Transforming Everyday Life into Extraordinary Ideas
Application of Text Data Mining Approaches to the Canadian Longitudinal Study on Aging (CLSA) Medication Dataset: Development and Validation

Benoit Cossette, BPharm, PhD
Lauren Griffith BSc, MSc, PhD
Lisa Dolovich BScPhm, PharmD, MSc

DSECT/DSEN Seminar Series
Nov 22, 2018
CLSA Medication Coding Group

Benoit Cossette
Lauren Griffith
Lisa Dolovich
Patrick Emond
Dee Mangin
Jinhui Ma
Medication Coders: Helga Weigelin and Bevonie Brown

With acknowledgment of Kathryn Nicholson, Joanne Ho and Kasia Makara
Overview

- Overview CLSA
- Challenges with medication collection
- Medication cleaning protocol development, results to date, validation, and lessons learned
What is the Canadian Longitudinal Study on Aging (CLSA)?

A research platform – infrastructure to enable state-of-the-art, interdisciplinary population-based research and evidenced-based decision-making that will lead to better health and quality of life for Canadians.
Canadian Longitudinal Study on Aging (CLSA)

- Strategic initiative of CIHR; on Canadian research agenda since 2001
- More than 160 researchers and collaborators – 26 institutions
- Multidisciplinary – biology, genetics, medicine, psychology, sociology, demography, economics, epidemiology, nutrition, health services
- Largest research platform of its kind in Canada for breadth and depth
- Following 50,000+ Canadians aged 45-85 at baseline for 20 years
CLSA Leads

Lead Principal Investigator
Parminder Raina (McMaster)

Co-principal Investigator
Christina Wolfson (McGill)

Co-principal Investigator
Susan Kirkland (Dalhousie)
CLSA Research Platform

50,000 women and men aged 45 - 85 at baseline

<table>
<thead>
<tr>
<th>Target: 20,000</th>
<th>Target: 30,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual: 21,241</td>
<td>Actual: 30,097</td>
</tr>
<tr>
<td>Randomly selected within provinces</td>
<td>Randomly selected within 25-50 km of 11 sites</td>
</tr>
</tbody>
</table>

Questionnaire
- By telephone (CATI)
- In person, in home (CAPI)

Clinical/physical tests
- Blood, urine
  - @ Data Collection Site

2010 - 2015  2015  2018

Participants aged 45 to 85 at baseline (51,338)

Active follow-up every 3 years
# Depth and Breadth of CLSA

## Physical & Cognitive Measurements
- Height & weight
- Waist and hip measurements
- Blood pressure
- Grip strength, timed up-and-go, chair raise, 4-m walk
- Standing balance
- Vision (retinal imaging, tonometer & visual acuity)
- Hearing (audiometer)
- Spirometry
- Body composition (DEXA)
- Bone density (DEXA)
- Aortic calcification (DEXA)
- ECG
- Carotid intima-media thickness (ultrasound)
- Cognitive assessment (30-minute battery)
- Biospecimen collection (blood and urine)

## Health Information
- Chronic disease symptoms (11 chronic conditions)
- Medication and supplement intake & compliance
- Women’s health
- Self-reported health-care utilization
- Oral health
- Administrative data linkage health services, drugs and other administrative databases

## Psychosocial
- Social participation
- Social networks and support
- Caregiving and care receiving
- Mood, psychological distress
- PTSD
- Injuries and consumer products
- Work-to-retirement transitions
- Personality traits
- Retirement planning
- Social inequalities
- Mobility-lifespace
- Built environments and contextual factors
- Income, wealth and assets

## Lifestyle & Sociodemographic
- Smoking
- Alcohol consumption
- Physical activity (PASE)
- Nutrition (nutrition risk and food frequency)
- Ethnicity/race/gender
- Birth location
- Marital status
- Education
Terminology

- Tracking Cohort
  - Target - 20,000 participants from all 10 provinces, followed through Computer-Assisted Telephone Interviews (60 minutes at baseline)
  - 21,241 recruited*

