APAC Community Call
Mini tutorials
May 19, 2022
Agenda

• OHDSI News
• Mini-tutorial #1 on Concept Set and Cohort Creation by Kyulee Jeon
• Mini-tutorial #2 on ATLAS Characterization by Yizhi Dong
Kyulee works as a researcher assistant under Dr. Seng Chan You for Biomedical Informatics in YUHS. She received B.S. in Public Health from Ewha Womans University, Republic of Korea in 2021. Her research interests are applying artificial intelligence technique to biomedical data with using Common Data Model.
APAC Community Call
Mini-tutorial 1
: Concept Set & Cohort creation

May 19, 2022
Kyulee Jeon (Yonsei University Health System)
Overview : Cohort in OHDSI
What is a “Cohort” in OHDSI?

• In OHDSI research we define a cohort as a set of persons who satisfy one or more inclusion criteria for a duration of time.
How to define a “Cohort” in OHDSI?

• OHDSI utilizes 3 building blocks as part of a cohort

1) Cohort entry event (initial event)
2) Inclusion criteria
3) Cohort exit
How to define a “Cohort” in OHDSI?

1) **Cohort entry event (initial event)**
   *define the time when people enter the cohort (cohort index date)*
How to define a “Cohort” in OHDSI?

2) **Inclusion criteria**

applied to the initial event cohort to further restrict the set of people

- OHDSI avoid using exclusion criteria deliberately as it often introduces double negatives. Instead, define everything as an inclusion criteria - e.g. include people with 0 occurrence.
How to define a “Cohort” in OHDSI?

3) **Cohort exit criteria**

signify when a person no longer qualifies for cohort membership
How to define a “Cohort” in OHDSI?

The standard components we use to assemble these criteria are:

• **Domain & Domain-specific attribute**
• **Concept set**
• **Temporal logic**

You need to assemble your concept sets to apply to your criteria.
Concept Set

[e.g.] Acute myocardial Infarction
A concept set consists of a list of concepts with attributes:

- **Exclude** this concept (and its descendants if selected) from the concept set
- Consider not only this concept, but also all of its **Descendants**
- **Mapped** to standard-concepts from non-standard concepts

[An example concept set expression]

- **With the Descendants option**, this concept set now includes nearly 100 concepts
Let’s create a Concept Set on ATLAS

• [https://atlas-demo.ohdsi.org/#/conceptsets](https://atlas-demo.ohdsi.org/#/conceptsets)
Cohort definition

[e.g.] Acute myocardial infarction (7Ps, 180Era)
Implementing a Cohort using ATLAS

• To begin in ATLAS, click on the **Cohort Definitions** module.
• When the module loads, click on **New Cohort**.
① Cohort entry event (initial event)

- define the time when people enter the cohort, called the cohort index date
- In this example, we want to find patients who diagnosed AMI for the first time

- Despite its appearance, the ❌ is not intended to mean “No” - It's a delete button
1 Cohort entry event (initial event)

- with continuous observation at least (α days) before and (β days) after event index date
② Inclusion criteria

- applied to the initial event cohort to further restrict the set of people
- In this case, required at least 1 occurrence of 365 days prior observation
③ Cohort exit criteria

- signify when a person no longer qualifies for cohort membership
- In this case, event will persist until 7 days relative to initial event
Let’s create a Cohort Definition on ATLAS

- [https://atlas-demo.ohdsi.org/#/cohortdefinitions](https://atlas-demo.ohdsi.org/#/cohortdefinitions)
Thanks
Yizhi works as a research assistant under Dr. Mengling Feng for Healthcare AI, Saw Swee Hock School of Public Health (SSHSPH) in NUS. Currently researching on healthcare event prediction using statistical and machine learning methods.
APAC Community Call
ATLAS Characterization Mini Tutorial

May 19, 2022
Characterization

Statistics of features regarding selected cohorts

- Create new characterization
- Create new feature analyses
- Generate report
- Export / import
Navigate to Characterization
New Characterization - Cohort Definition

Cohort characterization is defined as the process of generating cohort level descriptive summary statistics from person level covariate data. Summary statistics of these person level covariates may be count, mean, sd, var, min, max, median, range, and quantiles. In addition, covariates during a period may be stratified into temporal units of time for time-series analysis such as fixed intervals of time relative to cohort_start_date (e.g. every 7 days, every 30 days etc.), or in absolute calendar intervals such as calendar-week, calendar-month, calendar-quarter, calendar-year.

