



HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Open-Source Software for Observational Research

Martijn Schuemie
Epidemiology department, Johnson & Johnson
Department of Biostatistics, UCLA



What is HADES?

Health Analytics Data to Evidence Suite (HADES)

- Collection of 35 R packages used in almost every OHDSI study
 - OHDSI is an international collaborative for observational research
- Run directly against data in the OMOP Common Data Model (CDM)
 - Health care insurance claims
 - Electronic health records
- Perform observational analyses
 - Characterization
 - Causal effect estimation
 - Patient-level prediction

<https://ohdsi.github.io/Hades/>





HADES design principles

- Promote **open science** through open source
- Execute directly against the **OMOP CDM**
- Implement **best practices** as informed by methods research
- Provide **high quality software** (documented, maintained, tested, validated)
- Facilitate **large-scale analytics**, answering many questions at once
- Support **big data**, covering hundreds of millions of lives
- Enable **federated analyses**
- Run **across a wide variety of technical infrastructures**



HADES – supporting packages

- **DatabaseConnector + SqlRender:** write code once, run on all supported platforms (SQL Server, Oracle, Postgres, RedShift, BigQuery, DataBricks, Snowflake)
- **Andromeda:** Work with data objects too big to fit in memory
- **ParallelLogger:** extensive logging to facilitate remote debugging
- **Cyclops:** fit very large regression models (logistic, Poisson, Cox)
- **DataQualityDashboard:** evaluate data quality



HADES – cohort packages

We define a cohort as a set of persons who satisfy one or more inclusion criteria for a duration of time.

- Exposure cohorts (e.g. people exposed to warfarin)
 - Outcome cohorts (e.g. people experiencing GI bleeding)
 - Cohorts of special interest (e.g. pregnant women)
-
- **Capr**: Define cohorts using complex logic
 - **PhenotypeLibrary**: for storing OHDSI-approved cohort definitions
 - **CirceR**: for turning cohort definitions into SQL or human-readable text
 - **CohortGenerator**: for instantiating cohorts in a database
 - **CohortDiagnostics** and **PheValuator**: for evaluating cohorts



