Preparing Electronic Medical Record Data for Pharmaco-epidemiologic Research: Essential Data Validation

Tyler Schneider, PharmD, RPh, MSc eHealth Student
Canadian Association for Population Therapeutics
October 21, 2024

Schneider T, Punjani T, Silva JM, Garcia MC, Telford V, Mbuagbaw L, Gibson P, Holbrook AM

McMaster University
Division of Clinical Pharmacology and Toxicology

CLINICAL PHARMACOLOGY & TOXICOLOGY
Department of Medicine

Background

- Electronic medical records (EMRs) contain far more data useful for research than administrative health records
  - Demographics, medication administrations (as opposed to dispensations), vitals, labs, social situation, risk behaviours, etc.
- Epic has exploded in popularity in Canada becoming the 2nd most common hospital EMR since its first installation in 2017
  - St. Joseph's Healthcare Hamilton (SJHH) was the first Canadian academic adult hospital to install Epic
  - Leading EMR vendor in the United States
  - Highly structured EMR with > 18,000 tables but no guide or roadmap to demonstrate which tables or fields contain which data
    - Considered proprietary information

Ehrenstein et al., 2019;
PennDnA, n.d.;
St. Joseph's Healthcare Hamilton, n.d.;
Witowski, 2024

Background

- EMR data not collected or stored primarily for research
  - High quality, accessible, and organized data are important for any resulting analysis (*garbage in, garbage out*)
- Data must be validated to assess quality and accuracy
- Few studies assess the validity of EMR data
  - Canadian Primary Care Sentinel Surveillance Network (CPCSSN) – chronic conditions in their database
  - Hospital EMRs – completeness of problem list, cancer treatment and progression
  - Epic specifically – algorithms to identify diagnoses
- International scan of research-intensive institutions did not reveal any readily usable data validation metrics
  - SickKids, Stanford, and Harvard were creating research repositories with EMR data

Botsis et al., 2010;
Breeyear et al., 2024;
Williamson et al., 2014*;*
Wright et al., 2015

Objectives

1. Create an entity relationship diagram for the Epic EMR for key general research themes
   - Using SJHH Epic-Dovetale data – 6 years' worth of data
   - Demographics, providers, exposures, outcomes, comorbidities, diagnostics, timing, decision support, etc.
2. Validate EMR data for these key themes relevant to pharmacoepidemiologic research
   - Ongoing project on whether QT-prolonging (QTP) medications are/are not associated with major adverse cardiac events (MACE)
     - Aim to develop clinical prediction rule on who will have MACE while taking QTP medication(s)