- Comprehensive Cohort
  - Target - 30,000 participants living within 25 km (or 50 km) of a CLSA Data Collection Site (DCS)
  - Followed through in-home interviews (60 minutes) and physical assessments (2-3 hours) at a DCS
  - 30,097 recruited*
CLSA Data Collection

Data Collection Site

Physical Assessments:
- Height, Weight, BMI
- Bone Density, Body Composition, Aortic Calcification
- Blood Pressure
- ECG
- Carotid Intimal-Medial Thickness
- Pulmonary Function
- Vision & Hearing
- Performance testing

Biospecimen Collection:
- Blood
- Urine

Cognitive Assessments:
- Neuropsychological Battery
  - Memory
  - Executive function
  - Reaction time
Medication Data Collection
Always use the DIN number field when entering a drug.

If the DIN number isn’t available type the name of the medication in the DIN field, as you are typing, a list of drugs will appear, locate the correct one and click on it to select it.
Multivitamins and herbal supplements will not have DIN’s.

Enter the name of the supplement in the “Name of Medication” field as shown below
New Question Added (Dec. 2014)

Was this medication prescribed by a doctor or is it a non-prescription medication?

- Yes
- No
- Don’t know/No answer
- Refused

NB: This questions was added to try a determine why we have so many “medication” without DINs
Additional Information Collected

- Dosage (quantity, units)
- Frequency (e.g., once a day, once a week)
- Duration (< 6 mo, 6 mo-1 yr, >1 yr)
- Start date
- Reason(s) for use
Health Canada Drug Product Database

The Drug Product Database (DPD) is maintained by Health Canada.

Included drugs:
- Drugs for humans and animals
- Disinfectants
- Radiopharmaceutical drugs
- Biological drugs for humans

**Health Canada DPD - online query**

### Drug Product Database online query

**From Health Canada**

Due to the fact that the information originated with an organization that is not subject to the Official Languages Act, the document may only appear in the language in which it was written. Translations of the document are the responsibility of the sponsor involved.

### Search criteria

You may search by either:
- a) drug identification number (DIN),
- b) Anatomical Therapeutic Chemical (ATC) code, or
c) by company or one or more of the various other product characteristics listed. When typing inside fields, do not include punctuation marks such as hyphens, commas, colons, brackets and wildcard characters (%).

#### Search by drug identification number

**Drug identification number (DIN):**

#### Search by Anatomical Therapeutic Chemical

**Anatomical Therapeutic Chemical (ATC):**

#### Search by other criteria

---

### Search by other criteria

**Status:**

Select a status

**Company:**

**Product name:**

**Active ingredient(s):**

**Active ingredient group number:**

**Class(es):**

- Disinfectant
- Human
- Radiopharmaceutical
- Veterinary

**Route(s) of administration:**

- Intravenous
- Intramuscular
- Intrabursal
- Intracardiac
- Intracutaneous

**Dosage form(s):**

- 0-Unassigned
- Aerosol
- Aerosol泡沫
- Aerosol Metered Dose
- Ointment

**Schedule(s):**

- CDSA Recommended
- Ethical
- Homoeopathic
- Narcotic

**Search**  **Reset**
# Search results summary

From Health Canada

New search

## Search criteria
- Status: Marketed
- Product name: Lipitor
- Class(es): Select all
- Route(s) of administration: Select all
- Dosage form(s): Select all
- Schedule(s): Select all