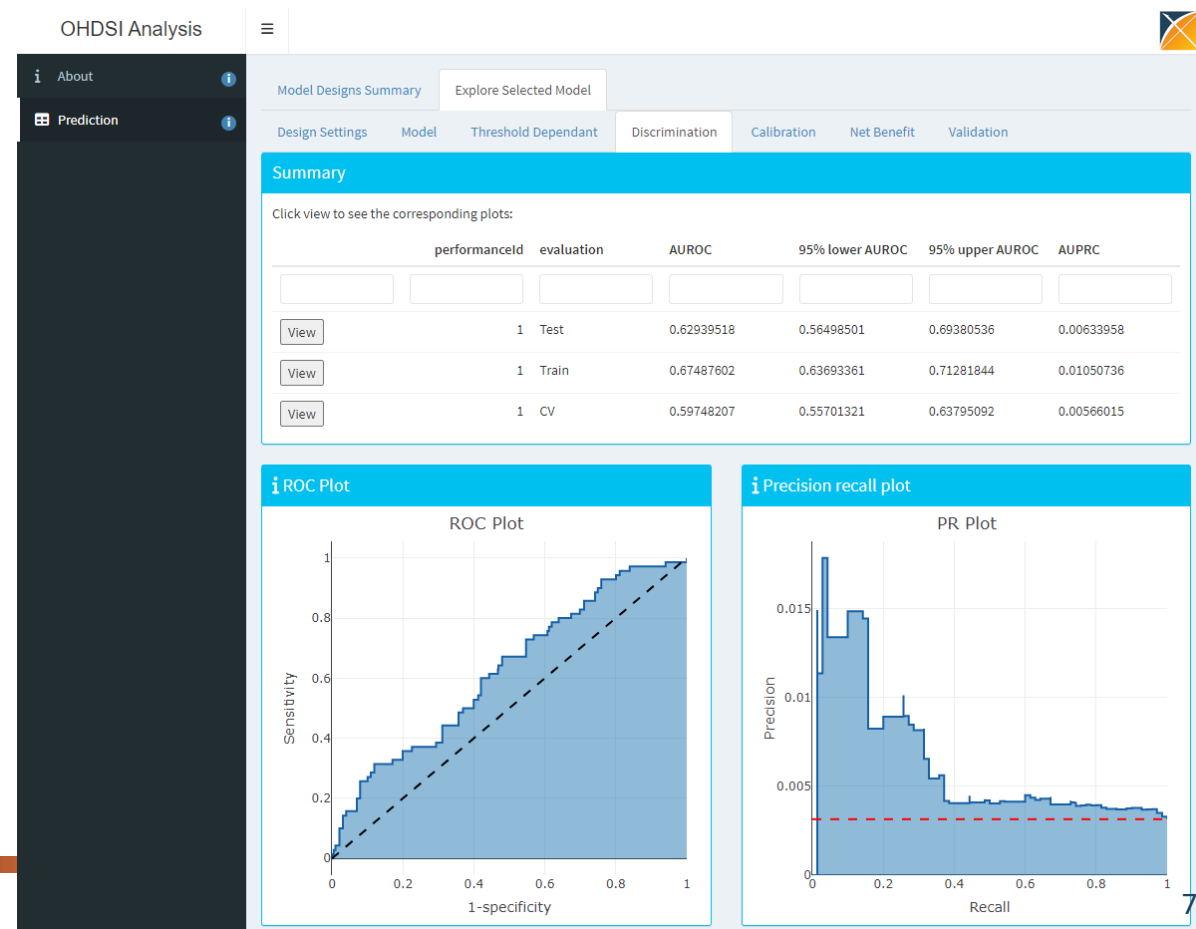
HADES – analytics packages

- **PatientLevelPrediction:** develop and evaluate prediction models
- **CohortMethod** and **SelfControlledCaseSeries:** estimate causal effects
- **EmpiricalCalibration:** Calibrate causal effect estimates based on negative controls
- **EvidenceSynthesis:** Combine causal effect estimates across databases without sharing patient-level data.



HADES analytics output

- R objects possibly containing patient-level data
- CSV files / database tables for sharing
- Shiny apps





Publications using HADES

- 38 peer-reviewed clinical research papers
- 29 methods research papers

THE LANCET

Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis



Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, Ruijun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, Patrick

Williams et al. *BMC Medical Research Methodology* (2022) 22:35
<https://doi.org/10.1186/s12874-022-01505-z>

Summary
Background Uncertainty in estimating any primary enzyme inhibitors, a calcium channel blocker.

Methods We develop and safety evaluation while minimising incohort design to estimate failure, and stroke) global network of sites residual confounding outcomes, and full c

Findings Using 4.9 comparing all classes; however, the enzyme inhibitors: 0.74-0.95), and str thiazide-like diuretic blockers were signifi

RESEARCH

BMC Medical Research Methodology

Open Access

Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network

Ross D. Williams^{1†}, Aniek F. Markus^{1†}, Cynthia Yang¹, Talita Duarte-Salles², Scott L. DuVall³, Thomas Falconer⁴, Jitendra Jonnagaddala⁵, Chungsoo Kim⁶, Yeunsook Rho⁷, Andrew E. Williams⁸, Amanda Alberga Machado⁹, Min Ho An¹⁰, María Aragón², Carlos Areia¹¹, Edward Burn^{2,12}, Young Hwa Cho¹³, Iannis Drakos¹⁴, Maria Tereza Fernandes Abrahão¹⁵, Sergio Fernández-Bertolín², George Hripcsak⁴, Benjamin Skov Kaas-Hansen^{16,17}, Prasanna L. Kandukuri¹⁸, Jan A. Kors¹, Kristin Kostka¹⁹, Siaw-Teng Liaw⁵,



Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study

Jennifer C E Lane*, James Weaver*, Kristin Kostka, Talita Duarte-Salles, Maria Tereza F Abrahao, Heba Alghoul, Osaid Alser, Thamir M Alshammari, Patricia Biedermann, Juan M Banda, Edward Burn, Paula Casaiust, Mitchell M Conover, Aedin C Cullhane.



