The CohortMethod R package & Review of code generated by ATLAS
CohortMethod package

GitHub repository:
- Installation instructions
- Package manual
- Vignettes

https://github.com/OHDSI/CohortMethod
CohortMethod package

What is it?
• Library of R commands
• Few lines of code $\Rightarrow$ full cohort study
• Flexible (if you know R)
• ‘Talks’ directly to database in CDM

What is it not?
• No point-and-click interface
• Won’t construct T-C-O cohorts
CohortMethod package

Two layers:
• ‘Single study’
• ‘Multiple analyses’

Fully described in the vignettes included in the package
Single studies approach

\[
\text{cmData} \leftarrow \text{getDbCohortMethodData}(...) \\
\text{studyPop} \leftarrow \text{createStudyPopulation(cmData, ...)} \\
\text{ps} \leftarrow \text{createPs(studyPop, cmData, ...)} \\
\text{matchedPop} \leftarrow \text{matchByPs(ps, ...)} \\
\text{model} \leftarrow \text{fitOutcomeModel(matchedPop, ...)}
\]
Running multiple analyses

Drug – Comparator - Outcomes

Analysis settings

CohortMethod

Estimates, Diagnostics
Multiple analyses approach

```r
cmdArgs <- createGetDbCohortMethodDataArgs(...)  
spArgs <- createCreateStudyPopulationArgs(...)  
psArgs <- createCreatePsArgs(...)  
matchArgs <- createMatchByPsArgs(...)  
modelArgs <- createFitOutcomeModelArgs(...)  

analysis1 <- createCmAnalysis(analysisId = 1,  
description = "1-on-1 matching",  
cmdArgs,  
spArgs,  
psArgs,  
matchArgs,  
modelArgs,  
...)```
ATLAS editor creates analysis object
Reproducing Graham et al. 2014 using the single study approach
Step 1: Getting the necessary data from the database

- Target and comparator cohorts of interest
  - Index date
  - End of cohort eligibility (e.g. end of exposure)
  - End of observation
- Outcome(s) of interest
  - Dates of occurrence
- Covariates

Take as-is from ATLAS
Covariates (Graham et al.)

Graham et al. used several covariates, including:

- Demographics
- Medical history
  - Various conditions
  - CHADS
  - HAS-BLED
- Medication use

These were hand-picked for this study
Connecting to the database

```r
connectionDetails <- createConnectionDetails(
  dbms = "pdw",
  server = "JRDUSAPSCONT01",
  port = 17001)

cdmDatabaseSchema <- "CDM_Truven_MDCR_V415.dbo"
cdmVersion <- "5"
```
Loading the list of selected covariate concept IDs

includedConcepts <- read.csv("inst/csv/GrahamIncludedConcepts.csv")$Concept.ID
```
covariateSettings <- createCovariateSettings(
  useCovariateDemographics = TRUE,
  useCovariateDemographicsGender = TRUE,
  useCovariateDemographicsRace = TRUE,
  useCovariateDemographicsEthnicity = TRUE,
  useCovariateDemographicsAge = TRUE,
  useCovariateDemographicsYear = FALSE,
  useCovariateDemographicsMonth = FALSE,
  ...
  useCovariateRiskScores = TRUE,
  useCovariateRiskScoresCharlson = FALSE,
  useCovariateRiskScoresDCSI = FALSE,
  useCovariateRiskScoresCHADS2 = TRUE,
  useCovariateRiskScoresCHADS2VASC = FALSE,
  excludedCovariateConceptIds = c(),
  includedCovariateConceptIds = includedConcepts,
  deleteCovariatesSmallCount = 100)
```

Including only the concept IDs we just loaded
cohortMethodData <- getDbCohortMethodData(
    connectionDetails = connectionDetails,
    cdmDatabaseSchema = cdmDatabaseSchema,
    targetId = 2649,
    comparatorId = 2650,
    outcomeIds = 2651,
    exposureTable = "cohort",
    outcomeTable = "cohort",
    cdmVersion = cdmVersion,
    excludeDrugsFromCovariates = FALSE,
    covariateSettings = covariateSettings)

saveCohortMethodData(cohortMethodData, "s:/GrahamStudy/cohortMethodData")
Loading all data

cohortMethodData <- getDbCohortMethodData(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDatabaseSchema,
  targetId = 2649,
  comparatorId = 2650,
  outcomeIds = 2651,
  exposureTable = "cohort",
  outcomeTable = "cohort",
  cdmVersion = cdmVersion,
  excludeDrugsFromCovariates = FALSE,
  covariateSettings = covariateSettings)

saveCohortMethodData(cohortMethodData, "s:/GrahamStudy/cohortMethodData")

Using the cohorts generated by ATLAS
Loading all data

cohortMethodData <- getDbCohortMethodData(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDatabaseSchema,
  targetId = 2649,
  comparatorId = 2650,
  outcomeIds = 2651,
  exposureTable = "cohort",
  outcomeTable = "cohort",
  cdmVersion = cdmVersion,
  excludeDrugsFromCovariates = FALSE,
  covariateSettings = covariateSettings)

saveCohortMethodData(cohortMethodData, 
  "s:/GrahamStudy/cohortMethodData")

The covariate settings we specified earlier
Inspecting the data

summary(cohortMethodData)

CohortMethodData object summary

Treatment concept ID: 2649
Comparator concept ID: 2650
Outcome concept ID(s): 2651

Treated persons: 19046
Comparator persons: 50918

Outcome counts:
  Event count Person count
2651  8063    8063

Covariates:
Number of covariates: 827
Number of non-zero covariate values: 2428813
Step 2: Defining the study population