## Search results
Showing 1 to 4 of 4 entries | Show [10] entries

### List of returned drug products

<table>
<thead>
<tr>
<th>Status</th>
<th>DIN</th>
<th>Company</th>
<th>Product</th>
<th>Class</th>
<th>PM</th>
<th>Schedule</th>
<th>#</th>
<th>A.I. name</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marketed</td>
<td>02230711</td>
<td>PFIZER CANADA INC</td>
<td>Lipitor</td>
<td>Human</td>
<td>Yes</td>
<td>Prescription</td>
<td>1</td>
<td>ATORVASTATIN (ATORVASTATIN CALCIUM)</td>
<td>10 MG</td>
</tr>
<tr>
<td>Marketed</td>
<td>02230713</td>
<td>PFIZER CANADA INC</td>
<td>Lipitor</td>
<td>Human</td>
<td>Yes</td>
<td>Prescription</td>
<td>1</td>
<td>ATORVASTATIN (ATORVASTATIN CALCIUM)</td>
<td>20 MG</td>
</tr>
<tr>
<td>Marketed</td>
<td>02230714</td>
<td>PFIZER CANADA INC</td>
<td>Lipitor</td>
<td>Human</td>
<td>Yes</td>
<td>Prescription</td>
<td>1</td>
<td>ATORVASTATIN (ATORVASTATIN CALCIUM)</td>
<td>40 MG</td>
</tr>
<tr>
<td>Marketed</td>
<td>02243097</td>
<td>PFIZER CANADA INC</td>
<td>Lipitor</td>
<td>Human</td>
<td>Yes</td>
<td>Prescription</td>
<td>1</td>
<td>ATORVASTATIN (ATORVASTATIN CALCIUM)</td>
<td>80 MG</td>
</tr>
<tr>
<td>ingredient</td>
<td>Brand_name</td>
<td>TC_ATC_NUMBER</td>
<td>TC_AHFS_NUMBER</td>
<td>strength</td>
<td>strength_unit</td>
<td>Drug_identification_number</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>----------</td>
<td>---------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16925 CEFAZOLIN (CEFAZOLIN SODIUM)</td>
<td>CEFAZOLIN FOR INJECTION, USP</td>
<td>J01DB04</td>
<td>08:12.06.04</td>
<td>500</td>
<td>MG</td>
<td>2237137</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16926 CEFAZOLIN (CEFAZOLIN SODIUM)</td>
<td>KEFZOL</td>
<td>J01DB04</td>
<td>08:12.06.04</td>
<td>1</td>
<td>G</td>
<td>322296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16927 CEFAZOLIN (CEFAZOLIN SODIUM)</td>
<td>KEFZOL</td>
<td>J01DB04</td>
<td>08:12.06.04</td>
<td>10</td>
<td>G</td>
<td>411450</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16928 CEFAZOLIN (CEFAZOLIN SODIUM)</td>
<td>KEFZOL</td>
<td>J01DB04</td>
<td>08:12.06.04</td>
<td>500</td>
<td>MG</td>
<td>322288</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16929 CEFAZOLIN (CEFAZOLIN SODIUM)</td>
<td>KEFZOL ADD-VANTAGE INJ 1GM/VIAL</td>
<td>J01DB04</td>
<td>08:12.06.04</td>
<td>1</td>
<td>G</td>
<td>411442</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16930 CEFAZOLIN (CEFAZOLIN SODIUM)</td>
<td>KEFZOL ADD-VANTAGE INJ 500MG/VIAL</td>
<td>J01DB04</td>
<td>08:12.06.04</td>
<td>500</td>
<td>MG</td>
<td>411434</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16931 CEFAZOLIN SODIUM</td>
<td>CEFAZOLIN FOR INJECTION USP</td>
<td>J01DB04</td>
<td>08:12.06.04</td>
<td>20</td>
<td>G</td>
<td>2318830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16932 CEFEPIM (CEFEPIME HYDROCHLORIDE)</td>
<td>CEFEPIME FOR INJECTION</td>
<td>J01DE01</td>
<td>08:12.06.16</td>
<td>2</td>
<td>G</td>
<td>2319039</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16933 CEFEPIM (CEFEPIME HYDROCHLORIDE)</td>