Lancet Rheumatol 2020; 2: e698-711
Published Online August 21, 2020
[https://doi.org/10.1016/S2665-9913\(20\)30276-9](https://doi.org/10.1016/S2665-9913(20)30276-9)
See Comment page e652

*Contributed equally
Centre for Statistics in Medicine, Nuffield Department

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Renin-angiotensin system blockers and susceptibility to COVID-19: an international, open science, cohort analysis

Daniel R Morales, Mitchell M Conover, Seng Chan You, Nicole Pratt, Kristin Kostka, Talita Duarte-Salles, Sergio Fernández-Bertolín, Maria Aragón, Scott L DuVall, Kristine Lynch, Thomas Falconer, Kees van Bochove, Cynthia Sung, Michael E Matheny, Christophe G Lambert, Fredrik Nyberg, Thamir M Alshammari, Andrew E Williams, Rae Woong Park, James Weaver, Anthony G Sena, Martijn J Schuemie, Peter R Rijnbeek, Ross D Williams, Jennifer C E Lane, Albert Prats-Urbe, Lin Zhang, Carlos Areia, Harlan M Krumholz, Daniel Prieto-Alhambra, Patrick B Ryan, George Hripcsak, Marc A Suchard

Summary

Background Angiotensin-converting enzyme inhibitors (ACEIs) and angiotensin receptor blockers (ARBs) have been postulated to affect susceptibility to COVID-19. Observational studies so far have lacked rigorous ascertainment adjustment and international generalisability. We aimed to determine whether use of ACEIs or ARBs is associated with an increased susceptibility to COVID-19 in patients with hypertension.



Lancet Digit Health 2021; 3: e98-114
Published Online December 17, 2020
<https://doi.org/10.1016/>



And now...

code for performing a comparative safety study
using the CohortMethod package



Download all required data from server

```
covSettings <- createDefaultCovariateSettings(  
  excludedCovariateConceptIds = c(diclofenacConceptId,  
                                   celecoxibConceptId),  
  addDescendantsToExclude = TRUE  
)  
cohortMethodData <- getDbCohortMethodData(  
  connectionDetails = connectionDetails,  
  cdmDatabaseSchema = cdmDatabaseSchema,  
  targetId = 1,  
  comparatorId = 2,  
  outcomeIds = 77,  
  exposureDatabaseSchema = cohortDatabaseSchema,  
  exposureTable = cohortTable,  
  outcomeDatabaseSchema = cohortDatabaseSchema,  
  outcomeTable = cohortTable,  
  covariateSettings = covSettings  
)
```

Create covariates
for all conditions,
drugs, procedures,
etc.

Assuming exposure
and outcome
cohorts have already
been created



Estimate causal effect

```
studyPop <- createStudyPopulation(  
  cohortMethodData = cohortMethodData,  
  outcomeId = 77,  
  removeDuplicateSubjects = "keep first",  
  removeSubjectsWithPriorOutcome = TRUE,  
  minDaysAtRisk = 1,  
  riskWindowStart = 0,  
  startAnchor = "cohort start",  
  riskWindowEnd = 30,  
  endAnchor = "cohort end "  
)  
ps <- createPs(cohortMethodData = cohortMethodData,  
  population = studyPop)  
matchedPop <- matchOnPs(ps, maxRatio = 1)  
outcomeModel <- fitOutcomeModel(matchedPop, modelType = "cox")
```

