International Journal of Population Data Science

Chronic Disease Case Definitions for Electronic Medical Records: A Canadian Validation Study

Lix, Lisa[1*], Singer, Alexander[1], Katz, Alan[1,2], Yogendran, Marina[2], and Al-Azazi, Saeed[1]

[1]University of Manitoba

[2]Manitoba Centre for Health Policy

Objectives

Canadians are investing heavily in electronic medical records (EMRs) to inform primary care practice improvements. The Canadian Primary Care Sentinel Surveillance Network (CPC-SSN) is a national practice-based network that has enrolled more than one million patients to date. Accurate CPCSSN EMR data are essential for unbiased research about chronic disease prevention and management. The study purpose was to test the accuracy of chronic disease case definitions in EMR data from one CPCSSN site.

Approach

This study linked CPCSSN EMR data, hospital records, physician billing claims, prescription drug records, and population registration files for the province of Manitoba. Individuals who had at least one encounter with a CPCSSN practice between 1998 and 2012, were at least 18 years of age, and had a minimum of two years of healthcare coverage before and after the study index date were included. Separate cohorts were defined for the following chronic diseases: chronic obstructive pulmonary disease (COPD), depression, diabetes, hypertension, and osteoarthritis. Validated case definitions based on diagnoses in physician and hospital records and prescription drug data were used estimate sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and kappa of each EMR chronic disease case definition.

Results

More than 74,000 individuals were included in each cohort, except for COPD which had 51,000. Approximately half of each cohort was comprised of urban residents. The average age ranged from 45.9 years for individuals with depression to 65.3 years for individuals with COPD. Hypertension had the highest prevalence (22.0%) in EMR data followed by depression (14.6%). Estimates of agreement (i.e., kappa) for EMR and administrative data ranged from 0.47 for COPD to 0.58 for diabetes. Sensitivity of the EMR data was lowest for COPD (37.4%; 95% CI 36.0-38.8) and highest for diabetes (57.6%; 95% confidence interval [CI] 56.6-58.6). PPV estimates were lowest for osteoarthritis (66.9%; 95% CI 66.0-67.8) and highest for hypertension (78.3%; 95% CI 77.7-78.9). Specificity estimates were consistently above 90% and NPV estimates were always greater than 80%. Validity estimates for the EMR case definitions were associated with demographic and comorbidity characteristics of the study cohorts.

Conclusion

Validity of EMR data, when compared to administrative health data, for ascertaining five different chronic diseases was fair to good; it varied with the disease under investigation. Further research is needed to identify methods for improving the accuracy of chronic disease case definitions in EMR data.

*Corresponding Author:
Email Address: Lisa.Lix@umanitoba.ca (L. Lix)