Data quality of OHDSI APAC: CDM Inspection study

2022-05-24

Community call

OHDSI APAC Study Team 4
Background

Figure is recreated from Maxim Moinat slides (21/11/10, at OHDSI community call)
Objectives

What is this study for?
• Collecting CDM Inspection reports from APAC community

Why this study is needed?
• To check the current status of CDMs, get insights from the CDMs, and improve their data quality

What is the final goal?
• Disclosure of current status of conversion, contents, and data distribution of CDMs of the OHDSI APAC community.
• To provide the basic statistics which can be used as references for future CDM conversion
Study package

• Data sources: CDM databases from OHDSI APAC community
• Collecting inspection reports from each site.
• R package for automatically creating inspection reports.

• Collectibles
  • Number of record, person
  • Number of unique concepts per person
  • Source-CDM mapping ratio
  • Proportion of standard concepts in mapped codes
  • Drug mapping level (granularity)
  • Frequent concept list in each domain
  • Achilles heel result (error / notification / warnings)
  • Sample cohort generation
Study participants

China
• 1 EHR (Wonders group)

Japan
• 2 Claims (IQVIA, JNJ)

South Korea
• 21 EHRs (OHDSI Korea)
• 1 Claims

Singapore
• 2 EHRs (OHDSI Singapore)

Australia
• 4 EHRs (JnJ, 3 OHDSI Australia)

5 Regions
27 EHRs
3 Claims

— Planned
— Done
Preliminary results

Total n of records = 36,096,359,491
Total n of persons = 48,127,420
Preliminary results

Table 2. Summary result of record mapping to the OMOP concept from common data model databases

<table>
<thead>
<tr>
<th>Domain</th>
<th>Mapping codes / source codes</th>
<th>Mapped records / total records</th>
<th>Mapped as standard / Mapped records</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median [Q1, Q3]</td>
<td>Median [Q1, Q3]</td>
<td>Median [Q1, Q3]</td>
</tr>
<tr>
<td>Condition occurrence</td>
<td>97.5 [87.6, 99.6]</td>
<td>99.5 [94.8, 100.0]</td>
<td>100.0 [99.0, 100.0]</td>
</tr>
<tr>
<td>Device exposure</td>
<td>56.6 [48.6, 78.2]</td>
<td>75.4 [64.7, 92.2]</td>
<td>79.3 [54.2, 100.0]</td>
</tr>
<tr>
<td>Drug exposure</td>
<td>85.0 [74.8, 90.6]</td>
<td>97.0 [96.0, 98.2]</td>
<td>98.3 [97.7, 99.1]</td>
</tr>
<tr>
<td>Measurement</td>
<td>50.1 [25.4, 87.1]</td>
<td>97.0 [69.1, 99.7]</td>
<td>100.0 [99.7, 100.0]</td>
</tr>
<tr>
<td>Measurement-unit</td>
<td>96.7 [29.7, 100.0]</td>
<td>100.0 [40.8, 100.0]</td>
<td>100.0 [100.0, 100.0]</td>
</tr>
<tr>
<td>Measurement-value</td>
<td>13.3 [4.4, 46.0]</td>
<td>7.5 [4.1, 49.4]</td>
<td>100.0 [100.0, 100.0]</td>
</tr>
<tr>
<td>Observation</td>
<td>100.0 [98.3, 100.0]</td>
<td>100.0 [93.0, 100.0]</td>
<td>100.0 [100.0, 100.0]</td>
</tr>
<tr>
<td>Observation-unit</td>
<td>100.0 [61.1, 100.0]</td>
<td>100.0 [50.4, 100.0]</td>
<td>97.8 [44.4, 100.0]</td>
</tr>
<tr>
<td>Observation-value</td>
<td>50.0 [50.0, 100.0]</td>
<td>92.1 [77.8, 100.0]</td>
<td>100.0 [100.0, 100.0]</td>
</tr>
<tr>
<td>Procedure occurrence</td>
<td>66.3 [57.5, 97.2]</td>
<td>38.7 [22.4, 95.1]</td>
<td>100.0 [85.1, 100.0]</td>
</tr>
<tr>
<td>Visit occurrence</td>
<td>100.0 [100.0, 100.0]</td>
<td>100.0 [100.0, 100.0]</td>
<td>100.0 [100.0, 100.0]</td>
</tr>
</tbody>
</table>

Table 3. Vocabulary granularity in drug exposure table

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Classification</th>
<th>N of records</th>
<th>Mapped records / total records, Median [Q1, Q3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RxNorm (Extension)</td>
<td>Branded Drug</td>
<td>607,710,428</td>
<td>40.1 [1.1, 57.1]</td>
</tr>
<tr>
<td></td>
<td>Clinical Drug</td>
<td>722,805,309</td>
<td>9.4 [2.3, 46.6]</td>
</tr>
<tr>
<td></td>
<td>Quant Branded Drug</td>
<td>266,331,050</td>
<td>17.3 [0.3, 30.6]</td>
</tr>
</tbody>
</table>
Thank you

Chungsoo Kim
ted9219@ajou.ac.kr