<td>MAXIPIME FOR INJECTION -PWS IV IM 1G</td>
<td>J01DE01</td>
<td>08:12.06.16</td>
<td>1</td>
<td>G</td>
<td>2163632</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16934 CEFEPIM (CEFEPIME HYDROCHLORIDE)</td>
<td>MAXIPIME FOR INJECTION-PWS IV 2GM</td>
<td>J01DE01</td>
<td>08:12.06.16</td>
<td>2</td>
<td>G</td>
<td>2163640</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16935 CEFIXIME</td>
<td>AURO-CEFIXIME</td>
<td>J01DD08</td>
<td>08:12.06.12</td>
<td>400</td>
<td>MG</td>
<td>2432773</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16936 CEFIXIME</td>
<td>SUPRAX</td>
<td>J01DD08</td>
<td>08:12.06.12</td>
<td>100</td>
<td>MG</td>
<td>868965</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16937 CEFIXIME</td>
<td>SUPRAX</td>
<td>J01DD08</td>
<td>08:12.06.12</td>
<td>100</td>
<td>MG</td>
<td>2195992</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16938 CEFIXIME</td>
<td>SUPRAX</td>
<td>J01DD08</td>
<td>08:12.06.12</td>
<td>200</td>
<td>MG</td>
<td>868973</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16939 CEFIXIME</td>
<td>SUPRAX</td>
<td>J01DD08</td>
<td>08:12.06.12</td>
<td>400</td>
<td>MG</td>
<td>868981</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16940 CEFIXIME</td>
<td>SUPRAX</td>
<td>J01DD08</td>
<td>08:12.06.12</td>
<td>400</td>
<td>MG</td>
<td>2195984</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16941 CEFIXIME</td>
<td>SUPRAX - TAB 200MG</td>
<td>J01DD08</td>
<td>08:12.06.12</td>
<td>200</td>
<td>MG</td>
<td>2195976</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16942 CEFOTAXIME (CEFOTAXIME SODIUM)</td>
<td>CEFOTAXIME SODIUM FOR INJECTION, BP</td>
<td>J01DD01</td>
<td>08:12.06.12</td>
<td>1.0</td>
<td>G</td>
<td>2261510</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16943 CEFOTAXIME (CEFOTAXIME SODIUM)</td>
<td>CEFOTAXIME SODIUM FOR INJECTION, BP</td>
<td>J01DD01</td>
<td>08:12.06.12</td>
<td>2.0</td>
<td>G</td>
<td>2261529</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16944 CEFOTAXIME (CEFOTAXIME SODIUM)</td>
<td>CEFOTAXIME SODIUM FOR INJECTION, BP</td>
<td>J01DD01</td>
<td>08:12.06.12</td>
<td>500</td>
<td>MG</td>
<td>2261502</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16945 CEFOTAXIME (CEFOTAXIME SODIUM)</td>
<td>CLAFORAN 0.5GM/VIAL</td>
<td>J01DD01</td>
<td>08:12.06.12</td>
<td>500</td>
<td>MG</td>
<td>546208</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16946 CEFOTAXIME (CEFOTAXIME SODIUM)</td>
<td>CLAFORAN 1GM/VIAL</td>
<td>J01DD01</td>
<td>08:12.06.12</td>
<td>1</td>
<td>G</td>
<td>546216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16947 CEFOTAXIME (CEFOTAXIME SODIUM)</td>
<td>CLAFORAN 2GM/VIAL</td>
<td>J01DD01</td>
<td>08:12.06.12</td>
<td>2</td>
<td>G</td>
<td>546224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16948 CEFOTAXIME (CEFOTAXIME SODIUM)</td>
<td>CLAFORAN ADD-VANTAGE VIALS 1GM/VIAL</td>
<td>J01DD01</td>
<td>08:12.06.12</td>
<td>1</td>
<td>G</td>
<td>1989758</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Licensed Natural Health Products Database (LNHPD)