Interested in the first 30
days after exposure

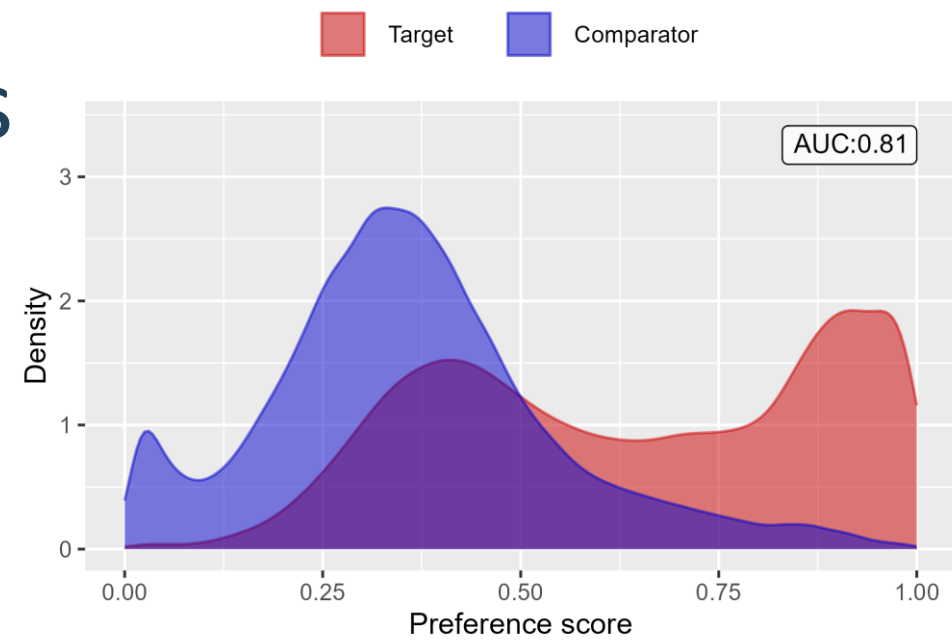
Fit a propensity
model using LASSO

1-1 PS match
Cox regression



Diagnostics

```
plotPs(ps, showAucLabel = TRUE)
```



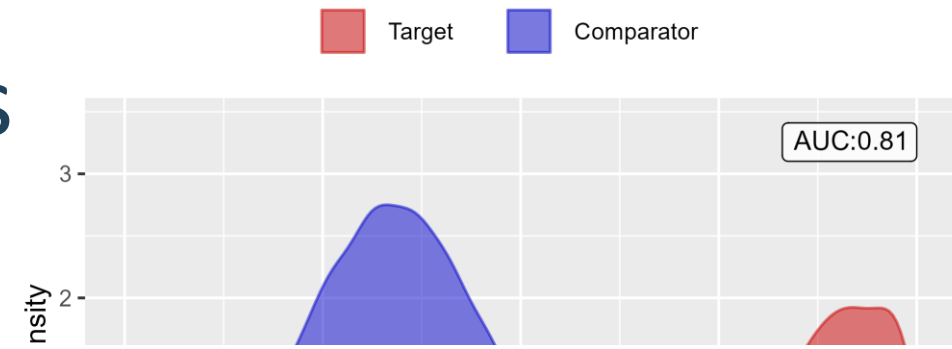


Diagnostics

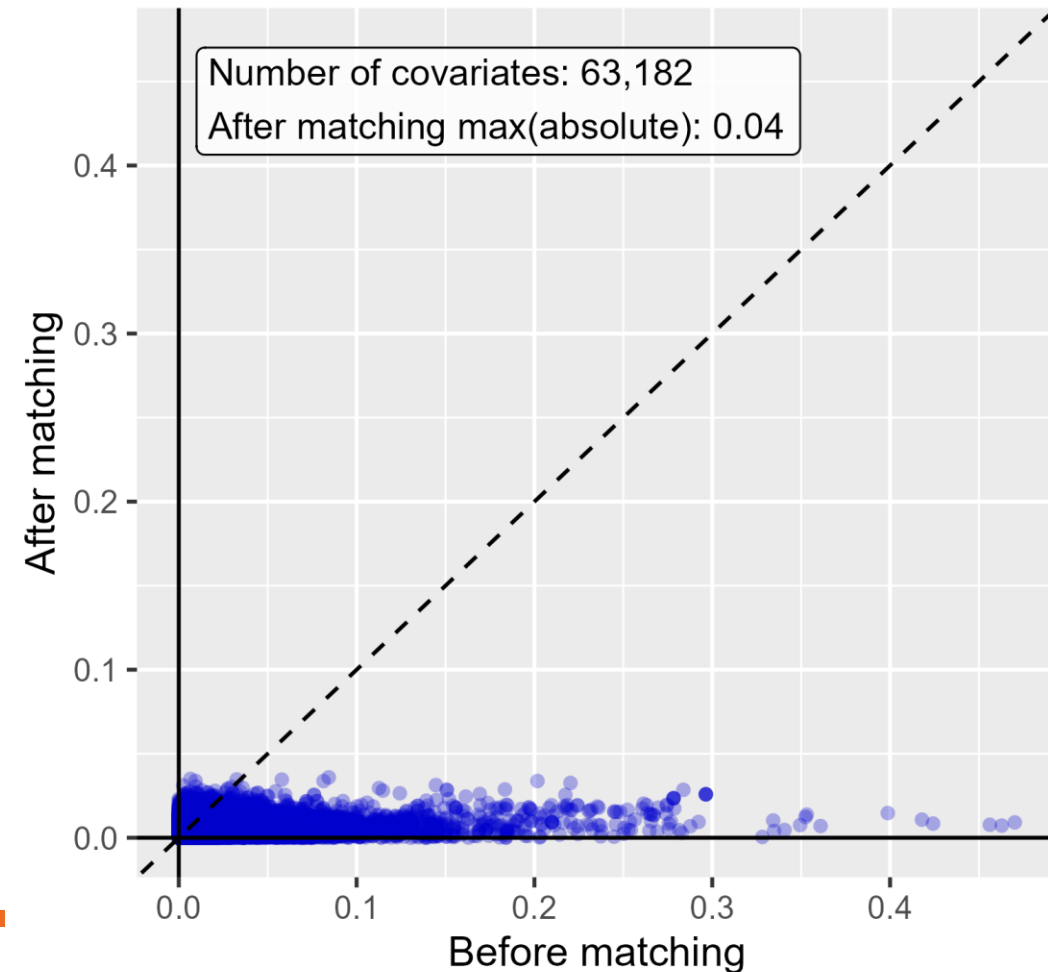
```
plotPs(ps, showAucLabel = TRUE)
```

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

```
plotCovariateBalanceScatterPlot(  
  balance,  
  showCovariateCountLabel = TRUE,  
  showMaxLabel = TRUE  
)
```