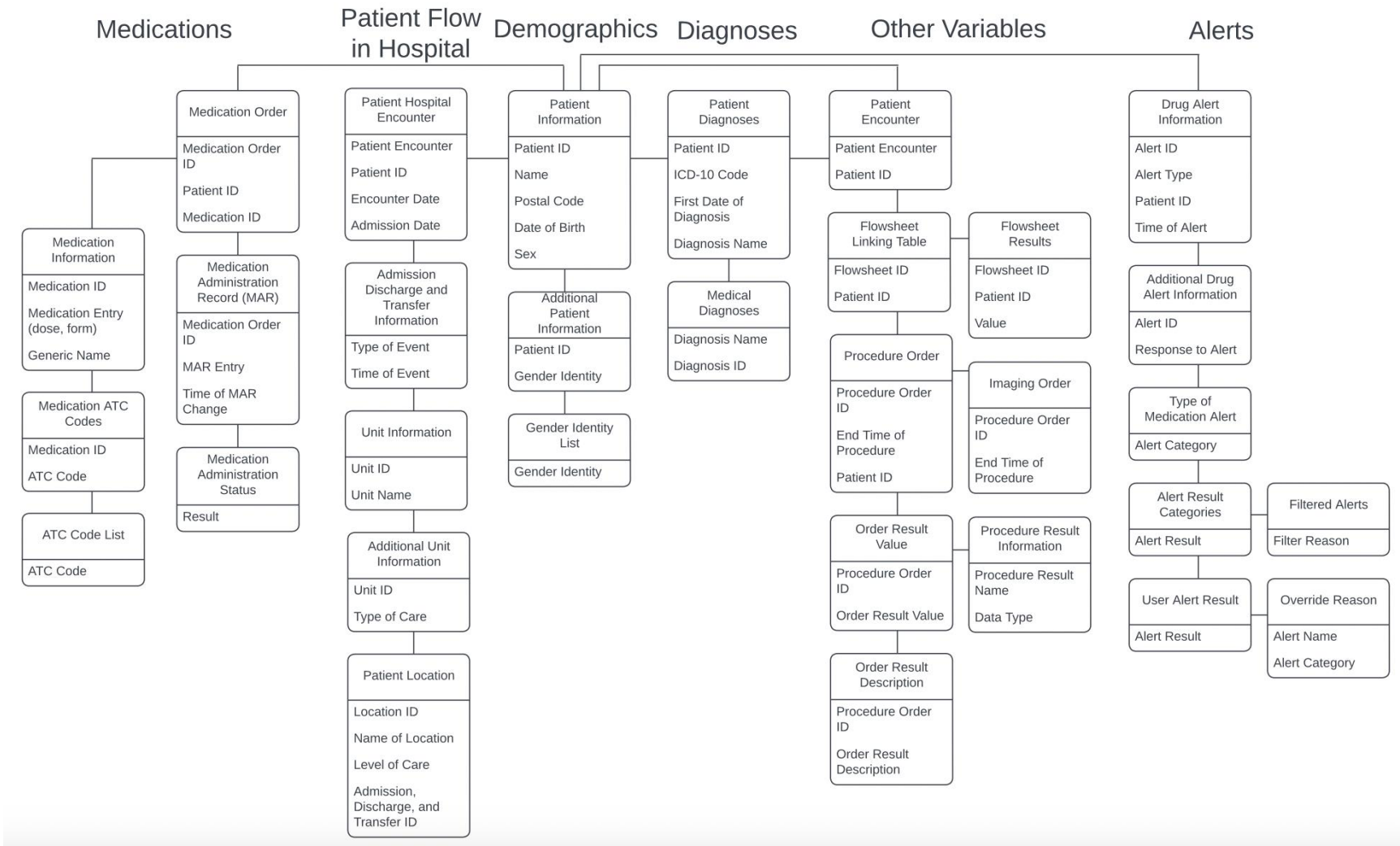
Methods

- Team of clinical pharmacologists, CMIO, medicine specialists, pharmacists, research data analysts, data scientists, informatics specialists, methodologists
- Data extracted from Epic Clarity database
- Relevant tables were found via the Epic Data Handbook, trial-and-error
- Data validated for completeness and correctness
  - Manual validation using chart review with random selection of cases
  - Computational validation comparing Epic-Dovetale to CIHI DAD data, Epic-Dovetale to Slicer Dicer

Methods

- All patients aged 18 years or older admitted to SJHH from December 2, 2017 to March 24, 2023 (N=70,079)
- CIHI DAD data available as a 'gold standard' for diagnoses if validated poorly
- Validation statistics
  - Correctness (agreement)
  - Diagnostic accuracy (positive predictive value (PPV), negative predictive value (NPV), sensitivity (Sn), specificity (Sp))

Results: Entity Relationship Diagram



- Only **31 of 20,764** tables in Epic-Dovetale required

Results: Validation Examples

1. Demographics
   - Computational validation of 70,079 unique patients from Epic-Dovetale vs 70,170 from SlicerDicer (**>99%** agreement); Age at admission vs CIHI - 9,472 patients (**100%** agreement)
   - Manual validation of sex and gender - 131 patients (**100%** agreement)
2. Exposures (ATC code + generic medication name)
   - 50,487 patients took a 'known' QTP medication
   - Manual validation of 335 charts - Diagnostic accuracy**: PPV 100%, NPV 95%, Sn 95%, Sp 100%**
3. MACE Outcomes (death, ventricular arrhythmia including Torsade de Pointes, non-fatal cardiac arrest, syncope)
   - Manual validation of 335 CIHI charts - Diagnostic accuracy**: PPV 93%, NPV 98%, Sn 99%, Sp 81%**

Results: Validation Examples (2)

4. Lab Values (potassium, magnesium, calcium, troponin)
   - Manual validation of 736 lab results for 50 patients (**100%** agreement)
5. EKGs: QTc and rhythm
   - Manual validation of 53 EKGs for 38 patients (**100%** agreement)
6. Comorbidities (diabetes, hypertension, heart failure)
   - Computational validation vs CIHI agreement - 519/560 (**93%**), 418/513 (**81%**), and 106/152 (**70%**), respectively
7. Timestamping
   - Computational validation of date of admission vs CIHI - 9,958/10,065 admissions (**>99%** agreement)
8. Alerts
   - 850,672 medication-related alerts **shown** to providers in 2023
     - **98% overridden**

Discussion

- The lack of an official entity relationship diagram wastes time, potentially creates errors
- Very small percentage of data tables needed for research
- Data from CIHI will be required for now for some fields
  - MACE outcomes (diagnoses) are not stored discretely in Epic-Dovetale unless added to medical history or the problem list
  - Natural language processing may help
- Common data models such as OMOP or i2b2 required for collaborative research across different EMRs
  - Epic has its own COSMOS network with 274 million patients