Cohort Definition using Criteria2Query (C2Q)

Cong Liu, PhD
Biomedical Informatics Department
Columbia University
Disclosure

- I have no relevant relationships with commercial interests to disclose.

Chi Yuan, MS
Columbia University

https://github.com/OHDSI/Criteria2Query
Outline

• Criteria2Query introduction
• Criteria2Query demo using RELY trial
• Q&A
• Exercise instruction
Hi Alex, Could you help me query some patients?

5 min later

How many patients who have type 1 diabetes are in our database?

32,125

3 min later

So how many patients took insulin for at least 1 year after that diagnosis?

30,130

What’s Type 1 diabetes? Our database used ICD9 code to encode diagnosis info, so what’s the ICD9 code of Type 1 diabetes? Let me check it.

Hi Alex, Could you help me query some patients?

Sure

What’s Type 1 diabetes?

Our database used ICD9 code to encode diagnosis info, so what’s the ICD9 code of Type 1 diabetes? Let me check it.

Which table is for querying patient’s drug?

Will I lose some patients if I only use string “insulin” as a query condition?

30,130
I can’t find HbA1c in our database. What does HbA1c stand for? Let me google it. Hemoglobin A1c?

Are they using the same database schema? Is our sql scripts reusable?

Among them how many people’s HbA1c is greater than 6.0% and less than 12%?

10 min later

8,125

Could you send your queries to other sites and get their results by the end of today?
The goal of Criteria2Query: clinician autonomy with minimal effort

Cohort & Pt Counts
The goal of Criteria2Query: clinician autonomy with minimal effort
The goal of Criteria2Query: clinician autonomy with minimal effort

• Use Criteria2Query to translate free-text eligibility criteria to database query

Cohort & Pt Counts
The modular pipeline for C2Q

Eligibility Criteria in free-text format

Information Extraction

Paragraph Segmentation \rightarrow Sentence Segmentation

Logic Extraction \rightarrow Relation Extraction

Negation Detection \rightarrow Named Entity Recognition

Eligibility Criteria Representation in JSON format

Query formulation

Entity Normalization \rightarrow Logic Translation

Temporal Normalization \rightarrow Numeric Normalization

Cohort Definition Query in JSON format

Query review & editing

Cohort Definition Visualization

Cohort Definition Query in SQL format
### Criteria2Query

#### Inclusion Criteria
- Age from 50 to 85 years.
- Rosen Modified Hachinski ischemic score less than or equal to 4.
- Fluency in English.

#### Exclusion Criteria
- History of autoimmune disease.
- History of stroke or seizure.

#### Configuration
- Machine Learning-based Model (CRF model)
- Rule-based Model (OHDSI Usagi)
- Abbreviation Extension (UMLS)

### Parsing

<table>
<thead>
<tr>
<th>#</th>
<th>Initial Events</th>
<th>EHR Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Diagnosis of AD</td>
<td>YES</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Inclusion Criteria</th>
<th>EHR Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age from 50 to 85 years</td>
<td>YES</td>
</tr>
<tr>
<td>2</td>
<td>Rosen Modified Hachinski ischemic score less than or equal to 4</td>
<td>YES</td>
</tr>
<tr>
<td>3</td>
<td>Fluency in English</td>
<td>NO</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Exclusion Criteria</th>
<th>EHR Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>History of autoimmune disease</td>
<td>YES</td>
</tr>
<tr>
<td>2</td>
<td>History of stroke or seizure</td>
<td>YES</td>
</tr>
</tbody>
</table>
The modular pipeline for C2Q
Entity Normalization (Concept Mapping)

- **Origin Concept**: AD (Condition)
- **Acronym?**
  - **Y**: **Unified Medical Language System®**
  - **N**: **Extended Concept**
    - **Standard Concept**: Alzheimer’s Disease (Condition)
      - **Concept name**: Alzheimer’s Disease
      - **Concept id**: 378419
      - **Vocabulary**: SNOMED
      - **Source code**: 26929004
      - **Concept Set**: AD
      - **Confidence Score**: 100
      - **Included concepts** (count = 29)
        - 36716558 Non-amnestic Alzheimer disease
        - 44782726 Delusions in Alzheimer's disease
        - ...
      - **Included source codes** (count = 74)
        - 44826537 331.0 Alzheimer's disease ICD9CM
        - 45538101 G30 Alzheimer disease ICD10
        - ...
- **OHDSI Standard Vocabularies**
  - **Concept Set**
The modular pipeline for C2Q
Snapshot

Initial Event Cohort

People having any of the following:

- a condition occurrence of [C2Q]Alzheimer's Disease [cs...]

with continuous observation of at least 0 days before and 0 days after event index date

Limit initial events to: all events per person.

