Validity checks for Cohort Definitions

Gowtham A. Rao

2023-09-22

Contents

PhenotypeLibrary is part of HADES

This document is a DRAFT guidance.

0.1 Cohort definition checks:

Concept set mismatch - No unused concept sets. Cohort Definitions should not have concept sets that are not used. - Match concept id domain to atlas criteria. E.g. if Atlas criteria 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 - Units should be specified. Since values may be on different scales units should be used. - Use valid, standard unit concept id. See Measurement_unit_concept_id options Athena - One criterion should have only one scale. - Exceptions: unit less measurements.

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: Justification is required to accept a cohort definition with non-standard concepts such as why a standard equivalent is not used, if use of the non standard concepts changes the operating characteristics of the cohort definition compared to the use of standard concepts, attempt to improve vocabulary mapping. - Concept optimization: we expect the concept set to be an expression that is as parsimonious in its representation. 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 should not have mapped selected. Mapped concepts are rarely needed.

Event cohorts - We expect cohorts to be event cohorts, i.e. a person may enter the cohort more than once in different time periods. If a cohort definition is modeled as a first occurrence/earliest event, we ask for an explanation. To model a cohort as an event cohort, rules for cohort end date should be thought through.