Population-level Estimation

Jimyung Park

Department of Biomedical Sciences, Ajou University School of Medicine


OHDSI's mission
To improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.
To generate reliable evidence for the benefit of patients, providers, researchers, health care systems, industry, and government agencies
HDSI KOREA Tutorial
2th(Thu) / 14th(Sat), 2019 VENUE Konjiam resort


OHDSI
KOREA OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS
OHDSI Korea Symposium
12~14th December 2019,
KONJIAM Resort, Gwangju, Gyeonggi-Do, Korea


Threats to the validity of study results
Martijn Schuemie PhD
Janssen Research and Development
OHDSI
OHDSI KOREA Tutorial
VENUE Konjiam resort


WIFI
WIFI ID : ohdsikorea2019_1
Password : ohdsikorea2019!@
OHDSI
QR Code or URL
QR Code :
URL : ohdsi.site
CDM_and_Vocabulary
Phenotyping_by_ATLAS
OHDSI KOREA Tutorial
VENUE Konjiam resort

Exercise: the Example from the **Book of OHDSI**

book.ohdsi.org

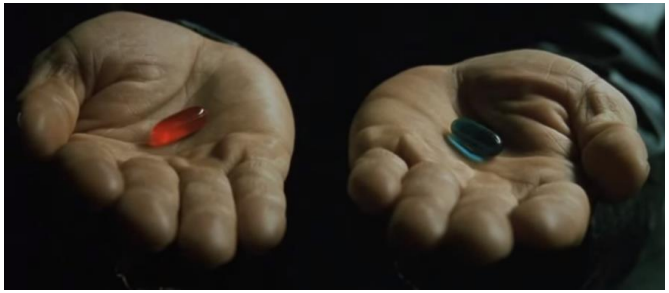
- Chapter 12. Population-Level Estimation
- 12.6 Designing a Hypertension Study

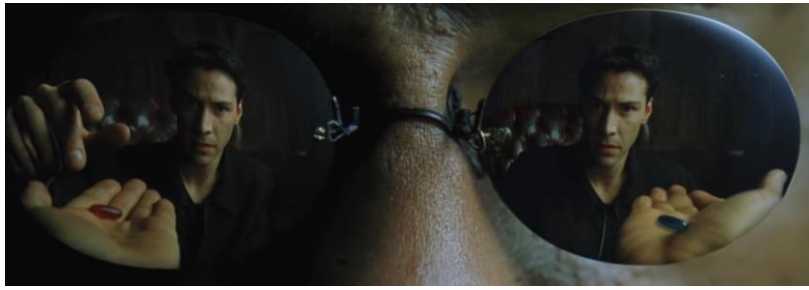
12.6 Designing a Hypertension Study

12.6.1 Problem Definition

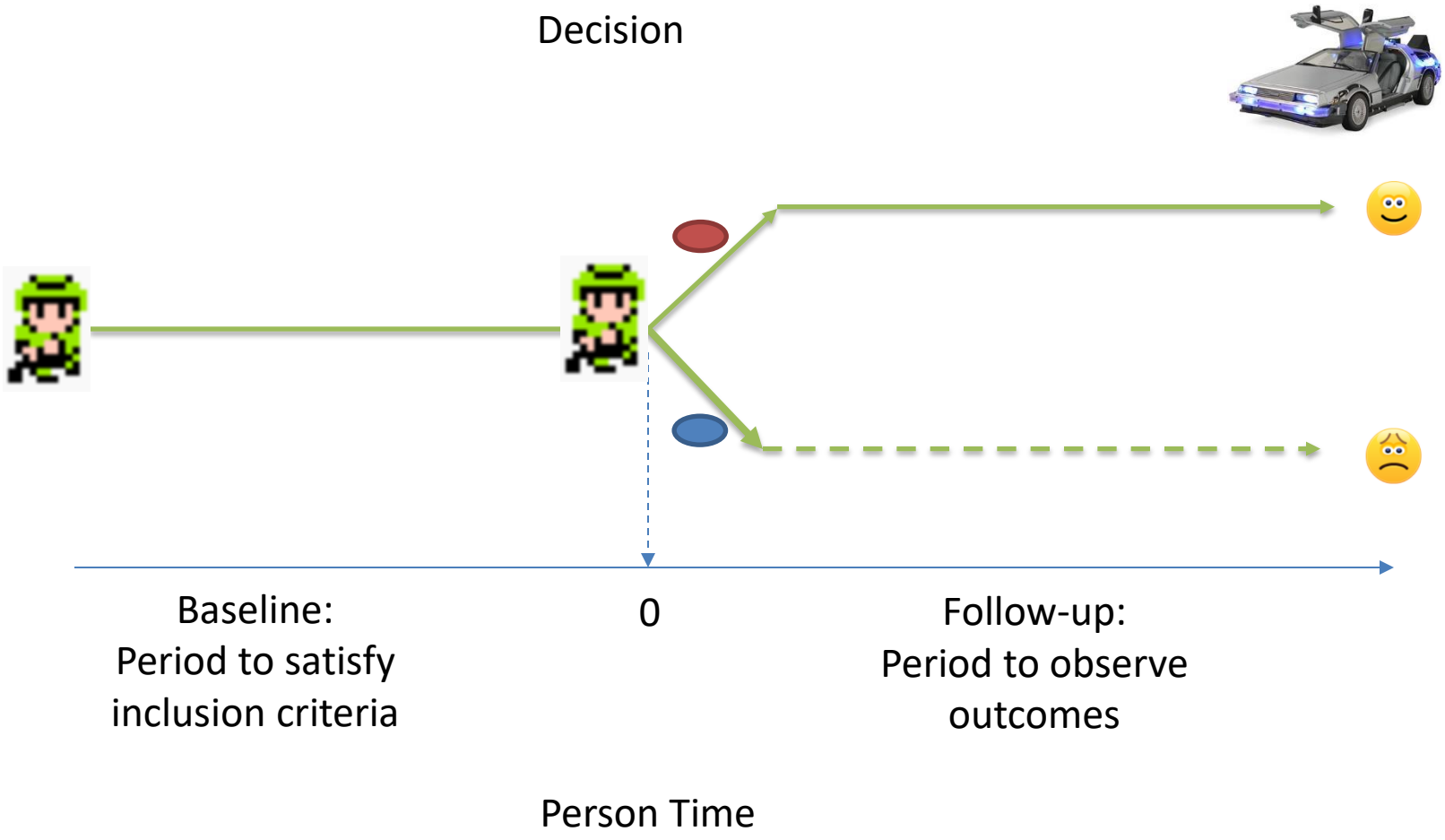
ACE inhibitors (ACEi) are widely used in patients with hypertension or ischemic heart disease, especially those with other comorbidities such as congestive heart failure, diabetes mellitus, or chronic kidney disease. (Zaman, Oparil, and Calhoun 2002) Angioedema, a serious and sometimes life-threatening adverse event that usually manifests as swelling of the lips, tongue, mouth, larynx, pharynx, or periorbital

A pop culture mash-up to explain counterfactual reasoning…

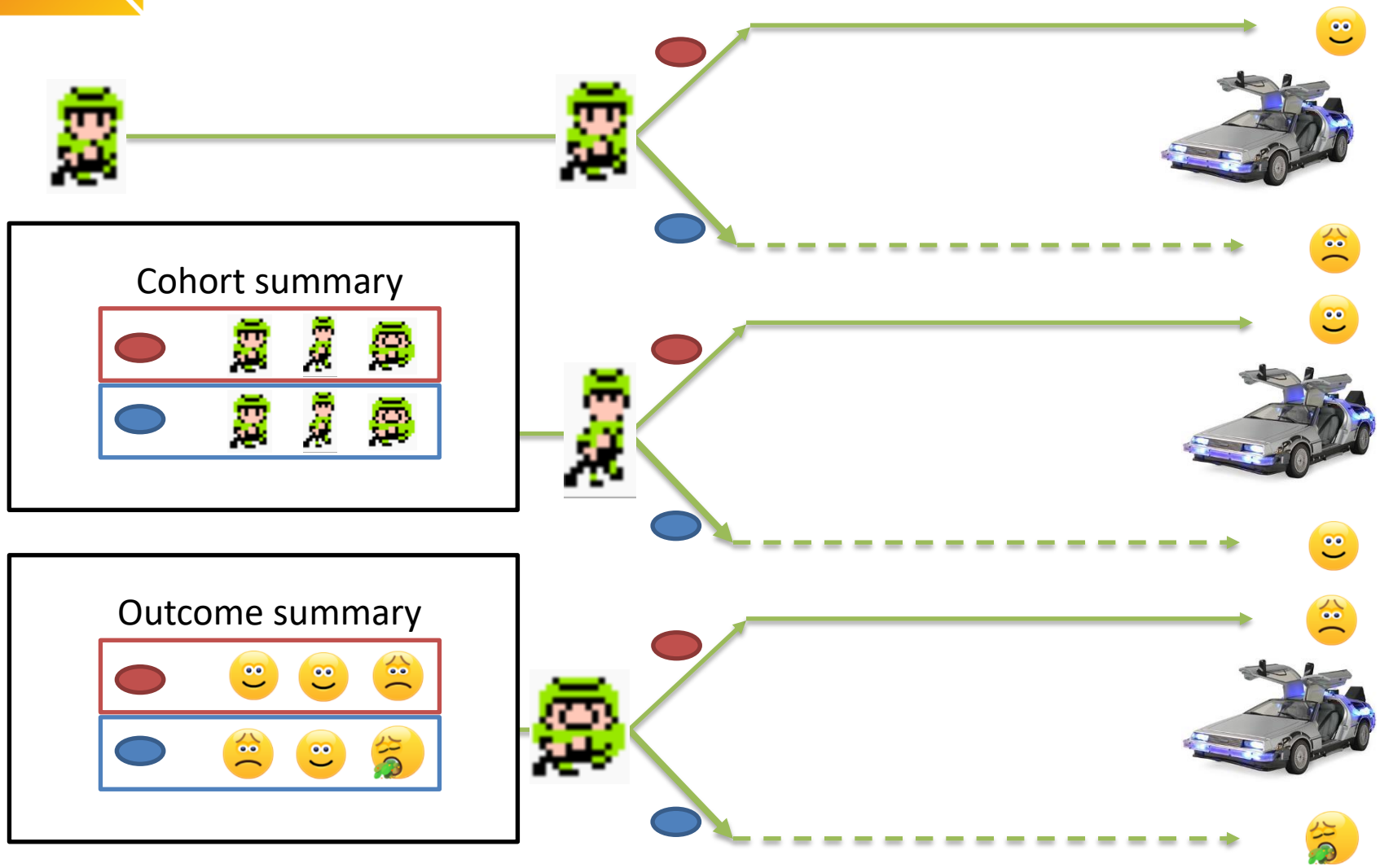




Counterfactual reasoning for one person



Counterfactual reasoning for a population



OHDSI's definition of 'cohort'

Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

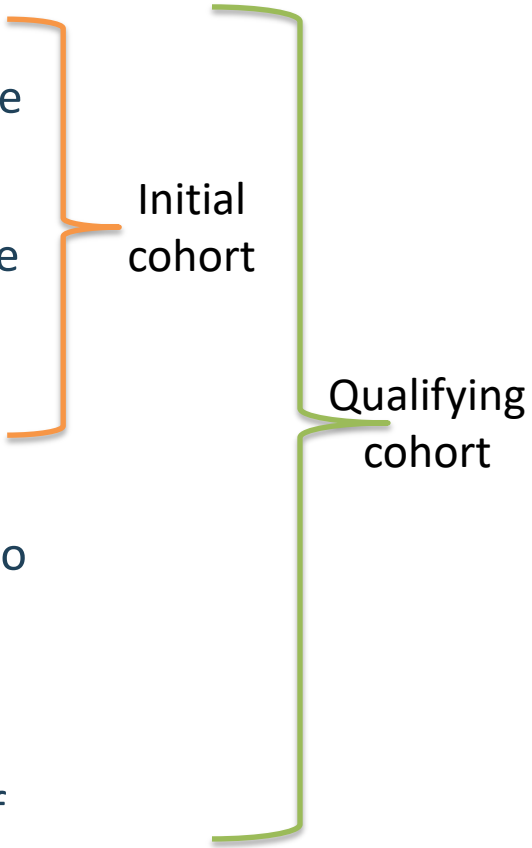
Objective consequences based on this cohort definition:

- One person may belong to multiple cohorts
- One person may belong to the same cohort at multiple different time periods
- One person may not belong to the same cohort multiple times during the same period of time
- One cohort may have zero or more members
- A codeset is NOT a cohort…

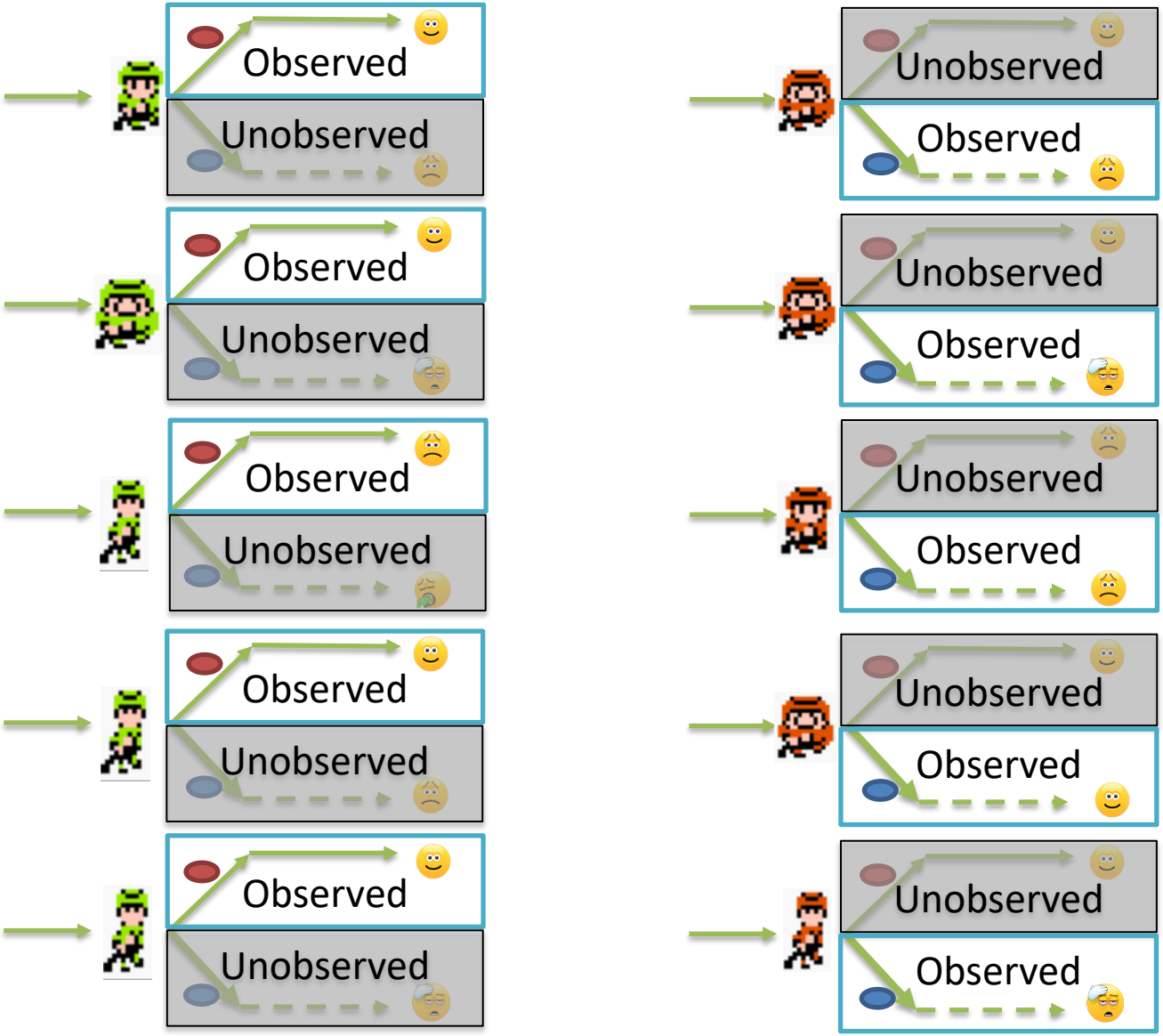
  …logic for how to use the codeset in a criteria is required

Process flow for formally defining a cohort in ATLAS

- Cohort entry criteria
  - Initial events
    - Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits.
    - All events have a start date and end date, though some events may have a start date and end date with the same value (such as procedures or measurements).
  - Initial event inclusion criteria
  - Additional qualifying inclusion criteria
    - The qualifying cohort will be defined as all persons who have an initial event, satisfy the initial event inclusion criteria, and fulfill all additional qualifying inclusion criteria.
    - Each qualifying inclusion criteria will be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort.
- Cohort exit criteria



An observational comparative cohort design to approximate counterfactual outcomes



Propensity score introduction

- Propensity score = probability of belonging to the target cohort vs. the comparator cohort, given the baseline covariates
- $e(x) = Pr(Z=1|x)$
  - Z is treatment assignment
  - x is a set of all covariates at the time of treatment assignment
- Propensity score can be used as a 'balancing score': if the two cohorts have similar propensity score distribution, then the distribution of covariates should be the similar (need to perform diagnostic to check)

Rubin Biometrika 1983

Methods for confounding adjustment using a propensity score

Regression adjustment	The PS is used as a covariable in an outcome regression model to adjust the as… assun… same… relationship between propensity score and outcome is correctly specified.
Matching	The PS is used to match exposed subjects to unexposed subjects with similar values of the PS. This method assumes that within the matched sample, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Stratification	The PS is used to stratify subjects into (often quintiles or deciles) strata. Treatment effects are estimated separately within each stratum and then combined into an overall estimate of treatment effect. This method assumes that within each stratum, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Inverse Probability Weighting	The PS is used to create weights based on the inverse probability which is defined as: E*/PS + (1-E)/(1-PS). This assumes that baseline characteristics are similar in the exposed and unexposed group.

