Observational study design

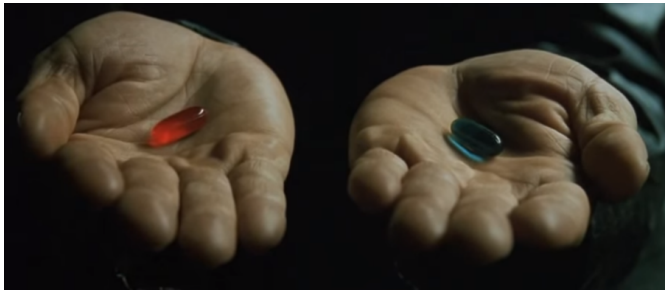
Patrick Ryan, PhD

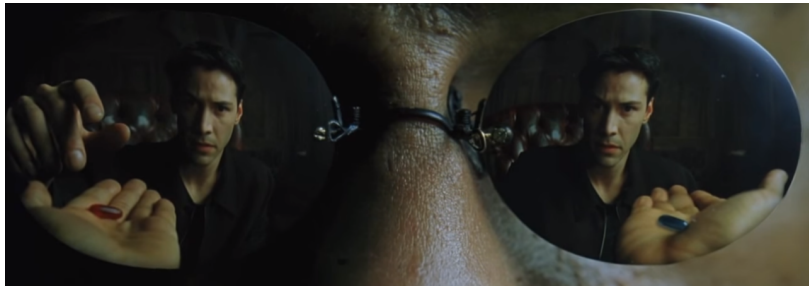
Columbia University

Janssen Research and Development

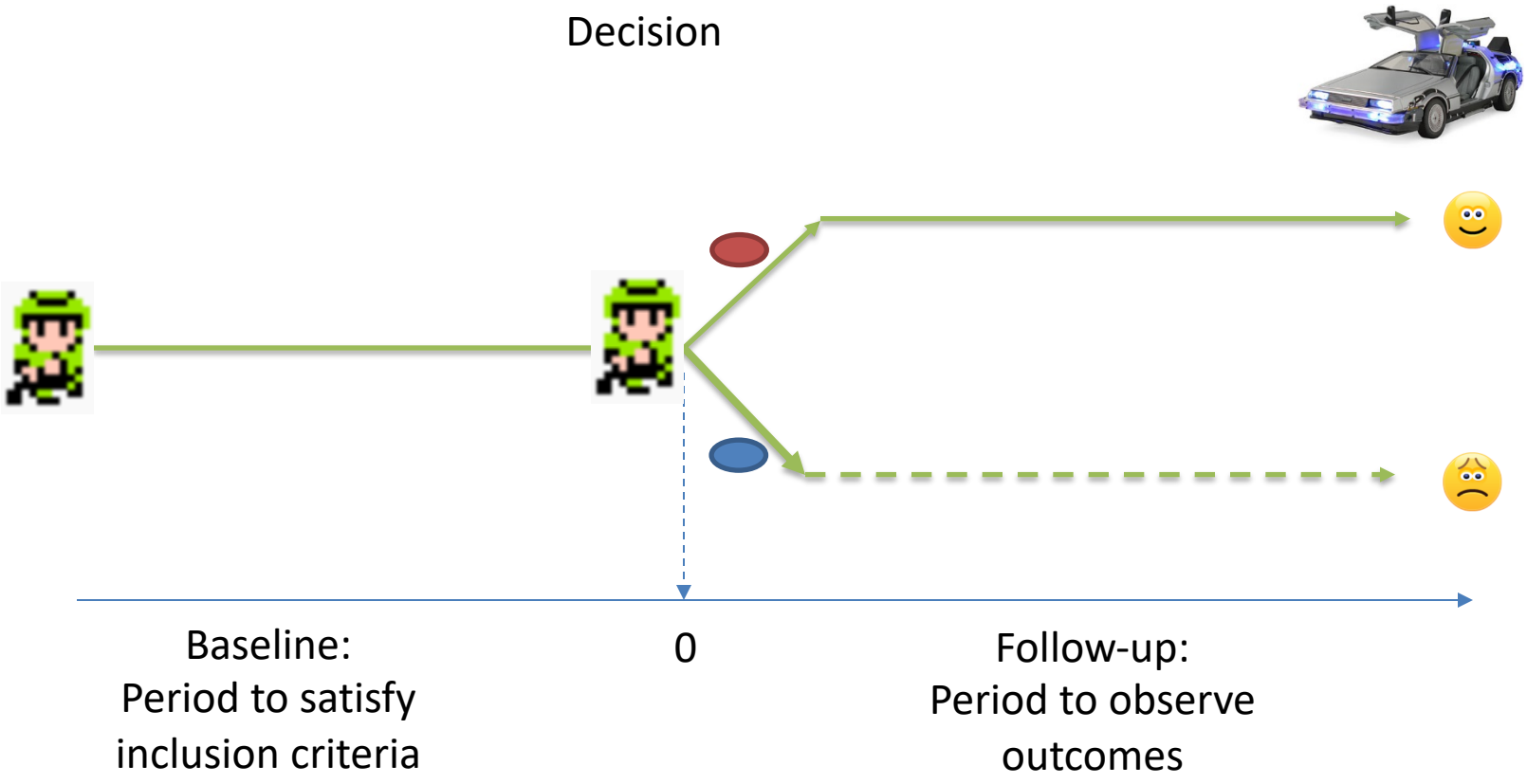
A little exercise:
choose your own adventure!