```r
studyPop <- createStudyPopulation(cohortMethodData = cohortMethodData, 
  outcomeId = 2651, 
  removeSubjectsWithPriorOutcome = TRUE, 
  minDaysAtRisk = 1, 
  riskWindowStart = 1, 
  addExposureDaysToStart = FALSE, 
  riskWindowEnd = 0, 
  addExposureDaysToEnd = TRUE)
```
Step 2: Defining the study population

```r
studyPop <- createStudyPopulation(cohortMethodData = cohortMethodData,
                                   outcomeId = 2651,
                                   removeSubjectsWithPriorOutcome = TRUE,
                                   minDaysAtRisk = 1,
                                   riskWindowStart = 1,
                                   addExposureDaysToStart = FALSE,
                                   addExposureDaysToEnd = TRUE)
```

Remove subjects with outcomes prior to index date
Step 2: Defining the study population

```r
studyPop <- createStudyPopulation(cohortMethodData = cohortMethodData,
  outcomeId = 2651,
  removeSubjectsWithPriorOutcome = TRUE,
  minDaysAtRisk = 1,
  riskWindowStart = 1,
  addExposureDaysToStart = FALSE,
  riskWindowEnd = 0,
  addExposureDaysToEnd = TRUE)
```

Define risk window: start on day after index date, end at end of cohort eligibility (end of exposure)

Graham: “Follow-up began on the day after the first qualifying anticoagulant prescription fill”
Step 3: Creating a propensity model

```r
ps <- createPs(cohortMethodData = cohortMethodData, population = studyPop)
saveRDS(ps, "s:/GrahamStudy/ps.rds")
```
Plot propensity score distribution

plotPs(ps, treatmentLabel = "Dabigatran", comparatorLabel = "Warfarin")
Step 4: Matching

matchedPop <- matchOnPs(ps,
  caliper = 0.25,
  caliperScale = "standardized",
  maxRatio = 1)

Select up to 1 comparator per target subject (1-on-1 matching)

Graham: “Dabigatran users were propensity score matched to warfarin users in a 1:1 ratio with the use of a greedy matching algorithm.”

No caliper was mentioned, but it is probably a good idea to use one
Attrition

drawAttritionDiagram(matchedPop, treatmentLabel = "Dabigatran", comparatorLabel = "Warfarin")
Attrition

drawAttritionDiagram(matchedPop, treatmentLabel = "Dabigatran", comparatorLabel = "Warfarin")

Original cohorts:
Treated: n = 19046
Comparator: n = 50918

No prior outcome

Have at least 1 days at risk

Matched on propensity score

Study population:
Treated: n = 17460
Comparator: n = 17460

Cohorts as selected by ATLAS
Study population that will go into Cox regression
Covariate balance

balance <- computeCovariateBalance(matchedPop, cohortMethodData)
plotCovariateBalanceScatterPlot(balance)

Most covariates are binary:

\[
\text{abs}(P_{\text{target group}} - P_{\text{comparator group}}) \over \text{standard deviation}
\]

Graham: “A standardized mean difference of \( \leq 0.1 \) indicates a negligible difference.”
Step 5: Fitting the outcome model

```r
outcomeModel <- fitOutcomeModel(population = matchedPop,
                               modelType = "cox",
                               stratified = FALSE,
                               useCovariates = FALSE)
```

Graham: “Cox proportional hazards regression was used to compare time to event in dabigatran compared with warfarin (reference) cohorts.

Ambiguous: Did they condition the model on the matched set (recommended)?
Inspect the outcome model