Standardized difference of mean





Diagnostics

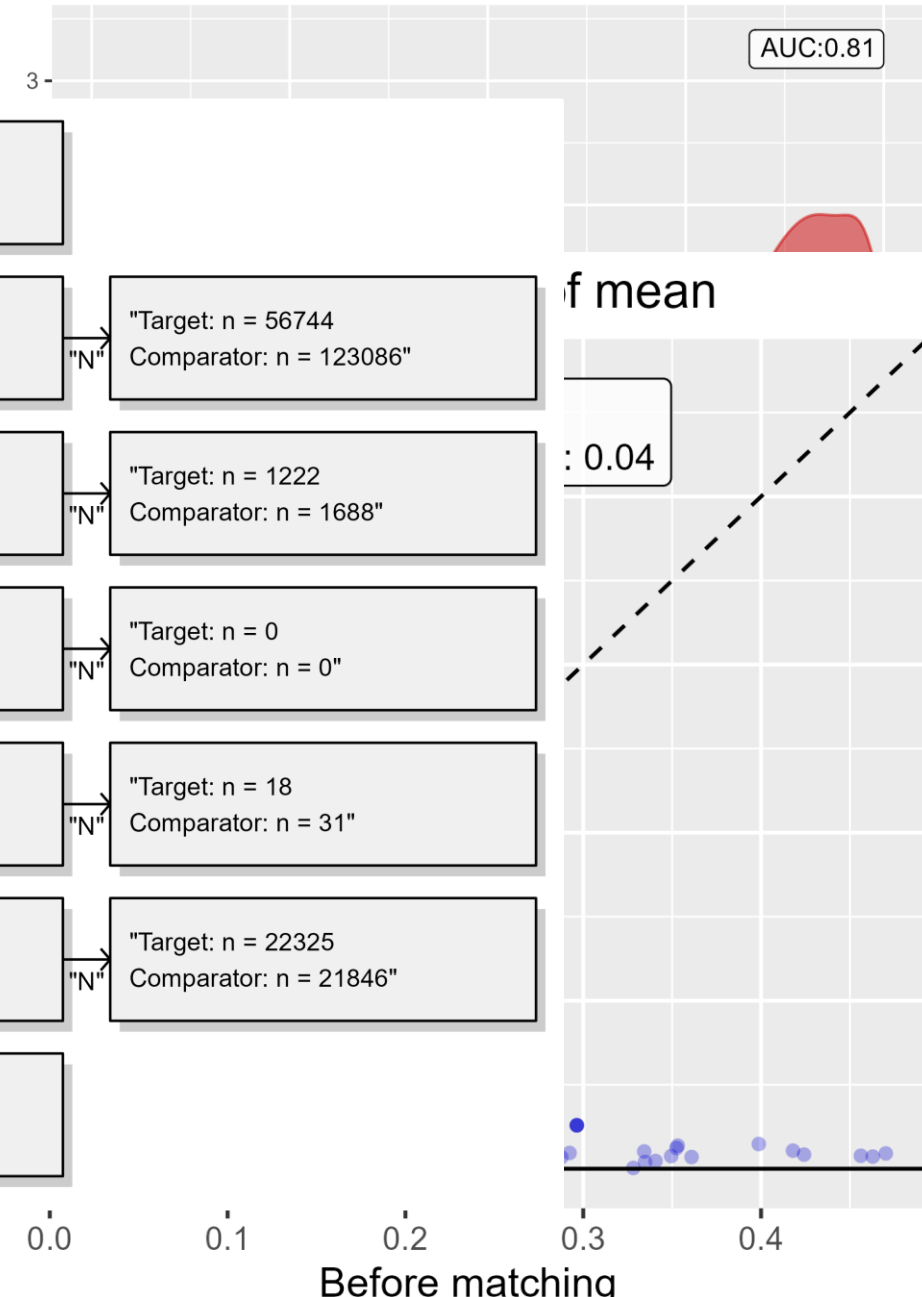