Initial event inclusion criteria: From among the initial events, Include:

- having all of the following criteria:

  - with at least 1 occurrence of:
    - a condition occurrence of [C2Q]Alzheimer's Disease [cs...]
    - starting between All days Before and All days After event index date and ending any time.
    - restrict to the same visit occurrence

Limit cohort of initial events to: earliest event per person.

Additional Qualifying Inclusion Criteria

New qualifying inclusion criteria

1. [INC]Age from 50 to 85 years.

2. [INC]Rosen Modified Hachinski ischemic score less than or equal to 4.

3. [INC]Fluency in English.


5. [EXC]History of stroke or seizure.

Limit qualifying cohort to: earliest event per person.
Demo

• Criteria2Query
  – http://www.ohdsi.org/web/criteria2query/
Demo

• RELY trial
  https://clinicaltrials.gov/ct2/show/NCT00262600)
RELY on NCT00262600

- **Initial Events:**
  - warfarin new users

- **Inclusion:**
  - Age >= 18
  - With history of non-valvular atrial fibrillation (AF)
  - Have any following risk factors: age >=75 years, previous ischemic stroke, left ventricular dysfunction, or age >=65 with either diabetes mellitus, history of coronary artery disease or hypertension

- **Exclusion:**
  - Severe, disabling stroke within the previous 6 months, or any stroke within the previous 14 days
  - Anaemia (haemoglobin <100g/L) within the previous 90 days
  - Thrombocytopenia (platelet count <100 x 109/L) within the previous 90 days
Q&A

• Thanks for joining the journey!
Get some hands-on experiences

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30am-9:00am</td>
<td>Welcome, get settled, get laptops ready</td>
<td></td>
</tr>
<tr>
<td>9:00am-10:30am</td>
<td>Presentation: Why do we need cohort definitions?</td>
<td>Patrick Ryan</td>
</tr>
<tr>
<td>10:30am-11:15am</td>
<td>Demo: Defining a cohort in ATLAS</td>
<td>Chris Knoll</td>
</tr>
<tr>
<td>11:15am-12:00am</td>
<td>Demo: Defining a cohort using Criteria2Query</td>
<td>Cong Liu</td>
</tr>
<tr>
<td>12:00pm-1:00 pm</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>1:00pm-3:00pm</td>
<td>Exercise: Hands-on experience using ATLAS and Criteria2Query</td>
<td>Cong Liu</td>
</tr>
<tr>
<td></td>
<td>You will be participating in formal evaluation of these phenotype definition tools, so your participation and feedback is greatly appreciated</td>
<td></td>
</tr>
<tr>
<td>3:00pm-3:15pm</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>3:15pm-4:30pm</td>
<td>Demo: Evaluating a phenotype using PheValuator</td>
<td>Joel Swerdel</td>
</tr>
<tr>
<td>4:30pm-5:00pm</td>
<td>Discussion: The journey ahead for phenotyping</td>
<td>All</td>
</tr>
</tbody>
</table>
Please sign up to form teams

- Each team will use C2Q and ATLAS to do **TWO** exercises, followed by a post-workshop survey.
  - We will pre-populate the concept set for you in the exercise
  - Assistants will be available on site
  - The expected duration for this exercise is **2 HOURS**

<table>
<thead>
<tr>
<th>team</th>
<th>trial</th>
<th>tool</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>DCP</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>TECOS</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>DCP</td>
<td>ATLAS</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>DCP</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>TECOS</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>DCP</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>2</td>
</tr>
</tbody>
</table>
We need your help in evaluating

• What I can get from the exercise?
  – Hands-on experiences on two OHDSI tools

• What is the goal of the evaluation?
  – Compare the **C2Q+ATLAS** process with the **ATLAS** system alone in order to identify strengths and weaknesses of the current C2Q prototype.

• Can I choose not to participate?
  – Yes, your participation is voluntary and you may choose to stop at any time.

