Using the measurement cohort covariate code

Jenna M. Reps

2020-10-13

Contents

1 Introduction

This vignette describes how one can use the function 'createMeasurementCohortCovariateSettings' to define measurement covariates that also require being in or not being in a cohort during the time period using the OMOP CDM. You will need:

1. A concept set for the measurements (a vector of measurement_concept_ids)
2. A function to standardise the measurements (e.g., filters out unlikely values or converts units)
3. A cohort database schema, a cohort table and cohort definition id for the cohort of interest
4. How to aggregate multiple measurement values (e.g., use most recent to index, max value, min value or mean value)

1.1 createMeasurementCohortCovariateSettings

This function contains the settings required to define the measurement cohort covariate. For a measurement cohort covariate, the code will check the measurement table in the OMOP CDM to find all rows where the measurement_concept_id is in the specified measurement concept set and will then restrict to patients in (or not in if 'type' == out) the 'cohortDatabaseSchema'.'cohortTable' with the cohort_definition_id of 'cohortId' between the index date plus the 'startDay' and the index date plus the 'endDay'. It will then check whether the measurement_date column calls between the index date plus the 'startDay' and the index date plus the 'endDay'. The 'scaleMap' will map the measurement values to a uniform scale - this standardises the values. If there are multiple measurements within the time period then the 'aggregateMethod' method with specify how to get a single value. The settings 'ageInteraction' and 'lnAgeInteraction' enable the user to create age/ln(age) interaction terms. The 'lnValue' enables the user to use the natural logarithm of the measurment value. Finally, the 'analysisId' is used to create the cohort covariateId as 1000*'measurementId' + 'analysisId'.

Table 1: The in

Input	Description
covariateName	The name of the covariate
covariateId	The id of the covariate - generally measurementId*1000+analysisId
cohortDatabaseSchema	The database schema with the cohort used to create a covariate
cohortTable	The table with the cohort used to create a covariate
cohortId	The cohort definition id for the cohort used to create a covariate
conseptSet	A vector of concept_ids corresponding to the measurement
type	in or out - in means the patients with a measurement must be in the cohort of interest during the

Input	Description
startDay	How many days prior to index to see whether the measurement occurs after
endDay	How many days relative to index to see whether the mesurement occurs before
scaleMap	A function that takes the covariate Amdromeda table as input and processes it - can include filter
aggregateMethod	How to pick a measurement value when there are more than 1 during the start and end dates - ca
imputationValue	A value to use if a person has no measurement during the start and end dates
ageInteraction	Include interaction with age
lnAgeInteraction	Include interaction with ln(age)
lnValue	Whether to us the natural log of the measurement value
analysisId	The analysis id for the covariate

1.2 Example

Assuming the concept set c(3004249, 3009395, 3018586, 3028737, 3035856, 4152194, 4153323, 4161413, 4197167, 4217013, 4232915, 4248525, 4292062, 21492239, 37396683, 44789315, 44806887, 45769778) corresponds to 'Systolic blood pressure'. The cohort in 'your database schema'.'cohort' with cohort_definition_id of 123 corresponds to periods of time where a patient is given an anti-hypertensive drug.

We create a function to map the covariate object (this contains the measurementConceptId, unitConceptId, rawValue and valueAsNumber columns) to standardise the measurement values. As the systolic blood pressure measurements often have no unit_cocept_id we cannot standarise based mapping the units. Instead, We remove unfeasible values such as any values less than 50 and greater than 250.

```
function(x){ x = dplyr::filter(x, rawValue >= 50 & rawValue <= 250 ); return(x)}
```

We include all measurement values with occured within 1 year prior to index and up to 60 days after, but use the value that occurred closest to index.

To create a treated systolic blood pressure covariate (a blood pressure measurement where the patient is in the anti-hypertensive cohort) using a measurement cohort covariate run:

```
cohortCov1 <- createCohortCovariateSettings(covariateName = 'Treated systolic blood pressure',
                                            covariateId = 1*1000+458,
                                            cohortDatabaseSchema = 'your database schema',
                                            cohortTable = 'cohort',
                                            cohortId = 123,
                                            conseptSet = c(3004249, 3009395, 3018586, 3028737, 3035856,
                                            type = 'in',
                                            startDay= -365,
                                            endDay=60,
                                            scaleMap = function(x){ x = dplyr::filter(x, rawValue >= 50
                                            aggregateMethod= 'recent',
                                            ageInteraction = FALSE,
                                            lnAgeInteraction = FALSE,
                                            lnValue = FALSE,
                                            analysisId = 458)
```

To create an untreated systolic blood pressure covariate (a blood pressure measurement where the patient is not in the anti-hypertensive cohort) using a measurement cohort covariate run:

```
cohortCov1 <- createCohortCovariateSettings(covariateName = 'Untreated systolic blood pressure',
                                            covariateId = 2*1000+458,
                                            cohortDatabaseSchema = 'your database schema',
                                            cohortTable = 'cohort',
                                            cohortId = 123,
                                            conseptSet = c(3004249, 3009395, 3018586, 3028737, 3035856,
```

```
                                                type = 'out',
                                                startDay= -365,
                                                endDay=60,
                                                scaleMap = function(x){ x = dplyr::filter(x, rawValue >= 50
                                                aggregateMethod= 'recent',
                                                ageInteraction = FALSE,
                                                lnAgeInteraction = FALSE,
                                                lnValue = FALSE,
                                                analysisId = 458)
```

You can use the ageInteraction, lnAgeInteraction and lnValue to do log mapping or include age interaction terms.

To include all the above as covariates, combine them into a list:

```
cohortCov <- list(cohortCov1,cohortCov2)
```