

## Association rule mining to identify potential under-coding of conditions in the problem list in primary care electronic medical records

Weaver, C<sup>1</sup>, Garies, S<sup>1</sup>, Williamson, T<sup>1</sup>, McBrien, K<sup>1</sup>, and Peng, M<sup>1</sup>

<sup>1</sup>University of Calgary

### Introduction

The problem list of a patient's primary care electronic medical record (EMR) generally reflects their important medical conditions. We will use association rule mining to assess between-provider and between-clinic variation in the coding of select conditions in the EMR problem list, in order to identify possible under-coding outliers.

### Objectives and Approach

EMR data from participating clinics in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) will be used, with a focus on three commonly-occurring conditions (hypertension, diabetes, and depression). Association rule mining will be used to develop association rules between these conditions and other clinical information available in the EMR, such as other diagnoses in the problem list, billing codes, medications, and laboratory results (e.g., a rule of "diabetic medication → diabetes" indicates that patients prescribed a diabetic medication are likely to have diabetes in the problem list). Under-coding outliers at the provider and clinic levels will be identified by comparing rule enforcement.

### Results

Results from this work in progress will be presented at the conference. An estimated 270 clinics, 1340 providers, and 1.8 million patients will be included from the CPCSSN database. Rule 'confidence' will be used to identify outliers; the confidence of a rule  $X \rightarrow Y$  is the proportion of individuals with  $X$  who also have  $Y$  ( $\Pr(Y|X)$ ). For example, we may find that on average 80% of patients prescribed a diabetic medication will also have a diagnosis of diabetes in the problem list (average confidence of 80%), but an outlier clinic may have a confidence of 40%; this low rule confidence may indicate under-coding of diabetes in the problem list. Confounding by patient demographics (e.g., age, sex, urban/rural) will be assessed and adjusted for, if necessary.

### Conclusion/Implications

This work examines a novel method to identify potential under-coding in the EMR problem list. Providers/clinics could use this information to update patients' problem list or inform quality improvement interventions. Researchers using primary care EMR data need to be aware of potential under-coding and take steps to mitigate the effects.

