**How the vocabulary change affects cohorts?**

**INTRO:**
- Vocabulary changes are drastic, for example, between Apr-2022 and Jan-2023 releases:
  - 19,965 concepts changed their standard status,
  - 80,021 concepts changed their mapping,
  - 58,327 concepts changed their immediate parent
- This affects the ETL output and concept sets resolution and thus the cohorts.
- And we must know what exactly changed for our studies!

**METHODS**
1. Concept set definitions are extracted from cohort definitions and resolved on two different vocabulary versions (including source concepts).
2. Output is the comparison of source and standard concepts included.
3. Results are written in Excel file with a tab for each metric.

**RESULTS**
The tool provides a recipe of what should be changed in a concept set expression accordingly to the vocabulary update:

<table>
<thead>
<tr>
<th>Comparison output</th>
<th>User’s action</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-standard concepts with their replacement mappings</td>
<td>Replace these concepts with their mapped equivalents</td>
</tr>
<tr>
<td>concepts above which the hierarchy is altered (peak concepts)</td>
<td>Add peak concepts to concept set definitions to preserve the hierarchy</td>
</tr>
<tr>
<td>Added or excluded source concepts with their old and new target concepts</td>
<td>Add or exclude new target concepts</td>
</tr>
<tr>
<td>domain changes in included concepts</td>
<td>Change the event table</td>
</tr>
</tbody>
</table>

- If detected changes are due to errors in the vocabulary update, submit a report for a vocabulary change.

**Example of the output and action required**

<table>
<thead>
<tr>
<th>Hierarchy change</th>
<th>level</th>
<th>cohortdefinitionid</th>
<th>cohortname</th>
<th>conceptsetname</th>
<th>node_concept_id</th>
<th>node_concept_name</th>
<th>isexcluded</th>
<th>includedescendants</th>
<th>action</th>
<th>peak_concept_id</th>
<th>peak_concept_name</th>
<th>peak_drc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>10657</td>
<td>Rhabdomyolysis</td>
<td>Trauma</td>
<td>440921</td>
<td>Traumatic injury</td>
<td>0</td>
<td>1</td>
<td>Added</td>
<td>436676</td>
<td>Posttraumatic stress disorder</td>
<td>167673919</td>
</tr>
</tbody>
</table>

**Explanation:** In this case, the ‘Posttraumatic stress disorder’ branch (peak concept) is the new descendant of “Traumatic injury” (Node concept) hierarchy. This is an important change, since the count of concepts in this branch in a data source is 167673919 (peak_drc).

**Action:** PTSD is not a part of physical trauma concept set, so it should be fixed by
- submitting the issue to the OMOP vocabulary.

**Figure 1. Medical event inclusion into cohort, simplified.**

**Numbers correspond to the metrics below.**

**Changes captured by the tool**
1. Related source concepts that were added or removed due to mapping change.
2. Included standard concepts changed their domain
3. Hierarchy changes that affect concept set resolution*
4. Non-standard concepts used in concept set definitions.

* The concept above which the hierarchy is altered is shown

Scan to access the github page