

# Populating the study package

Jenna M. Reps

2020-03-27

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Step 1: Populate skeleton settings . . . . .	1
1.2	Step 2: Build the study package . . . . .	3
1.3	Step 3: Execute the study to validate an existing model . . . . .	3

## 1 Introduction

This vignette describes how one can populate the `SkeletonExistingModelStudy` package with the target cohort, outcome cohorts and model settings.

First make sure to open the `Skeleton` R project in R studio, this can be done by finding the `SkeletonExistingModelStudy.Rproj` file in the folder. Once the package project is opened in R studio there are 3 steps that must be followed:

1. Run the function: `populatePackage` (found in `extras/populatePackage.R` on line 51) to add all cohorts and settings into the study package
2. Build the study package
3. Run the study package execute function

### 1.1 Step 1: Populate skeleton settings

All the settings can be added to the study package by using the function `'populatePackage()'` that is found in `extras/populatePackage.R`.

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

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

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

The `'populatePackage()'` function requires users to specify:

- `targetCohortId` - The ATLAS id for the target cohort
- `targetCohortName` - A string with a sharable name for the target cohort
- `outcomeId` - The ATLAS id for the outcome cohort
- `outcomeName` - A string with a sharable name for the outcome cohort
- `standardCovariates` - A data.frame with the columns: `covariateId` (for standard features using Feature-Extraction), `covariateName` and points to assign points for standard covariates
- `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
- points - a double or vector of doubles specifying the points corresponding to each variable

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

```
populatePackage(targetCohortId = 10845,
               targetCohortName = 'neg mammo',
               outcomeId = 10082,
               outcomeName = 'breast cancer',
               standardCovariates = data.frame(covariateId = c(0003, 1003,
                                                                2003, 3003,
                                                                4003, 5003,
                                                                6003, 7003,
                                                                8003, 9003,
                                                                10003, 11003,
                                                                12003, 13003,
                                                                14003, 15003,
                                                                16003, 17003,
                                                                8507001),
                                               covariateName = c('Age 0-4', 'Age 5-9',
                                                             'Age 10-14', 'Age 15-19',
                                                             'Age 20-24', 'Age 25-30',
                                                             'Age 30-34', 'Age 35-40',
                                                             'Age 40-44', 'Age 45-50',
                                                             'Age 50-54', 'Age 55-60',
                                                             'Age 60-64', 'Age 65-70',
                                                             'Age 70-74', 'Age 75-80',
                                                             'Age 80-84', 'Age 85-90',
                                                             'Male'),
                                               points = c(rep(0,19))),
               baseUrl = 'https://yourWebAPI',
               atlasCovariateIds = c(14709,14709, 14710),
               atlasCovariateNames = c('smoking anytime', 'smoking recent', 'traumatic brain injury'),
               startDays = c(-999,-30,-999),
               endDays = c(0,0,0),
               points = c(1,2,1))
```

The code above extracts the target and outcome cohorts and two ATLAS cohort (14709, 14710) to create three covariates:

- covariate 1: The ATLAS cohort with the id of 14709 named ‘smoking anytime’ looks for patients who have a smoking anytime cohort\_start\_date between (index date-999 days) and (index date). E.g., If a patient is in the smoking anytime 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 smoking anytime cohort between 999 days before index and the day of index then they will have a value of 0 for the custom covariate.
- covariate 2: The ATLAS cohort with the id of 14709 named ‘smoking recent’ looks for patients who have a smoking recent cohort\_start\_date between (index date-30 days) and (index date). E.g., If a patient is in the smoking recent 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 smoking recent cohort between 30 days before index and the day of index then they will have a value of 0 for the custom covariate.
- covariate 3: The ATLAS cohort with the id of 14710 named ‘traumatic brain injury’ looks for patients

who have a traumatic brain injury cohort\_start\_date between (index date-999 days) and (index date). E.g., If a patient is in the traumatic brain injury cohort 200 days before the index date then they will have a value of 1 for the custom covariate. If they are not in the traumatic brain injury cohort before index then they will have a value of 0 for the custom covariate.

It also creates three csv files in the inst/settings directory named:

- CohortsToCreate.csv - specifying the target and outcome cohorts
- CustomCovariates.csv - specifying the custom covariates
- SimpleModel.csv - settings specifying the simple prediction model

## 1.2 Step 2: Build the study package

After adding the settings 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 to validate an existing model

```
library(SkeletonExistingPredictionModelStudy)
options(fftempdir = "location with space to save big data")

# The folder where the study intermediate and result files will be written:
outputFolder <- "./SkeletonExistingPredictionModelStudyResults"

# Details for connecting to the server:
dbms <- "you dbms"
user <- 'your username'
pw <- 'your password'
server <- 'your server'
port <- 'your port'

connectionDetails <- DatabaseConnector::createConnectionDetails(dbms = dbms,
                                                                server = server,
                                                                user = user,
                                                                password = pw,
                                                                port = port)

# Add the database containing the OMOP CDM data
cdmDatabaseSchema <- 'cdm database schema'
# Add a database with read/write access as this is where the cohorts will be generated
cohortDatabaseSchema <- 'work database schema'

oracleTempSchema <- NULL

# table name where the cohorts will be generated
cohortTable <- 'SkeletonExistingPredictionModelStudyCohort'

# TAR settings
sampleSize <- NULL
riskWindowStart <- 1
startAnchor <- 'cohort start'
riskWindowEnd <- 365
```

