Validity checks for Cohort Definitions

Gowtham A. Rao

2022-09-06

Contents

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This document is a DRAFT guidance.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 does 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.