Examining the Differences in Baseline Characteristics of One-code and Two-code Phenotype Algorithms

**Presenters:** Pranav Bhimani, Raechel Davis

**BACKGROUND**

- The guidance and implications regarding broad and narrow phenotype algorithm (PA) use remain unclear.
- Broad PAs requiring one diagnostic code identify a greater number of subjects, often producing higher sensitivities, albeit with lower positive predictive values (PPVs).
- Narrow PAs that require a second diagnostic code during some timeframe after the first diagnostic code, are often accompanied by lower sensitivities but produce higher PPVs.
- The objective of this study was to compare the similarity of baseline characteristics for phenotype algorithms requiring one and two diagnostic codes for health outcomes in therapeutic areas of neurology, immunology, oncology, and cardiology using six real-world databases.

**METHODS**

- A network of six US observational databases that were transformed to the Observational Medical Outcomes Partnership (OMOP) Common Data Model version 5.3.1 were used [1].
- PAs for associated outcomes within each therapeutic area were analyzed.

### Table 1. US Databases with Inpatient and Outpatient Visit Types

<table>
<thead>
<tr>
<th>Database</th>
<th>Years</th>
<th>Number of Persons (millions)</th>
<th>Median Follow-up (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAE</td>
<td>2000-2021</td>
<td>162</td>
<td>1.56</td>
</tr>
<tr>
<td>MDCD</td>
<td>2006-2020</td>
<td>33</td>
<td>1.52</td>
</tr>
<tr>
<td>MDCR</td>
<td>2000-2021</td>
<td>10</td>
<td>2.46</td>
</tr>
<tr>
<td>Optum DOD</td>
<td>2007-2021</td>
<td>92</td>
<td>1.48</td>
</tr>
<tr>
<td>Optum EHR</td>
<td>2007-2021</td>
<td>105</td>
<td>2.63</td>
</tr>
<tr>
<td>Pharmatics</td>
<td>2013-2021</td>
<td>162</td>
<td>3.25</td>
</tr>
</tbody>
</table>

**RESULTS**

- In six of the nine outcomes there was minimal variability in the comparison of baseline covariates.
- Comparisons between the Ulcerative Colitis and Crohn’s Disease one-code and two-code PAs showed the greatest variability, while the Multiple Sclerosis comparison showed a moderate level of variability (Figure 1).
- Sensitivities of one-code algorithms were observed to be higher and less variable across databases than two-code algorithms (Figure 2).

**CONCLUSIONS**

- Comparisons of baseline characteristics in 1-code and 2-code PAs in most (6 of the 9) outcomes showed minimal variability.
- For outcomes in specific therapeutic areas such as immunology, greater variability in baseline covariates may be present when comparing 1 and 2 code algorithms.
- Comparison of the similarity of baseline covariates between phenotype algorithms provides a more complete understanding of algorithm differences.

**REFERENCES**