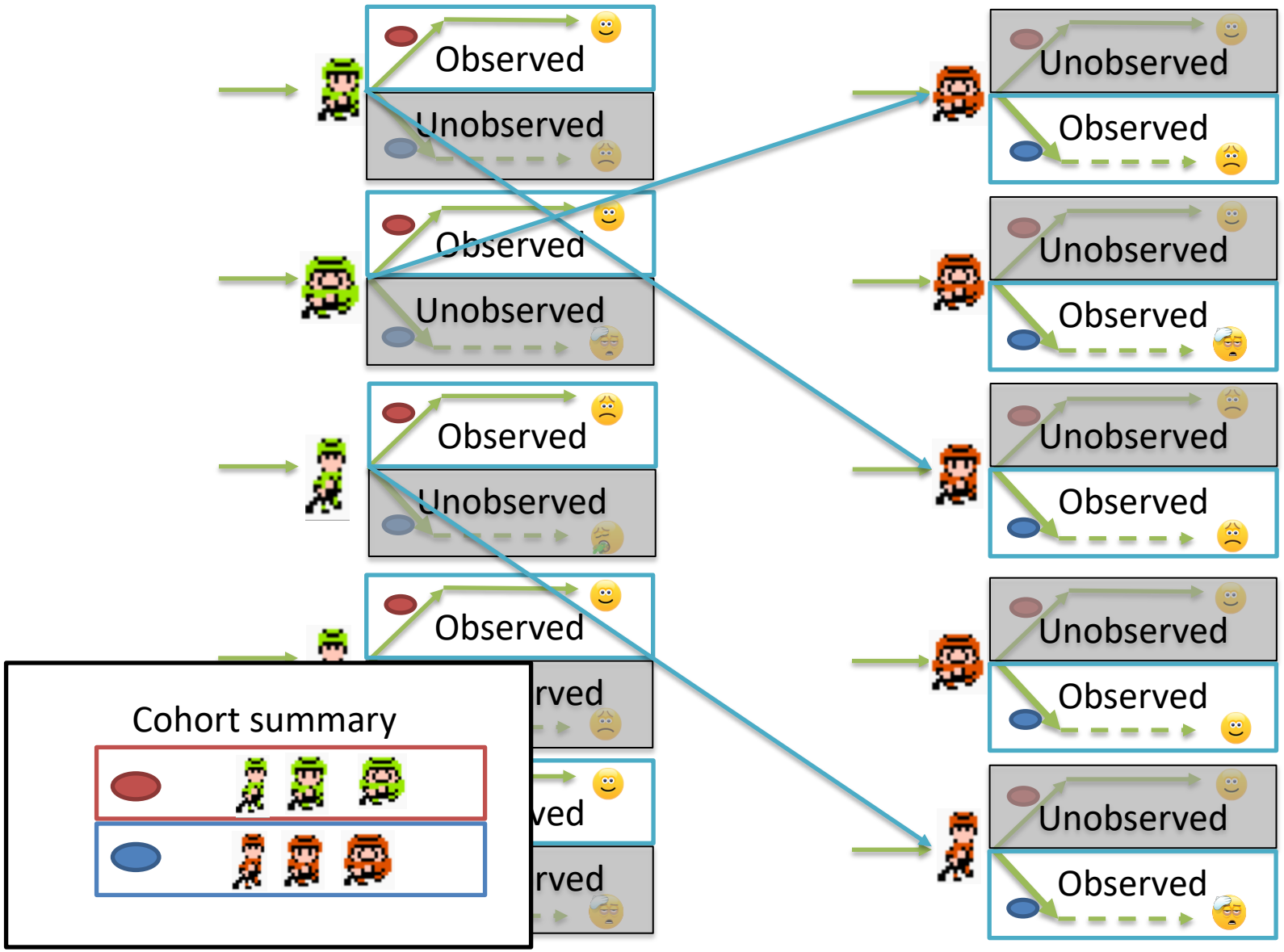
* E: exposure

Not generally recommended

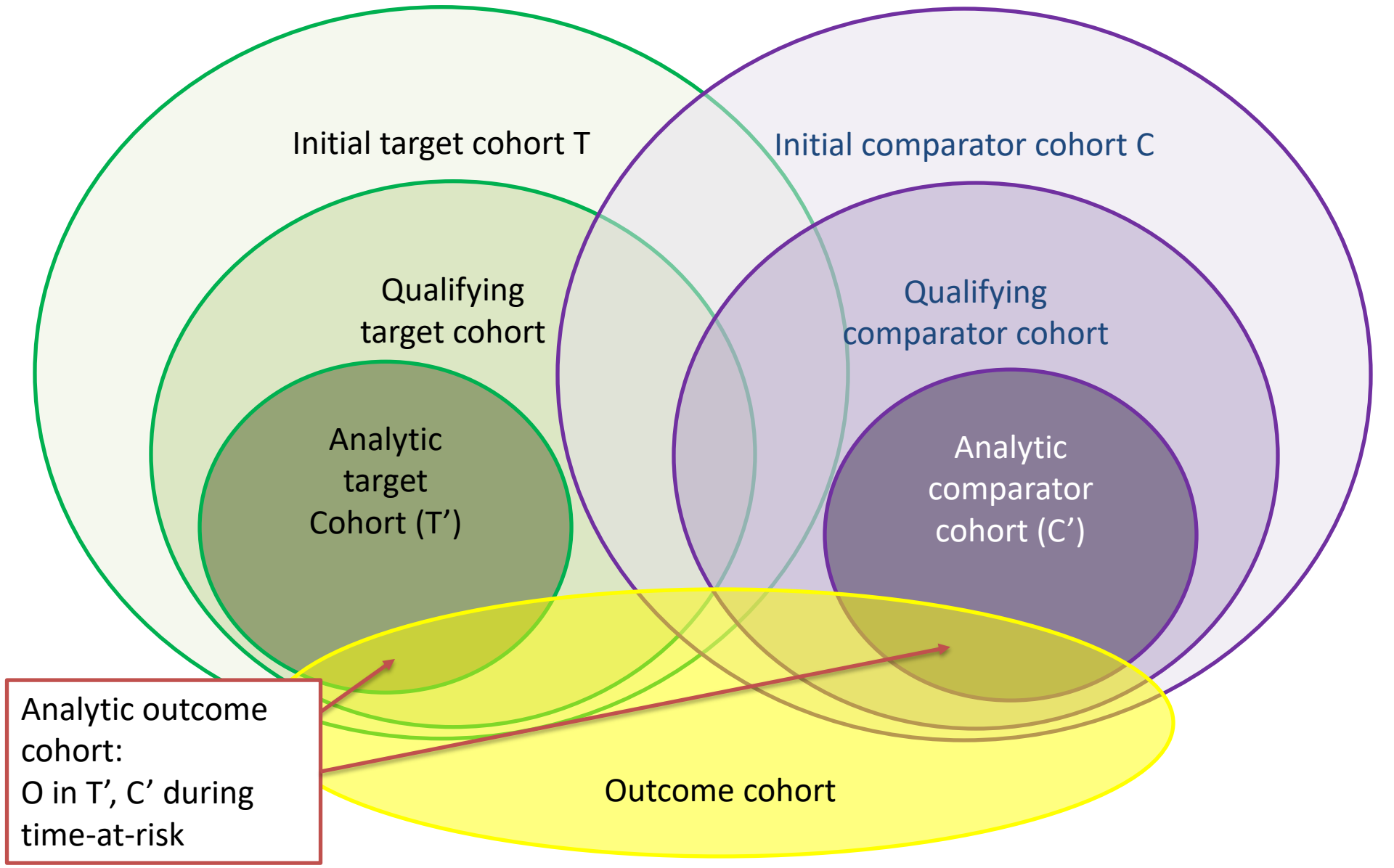
Fully implemented in OHDSI CohortMethod R package

Garbe et al, Eur J Clin Pharmacol 2013, http://www.ncbi.nlm.nih.gov/pubmed/22763756

Matching as a strategy to adjust for baseline covariate imbalance



Cohort restriction in comparative cohort analyses



The choice of the outcome model defines your research question

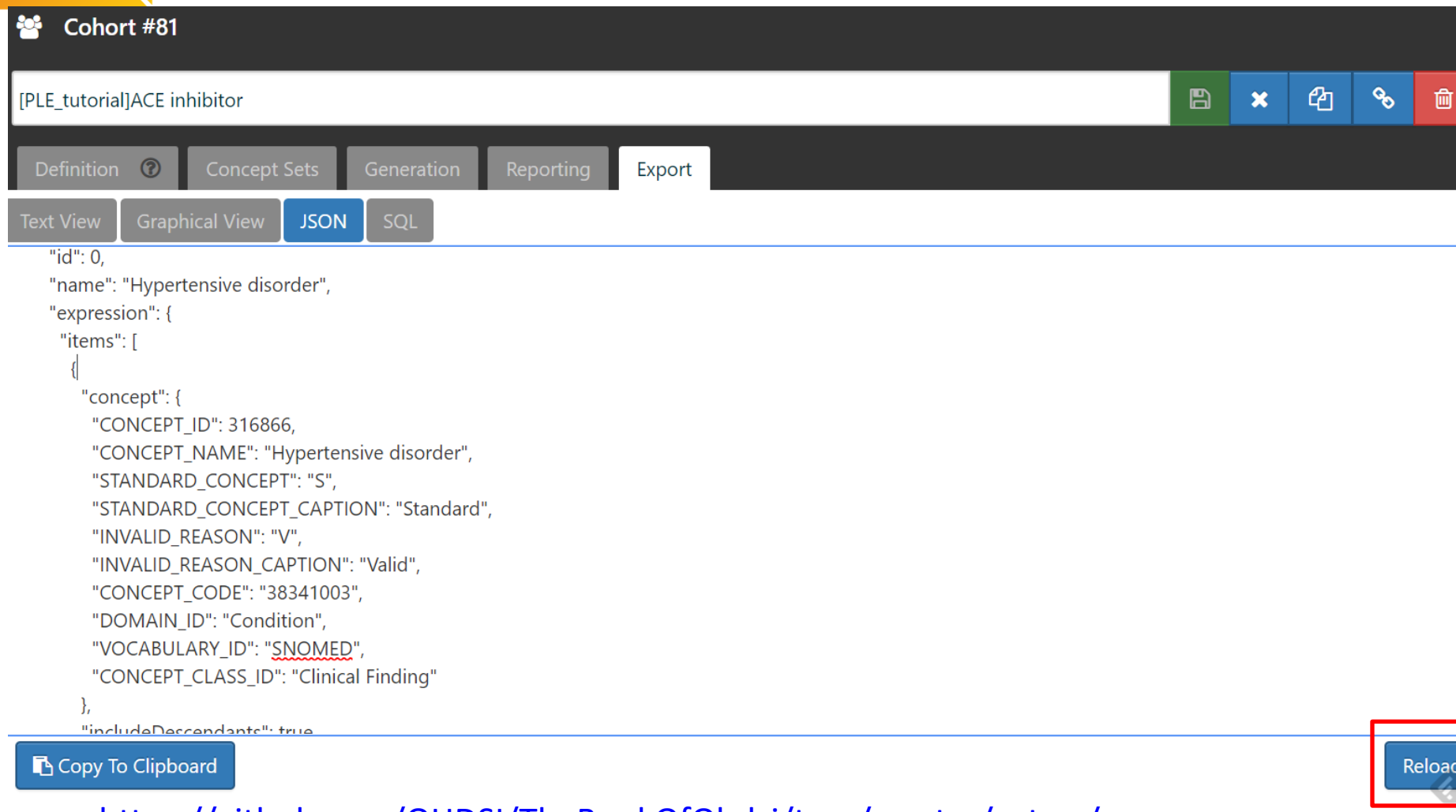
	Logistic regression	Poisson regression	Cox proportional hazards
How the outcome cohort is used	Binary classifier of presence/ absence of outcome during the fixed time-at-risk period	Count the number of occurrences of outcomes during time-at-risk	Compute time-to-event from time-at-risk start until earliest of first occurrence of outcome or time-at-risk end, and track the censoring event (outcome or no outcome)
'Risk' metric	Odds ratio	Rate ratio	Hazard ratio
Key model assumptions	Constant probability in fixed window	Outcomes follow Poisson distribution with constant risk	Proportionality – constant relative hazard

