

# Adding custom covariates based on ATLAS cohorts

Jenna M. Reps

2020-03-15

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Step 1: Populate custom covariate cohorts . . . . .	1
1.2	Step 2: Build the study package . . . . .	2
1.3	Step 3: Execute the study with cohortVariableSetting . . . . .	2
1.4	Extras . . . . .	3

## 1 Introduction

This vignette describes how one can add custom covariates using ATLAS cohorts (more complex covariates than simple concepts - you can add logic) into the study package. This will enable users to develop models that incorporate advanced covariates. This vignette assumes you have already created the covariate cohorts in ATLAS. In general, you want to make sure the ATLAS cohorts used for covariates use all events rather than restricted to first of last event.

First make sure to open the R project in R studio, this can be done by finding the [atlas package name].Rproj file in the folder downloaded via ATLAS (you need to extract this from the zipped file downloaded). Once the package project is opened in R studio there are 3 steps that must be followed:

1. Run the function: populateCustomCohortCovariates (found in extras/PackageMaintenance.R on line 51) to extract the atlas cohorts used by the custom covariates into the study package
2. Build the study package
3. Run the study package execute function with the correct setting for the input 'cohortVariableSetting'.

### 1.1 Step 1: Populate custom covariate cohorts

The custom covariates that use ATLAS cohorts can be added to the study package by using the function 'populateCustomCohortCovariates()' that is found in extras/PopulateCustomCovariate.R.

To add the function to your environment, make sure the package R project is open in R studio and run:

```
source('./extras/PopulateCustomCovariate.R')
```

This will make the function 'populateCustomCohortCovariates()' available to use within your R session.

The 'populateCustomCohortCovariates()' function requires users to specify:

- settingsName - This is a string ending in 'csv' that specifies the name of the settings file defining the custom cohort covariate settings that will be generated into the study package
- settingsLocation - This is the directory where the custom cohort covariate settings will be saved. This should be the 'inst/settings' directory within the package.
- baseUrl - The url for the ATLAS webapi (this will be used to extract the ATLAS cohorts)

- atlasIds - an integer or vector of integers specifying the atlas cohort Ids that are used by the custom cohort covariates
- atlasNames - a string or vector of strings specifying the names of the atlas ids (must be the same length as atlasIds)
- startDays - a negative integer or vector of negative integers specifying the days relative to index to start looking for the patient being in the covariate cohort
- endDays - a negative integer (or zero) or vector of negative integers (or zero) specifying the days relative to index to stop looking for the patient being in the covariate cohort

For example, to create two custom cohort covariates into the package I can run:

```
populateCustomCohortCovariates(settingsName = 'customCohortCov.csv',
                               settingsLocation = ".inst/settings",
                               baseUrl = 'https://atlas_webapi',
                               atlasIds = c(1,109),
                               atlasNames = c('Testing 1', 'Testing 109'),
                               startDays = c(-999,-30),
                               endDays = c(-1,0))
```

The code above extracts two ATLAS cohort covariates: \* covariate 1: The ATLAS cohort with the id of 1 named 'Testing 1' looks for patients who have a Testing 1 cohort\_start\_date between (index date-999 days) and (index date-1 days). If a patient is in the Testing 1 cohort 50 days before the index date then they will have a value of 1 for the custom covariate. If they are not in the Testing 1 cohort between 999 days before index and 1 day before index then they will have a value of 0 for the custom covariate. \* covariate 2: The ATLAS cohort with the id of 109 named 'Testing 109' looks for patients who have a Testing 109 cohort\_start\_date between (index date-30 days) and (index date). If a patient is in the Testing 109 cohort 20 days before the index date then they will have a value of 1 for the custom covariate. If they are not in the Testing 1 cohort between 30 days before index and the day of index then they will have a value of 0 for the custom covariate.

## 1.2 Step 2: Build the study package

After adding the custom cohort covariates into the package, you now need to build the package. Use the standard process (in R studio press the 'Build' tab in the top right corner and then select the 'Install and Restart' button) to build the study package so an R library is created.

## 1.3 Step 3: Execute the study with cohortVariableSetting

To include the custom covariate that uses ATLAS cohorts into the model set the input 'cohortVariableSetting' to the value you chose for 'settingsName' in step 1 (e.g., in my example I specified settingsName = 'customCohortCov.csv' so in execute() I need to set cohortVariableSetting = 'customCohortCov.csv'):

```
execute(connectionDetails = connectionDetails,
        cdmDatabaseSchema = cdmDatabaseSchema,
        cdmDatabaseName = cdmDatabaseName,
        cohortDatabaseSchema = cohortDatabaseSchema,
        cohortTable = cohortTable,
        oracleTempSchema = oracleTempSchema,
        outputFolder = outputFolder,
        createProtocol = F,
        createCohorts = T,
        runAnalyses = T,
        createResultsDoc = F,
        packageResults = F,
```

```
createValidationPackage = F,  
minCellCount= 5,  
cohortVariableSetting = 'customCohortCov.csv'  
)
```

This will now run the study but will include the additional covariates you specified using ATLAS cohorts. The study package will create the cohorts used for covariates when `createCohorts = T`, so the cohort creation step will take longer due to additional cohorts.

## 1.4 Extras

You can create multiple custom ATLAS covariate settings using `populateCustomCohortCovariates()` with different `settingsName` and pick the one you want when you execute the study.