

# Validity checks for Cohort Definitions

Gowtham A. Rao

2023-02-13

## Contents

|     |                                     |   |
|-----|-------------------------------------|---|
| 0.1 | Name . . . . .                      | 1 |
| 0.2 | Cohort definition checks: . . . . . | 1 |

PhenotypeLibrary is part of HADES

*This document is a DRAFT guidance.*

### 0.1 Name

- *Length* - cohort definition name should not be longer than 90 characters
- *Name clash* - please check if there is another valid cohort with the exact same name. If yes, then we ask that you come up with another name. You can use the name of a deprecated cohort.
- *Reserved characters* - should not have '[' or ']' in cohort name, as it is reserved for use in Phenotype library.

### 0.2 Cohort definition checks:

- *Concept set mismatch*
  - We do not allow cohort definitions with unused concept sets
  - At least one of the concept id(s) in the concept set should belong the domain of the rule. E.g. if the rule says 'at least 1 occurrence of 'condition\_occurrence' - but the conceptId's do not even have one conceptId belonging to the condition domain.
- *Measurement units*
  - If using measurement - make sure units are specified. Cohort Definitions using measurement domains that do not have units will be automatically flagged for rejection. Reason: values may be on various scales and so value as number may not be meaningful.
  - Measurement\_unit\_concept\_id should be standard omop vocabulary that belongs to the domain Unit and should be valid. See Athena
  - If measurement\_unit\_concept\_id are used then we expect that each rule only combines units that are on the same scale i.e. the unit\_concept\_id should have a conversion factor = 1.
  - Exceptions
    - \* unit less measurements and if there is only one unit.
- *Concept Sets used in Cohort Definitions*

- Concept set names: We expect the names of the concept sets to be meaningful to a broad audience i.e. please remove any internal identifiers that are specific to an organization or project e.g. '[raos] Hypertension' is not acceptable.
- Use of Non-standard concept codes: The OHDSI Phenotype library will flag cohort definitions that have non-standard concepts in the concept set expression for additional review. Additional justification will be required to accept a cohort definition for peer review if it has non-standard concepts. The justification will have to address, why a standard equivalent is not used and if the use of the non standard concepts changes the operating characteristics of the cohort definition compared to the use of standard concepts.
- Concept optimization: we expect the concept set to be an expression that is as parsimonious in its representation as possible. At a minimum, the concept set should be optimized, should not have duplicate concepts with same selection of included and exclude.
- Use of mapped concepts: concept set expression allows for selecting mapped concepts. When mapped concepts is selected in the concept set expression, it will be flagged for additional review. Mapped concepts are rarely needed and so will need additional justification.
- – Event cohorts\*
- If in your phenotype a person may reasonably be expected to not be in the phenotype after starting in the phenotype: we expect an attempt to model the cohort definition as an event cohort i.e. cohort end date should be modeled using exit criteria.