Population-level Estimation <Exercise>

Jimyung Park

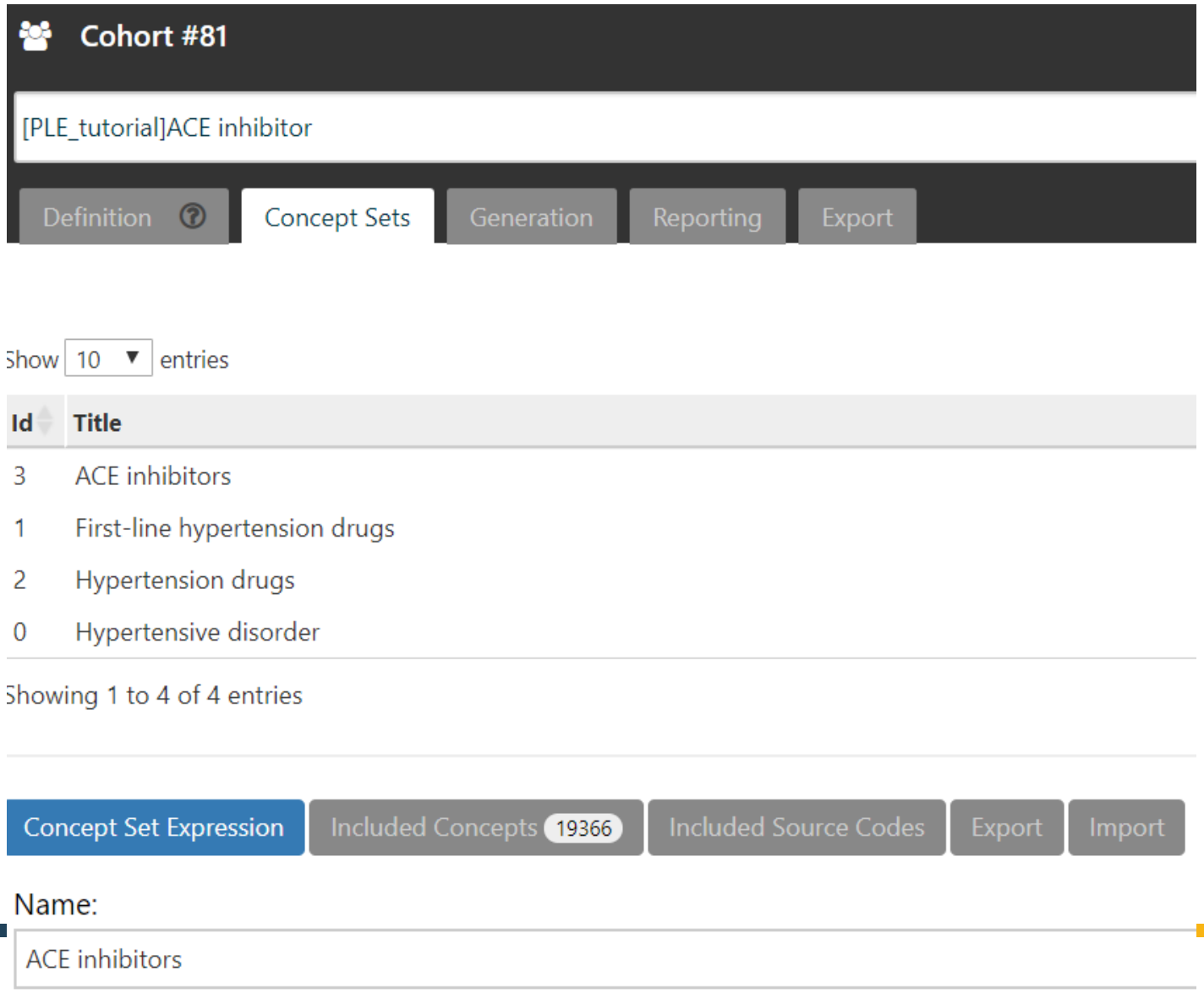
Department of Biomedical Sciences, Ajou University School of Medicine

Load Cohort Definitions



https://github.com/OHDSI/TheBookOfOhdsi/tree/master/extras/CohortMethodAceiVsThz/inst/cohorts

Target Cohort



[PLE_tutorial]ACE inhibitor

Definition | Concept Sets | Generation | Reporting | Export

enter a cohort definition description here

Cohort Entry Events

Events having any of the following criteria:

a drug exposure of ACE inhibitors

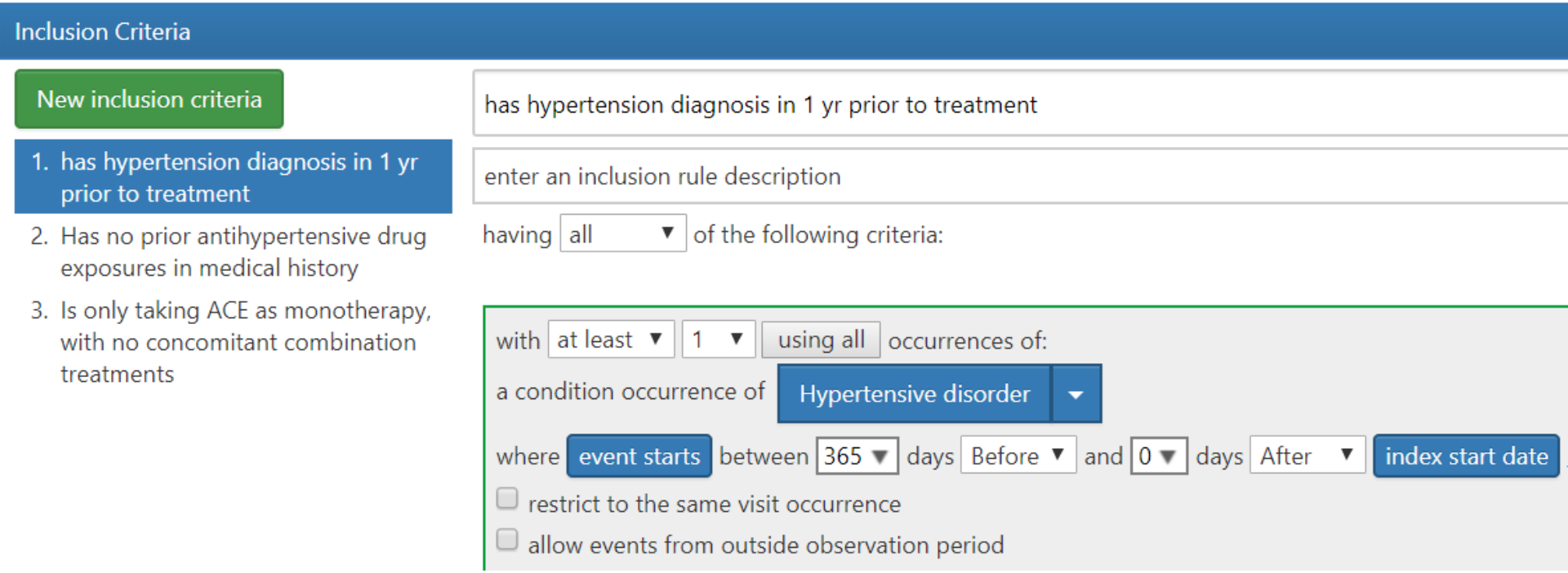
✖ for the first time in the person's history

with continuous observation of at least 365 days before and 0 days after event index date

Limit initial events to: earliest event per person.

Restrict initial events

Target Cohort



Target Cohort

Inclusion Criteria

New inclusion criteria

1. has hypertension diagnosis in 1 yr prior to treatment
2. Has no prior antihypertensive drug exposures in medical history
3. Is only taking ACE as monotherapy, with no concomitant combination treatments

Has no prior antihypertensive drug exposures in medical history

enter an inclusion rule description

having all ▼ of the following criteria:

with exactly ▼ 0 ▼ using all occurrences of:

a drug exposure of Hypertension drugs ▼

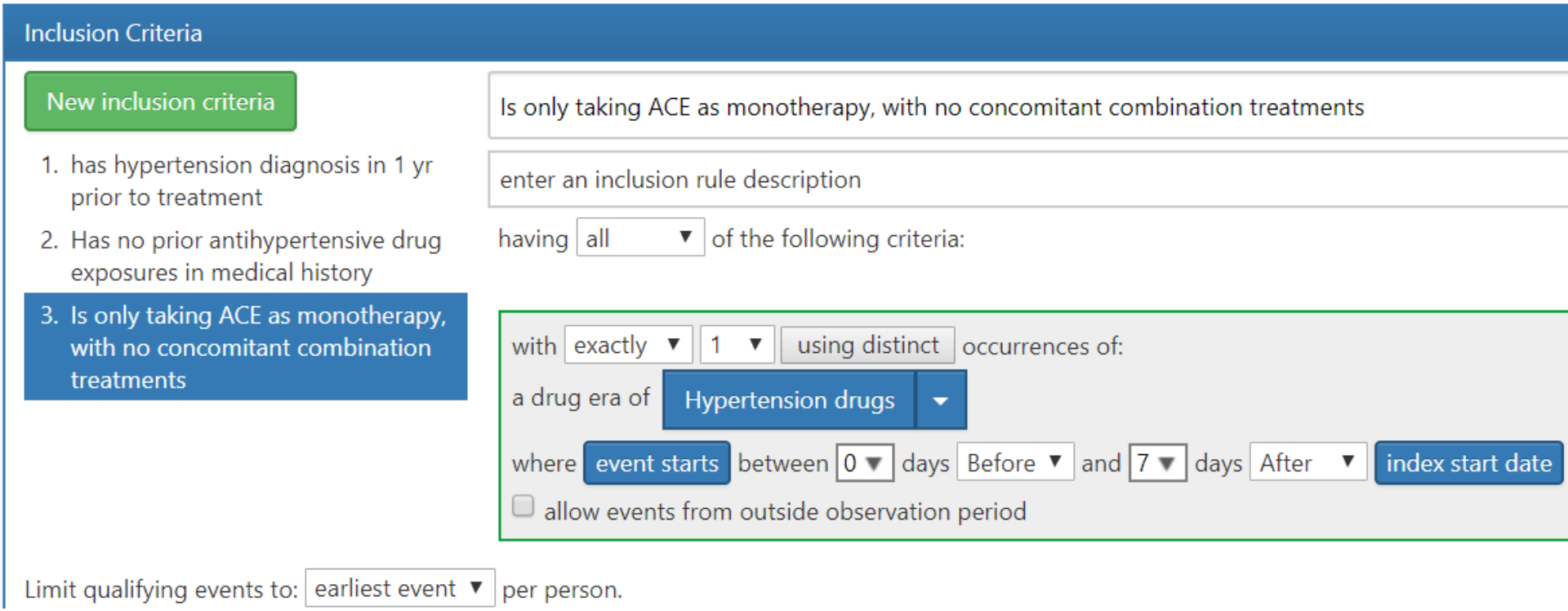
where event starts between All ▼ days Before ▼ and 1 ▼ days Before ▼ index start date

☐ restrict to the same visit occurrence

☐ allow events from outside observation period

Limit qualifying events to: earliest event ▼ per person.

