

Improving the quality of prescription information in primary care EMR data using a machine learning approach



M. Taylor,¹ S. Garies,² N. Drummond,^{1,2,3,4} M. Cummings,¹ B. Soos,^{2,3} C. Weaver,³ S. Aponte-Hao,³ R. Miyagishima,¹ A. Pham,¹ T. Williamson³

¹Dep't of Family Medicine, University of Alberta; ²Dep't of Family Medicine, University of Calgary; ³Dep't Community Health Sciences, University of Calgary;

⁴School of Public Health, University of Alberta

CONTEXT

- Prescribed medications from community-based primary care settings are important for many secondary purposes:
 - Med recall lists
 - Prescribing surveillance
 - Policy decision-making
 - Pharmacovigilance
- Prescribing data from EMRs can be difficult to use due to poor quality, high variation, and unstructured free text.

Objective

- To build a machine learning tool to clean unstructured prescribing information in primary care EMR data into consistent, standardized terms.

DATABASE

- Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a collaboration of 13 primary care practice-based research networks across Canada
- Family physicians & nurse practitioners contribute de-identified patient data from their EMR.
- Approx. **41 million prescribing records** from >1.8 million patients

APPROACH

- Health Canada's Drug Product Database used as reference standard
- A FastText linear classifier was developed by training a model with a set of labeled data derived from the Drug Product Database, and a learned text representation based on CPCSSN's collection of medication strings.
- Manual review of a random sample of 5,000 prescription records to check for accuracy

Example of uncleaned text & coded text

Original EMR text	Cleaned Code	Cleaned Name
oxycodone HCL 30 mg Oral Tablet, Oral Only, Extended Release 12 Hr	N02AA05	Oxycodone
Fluticasone-salmeterol (Advair 250 250 mcg-25 mcg AERO) i INH BID	R03AK06	Salmeterol and Fluticasone
VICTOZA MULTIDOSE PEN-INJECTOR. 30DOSES(0.6MG)-15DOSES(1.2MG)-10DOSES(1.8MG). 1/2/3	A10BJ02	Liraglutide
CONCERTA 36MG 1 po od total 56 in lots of 14 q14 days x 4 Qty:14 Repeats:3	N06BA04	Methylphenidate
fluticasone propionate 125 mcg Inhalation Aerosol With Adapter (Gram)	R03BA05	Fluticasone
budesonide/formoterol fumarate 200 mcg-6 mcg/actuation Inhalation Aerosol Powder, Breath Activated (Ea)	R03AK07	Formoterol and Budesonide
rabeprazole sodium 10 mg Oral Tablet, Delayed Release (Enteric Coated)	A02BC04	Rabeprazole
Codeine/guaiFENesin/pseudoephedrine (Ratio-Cotridin Expectorant 10 mg-100 mg-30 mg/5 mL LIQ) 0.5 tsp i PO BID prn	R05DA04	Codeine
ALUPENT SYR 10MG/5.0ML 2 TSP TID for COUGH Qty:120 Repeats:1	R03CB03	Orciprenaline
cholera vaccine (whole cell-cholera toxin B subunit, recomb) Oral Suspension, Reconstituted, Oral (mL)	J07AE01	Cholera, Inactivated, Whole Cell
HYDROCORTISONE CREAM 0.5% USP CREAM TOPICAL (JAMP PHARMA CORPOR...)	D07AA02	Hydrocortisone
acetylsalicylic acid 81 mg Oral Tablet, Delayed Release (Enteric Coated)	B01AC06	Acetylsalicylic Acid
pantoprazole magnesium 40 mg Oral Tablet, Delayed Release (Enteric Coated)	A02BC02	Pantoprazole
acetaminophen/codeine phosphate/caffeine 300 mg-30 mg-15 mg Oral Tablet	N02AJ06	Codeine and Paracetamol
zoster vaccine live/Pf 19,400 unit/0.65 ml Subcutaneous Vial (Ea)	J07BK01	Varicella, Live Attenuated
fluticasone propionate/salmeterol xinafoate 250 mcg-25 mcg/actuation Inhalation Hfa Aerosol With Adapter (Gram)	R03AK06	Salmeterol and Fluticasone
ciprofloxacin/ciprofloxacin HCL 500 mg Oral Tablet, Extended Release Multiphase 24 Hr	J01MA02	Ciprofloxacin
VARENICLINE TARTRATE 0.5 mg (11)-1 mg (3x14) TABLET, DOSE PACK (0.5(11)-1)	N07BA03	Varenicline
acetaminophen/codeine phosphate/caffeine 300 mg-30 mg-15 mg Oral Tablet	N02AJ06	Codeine and Paracetamol
VIMOVO MODIFIED RELEASE TABLET 20MG/500MG 1 po bid prn Qty:60 Repeats:0	M01AE52	Naproxen and Esomeprazole

PRELIMINARY OUTCOMES & NEXT STEPS

- Preliminary observations comparing CPCSSN's previous medication coding process to the new machine learning method:
 - The error rate has decreased by an estimated order of magnitude;
 - Any identified coding errors seem to be relatively close to the correct code, which is helpful when related medication codes are part of a clinical query or case definition;
 - Updating the ML model with newly approved drugs and corrections appears to have less stability issues in how it may affect other aspects of CPCSSN coding and processing.
- We are currently conducting a manual review to quantify the accuracy of the algorithm (sensitivity, specificity, PPV, NPV).



Contact: Stephanie Garies (sgaries@ucalgary.ca)

This work is funded by the Alberta Children's Hospital Research Institute (Calgary, Alberta).