

Methods for Defining a Patient Denominator in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) Database: Recommendations for Best Practices

Stephanie Garies,¹ Michael Cummings,² D. Himasara Marasinghe,¹ John Queenan³

¹Dept of Family Medicine, University of Calgary; ²Dept of Family Medicine, University of Alberta; ³Centre for Studies in Primary Care, Queen's University

CONTEXT

- Disease reporting & prevalence studies require definition of a:
 - **numerator** (number of affected individuals)
 - **denominator** (number of individuals at risk in the population)
- Calculating the denominator or patient population from primary care electronic medical record (EMR) data can be challenging, as various types of biases & limitations exist within these data.

Objectives

- To describe different methods used to define patient denominators in primary care EMR databases globally;
- To assess different patient denominators in Canadian primary care EMR data.

CPCSSN DATA

- Collaboration of 12 primary care practice-based research networks across Canada.
- Family physicians & nurse practitioners contribute de-identified patient data from their EMR.
- EMR data are cleaned, coded, and standardized by CPCSSN to use for secondary purposes.

APPROACH

- Part 1: Scoping review of published & grey literature (2000-2020) to understand how other international primary care EMR databases define their patient denominators.
- Part 2: Descriptive analysis of national CPCSSN data (up to Dec 31, 2019).
- EMR data for over 1.8 million patients from 7 provinces, with up to 10 years of longitudinal data, including:
 - 43 million billing records
 - 52 million encounter records
 - 28 million encounter dx records
 - 41 million prescribed medication records