Target Cohort



Target Cohort

Cohort Exit

Event Persistence:

Event will persist until: end of a continuous drug exposure ▼

Continuous Exposure Persistence:

Specify a concept set that contains one or more drugs. A drug era will be derived from all drug exposure events for any of the drugs within the conce
adding a specified surveillance window to the final exposure event. If no exposure event end date is provided, then an exposure event end date is infe
persistence assures that the cohort end date will be no greater than the drug era end date.

Concept set containing the drug(s) of interest: ACE inhibitors ▼

- Persistence window: allow for a maximum of 30 ▼ days between exposure records when inferring the era of persistence exposure
- Surveillance window: add 0 ▼ days to the end of the era of persistence exposure as an additional period of surveillance prior to cohort exit.

Censoring Events:

Exit Cohort based on the following criteria:

No censoring events selected.

Result from generation

Available CDM Sources

	Source Name	Generation Status	People	Records
▶ Generate	CMSDESynPUF100k	COMPLETE	261	261
▶ Generate	CMSDESynPUF1k	COMPLETE	21	21
▶ Generate	CMSDESynPUF23m	COMPLETE	5,124	5,124

		Match Rate	Matches	Total Events
	Summary Statistics:	1.62%	5,124	316,688

	Inclusion Rule	N	% Remain	% Diff
1.	has hypertension diagnosis in 1 yr prior to treatment	25,979	8.20%	91.80%
2.	Has no prior antihypertensive drug exposures in medical history	5,943	1.88%	6.33%
3.	Is only taking ACE as monotherapy, with no concomitant combination treatments	5,124	1.62%	0.26%

Attrition Visualization

Switch to intersect view

Comparison setting

Comparison
Add or update the target, comparator, outcome(s) cohorts and negative control outcomes

Choose your target cohort:

New users of ACE inhibitors as first-line monotherapy for hypertension

Choose your comparator cohort:

New users of Thiazide-like diuretics as first-line monotherapy for hypertension

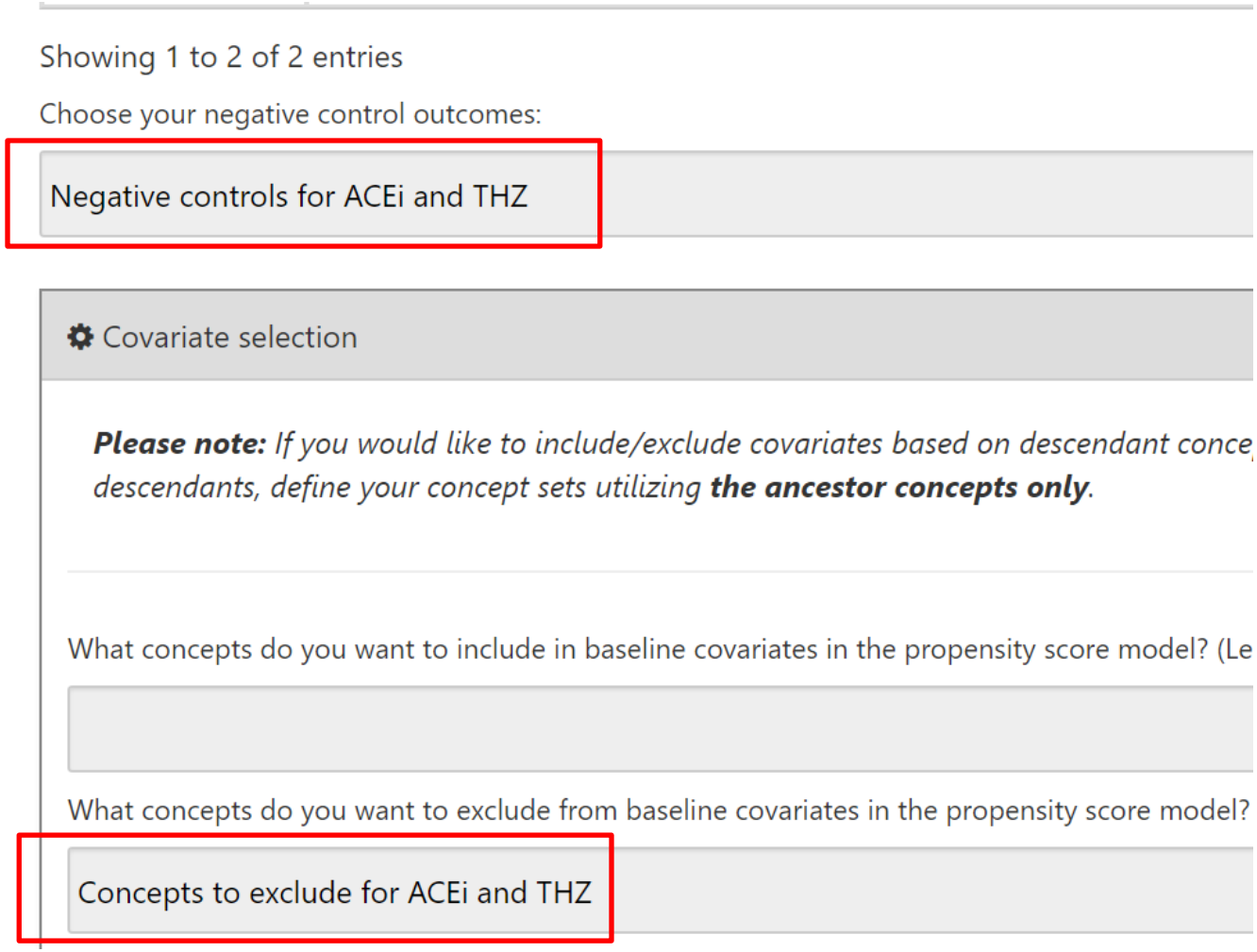
Choose your outcome cohorts:

Add Outcome

Show 10 entries Search:

ID	Name		
90	Angioedema outcome	Edit cohort	Remove
91	Acute myocardial infarction outcome	Edit cohort	Remove

Add negative controls and Concepts to exclude



Analysis Settings

Should only the first exposure per subject be included?

No

Remove subjects that are in both the target and comparator cohort?

Remove All

Restrict the analysis to the period when both exposures are observed?

No

The mininum required continuous observation time prior to index date for a person to be included in the cohort.

0

If either the target or the comparator cohort is larger than this number it will be sampled to this size. (0 for this value i

0

Remove subjects that have the outcome prior to the risk window start?

Yes

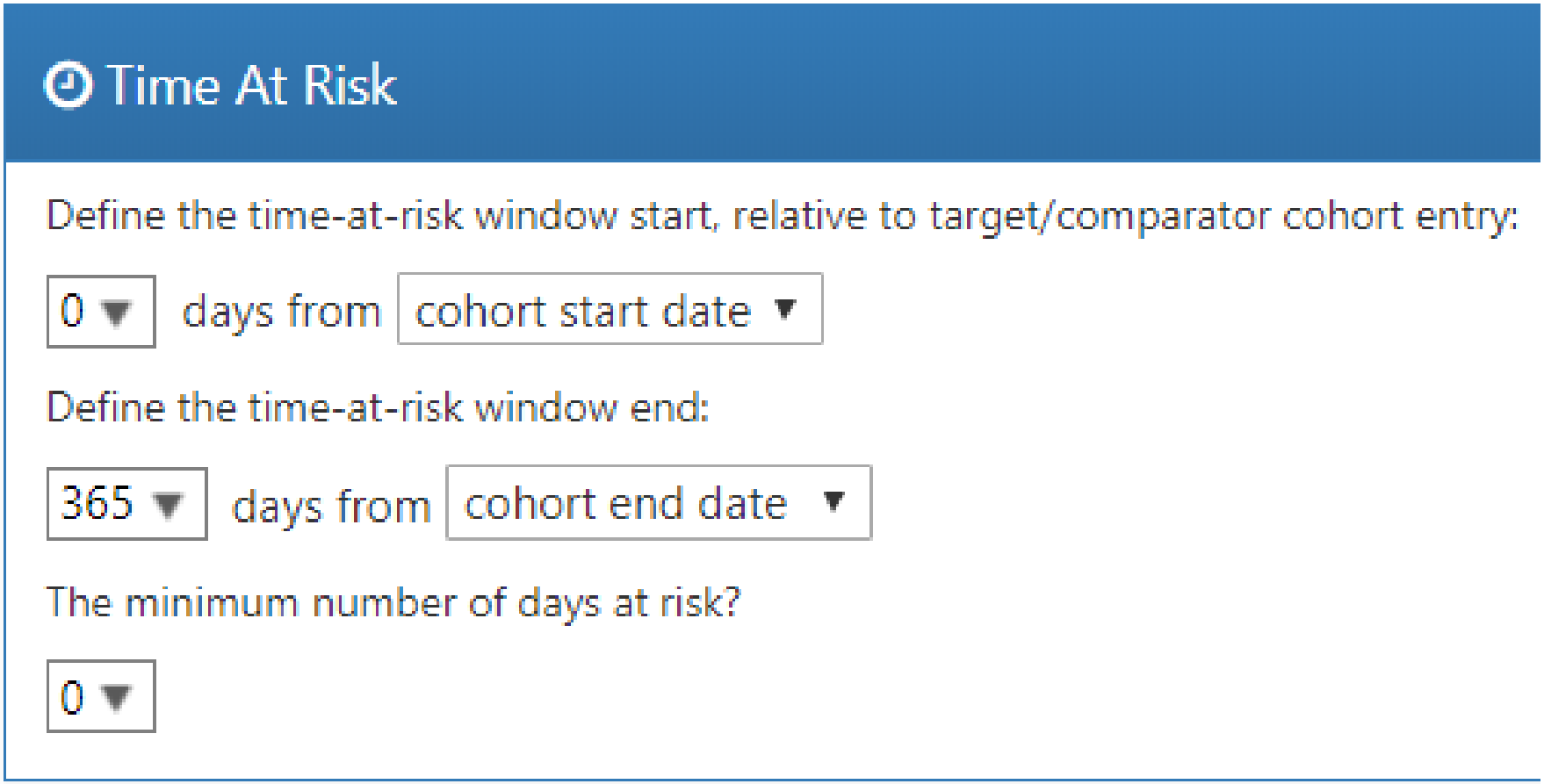
How many days should we look back when identifying prior outcomes?