Managed by Health Canada and includes information on licensed natural health products, including:

• vitamin and mineral supplements
• herb and plant-based remedies
• traditional medicines like Traditional Chinese Medicines or Ayurvedic (Indian) Medicines
• omega 3 and essential fatty acids
• probiotics
• homeopathic medicines
• many everyday consumer products, like certain toothpastes, antiperspirants, shampoos, facial products and mouthwashes

Findings from the CLSA baseline evaluation

• 42% of medications were coded by the interviewers using a medication name / DIN proposed by the CLSA software linked to an extract from the Health Canada Drug Product Database (DPD).

• The remaining entries were entered in the ‘name of medication’ field in free text.

• Based on a sample of 100 participants, 73% of the free text entries were for natural products.
Start date / duration

Decision to drop the start date

• Start date:
  ▪ Selected from calendar
  ▪ 28% in January
  ▪ 5% to 8% in February – December

• Decision to drop the start date Duration:
  ▪ 6 months to one year
  ▪ More than one year
  ▪ …
Protocol: Computer algorithm development

Objectives

• To describe the process used to develop an original computer generated approach to clean medication name data that have been recorded in the CLSA database.
  - Agreement for same medication name is not the same as agreement for same DIN (DIN match).

• To conduct manual re-coding on the remaining uncleaned medication data to inform the refinement of the computer generated approach.

• To identify common medication data entry errors that can be used to support more accurate recording of medications data in the CLSA data in future stages of data collection.
Protocol: Computer algorithm development

Methods

• Computer generated approach was applied to all CLSA input entries that required cleaning. The intent of this application was to match or clean the prescribed medication or natural product number entered during a participant’s interview, (an input) to one or more product(s) in the DPD or NHP.

• The approach uses multiple distinct algorithms, known hereafter as a test, for matching inputs to drugs or naturals.

• All tests are ordered sequentially such that once an input is matched it is no longer considered in the remaining tests.
Protocol: Computer algorithm development
Methods

• Programming conducted using SQL (a database scripting language) and PHP (a general programming language).

• The Health Canada and the CLSA data repository (Opal) data were loaded into a MySQL database using SQL.

• Some pre-processing was conducted on these databases before using PHP to enhance performance and increased speed of matching.
  - For instance, the "simple" test compares the free text to drug names by ignoring non-alpha-numeric characters.
Results so far: Applying computer based algorithm tests
Algorithms

- **Direct**: matches the brand or product name (not case sensitive)
- **Code**: matches 6 or more consecutive digits which matches a DIN or NPN (with or without leading 0's)
- **Word**: matches the brand or product name as a sub-string of the input
- **Simple**: matches ignoring non alpha-numeric characters
- **No parentheses**: matches ignoring any words found in (parentheses)
- **No-units**: matches ignoring any numerical units
Algorithm - direct

**Direct**: matches the brand or product name (not case sensitive)

**Examples:**

<table>
<thead>
<tr>
<th>Input</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetaminophen 500mg</td>
<td>ACETAMINOPHEN 500MG</td>
</tr>
<tr>
<td>acetaminophene</td>
<td>ACETAMINOPHENE</td>
</tr>
<tr>
<td>Acid Reducer</td>
<td>ACID REDUCER</td>
</tr>
<tr>
<td>actonel</td>
<td>ACTONEL</td>
</tr>
<tr>
<td>adalat</td>
<td>ADALAT</td>
</tr>
</tbody>
</table>
Algorithm - code

**Code**: matches 6 or more consecutive digits which matches a DIN or NPN (with or without leading 0's)

**Examples:**

<table>
<thead>
<tr>
<th>Input</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>00026158 Senokot Sennosides 8.6 mg</td>
<td>Senokot Tablets</td>
</tr>
<tr>
<td>02237556 Eur-Fer 300mg</td>
<td>Euro-Fer</td>
</tr>
<tr>
<td>02367335 sandoz latanoprost 1 drop</td>
<td>SANDOZ LATANOPROST</td>
</tr>
<tr>
<td>02371987 Mar-Atenolol 50mg</td>
<td>MAR-ATENOLOL</td>
</tr>
<tr>
<td>02374420 Apo-Anastrozole 1mg</td>
<td>APO-ANASTROZOLE</td>
</tr>
</tbody>
</table>
**Algorithm - word**