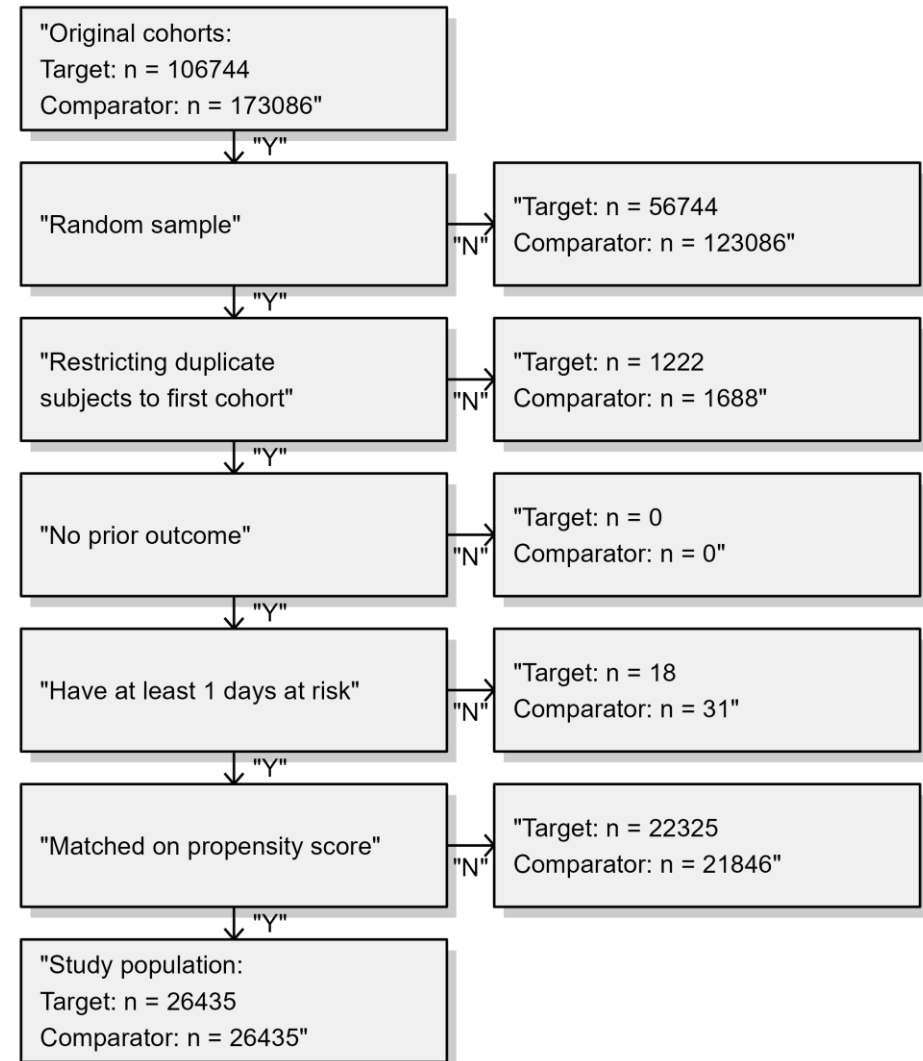
Target Comparator

```
plotPs(ps, showAucLabel = TRUE)
```

