Reproducing Graham et al. 2014 using the CohortMethod package

Martijn Schuemie
CohortMethod package

GitHub repository:
- Installation instructions
- Package manual
- Vignettes

https://github.com/OHDSI/CohortMethod
R script implementing the study

```
# Define CDM version, results database schema, and server
cdmVersion <- "V415"
resultsDatabaseSchema <- "CDM_Truven_MDCR"
server <- "JRDUSAPSCTL01"

# Set options for Cohort Method Study: OHDSI estimation tutorial
install.packages("SqlRender")

# Cohort Method Data
cohortMethodData <- read.csv("/s:/temp/GrahamStudy/cohort/IncludedConcepts.csv")

# Define study parameters
studyStartDate <- "2015-01-01"
outcomeTable <- "Target"
comparatorLabel <- "Warfarin"
outcomeIds <- c("ischemicStoke";

# Create study population
studyPop <- createStudyPopulation(cohortMethodData,
  ps = getPsModel(studyPop),
  studyStartDate = studyStartDate,
  outcomeTable = outcomeTable,
  comparatorLabel = comparatorLabel,
  outcomeIds = outcomeIds)

# Display study population
summary(studyPop)

# Display propensity scores
summary(studyPop["ps.rds"])

# Add washout period
washoutPeriod <- 30

# Create matched population
matchedPop <- matching(studyPop,
  ps = studyPop["ps.rds"],
  washoutPeriod = washoutPeriod,
  firstExposureOnly = TRUE)

# Display matched population
summary(matchedPop)

# Get all included concept IDs
includedConcepts <- getIncludedConcepts(cohortMethodData)
excludedConcepts <- getExcludedConcepts(cohortMethodData)

# Display included and excluded concept IDs
summary(includedConcepts)
summary(excludedConcepts)

# Display covariate balance
plotCovariateBalanceOfTopVariables(studyPop)

# Display propensity score distribution
plotPs(studyPop)
```

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 = "JRDUSAPSCCTL01",
                                           port = 17001)

cdmDatabaseSchema <- "CDM_Trufen_MDCR_V415.dbo"
cdmVersion <- "5"
```
Loading the list of selected covariate concept IDs

```r
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, "~/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

```
studyPop <- createStudyPopulation(cohortMethodData = cohortMethodData,
  outcomeId = 2651,
  removeSubjectsWithPriorOutcome = TRUE,
  minDaysAtRisk = 1,
  riskWindowStart = 1,
  riskWindowEnd = 0,
  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
drawAttritionDiagram(matchedPop, treatmentLabel = "Dabigatran", comparatorLabel = "Warfarin")

Attrition

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

No prior outcome
- Y
- N
  - No prior outcome
    - Y
    - N
      - Treated: n = 1198
        - Comparator: n = 4809
      - Treated: n = 82
        - Comparator: n = 447
    - Y
      - Treated: n = 306
        - Comparator: n = 28202
  - Matched on propensity score

Study population:
- Treated: n = 17460
- Comparator: n = 17460
drawAttritionDiagram(matchedPop, treatmentLabel = "Dabigatran", comparatorLabel = "Warfarin")

Attrition

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
Attrition

drawAttritionDiagram(matchedPop, treatmentLabel = "Dabigatran", comparatorLabel = "Warfarin")
balance <- computeCovariateBalance(matchedPop, cohortMethodData)
plotCovariateBalanceScatterPlot(balance)

Most covariates are binary:

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

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

```
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

```
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>164</td>
<td>155</td>
<td>164</td>
<td>164</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>
```

Point estimate and 95% confidence interval
Inspect the outcome model

```r
summary(outcomeModel)
```

<table>
<thead>
<tr>
<th>Model type: cox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stratified: FALSE</td>
</tr>
<tr>
<td>Use covariates: FALSE</td>
</tr>
<tr>
<td>Status: OK</td>
</tr>
</tbody>
</table>

<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>

<table>
<thead>
<tr>
<th>Population counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatedPersons</td>
</tr>
<tr>
<td>Count</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatedPersons</td>
</tr>
<tr>
<td>Count</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatedDays</td>
</tr>
<tr>
<td>Days</td>
</tr>
<tr>
<td>4912947</td>
</tr>
</tbody>
</table>
summary(outcomeModel)

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

Population counts
- treatedPersons: Count = 17460
- comparatorPersons: Count = 17460

Outcome counts
- treatedPersons: Count = 164
- comparatorPersons: Count = 155

Time at risk
- treatedDays: Days = 4912947
- comparatorDays: Days = 3954046

Graham:
- IR$_{dabigatran}$ = 11.3
- IR$_{warfarin}$ = 13.9
- HR$_{adjusted}$ = 0.80 (0.67–0.96)

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

plotKaplanMeier(matchedPop,
    treatmentLabel = "Dabigatran",
    comparatorLabel = "Warfarin")
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
Handpicking covariates is not recommended!