99999

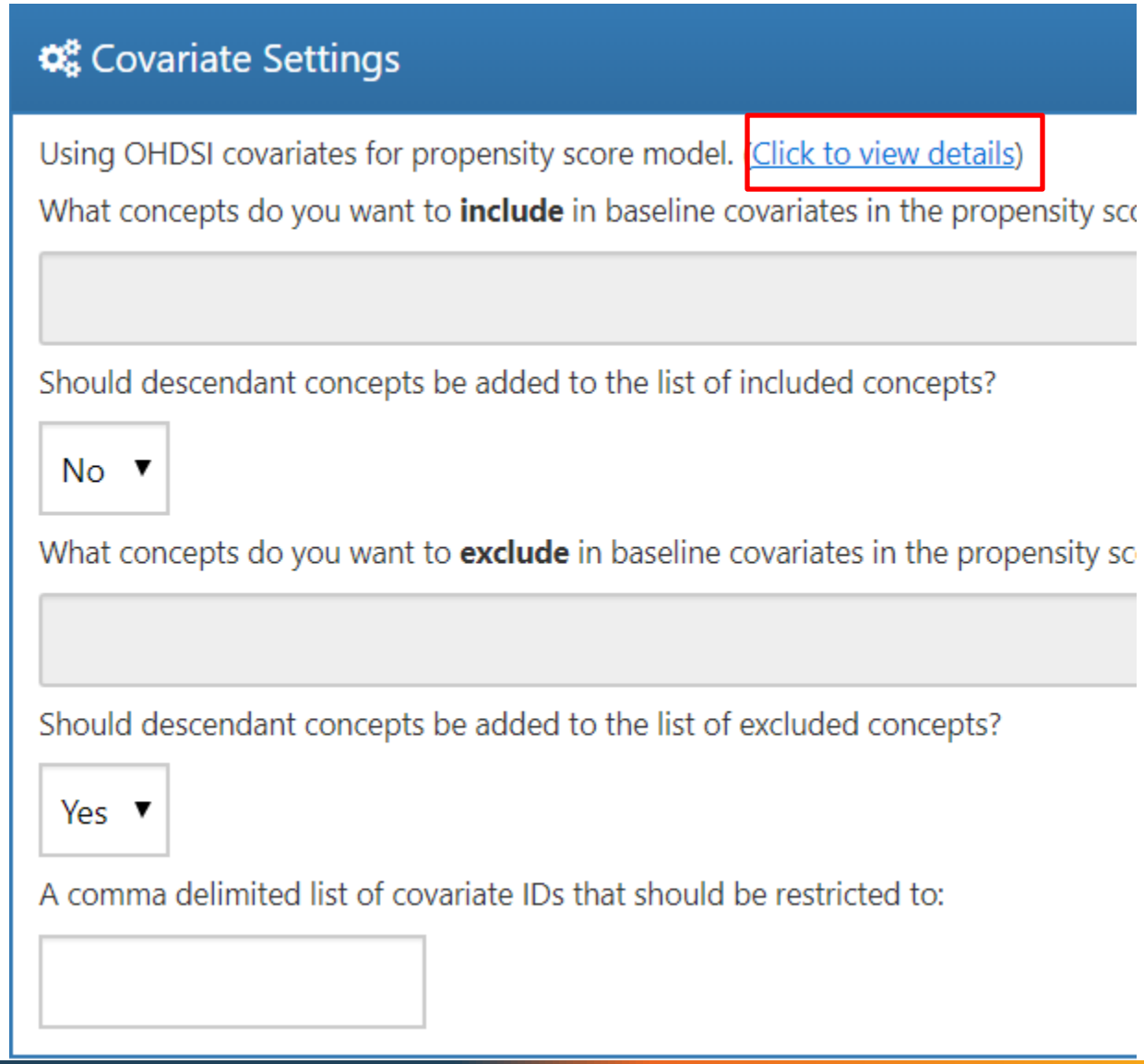
If a subject is in multiple cohorts, should time-at-risk be censored when the new time-at-risk start to prevent overlap?

No

Changing Time At Risk



Covariate Settings



Propensity Score Adjustment

How do you want to trim your cohorts based on the propensity score distribution?

None ▼

Do you want to perform matching or stratification?

Match on propensity score ▼

What is the maximum number of persons in the comparator arm to be matched to each person in the target arm within the defined caliper? (0 = means no maximum

100 ▼

What is the caliper for matching:

0.2

What is the caliper scale:

Standardized Logit ▼

What is the maximum number of people to include in the propensity score model when fitting? Setting this number to 0 means no down-sampling will be applied:

250000 ▼

Test each covariate for correlation with the target assignment? If any covariate has an unusually high correlation (either positive or negative), this will throw an error.

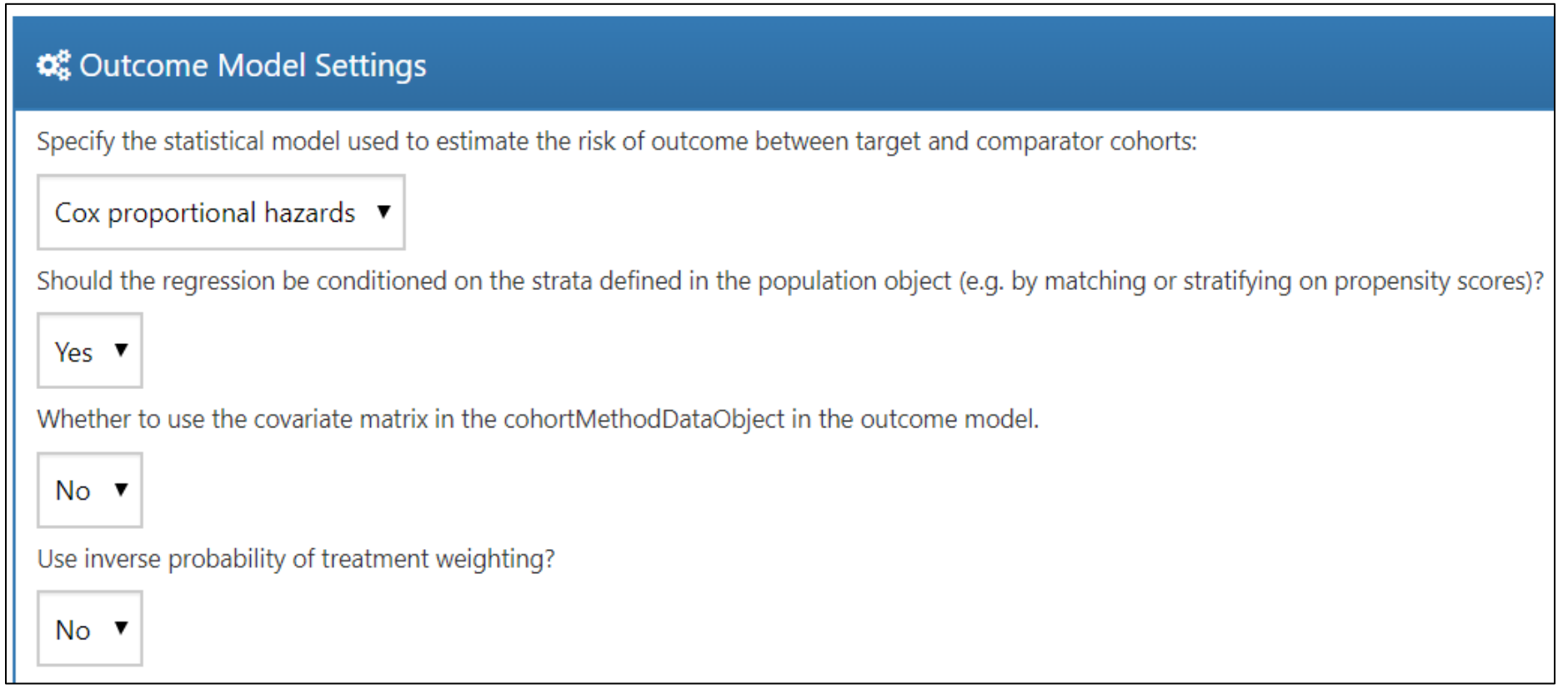
Yes ▼

If an error occurs, should the function stop? Else, the two cohorts will be assumed to be perfectly separable.

Yes ▼

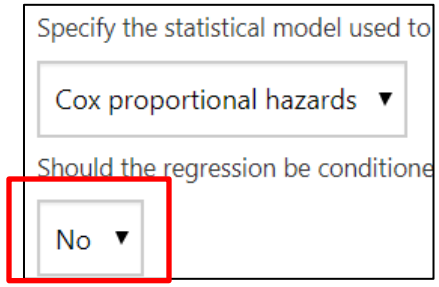
Outcome Model Settings

Choice	Value
Model	Cox proportional hazards model using variable-ratio matching.