**Word**: matches the brand or product name as a sub-string of the input

**Examples:**

<table>
<thead>
<tr>
<th>Input</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>adalat XL</td>
<td>ADALAT</td>
</tr>
<tr>
<td>acetaminophen extra fort</td>
<td>ACETAMINOPHEN</td>
</tr>
<tr>
<td>Alendronate sodium (Fosamax)</td>
<td>ALENDRONATE</td>
</tr>
<tr>
<td>Altace HCL</td>
<td>ALTACE</td>
</tr>
<tr>
<td>apo metoprolol</td>
<td>METOPROLOL</td>
</tr>
</tbody>
</table>
**Algorithms - simple**

**Simple**: matches ignoring non alpha-numeric characters (including spaces)

**Examples:**

<table>
<thead>
<tr>
<th>Input</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Act-Atorvastatin</td>
<td>ACT ATORVASTATIN</td>
</tr>
<tr>
<td>advil liqui gels</td>
<td>ADVIL LIQUI-GELS</td>
</tr>
<tr>
<td>apo hydro</td>
<td>APO-HYDRO</td>
</tr>
<tr>
<td>Apo- Hydro</td>
<td>APO-HYDRO</td>
</tr>
<tr>
<td>apometoprolol sr</td>
<td>APO-METOPROLOL SR</td>
</tr>
</tbody>
</table>
## Algorithm – no parentheses

**No parentheses**: matches ignoring any words found in (parentheses)

**Examples:**

<table>
<thead>
<tr>
<th>Input</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVISTA (60MG)</td>
<td>evista</td>
</tr>
<tr>
<td>MYLAN-METOPROLOL (TYPE L)</td>
<td>mylan-metoprolol</td>
</tr>
<tr>
<td>Vitalux (areds)</td>
<td>Vitalux</td>
</tr>
<tr>
<td>Whey Powder (Vanilla flavour/Chocolate Flavour)</td>
<td>Whey powder</td>
</tr>
</tbody>
</table>
Algorithm – no unit

No-units: matches ignoring any numerical units

Examples:

<table>
<thead>
<tr>
<th>Input</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>APO-SALVENT 5MG/ML</td>
<td>apo-salvent</td>
</tr>
<tr>
<td>ASPIRIN 81MG</td>
<td>aspirin</td>
</tr>
<tr>
<td>calcium/magnésium 500 mg</td>
<td>Calcium/Magnesium</td>
</tr>
<tr>
<td>synthroid.05 mg</td>
<td>SYNTHROID</td>
</tr>
<tr>
<td>synthroid.75mg</td>
<td>SYNTHROID</td>
</tr>
</tbody>
</table>
## Table

<table>
<thead>
<tr>
<th>Method</th>
<th>Match Type</th>
<th>Drug</th>
<th>%</th>
<th>Natural</th>
<th>%</th>
<th>Drug+Nat</th>
<th>%</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single</strong></td>
<td>Direct</td>
<td>3079</td>
<td>32.8%</td>
<td>1907</td>
<td>20.3%</td>
<td>4986</td>
<td>53.1%</td>
<td>5076</td>
<td>54.1%</td>
</tr>
<tr>
<td></td>
<td>Code</td>
<td>90</td>
<td>1.0%</td>
<td>0</td>
<td>0.0%</td>
<td>90</td>
<td>1.0%</td>
<td>51</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>Word</td>
<td>1066</td>
<td>11.4%</td>
<td>441</td>
<td>4.7%</td>
<td>1507</td>
<td>16.1%</td>
<td>2106</td>
<td>22.4%</td>
</tr>
<tr>
<td></td>
<td>Simple</td>
<td>282</td>
<td>3.0%</td>
<td>317</td>
<td>3.4%</td>
<td>599</td>
<td>6.4%</td>
<td>224</td>
<td>2.4%</td>
</tr>
<tr>
<td></td>
<td>No-Parens</td>
<td>4</td>
<td>0.0%</td>
<td>4</td>
<td>0.0%</td>
<td>8</td>
<td>0.1%</td>
<td>12</td>
<td>0.1%</td>
</tr>
<tr>
<td></td>
<td>No-Units</td>
<td>186</td>
<td>2.0%</td>
<td>1</td>
<td>0.0%</td>
<td>187</td>
<td>2.0%</td>
<td>323</td>
<td>3.4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Single</td>
<td>4669</td>
<td>49.8%</td>
<td>2788</td>
<td>29.7%</td>
<td>7457</td>
<td>79.5%</td>
<td>7457</td>
<td>79.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Multiple</td>
<td>1926</td>
<td>20.5%</td>
<td>7457</td>
<td>79.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Total</td>
<td>9383</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Method Descriptions