```r
summary(outcomeModel)
```

Model type: cox  
Stratified: FALSE  
Use covariates: FALSE  
Status: OK

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>lower .95</th>
<th>upper .95</th>
<th>logRr</th>
<th>seLogRr</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>0.89626</td>
<td>0.71863</td>
<td>1.11829</td>
<td>-0.10952</td>
<td>0.1128</td>
</tr>
</tbody>
</table>

Population counts

<table>
<thead>
<tr>
<th></th>
<th>treatedPersons</th>
<th>comparatorPersons</th>
<th>treatedExposures</th>
<th>comparatorExposures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>17460</td>
<td>17460</td>
<td>17460</td>
<td>17460</td>
</tr>
</tbody>
</table>

Outcome counts

<table>
<thead>
<tr>
<th></th>
<th>treatedPersons</th>
<th>comparatorPersons</th>
<th>treatedExposures</th>
<th>comparatorExposures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>164</td>
<td>155</td>
<td>164</td>
<td>155</td>
</tr>
</tbody>
</table>

Time at risk

<table>
<thead>
<tr>
<th></th>
<th>treatedDays</th>
<th>comparatorDays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>4912947</td>
<td>3954046</td>
</tr>
</tbody>
</table>
Inspect the outcome model

```r
summary(outcomeModel)
```

Model type: cox
Stratified: FALSE
Use covariates: FALSE
Status: OK

<table>
<thead>
<tr>
<th>Estimate</th>
<th>lower .95</th>
<th>upper .95</th>
<th>logRr</th>
<th>seLogRr</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>0.89626</td>
<td>0.71863</td>
<td>1.11829</td>
<td>-0.10952</td>
</tr>
</tbody>
</table>

Population counts

<table>
<thead>
<tr>
<th></th>
<th>treatedPersons</th>
<th>comparatorPersons</th>
<th>treatedExposures</th>
<th>comparatorExposures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>164</td>
<td>155</td>
<td>164</td>
<td>155</td>
</tr>
</tbody>
</table>

Time at risk

<table>
<thead>
<tr>
<th></th>
<th>treatedDays</th>
<th>comparatorDays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>4912947</td>
<td>3954046</td>
</tr>
</tbody>
</table>

Point estimate and 95% confidence interval
Inspect the outcome model

```
summary(outcomeModel)
```

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>lower .95</th>
<th>upper .95</th>
<th>logRr</th>
<th>seLogRr</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>0.89626</td>
<td>0.71863</td>
<td>1.11829</td>
<td>-0.10952</td>
<td>0.1128</td>
</tr>
</tbody>
</table>

**Model type:** cox  
**Stratified:** FALSE  
**Use covariates:** FALSE  
**Status:** OK

**Population counts**

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatedPersons</td>
<td>17460</td>
</tr>
<tr>
<td>comparatorPersons</td>
<td>17460</td>
</tr>
</tbody>
</table>

**Outcome counts**

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatedPersons</td>
<td>164</td>
</tr>
<tr>
<td>comparatorPersons</td>
<td>155</td>
</tr>
</tbody>
</table>

**Time at risk**

<table>
<thead>
<tr>
<th></th>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatedDays</td>
<td>4912947</td>
</tr>
<tr>
<td>comparatorDays</td>
<td>3954046</td>
</tr>
</tbody>
</table>

**Target group (dabigatran) has more outcomes, but also more time at risk**
Inspect the outcome model

```
summary(outcomeModel)
```

Model type: cox
Stratified: FALSE
Use covariates: FALSE
Status: OK

- **Estimate**
  - treatment: 0.89626
  - lower .95: 0.71863
  - upper .95: 1.11829

- **Population counts**
  - treatedPersons: 17460
  - comparatorPersons: 17460

- **Outcome counts**
  - treatedPersons: 164
  - comparatorPersons: 155

- **Time at risk**
  - treatedDays: 4912947
  - comparatorDays: 3954046

**Graham:**

- IR\textsubscript{dabigatran} = 11.3
- IR\textsubscript{warfarin} = 13.9
- HR\textsubscript{adjusted} = 0.80 (0.67–0.96)

- IR\textsubscript{dabigatran} = 12.2
- IR\textsubscript{warfarin} = 14.3
- HR\textsubscript{adjusted} = 0.90 (0.72 – 1.12)
Kaplan Meier plot