A pop culture mash-up to explain counterfactual reasoning…

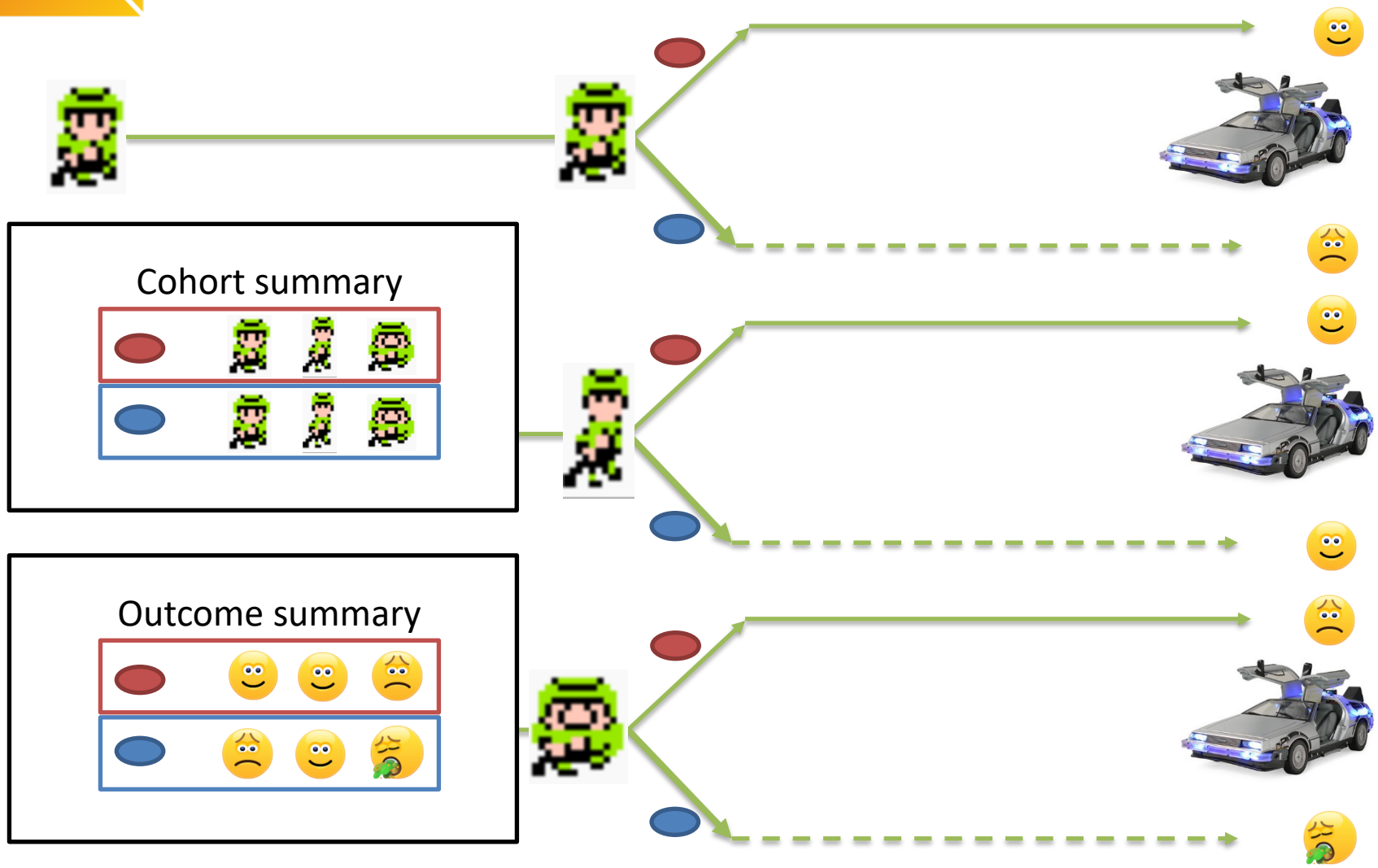




Counterfactual reasoning for one person



Counterfactual reasoning for a population

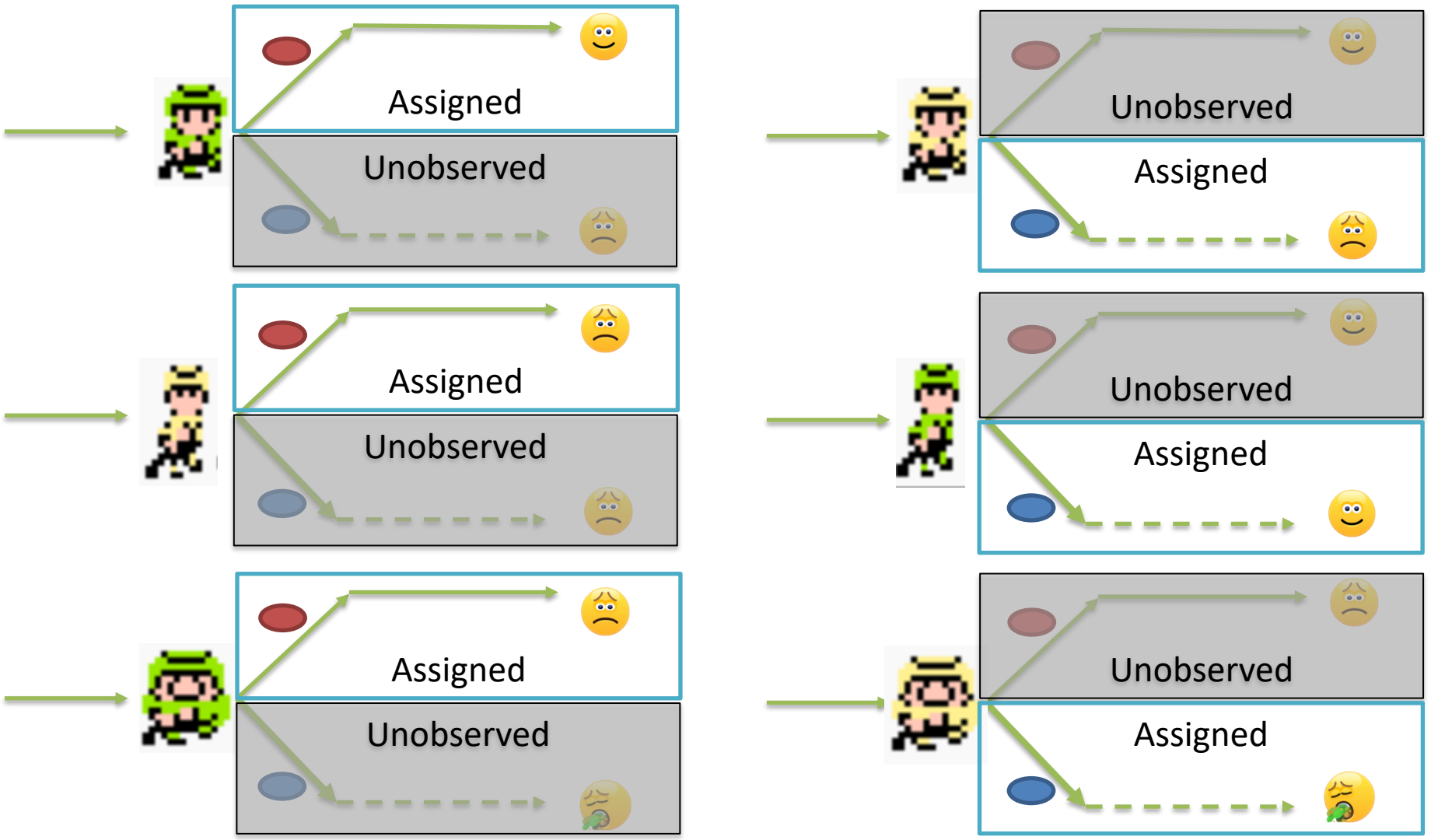


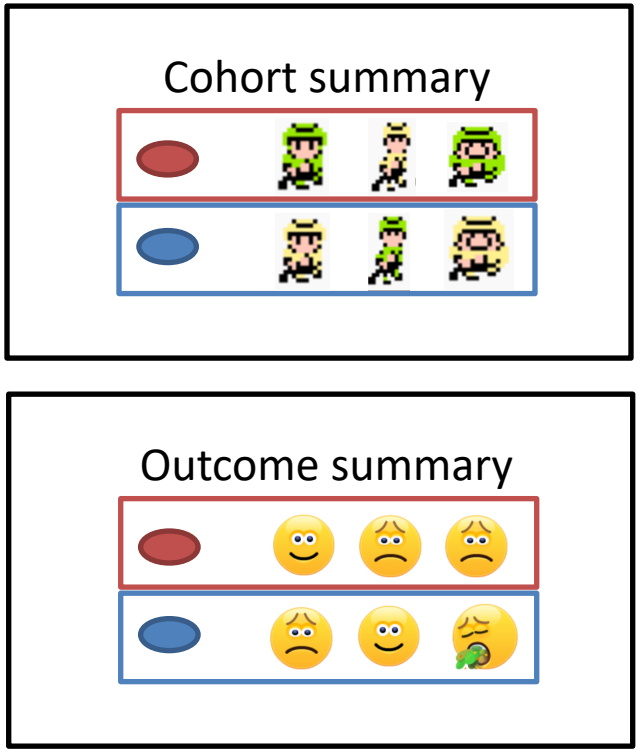
Alas, we don't have a Delorean…

- What is our *next* best approximation?

- Instead of studying the same population under both decision options, let's define a larger population and randomly assign one treatment to each person, then compare outcomes between the two cohorts…

Randomized treatment assignment to approximate counterfactual outcomes





- Randomization allows for assumption that persons assigned to target cohort are exchangeable at baseline with persons assigned to comparator cohort

Alas, we can't randomize…

- What is our *next, next* best approximation?

- Define a larger population, observe the treatment choices that were made, then compare outcomes:
  - Between persons who made different choices (comparative cohort design)

  OR

  - Within persons during time periods with different exposure status (self-controlled designs)

How does Epidemiology define a comparative cohort study?

…it depends on what Epidemiology textbook you read…

"In a retrospective cohort study…the investigator identified the cohort of individu
their su
recent

"Cohort studies are studies that identify subsets of a defined population and follow them over time, looking for differences in their outcome. Cohort studies generally compare exposed patients to unexposed patients, although they can also be used to compare one exposure to another."
--Strom, Pharmacoepidemiology, 2005

"In
has experienced the outcome of interest, but all of whom could experience
it…On
charac
people

"In the coho
identified. I
incidence of
ascertained
--S

"In the paradigmatic cohort study, the investigator defines two or more groups of people that are free of disease and that differ according to the extent of their exposure to a potential cause of disease. These groups are referred to as the study cohorts. When two groups are studies, one is usually though of as the exposed or index cohort – those individuals who have experienced the putative causal event or condition – and the other is then thought of as the unexposed or reference cohort."
--Rothman, Modern Epidemiology, 2008