• Might miss an important confounder (proxy)
• Subjective (non-reproducible)

Better:

• Include all pre-defined covariates
• Let data (and regularization) decide which ones are important
• Bonus: more work for the computer, but less work for you!
Covariate settings

covariateSettings <- createCovariateSettings(
  useCovariateDemographics = TRUE,
  useCovariateConditionOccurrence = TRUE,
  useCovariateConditionOccurrence365d = TRUE,
  useCovariateConditionOccurrence30d = TRUE,
  useCovariateConditionOccurrenceInpt180d = TRUE,
  ...
  useCovariateRiskScores = TRUE,
  useCovariateRiskScoresCharlson = TRUE,
  useCovariateRiskScoresDCSI = TRUE,
  useCovariateRiskScoresCHADS2 = TRUE,
  useCovariateInteractionYear = FALSE,
  useCovariateInteractionMonth = FALSE,
  excludedCovariateConceptIds = excludeConceptIds,
  deleteCovariatesSmallCount = 100)
Covariate settings

constructedCovariates <- createCovariateSettings(
  useCovariateDemographics = TRUE,
  useCovariateConditionOccurrence = TRUE,
  useCovariateConditionOccurrence365d = TRUE,
  useCovariateConditionOccurrence30d = TRUE,
  useCovariateConditionOccurrenceInpt180d = TRUE,
  ...
  useCovariateRiskScores = TRUE,
  useCovariateRiskScoresCharlson = TRUE,
  useCovariateRiskScoresDCSI = TRUE,
  useCovariateRiskScoresCHADS2 = TRUE,
  useCovariateInteractionYear = FALSE,
  useCovariateInteractionMonth = FALSE,
  excludedCovariateConceptIds = excludeConceptIds,
  deleteCovariatesSmallCount = 100)

Constructing covariates for
- Demographics
- All drugs & classes
- All conditions & groups
- All procedures
- All measurements
- All observations
- Risk scores
Covariate settings

```r
covariateSettings <- createCovariateSettings(
  useCovariateDemographics = TRUE,
  useCovariateConditionOccurrence = TRUE,
  useCovariateConditionOccurrence365d = TRUE,
  useCovariateConditionOccurrence30d = TRUE,
  useCovariateConditionOccurrenceInpt180d = TRUE,
  ...
  useCovariateRiskScores = TRUE,
  useCovariateRiskScoresCharlson = TRUE,
  useCovariateRiskScoresDCSI = TRUE,
  useCovariateRiskScoresCHADS2 = TRUE,
  useCovariateInteractionYear = FALSE,
  useCovariateInteractionMonth = FALSE,
  excludedCovariateConceptIds = excludeConceptIds,
  deleteCovariatesSmallCount = 100)
```

Have to explicitly exclude dabigatran and warfarin
Effect on propensity score distribution

Using hand-picked covariates

Using all covariates

Whole subgroup of people likely to get warfarin was not identified by Graham
Effect on matching

Using hand-picked covariates

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

Using all covariates

Original cohorts:
- Dabigatran: n = 19046
- Warfarin: n = 50918

No prior outcome

Have at least 1 days at risk

Matched on propensity score

Study population:
- Dabigatran: n = 2415
- Warfarin: n = 30311

Fewer people left after matching
Effect on balance

Using hand-picked covariates

Using all covariates

Balanced on every covariate, including those hand-picked by Graham
Effect on hazard ratio

Using hand-picked covariates

Using all covariates

In this case, effect on estimate is small

HR = 0.90 (0.72 – 1.12)

HR = 0.95 (0.75 - 1.21)
In conclusion

• OHDSI tools can replicate the Graham study

• Diagnostics are an important part of both design and execution
  – Show impact of adjustments: lots of imbalance before matching
  – Lot of attrition: how generalizable are our results?

• Hand picking covariates is not recommended