```r
plotKaplanMeier(matchedPop,
    treatmentLabel = "Dabigatran",
    comparatorLabel = "Warfarin")
```

Number at risk
- **Dabigatran**: 17,460, 6,410, 3,308
- **Warfarin**: 17,460, 5,248, 2,172
Kaplan Meier plot

plotKaplanMeier(matchedPop, 
  treatmentLabel = "Dabigatran",
  comparatorLabel = "Warfarin")

Graham:
Execution steps

1. Getting the necessary data from the database
2. Defining the study population
3. Creating a propensity model
4. Matching
5. Fitting the outcome model

+ generating various diagnostics
Review of the R study package generated by ATLAS
Anatomy of the study package

- **Execution code**
  - Main execution code
  - Shiny app to view results

- **Cohorts**
  - JSON definitions
  - SQL definitions

- **Negative controls**
  - Negative control concepts
  - SQL template

- **Positive controls**
  - Positive control settings

- **Analysis settings**
  - Target-comparator-outcomes
  - Analysis definitions
How to install?

• Build and run in R-Studio
  – Open package
    • Own machine: unzip, double-click .Rproj file
    • R-Studio Server: upload zip, click .Rproj file
  – Install dependencies (see readme)
  – ‘Build’ → ‘Install and Restart’

• Build from GitHub
  – Unzip and put in GitHub repo
  – Install dependencies (see readme)
  – Install with devtools::install_github
Installing dependencies

How to run

1. In R, use the following code to install the dependencies:

```r
install.packages("devtools")
library(devtools)
install_github("ohdsi/SqlRender", ref = "v1.5.2")
install_github("ohdsi/DatabaseConnector", ref = "v2.2.0")
install_github("ohdsi/OhdsiSharing", ref = "v0.1.3")
install_github("ohdsi/FeatureExtraction", ref = "v2.1.5")
install_github("ohdsi/CohortMethod", ref = "v3.0.1")
install_github("ohdsi/EmpiricalCalibration", ref = "v1.3.6")
install_github("ohdsi/MethodEvaluation", ref = "v0.3.1")
```

Source: readme.md
Running package

3. Once installed, you can execute the study by modifying and using the following code:

```r
library(Graham)

# Optional: specify where the temporary files (used by the ff package) will be created:
options(fftempdir = "c:/FFtemp")

# Maximum number of cores to be used:
maxCores <- parallel::detectCores()

# Minimum cell count when exporting data:
minCellCount <- 5

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

# Details for connecting to the server:
# See ?DatabaseConnector::createConnectionDetails for help
connectionDetails <- DatabaseConnector::createConnectionDetails(dbms = "postgresql",
server = "some.server.com/ohdsi",
user = "joe",
password = "secret")

# The name of the database schema where the CDM data can be found:
cdmDatabaseSchema <- "cdm_synpuf"
```
Export data model

Can be shared:
- Aggregated, so no patient-level data
- Minimum cell count enforced
- Saved as CSV, so easily reviewable
- Zipped for convenience

See vignette at https://github.com/ohdsi/SkeletonComparativeEffectStudy for details
Launching the Shiny app

5. To view the results, use the Shiny app:

```r
prepareForEvidenceExplorer("Result<databaseId>.zip", "/shinyData")
launchEvidenceExplorer("/shinyData", blind = TRUE)
```
### Evidence Explorer

#### Target
- mscuemni graham dabigatran

#### Comparator
- mscuemni graham warfarin

#### Outcome
- mscuemni graham gbled

#### Data source
- MDCR

#### Analysis
- Primary

#### Table: Analysis Data

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Data source</th>
<th>HR</th>
<th>LB</th>
<th>UB</th>
<th>P</th>
<th>Cal.HR</th>
<th>Cal.LB</th>
<th>Cal.UB</th>
<th>Cal.P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>MDCR</td>
<td>1.12</td>
<td>0.85</td>
<td>1.48</td>
<td>0.40</td>
<td>1.18</td>
<td>0.86</td>
<td>1.66</td>
<td>0.32</td>
</tr>
</tbody>
</table>

#### Chart: Preference Score Distribution

Figure 2. Preference score distribution. The preference score is a transformation of the propensity score that adjusts for differences in the sizes of the two treatment groups. A higher overlap indicates subjects in the two groups were more similar in terms of their predicted probability of receiving one treatment over the other.

[Download plot]
Concluding remarks

• CohortMethod package + R offer large flexibility
• 80% of studies are ‘cookie-cutter’ design, supported by ATLAS
• For remaining 20%, will need to modify code generated by ATLAS