studies can benefit from the availability of simulated data, which do not require ethical approvals and data access permissions. We simulated AHRs using both the Observational Medical Dataset Simulator II (OSIM2) proposed by the Observational Medical Outcomes Partnership, and a modified OSIM (ModOSIM) method developed by the Canadian Network for Observational Drug Effect Studies (CNODES). Our objective was to compare the simulated data to real-world AHR data to assess the representativeness of the simulated data.

Methods: The real-world data comprised prescription drug records for all individuals with healthcare coverage at any point in a 10-year period (2008 – 2017) from the Manitoba Population Research Data Repository (MPRDR) in the province of Manitoba, Canada. OSIM2 and ModOSIM, which are empirical simulation models for longitudinal patient data, were used to simulate AHRs. The data were described using frequencies and percentages. We estimated agreement of prescription drug use measures in MPRDR, OSIM2 and ModOSIM using the concordance coefficient.

Results: The MPRDR cohort included 169,586,633 drug records and 1,395 drug types for 1,604,734 individuals. Data for 50,000 individuals were simulated using OSIM2 and ModOSIM. Sex and age group distributions were similar in the real-world and simulated data. There were significant differences in the total number of drug records and number of unique drugs for OSIM2 and ModOSIM when compared with MPRDR; the median number of unique drugs in MPRDR, OSIM2 and ModOSIM was 9.0, 6.0 and 10.0, respectively. For average number of days of drug use, concordance was 16% (95% confidence interval [CI]: 12% – 19%) for MPRDR and OSIM2 and 88% (95% CI: 87%-90%) for MPRDR and ModOSIM.

Conclusions: ModOSIM data were more similar to MPRDR than OSIM2 data on many measures of prescription drug use. Simulated AHRs that are consistent with those found in real-world settings can be generated using ModOSIM; these simulated data will benefit methodological studies and data analyst training.

Full-text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed.

However, the manuscript can be downloaded and accessed as a PDF.

Figures

Figure 1

Average number of days of prescription drug use for real-world data from the 392 Manitoba Population Research Data Repository (MPRDR) and simulated data from OSIM2 393 and ModOSIM

Figure 2

Number of prescription drug records, stratified by sex, for real-world data from the 397 Manitoba Population Research Data Repository (MPRDR) and simulated data from OSIM2 398 and ModOSIM

Figure 3

Number of prescription drug records, stratified by age group, for real-world data 401 from the Manitoba Population Research Data Repository (MPRDR) and simulated data from 402 OSIM2 and ModOSIM

Figure 4

Prescription drug use measures, stratified by health condition, for real-world data 407 from the Manitoba Population Research Data Repository (MPRDR) and simulated data from 408 OSIM2 and ModOSIM