- **direct**: Matches the brand or product name (not case sensitive)
- **code**: Matches 6 or more consecutive digits which matches a DIN or NPN (with or without leading 0's)
- **word**: Matches the brand or product name as a sub-string of the input
- **simple**: Matches ignoring non alpha-numeric characters
- **no-parens**: Matches ignoring any words found in (parentheses)
- **no-units**: Matches ignoring any numerical units

### Match Type Descriptions

- **single**: The input matches a single DIN or NPN
- **multiple**: The input matches more than one DIN or NPN
Spelling errors

• Acetaminophen
  ▪ Acetaminaphen
  ▪ Acetomenophen

• Actonel
  ▪ Actanel
  ▪ Actanol
  ▪ Actonal

• Timulol (timolol)

• Exium (Nexium)

• Tamsulosin hydrochlorothiazide (hydrochloride)

• Apothyazide
  ▪ Apo-triazide (triamterene / hydrochlorothiazide)
    OR
  ▪ Apo-hydro (Hydrochlorothiazide)
Examples of manual coding decision rules to code for DRUG NAME—based on iterative coding exercises (not all shown)

<table>
<thead>
<tr>
<th>Coding Issue</th>
<th>Manual Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXACT MATCH (ie has DIN there, exact product name, if dosage available this might help)</td>
<td>code exact DIN from DPD/NHPD</td>
</tr>
<tr>
<td>BRAND NAME but no dosage</td>
<td>code DIN from DPD that is the lowest number DIN</td>
</tr>
<tr>
<td>Generic NAME but multiple manufacturers</td>
<td>code DIN from DPD that is the lowest number DIN</td>
</tr>
<tr>
<td>Incorrect Spelling</td>
<td>Put in correct spelling and then choose lowest number DIN</td>
</tr>
<tr>
<td>Drug class do not code further</td>
<td>Drug class do not code further</td>
</tr>
<tr>
<td>Drug class – code further</td>
<td>Drug class – code further</td>
</tr>
<tr>
<td>Not a drug</td>
<td>Not a drug</td>
</tr>
<tr>
<td>Not enough information</td>
<td>Not enough information</td>
</tr>
</tbody>
</table>
Study Two: Validation (in progress)

• To determine the extent of agreement between the original and refined computer generated approach (test) and the manual re-coding (gold standard) for cleaning the medication data in the CLSA database.

• To assess the extent of agreement when comparing the original computer generated approach (test) with the refined computer generated approach (gold standard), as well as when the order of the algorithms are applied in different sequences.

• To assess whether the extent of agreement changes when re-coding prescribed medication vs. natural products or when re-coding among participants with specific chronic diseases.
Sample Size

- Based on 1) the presence of free text medication data; and 2) the presence of at least one prescribed medication or natural product within the alphanumeric data.