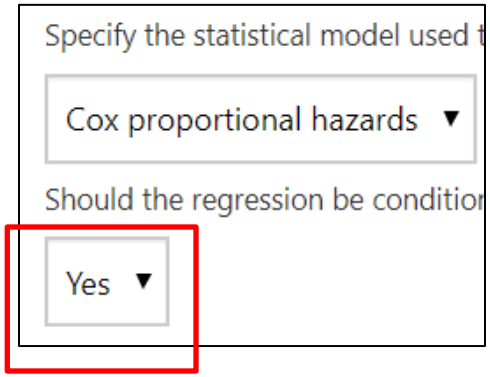


Recommended statistical model settings

- 1:1 PS matching and un-conditioned Cox regression

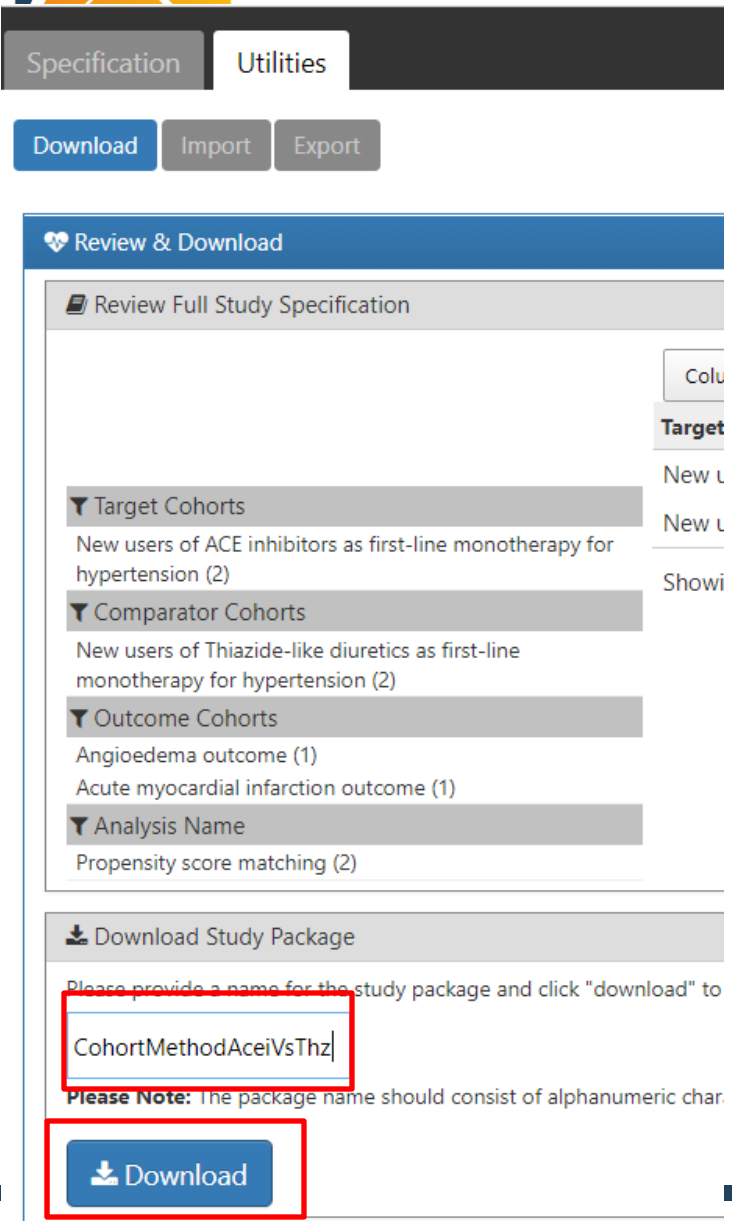


- Variable-ratio (1:100) PS matching and conditioned Cox regression



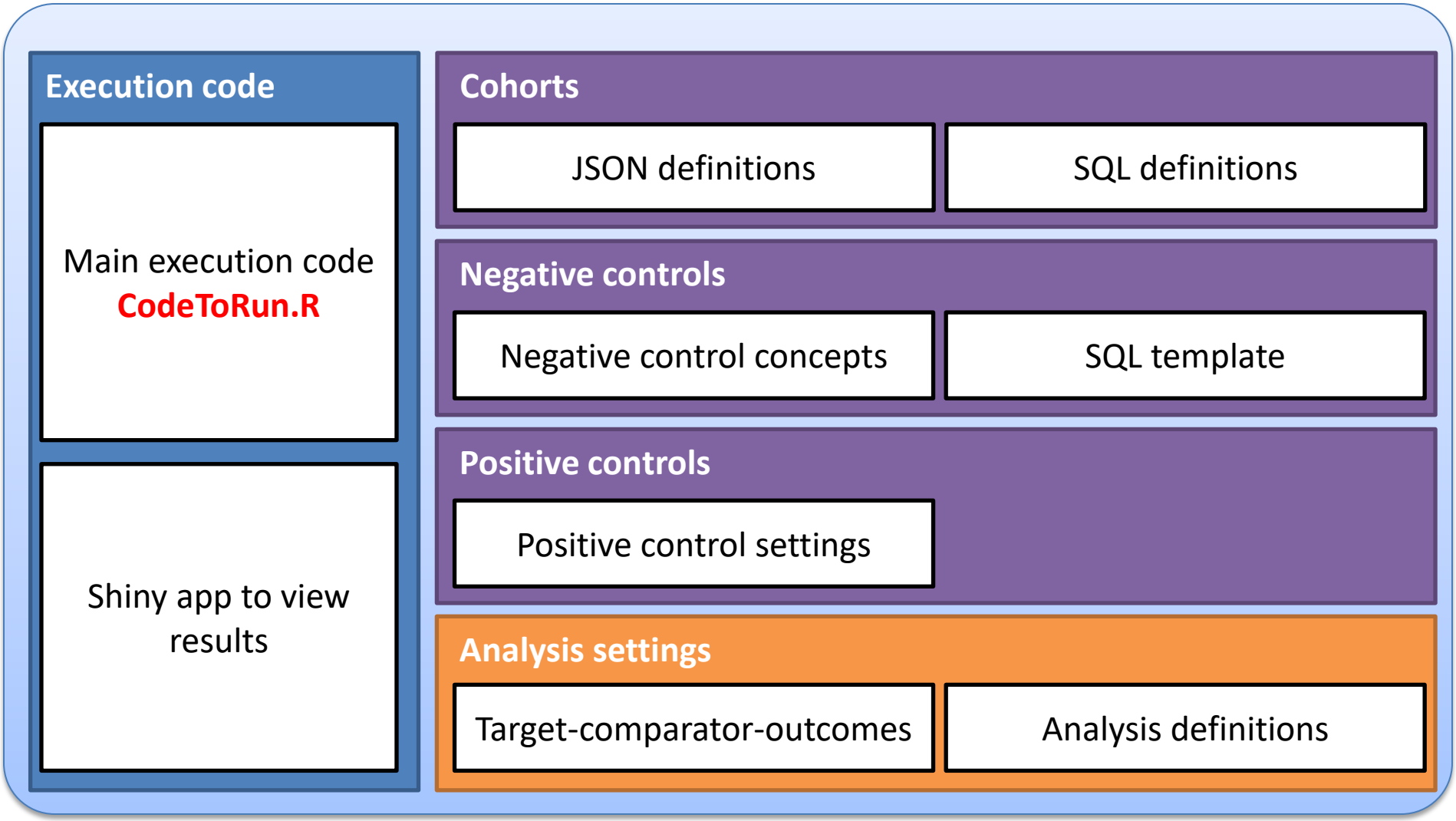
- PS stratification and conditioned Cox regression

