

Submission Id: 4964

Title

CPCSSN Data Quality: An Opportunity for Enhancing Canadian Primary Care Data

Priority 1 (Research Category)

Healthcare informatics

Presenters

Rachael Morkem, MSc, Richa Shah, Ayat Salman, David Barber, MD, BSc, Sabrina Wong, PhD, RN

Abstract

Context: Building a source for pan-Canadian EMR data, which has a complex and geographically varied healthcare system, is challenging. For more than a decade, the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) has been working to develop and standardize primary care data to ensure it is sufficient quality to be a valuable source for clinicians, researchers, and policy makers. A data quality (DQ) framework was developed to evaluate the CPCSSN database. Objective: to assess two DQ dimensions (1) accuracy and reliability; and (2) comparability and coherence, using evidence-based indicators. Study Design and Analysis: Three indicators were used: (a) element presence- the completeness of common data elements expected to be present or 'not null'; (b) data source agreement-how information derived from CPCSSN compared to other sources of information; and (c) data across jurisdictions and sources- the prevalence of common data elements across sites, EMR type and province. We used data that included records up until June 30, 2022. Outcome measure: (a) % present of common data elements within the database; (b) prevalence of common chronic diseases; and (c) prevalence of common ICD-9 codes, medication codes and lab codes. Results: Coded fields within CPCSSN are $\geq 93\%$ complete for demographic elements. Diagnostic data is highly present in uncoded fields (<6% null) but shows some missingness in coded fields (~75% present). Medication and lab names are well captured (> 99% present) but medication specifications (ex. duration, frequency) need standardization. The prevalence of common chronic diseases estimated using CPCSSN data are reasonable and comparable to estimates from administrative and survey data. Comparing common diagnostic, medication, and lab codes across site, EMR type and province shows that there is a great degree of variation in the use of these common codes at each site, which is influenced by EMR type and province. Conclusions: The CPCSSN database has reasonable DQ in terms of accuracy and reliability, and comparability and coherence when it is used for epidemiological research. The indicators highlight the extensive work CPCSSN has done to create coded, standardized information. We recommend CPCSSN operations continues to develop cleaning and processing tools to reduce missingness in coded fields. It is recommended that users request identification of site, EMR and province so that clustering can be accounted for in the analysis.