<table>
<thead>
<tr>
<th></th>
<th>Agreement</th>
<th>Precision</th>
<th>Confidence Interval</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Scenario</td>
<td>0.95</td>
<td>0.02</td>
<td>0.95</td>
<td>457</td>
</tr>
<tr>
<td>Worst Scenario</td>
<td>0.85</td>
<td>0.03</td>
<td>0.95</td>
<td>545</td>
</tr>
</tbody>
</table>
Lessons learned (so far)

• Manual entry by interviews derives many varied data entry issues (despite training)

• Many entries do not match to DPD or NHPD and so manual coding will always be needed
  ▪ Many international products
  ▪ Complex combination products, esp NHPD, vitamins
  ▪ Formatting such as parentheses, brackets cannot always be dealt with by computer
  ▪ Nondrug products

• If more than 1 manual coder they all need to use the exact same drug/NHP database

• Will always be behind with new products unless DPD/NHPD can be accessed in real time (versus updated in the back end every 3-6 months)
Conclusions

• The process of developing and validating the algorithm is challenging

• A well working computer algorithm will be a great asset to CLSA for retrospective and prospective medication data coding

• A clean CLSA medications database will be a great asset to Canadian researchers
Extra Slides re Data Access
Data Access
CLSA Data Access

- Designed as a research study, funded as a research platform
- Data available to researchers and trainees based in academic settings and research institutes in Canada and elsewhere
- 2019 application deadlines:
  - February 25
  - June 3
  - September 23
CLS A Data Access Guiding Principles

- The rights, privacy and consent of participants must be protected
- Confidentiality and security of data and biospecimens must be safeguarded
- CLSA data and biospecimens are resources to be used optimally to support research to benefit all Canadians
Data Access Timeline

Submission

6-9 weeks

≤ 12 weeks

Final Report

5-7 days

- Data and Sample Access Committee & Scientific Management Team Review
- Dataset Preparation & Release
- Administrative & Statistical Review
- Access Agreement Signed; Ethics Approval

Applicants advised to plan on receiving data 6 months after submission deadline
Data Access Fees

- **Partial Cost Recovery Model**
  - **Alphanumeric data**
    - $3,000 for a straightforward alphanumeric dataset
    - Graduate students using data solely for thesis research & Postdoctoral fellows using data solely for the postdoctoral project are eligible for a fee-waiver. Trainees must be enrolled at a Canadian institution or be supported by Canadian funds if working outside Canada.
  - **Images & raw data**
    - Additional fees of $1,000 per application are associated with the request for images & raw data.
Approved Applications

More than 130 applications approved since 2012

Year | Researcher | Trainee | Total
--- | --- | --- | ---
2012 | 2 | 3 | 5
2014 | 3 | 0 | 3
2015 | 9 | 0 | 9
2016 | 28 | 17 | 45
2017 | 31 | 15 | 46
2018 | 13 | 8 | 21

1Applications received Sep 24, 2018

Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement
Resources: www.clsa-elcv.ca
Information for Researchers & Trainees

Researchers

The CLSA provides documents online to facilitate understanding of the study and how we are gathering and managing the data.

- Protocols

  - Executive Summary
  - CLSA Protocol - Full Study Design and Baseline (2008)
  - CLSA Protocol - First Follow-up (2010)

The protocols listed are based on the applications CLSA submits to CIHR for each funding cycle. As the CLSA data collection progresses, occasionally, some measurements are changed. Updated versions of the protocols will be posted as necessary. Please refer to the Data Collection Tools section to review the specific questions and measurements gathered at each phase of the study.

- Data Collection Tools

- Physical Assessments

To ensure that physical assessment data are collected, processed, and stored in a consistent, professional, and structured manner at all CLSA sites across the country, Standard Operating Procedures (SOPs) help maintain the integrity of the data collection and data management...
FAQs

Data Access

- DataPreview Portal
- Data Release Timelines
- Data Access Application Process
- Data Access Application Documents
- Data and Biospecimens

FAQs

- Data Access Questions
- DataPreview Portal Questions
- Application Questions

How do I get access to the data?

Which data formats are available?

What do I do if I would like to obtain biospecimens?

What if there appears to be an error or omission in my data?

Can I apply for data as an international researcher?

What are the fees for access to CLSA data?