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

```
plotCovariateBalanceScatterPlot(  
  balance,  
  showCovariateCountLabel = TRUE,  
  showMaxLabel = TRUE  
)
```

```
drawAttritionDiagram(outcomeModel)
```





Diagnostics

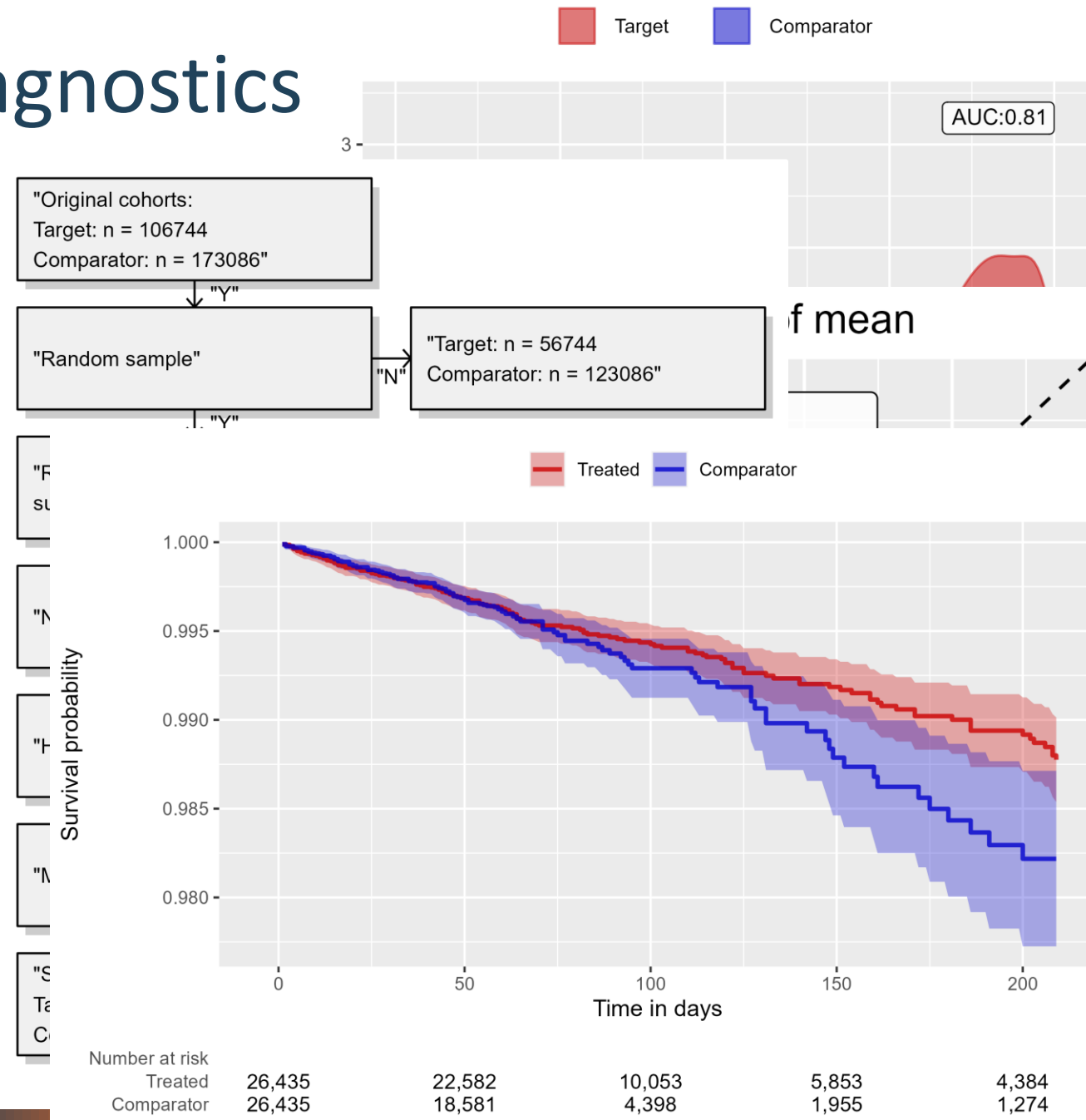
```
plotPs(ps, showAucLabel = TRUE)
```

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

```
plotCovariateBalanceScatterPlot(  
  balance,  
  showCovariateCountLabel = TRUE,  
  showMaxLabel = TRUE  
)
```

```
drawAttritionDiagram(outcomeModel)
```

```
plotKaplanMeier(outcomeModel)
```





Conclusions

- HADES is a suite of R packages for analyzing observational healthcare data
- Thanks to the OMOP Common Data Model, HADES runs on a wide variety of data sources across the world
- Open source, to promote open science (all analytics code can be shared as part of publication)
- Supports federated networks, where data stay locally, and results are shared



Thank you!



```
library(DatabaseConnector)
Library(keyring)
connectionDetails <- createConnectionDetails(
  dbms = "redshift",
  connectionString = key_get("redShiftConnectionString"),
  user = key_get("redShiftUserName"),
  password = key_get("redShiftPassword")
)
cdmDatabaseSchema <- "cdm_truven_mdcr_v2322"
cohortDatabaseSchema <- "scratch_mschuemi"
cohortTable <- "cm_vignette"
```



```
library(Capr)
library(CirceR)
osteoArthritisOfKneeConceptId <- 4079750
celecoxibConceptId <- 1118084
diclofenacConceptId <- 1124300
osteoArthritisOfKnee <- cs(descendants(osteoArthritisOfKneeConceptId),
                          name = "Osteoarthritis of knee")
attrition = attrition("prior osteoarthritis of knee" = withAll(
  atLeast(1, condition(osteoArthritisOfKnee), duringInterval(eventStarts(-Inf, 0))))