OHDSI's definition of 'cohort'

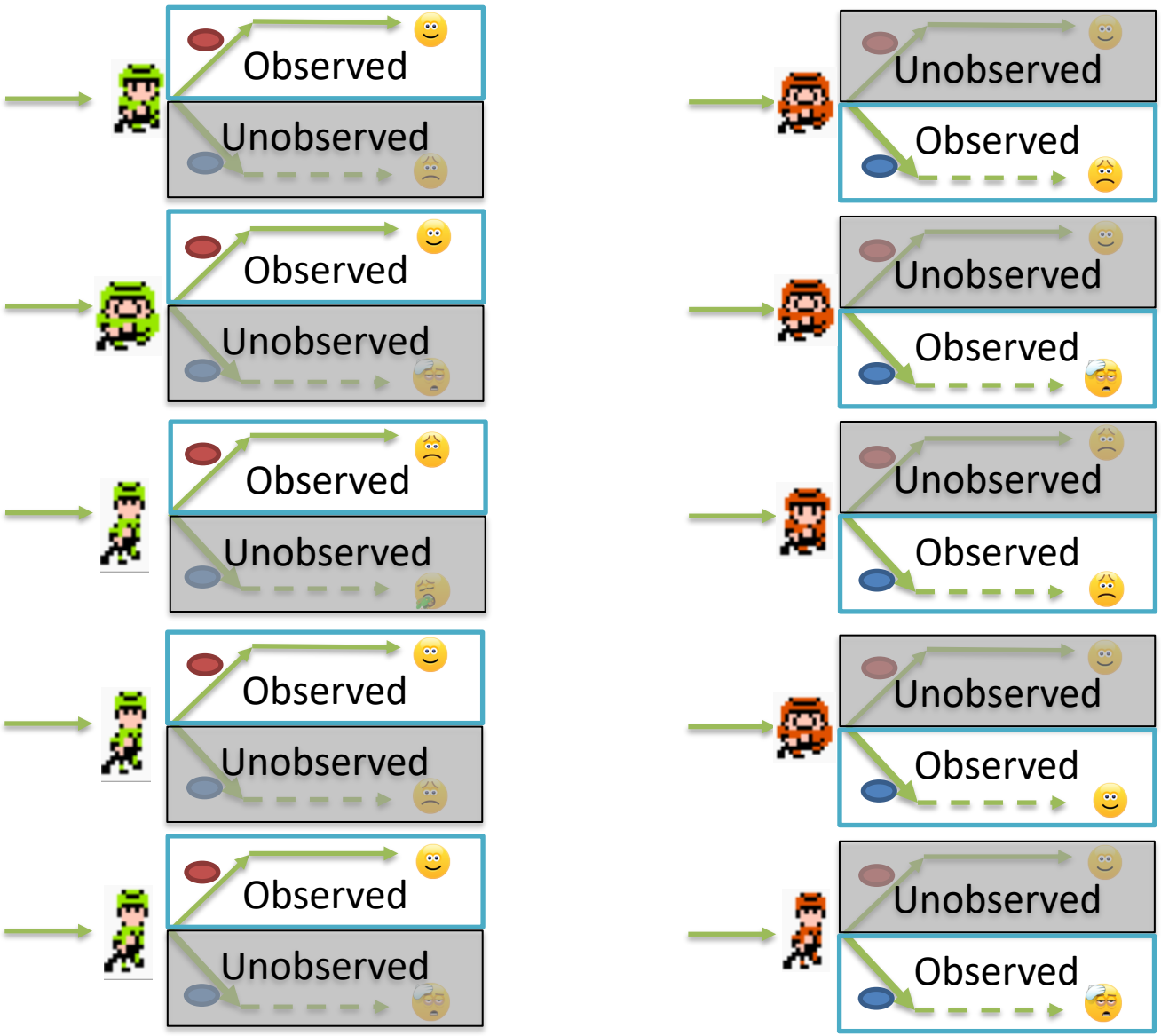
Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