Download the package



- Name the package and click ‘Download’ in the Utilities tab

Anatomy of the study package



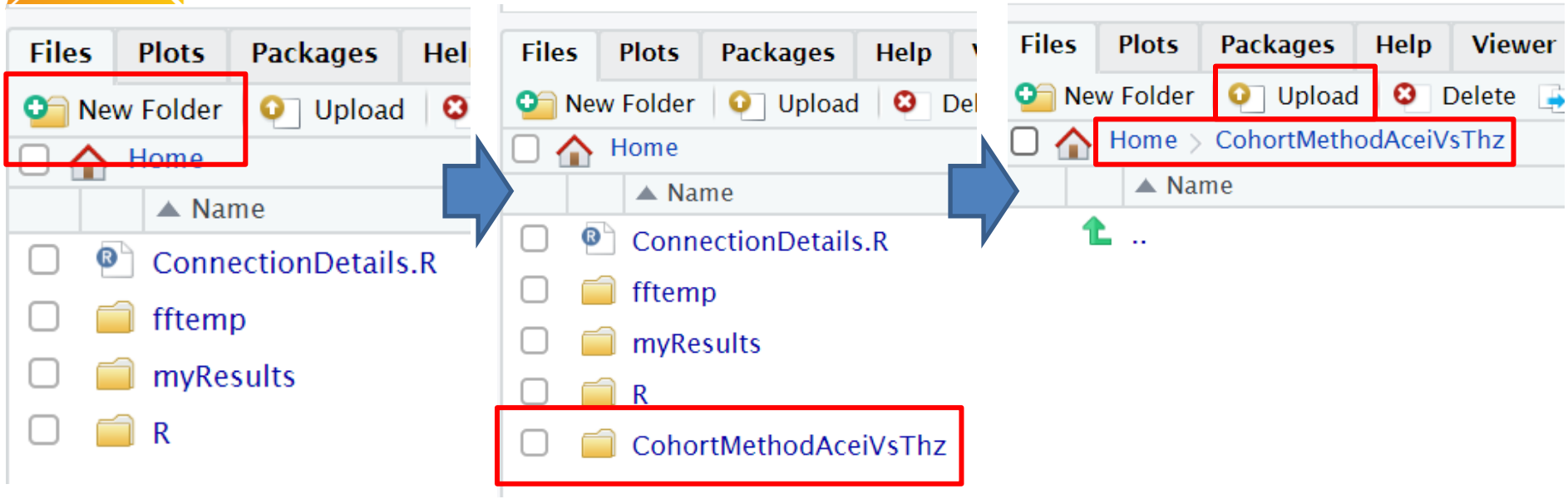
Running package

Source: readme.md

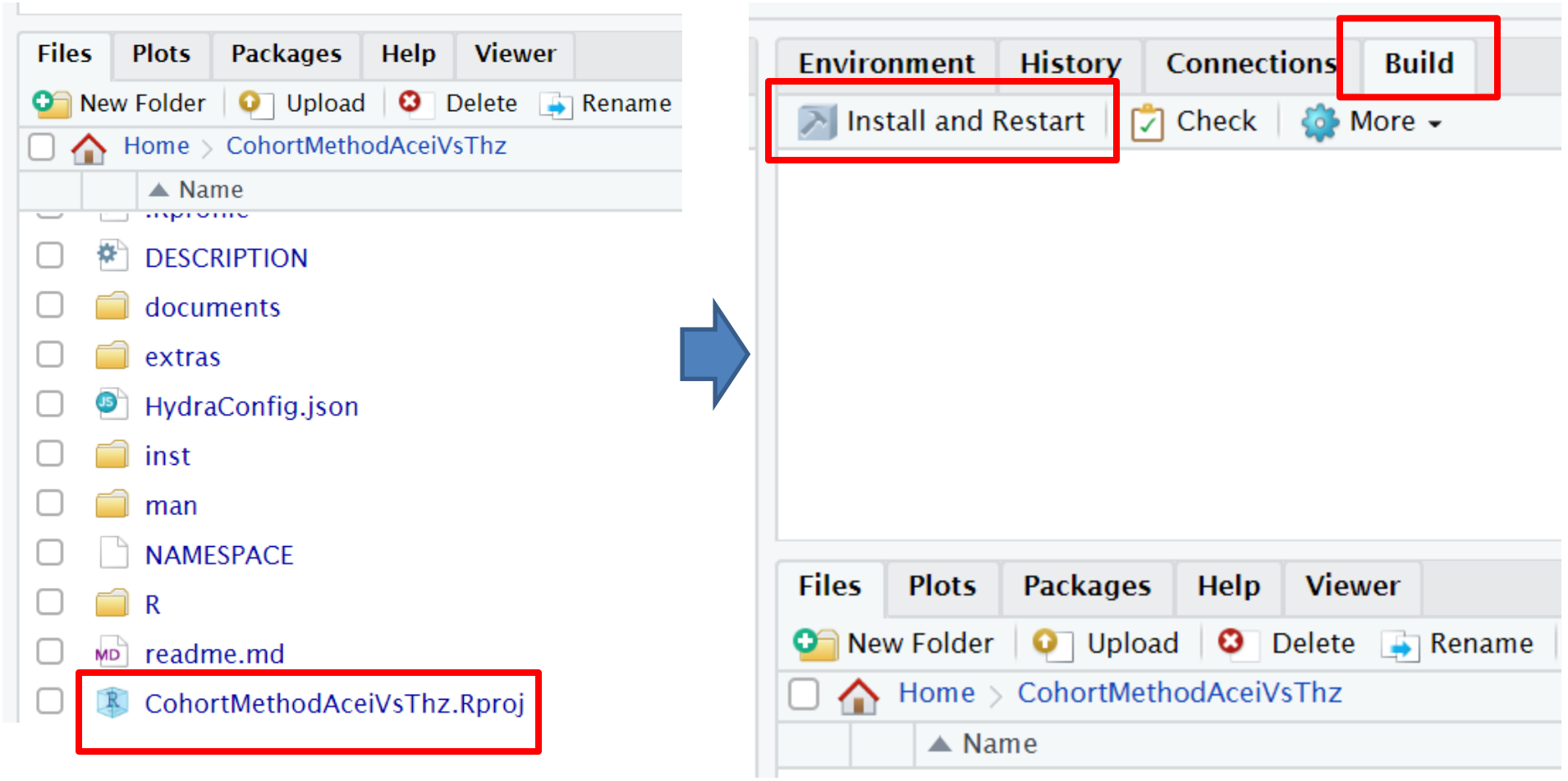
3. Once installed, you can execute the study by modifying and using the following code:

```
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"
```

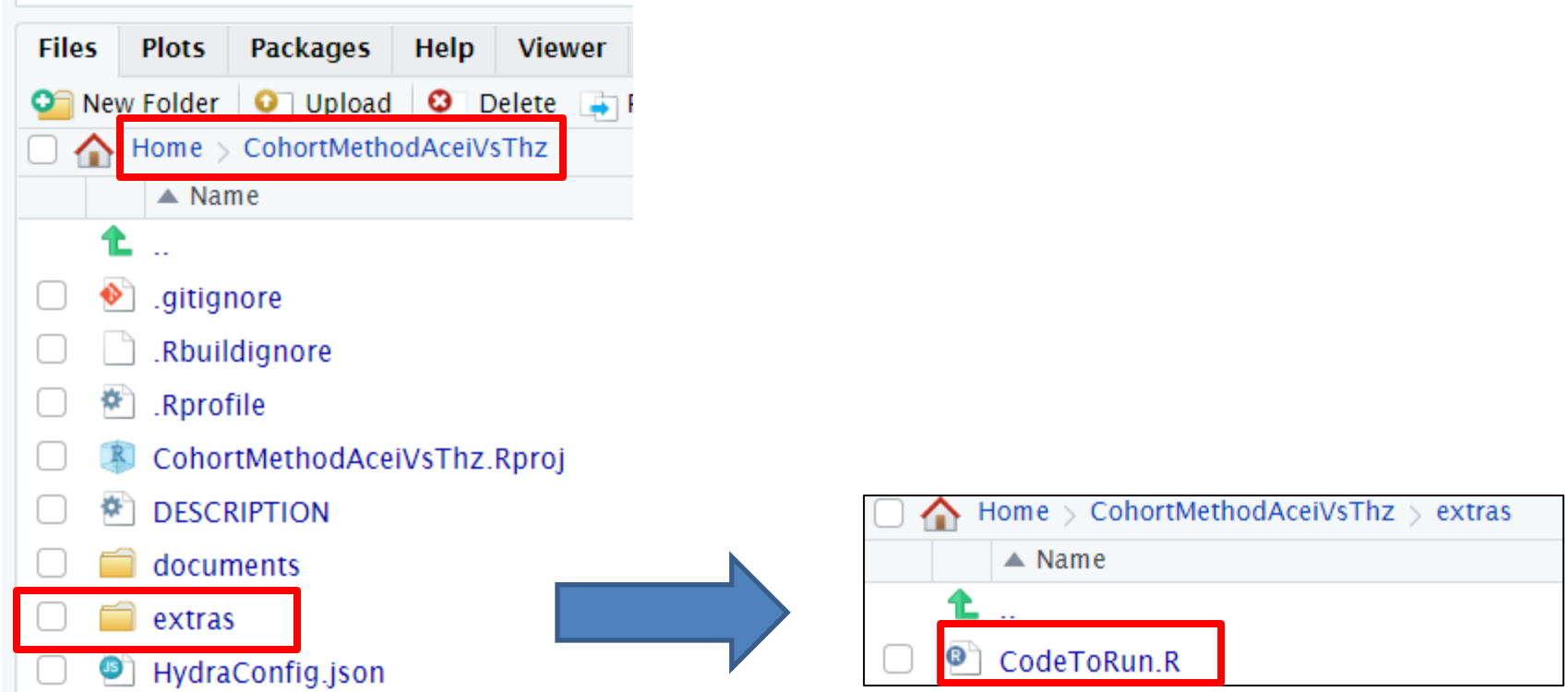
Upload the package



Install R package



CodeToRun.R


Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Home
CohortMethodAceiVsThz
Name
..
.gitignore
.Rbuildignore
.Rprofile
CohortMethodAceiVsThz.Rproj
DESCRIPTION
documents
extras
HydraConfig.json
inst
man
NAMESPACE
Home
CohortMethodAceiVsThz
extras
Name
..
CodeToRun.R

Execute the Code

```
cohortTable <- "mschuemi_skeleton"

#Connection string for the OMOP database on Redshift
connectionDetails <- DatabaseConnector::createConnectionDetails(dbms = "redshift",
                                                                server = "ajou-ohdsi-dat
                                                                user = "master",
                                                                password = "Ajoumed01",
                                                                port = "5439")


options(fftempdir = "~/fftemp")
outputFolder <- "myResults"

cdmDatabaseSchema <- "CMSDESynPUF100k"
cohortDatabaseSchema <- "CMSDESynPUF100kresults"

# Some meta-information that will be used by the export function:
databaseId <- "Synpuf"
databaseName <- "Medicare Claims Synthetic Public Use Files (SynPUFs)"
databaseDescription <- "Medicare Claims Synthetic Public Use Files (SynPUFs) were create

# For Oracle: define a schema that can be used to emulate temp tables:
oracleTempSchema <- NULL
```

Execute the Code

```
execute(connectionDetails = connectionDetails,
        cdmDatabaseSchema = cdmDatabaseSchema,
        cohortDatabaseSchema = cohortDatabaseSchema,
        cohortTable = cohortTable,
        oracleTempSchema = oracleTempSchema,
        outputFolder = outputFolder,
        databaseId = databaseId,
        databaseName = databaseName,
        databaseDescription = databaseDescription,
        createCohorts = TRUE,
        synthesizePositiveControls = TRUE,
        runAnalyses = TRUE,
        runDiagnostics = TRUE,
        packageResults = TRUE,
        maxCores = maxCores)

resultsZipFile <- file.path(outputFolder, "export", paste0("Results", databaseId, ".zip"))
dataFolder <- file.path(outputFolder, "shinyData")

prepareForEvidenceExplorer(resultsZipFile = resultsZipFile, dataFolder = dataFolder)
```

Check the result by ShinyViewer

```
69  launchEvidenceExplorer(dataFolder = dataFolder, blind = FALSE, launch.browser = FALSE)
```

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

Thank you

Jimyung Park

cory66421@gmail.com