)
celecoxibCs <- cs(descendants(celecoxibConceptId, name = "Celecoxib")
diclofenacCs <- cs(descendants(diclofenacConceptId), name = "Diclofenac")
celecoxibCohort <- cohort(entry = entry(drug(celecoxibCs, firstOccurrence()),
                                     observationWindow = continuousObservation(priorDays = 365)),
  attrition = attrition,
  exit = exit(endStrategy = drugExit(celecoxibCs, persistenceWindow = 30, surveillanceWindow = 0)))
diclofenacCohort <- cohort(entry = entry(drug(diclofenacCs, firstOccurrence()),
                                     observationWindow = continuousObservation(priorDays = 365)),
  attrition = attrition,
  exit = exit(endStrategy = drugExit(diclofenacCs, persistenceWindow = 30, surveillanceWindow = 0)))
```



```
exposureCohorts <- tibble(  
  cohortId = c(1, 2),  
  cohortName = c("Celecoxib", "Diclofenac"),  
  json = c(as.json(celecoxibCohort),  
           as.json(diclofenacCohort))  
)  
exposureCohorts$sql <- sapply(exposureCohorts$json,  
                              buildCohortQuery,  
                              options = createGenerateOptions())  
  
library(PhenotypeLibrary)  
outcomeCohorts <- getPlCohortDefinitionSet(77) # GI bleed
```



```
library(CohortGenerator)
allCohorts <- bind_rows(outcomeCohorts,
                        exposureCohorts)
cohortTableNames <- getCohortTableNames(cohortTable = cohortTable)
createCohortTables(
  connectionDetails = connectionDetails,
  cohortDatabaseSchema = cohortDatabaseSchema,
  cohortTableNames = cohortTableNames)
generateCohortSet(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDatabaseSchema,
  cohortDatabaseSchema = cohortDatabaseSchema,
  cohortTableNames = cohortTableNames,
  cohortDefinitionSet = allCohorts
)
```