Cohort definitions

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>[PhenotypePhebruary] Persons with new type 2 diabetes mellitus at first diagnosis</td>
</tr>
<tr>
<td>6</td>
<td>[PhenotypePhebruary] Persons with new type 2 diabetes mellitus at first dx rx or lab</td>
</tr>
</tbody>
</table>

Edit cohort  Remove
<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>Charlson Index</td>
<td>The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end.</td>
</tr>
<tr>
<td>70</td>
<td>Demographics Ethnicity</td>
<td>Ethnicity of the subject.</td>
</tr>
<tr>
<td>71</td>
<td>Demographics Age Group</td>
<td>Age of the subject on the index date (in 5 year age groups)</td>
</tr>
<tr>
<td>72</td>
<td>Demographics Race</td>
<td>Race of the subject.</td>
</tr>
<tr>
<td>73</td>
<td>Demographics Prior Observation Time</td>
<td>Number of continuous days of observation time preceding the index date.</td>
</tr>
<tr>
<td>74</td>
<td>Demographics Gender</td>
<td>Gender of the subject.</td>
</tr>
<tr>
<td>77</td>
<td>Demographics Age</td>
<td>Age of the subject on the index date (in years).</td>
</tr>
<tr>
<td>79</td>
<td>Demographics Time in Cohort</td>
<td>Number of days of observation time during cohort period.</td>
</tr>
<tr>
<td>80</td>
<td>Demographics Index Year</td>
<td>Year of the index date.</td>
</tr>
<tr>
<td>105</td>
<td>Drugs for diabetes except for insulin</td>
<td></td>
</tr>
</tbody>
</table>
New Characterization – Feature Analyses

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>Chads 2</td>
<td>The CHADSZ score using all conditions prior to the window end.</td>
</tr>
<tr>
<td>82</td>
<td>Chads 2 Vasc</td>
<td>The CHADS2VASc score using all conditions prior to the window end.</td>
</tr>
<tr>
<td>49</td>
<td>Charlson Index</td>
<td>The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end.</td>
</tr>
<tr>
<td>10</td>
<td>Condition Era Any Time Prior</td>
<td>One covariate per condition in the condition-era table overlapping with any time prior to index.</td>
</tr>
<tr>
<td>23</td>
<td>Condition Era Long Term</td>
<td>One covariate per condition in the condition-era table overlapping with any part of the long term window.</td>
</tr>
<tr>
<td>16</td>
<td>Condition Era Medium Term</td>
<td>One covariate per condition in the condition-era table overlapping with any part of the medium term window.</td>
</tr>
<tr>
<td>17</td>
<td>Condition Era Overlapping</td>
<td>One covariate per condition in the condition-era table overlapping with the end of the risk window.</td>
</tr>
<tr>
<td>4</td>
<td>Condition Era Short Term</td>
<td>One covariate per condition in the condition-era table overlapping with any part of the short term window.</td>
</tr>
<tr>
<td>9</td>
<td>Condition Era Start Long Term</td>
<td>One covariate per condition in the condition-era table starting in the long term window.</td>
</tr>
</tbody>
</table>
New Characterization – Subgroup (Optional)

Subgroup analyses

- New subgroup

Age group

- Calculate subgroup analyses only

greater than 60

having all of the following criteria:

- with the following event criteria:
  - with age Greater Than 60
## Navigate to Feature Analyses

### Cohort Characterizations

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Condition Group Era Start Medium Term</td>
<td>One covariate per condition era rolled up to groups in the condition-era table starting in the medium term window.</td>
</tr>
<tr>
<td>12</td>
<td>Drug Exposure Long Term</td>
<td>One covariate per drug in the drug_exposure table starting in the long term window.</td>
</tr>
<tr>
<td>13</td>
<td>Measurement Range Group Long Term</td>
<td>Covariates indicating whether measurements are below, within, or above normal range in the long term window.</td>
</tr>
<tr>
<td>14</td>
<td>Measurement Range Group Medium Term</td>
<td>Covariates indicating whether measurements are below, within, or above normal range in the medium term window.</td>
</tr>
<tr>
<td>15</td>
<td>Drug Group Era Any Time Prior</td>
<td>One covariate per drug rolled up to ATC groups in the drug-era table overlapping with any time prior to index.</td>
</tr>
<tr>
<td>16</td>
<td>Condition Era Medium Term</td>
<td>One covariate per condition in the condition-era table overlapping with any part of the medium term window.</td>
</tr>
<tr>
<td>17</td>
<td>Condition Era Overlapping</td>
<td>One covariate per condition in the condition-era table overlapping with the end of the risk window.</td>
</tr>
<tr>
<td>18</td>
<td>Condition Era Start Short Term</td>
<td>One covariate per condition in the condition-era table starting in the short term window.</td>
</tr>
</tbody>
</table>
New Feature Analyses

Analysis type: Prevalence

Criteria: Hyperglycemia

- with at least 1 using all occurrences of a condition occurrence of Hyperglycemia
- event starts between 0 to N days before and all days after
- restrict to the same visit occurrence
- allow events from outside observation period

Import from concept sets
Generate Report
### ATLAS

- **Home**
- **Data Sources**
- **Search**
- **Concept Sets**
- **Cohort Definitions**
- **Characterizations**
- **Cohort Pathways**
- **Incidence Rates**
- **Profiles**
- **Estimation**
- **Prediction**
- **Jobs**
- **Configuration**
- **Feedback**

### Report

#### Characterization #2

- **Mini Tutorial**

#### Executions ➤ Reports for SynPUF 110k

- **Date:** 05/13/2022 1:45 PM  
  - **Design:** 1042662361  
  - **Results:** 12 reports