Fig 1. Variation in % patients included by data table & EMR system

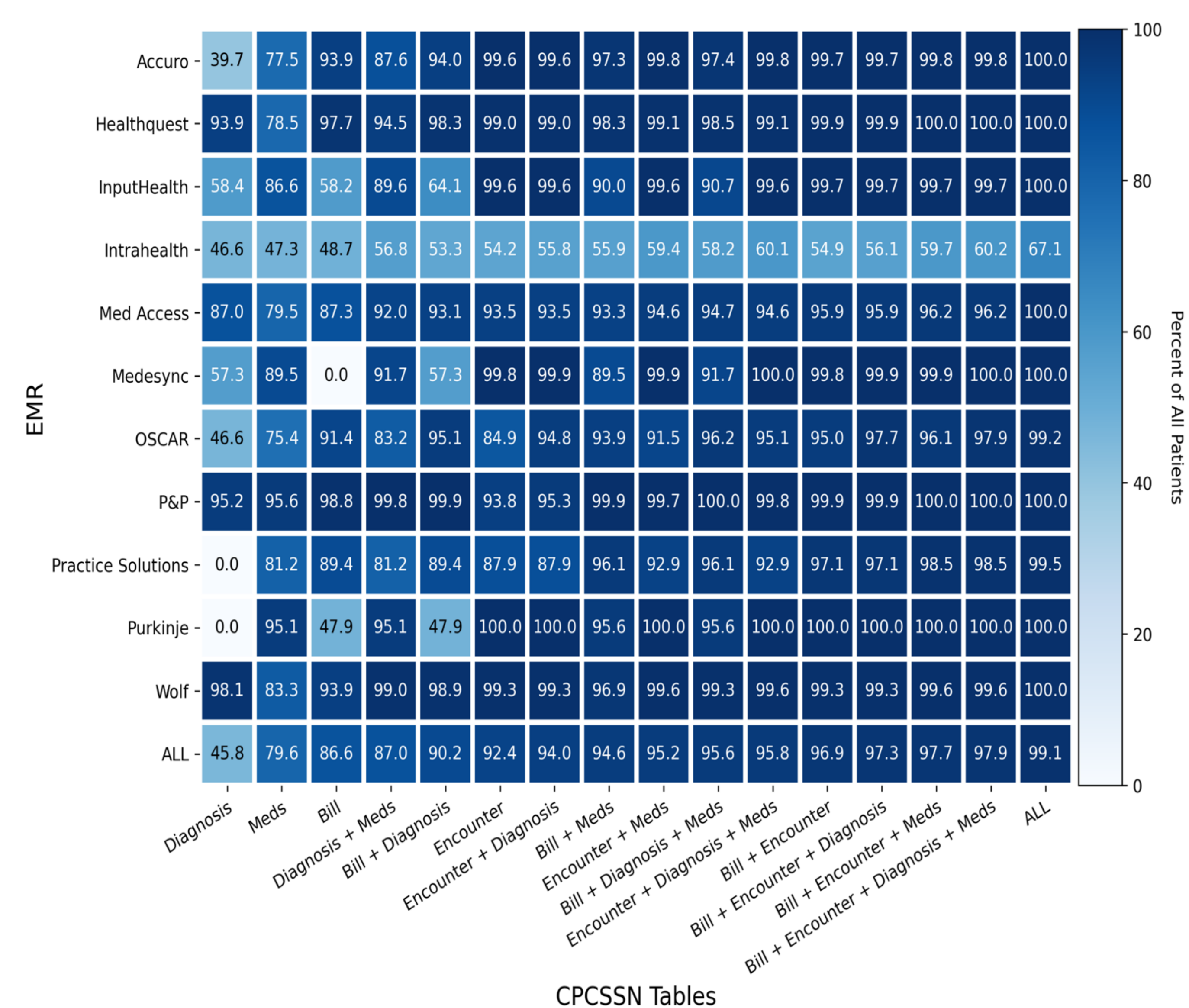
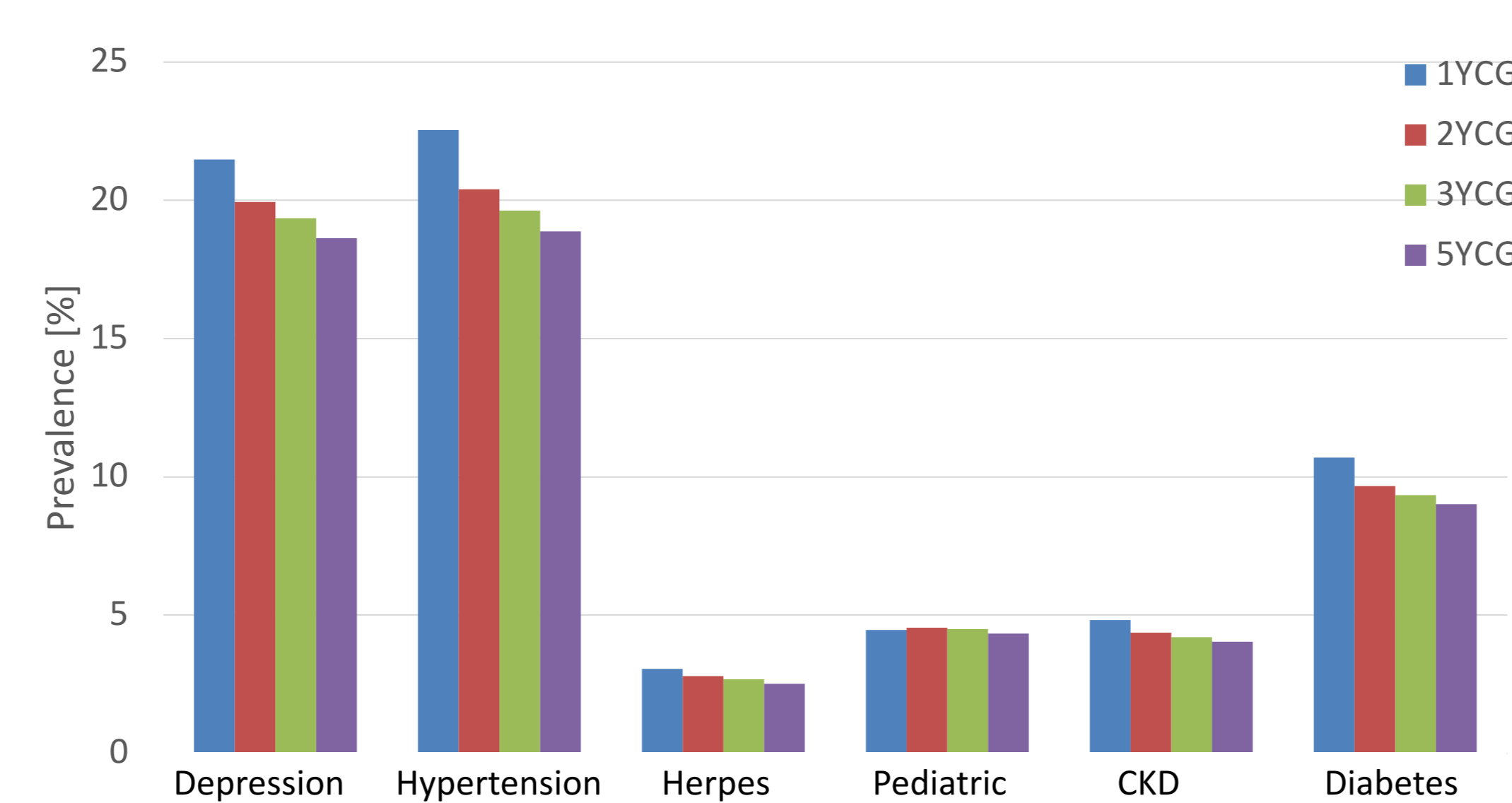


Fig 2. Variation in disease prevalence by different contact group time periods



FINDINGS & RECOMMENDATIONS

From scoping review: International studies calculate denominators for EMR data primarily using external registries (claims, insurance, census) or statistical modelling to standardize for age, sex, SES.

Denominator in CPCSSN Data:

- Use "Date Created" variable across 4 tables (encounter, diagnosis, billing & medications) to define a contact; this mitigates the gaps in data resulting from different EMR systems.
- Use one contact with a 2-year time period for a CPCSSN denominator ("2YCG") for most chronic disease epidemiology.
 - Includes more 'active' patients (vs. longer time period)
 - Balanced % male/female (more females in 1YCG)
 - Balanced median age (older patients in 1YCG)
 - Reasonably expected disease prevalence (gradient, with highest in 1YCG, lowest in 5YCG)
- Study of rare diseases may require a larger contact group to capture more patients (e.g. 5YCG)