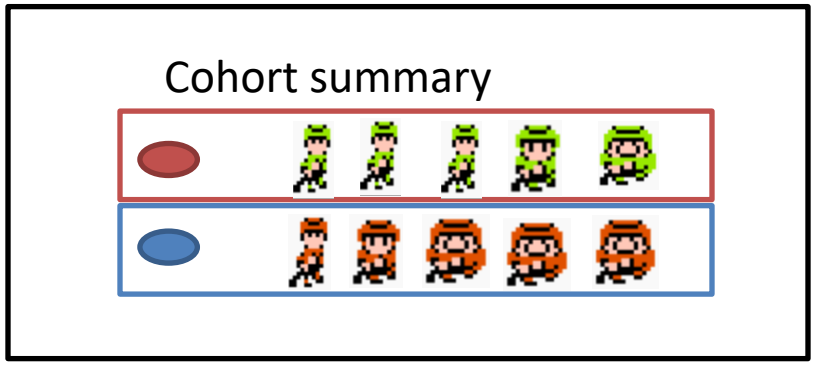
Objective consequences based on this cohort definition:

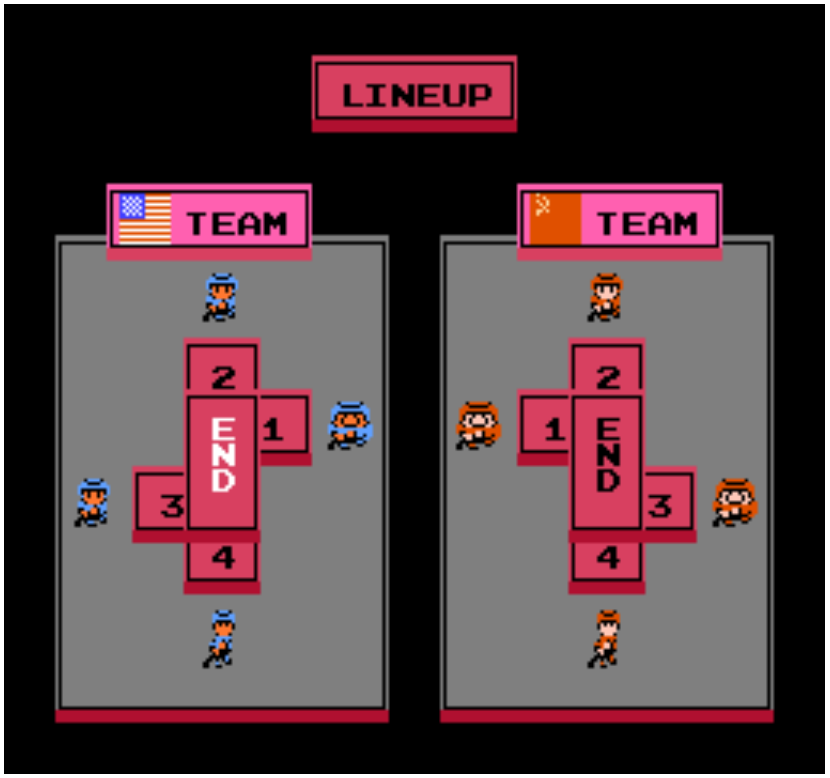
- One person may belong to multiple cohorts
- One person may belong to the same cohort at multiple different time periods
- One person may not belong to the same cohort multiple times during the same period of time
- One cohort may have zero or more members
- A codeset is NOT a cohort…

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

An observational comparative cohort design to approximate counterfactual outcomes