#### Filter panel

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>Analyses</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>[PhenotypePhebruary] Persons with new type 2 diabe</td>
<td>12 items selected</td>
<td>3 items selected</td>
</tr>
</tbody>
</table>

#### All prevalence covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Explore</th>
<th>Concept ID</th>
<th>[PhenotypePhebruary] Persons with new type 2 diabetes mellitus at first diagnosis</th>
<th>[PhenotypePhebruary] Persons with new type 2 diabetes mellitus at first dx rx or lab</th>
<th>Std diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Pct</td>
<td>Count</td>
<td>Pct</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Search:**
### All prevalence covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Explore</th>
<th>Concept ID</th>
<th>Person with new type 2 diabetes mellitus at first diagnosis</th>
<th>Person with new type 2 diabetes mellitus at first dx rx or lab</th>
<th>Std diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>diabetes drug except for insulin</td>
<td>Explore</td>
<td>0</td>
<td>Count 4,257 Pct 29.17%</td>
<td>Count 1,779 Pct 20.38%</td>
<td>-0.1248</td>
</tr>
<tr>
<td>insulin used</td>
<td>Explore</td>
<td>0</td>
<td>Count 286 Pct 1.96%</td>
<td>Count 49 Pct 0.56%</td>
<td>-0.0880</td>
</tr>
<tr>
<td>2010</td>
<td>N/A</td>
<td>0</td>
<td>Count 4,622 Pct 31.67%</td>
<td>Count 2,532 Pct 29.01%</td>
<td>-0.0341</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>N/A</td>
<td>38003563</td>
<td>Count 349 Pct 2.39%</td>
<td>Count 160 Pct 1.83%</td>
<td>-0.0271</td>
</tr>
<tr>
<td>MALE</td>
<td>N/A</td>
<td>8507</td>
<td>Count 6,498 Pct 44.52%</td>
<td>Count 4,086 Pct 46.81%</td>
<td>0.0240</td>
</tr>
<tr>
<td>65-69</td>
<td>N/A</td>
<td>0</td>
<td>Count 2,688 Pct 17.73%</td>
<td>Count 1,672 Pct 19.16%</td>
<td>0.0235</td>
</tr>
<tr>
<td>2009</td>
<td>N/A</td>
<td>0</td>
<td>Count 9,971 Pct 68.31%</td>
<td>Count 6,194 Pct 70.97%</td>
<td>0.0225</td>
</tr>
<tr>
<td>FEMALE</td>
<td>N/A</td>
<td>8532</td>
<td>Count 8,098 Pct 55.48%</td>
<td>Count 4,642 Pct 53.19%</td>
<td>-0.0220</td>
</tr>
<tr>
<td>40-44</td>
<td>N/A</td>
<td>0</td>
<td>Count 181 Pct 1.24%</td>
<td>Count 84 Pct 0.96%</td>
<td>-0.0187</td>
</tr>
<tr>
<td>95-99</td>
<td>N/A</td>
<td>0</td>
<td>Count 278 Pct 1.90%</td>
<td>Count 140 Pct 1.60%</td>
<td>-0.0160</td>
</tr>
</tbody>
</table>

[Graph showing correlation between two variables]
### DEMOGRAPHICS / Demographics Time In Cohort

#### Strata

<table>
<thead>
<tr>
<th>Strata</th>
<th>[Phenotype:Febuary] Persons with new type 2 diabetes mellitus at first diagnosis</th>
<th>[Phenotype:Febuary] Persons with new type 2 diabetes mellitus at first dx rx or lab</th>
<th>Std diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Avg</td>
<td>Std Dev</td>
</tr>
<tr>
<td>All stratas</td>
<td>14,596</td>
<td>369.35</td>
<td>188.91</td>
</tr>
</tbody>
</table>

Showing 1 to 1 of 1 entries

### CONDITION / Charlson Index

#### Strata

<table>
<thead>
<tr>
<th>Strata</th>
<th>[Phenotype:Febuary] Persons with new type 2 diabetes mellitus at first diagnosis</th>
<th>[Phenotype:Febuary] Persons with new type 2 diabetes mellitus at first dx rx or lab</th>
<th>Std diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Avg</td>
<td>Std Dev</td>
</tr>
<tr>
<td>All stratas</td>
<td>14,596</td>
<td>4.66</td>
<td>3.38</td>
</tr>
</tbody>
</table>

Showing 1 to 1 of 1 entries
Export / Import

```json
{
  "name": "Mini Tutorial",
  "cohorts": [
    {
      "id": 4,
      "name": "[PhenotypeFebruary] Persons with new type 2 diabetes mellitus at first diagnosis",
      "description": null,
      "createdDate": "2022-04-12 09:17",
      "modifiedDate": "2022-05-13 05:38",
      "expression": {
        "cdmVersionRange": ">=5.0",
        "PrimaryCriteria": {
          "CriterionList": [
            {
              "ConditionOccurrence": {
                "CodesetId": 0,
                "ConditionTypeExclude": false
              }
            }
          ]
        }
      }
    }
  ]
}
```
Thank you!