• What data is collected?
  – No personally identifiable information will be collected
  – time spent, cohort, user experiences, user preferences, questions asked about the systems and background information regarding prior cohort building experience.
**Please report to TA at checkpoints**

<table>
<thead>
<tr>
<th>ATLAS</th>
<th>Checkpoint1: Start ATLAS along exercise</th>
<th>Time:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Checkpoint2: Cohort definition is</td>
<td>Cohort:</td>
</tr>
<tr>
<td></td>
<td>finished and saved</td>
<td>Time:</td>
</tr>
<tr>
<td></td>
<td>Checkpoint3: Start C2Q + ATLAS exercise</td>
<td>Time:</td>
</tr>
<tr>
<td></td>
<td>Checkpoint4: The initial cohort is</td>
<td>Cohort:</td>
</tr>
<tr>
<td></td>
<td>loaded in the ATLAS and changed the</td>
<td>Time:</td>
</tr>
<tr>
<td></td>
<td>cohort name</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Checkpoint5: Copy the initial cohort</td>
<td>Cohort:</td>
</tr>
<tr>
<td></td>
<td>and create a new ATLAS working cohort</td>
<td>Time:</td>
</tr>
<tr>
<td></td>
<td>Checkpoint6: When the final cohort is</td>
<td>Cohort:</td>
</tr>
<tr>
<td></td>
<td>finished and saved</td>
<td>Time:</td>
</tr>
</tbody>
</table>
Please save your cohorts accordingly

• [TEAM-01-ATLAS-F]

Change your cohort name here

Make sure you click save
<table>
<thead>
<tr>
<th>team</th>
<th>trial</th>
<th>tool</th>
<th>sequence</th>
<th>team</th>
<th>trial</th>
<th>tool</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>1</td>
<td>13</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>DCP</td>
<td>C2Q</td>
<td>2</td>
<td>13</td>
<td>DCP</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>TECOS</td>
<td>C2Q</td>
<td>1</td>
<td>14</td>
<td>TECOS</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>DCP</td>
<td>ATLAS</td>
<td>2</td>
<td>14</td>
<td>DCP</td>
<td>ATLAS</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>DCP</td>
<td>ATLAS</td>
<td>1</td>
<td>15</td>
<td>DCP</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>TECOS</td>
<td>C2Q</td>
<td>2</td>
<td>15</td>
<td>TECOS</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>DCP</td>
<td>C2Q</td>
<td>1</td>
<td>16</td>
<td>DCP</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>2</td>
<td>16</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>1</td>
<td>17</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>DCP</td>
<td>C2Q</td>
<td>2</td>
<td>17</td>
<td>DCP</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>TECOS</td>
<td>C2Q</td>
<td>1</td>
<td>18</td>
<td>TECOS</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>DCP</td>
<td>ATLAS</td>
<td>2</td>
<td>18</td>
<td>DCP</td>
<td>ATLAS</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>DCP</td>
<td>ATLAS</td>
<td>1</td>
<td>19</td>
<td>DCP</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>TECOS</td>
<td>C2Q</td>
<td>2</td>
<td>19</td>
<td>TECOS</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>DCP</td>
<td>C2Q</td>
<td>1</td>
<td>20</td>
<td>DCP</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>2</td>
<td>20</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>1</td>
<td>21</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>DCP</td>
<td>C2Q</td>
<td>2</td>
<td>21</td>
<td>DCP</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>TECOS</td>
<td>C2Q</td>
<td>1</td>
<td>22</td>
<td>TECOS</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>DCP</td>
<td>ATLAS</td>
<td>2</td>
<td>22</td>
<td>DCP</td>
<td>ATLAS</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>DCP</td>
<td>ATLAS</td>
<td>1</td>
<td>23</td>
<td>DCP</td>
<td>ATLAS</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>TECOS</td>
<td>C2Q</td>
<td>2</td>
<td>23</td>
<td>TECOS</td>
<td>C2Q</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>DCP</td>
<td>C2Q</td>
<td>1</td>
<td>24</td>
<td>DCP</td>
<td>C2Q</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>2</td>
<td>24</td>
<td>TECOS</td>
<td>ATLAS</td>
<td>2</td>
</tr>
</tbody>
</table>
• **Initial Events:**
  – sitagliptin new users

• **Inclusion:**
  – Age >=50
  – Has T2DM
  – Has HbA1c between 6.5% (48 mmol/mol) and 8.0% (64 mmol/mol) on stable dose(s) of antihyperglycemic agent(s), including insulin
  – Has pre-existing cardiovascular disease

• **Exclusion:**
  – Has a history of type 1 diabetes mellitus or ketoacidosis.
  – has not taken sitagliptin
• Initial Events:
  – Hydrochlorothiazide (HCTZ)/Chlorthalidone

• Inclusion:
  – Are over age 65 years
  – Are receiving hydrochlorothiazide in the previous 60 days.
  – Have a most recent systolic blood pressure (SBP) in CPRS greater than or equal to 120 mm Hg, with no SBP less than 120 mm Hg recorded in CPRS in the previous 90 days

• Exclusion:
  – K<3.1 meq/L in the past 90 days or K<3.5 meq/L if on digoxin
  – Na<130 meq/L in the past 90 days