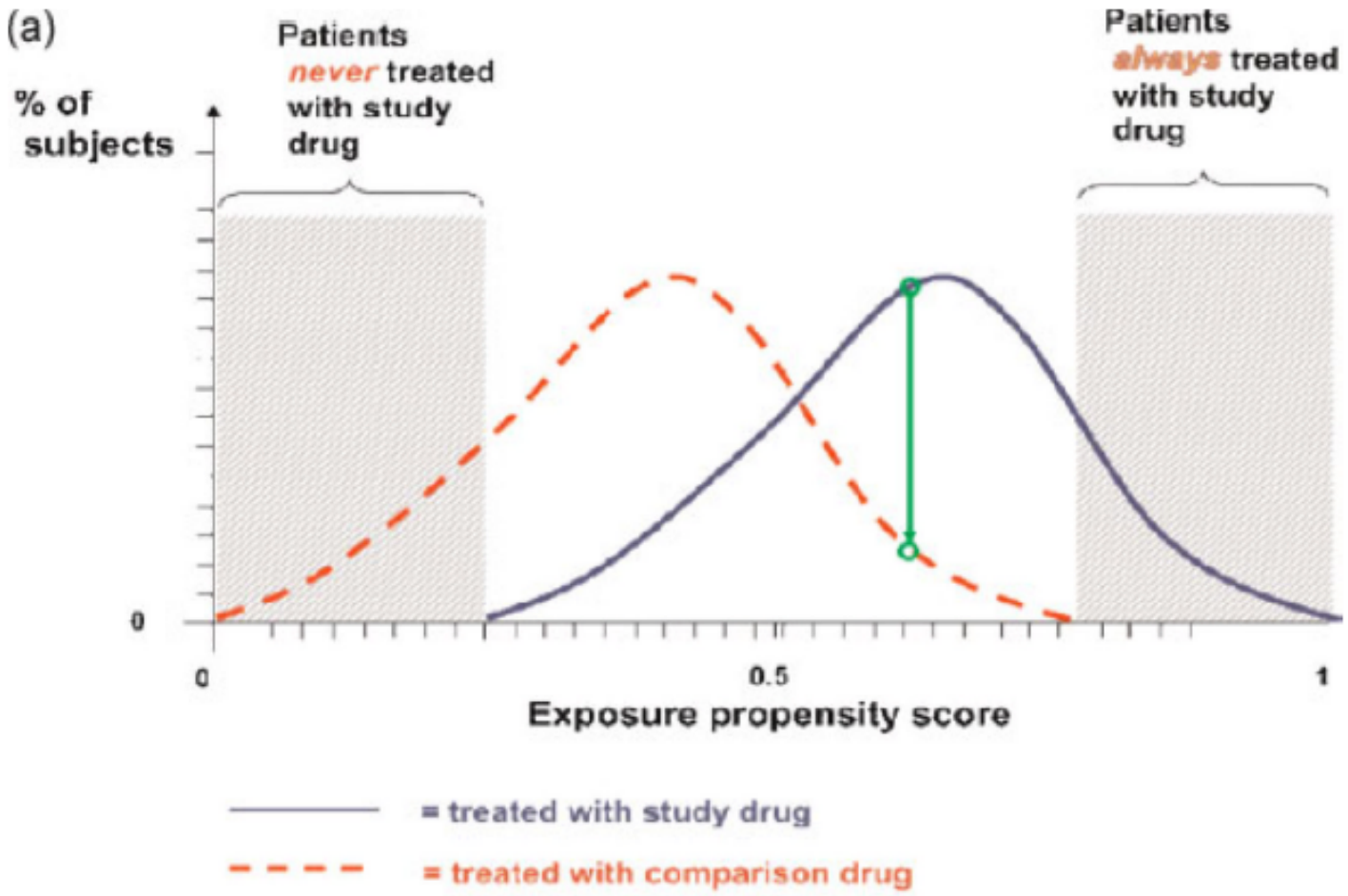
- Exchangeability assumption may be violated if there is reason for treatment choice...and there often is

Propensity score introduction

- $e(x) = Pr(Z=1|x)$
  - Z is treatment assignment
  - x is a set of all covariates at the time of treatment assignment
- Propensity score = probability of belonging to the target cohort vs. the comparator cohort, given the baseline covariates
- 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

Intuition around propensity score balance



Schneeweiss. PDS 2011

"Five reasons to use propensity score in pharmacoepidemiology"

- Theoretical advantages
  - Confounding by indication is the primary threat to validity, PS focuses directly on indications for use and non-use of drug under study
- Value of propensity scores for matching or trimming the population
  - Eliminate 'uncomparable' controls without assumptions of linear relationship between PS and outcome
- Improved estimation with few outcomes
  - PS allows matching on one scalar value rather than needing degrees of freedom for all covariates
- Propensity score by treatment interactions
  - PS enables exploration of patient-level heterogeneity in response
- Propensity score calibration to correct for measurement error

Glynn et al, BCPT 2006

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.

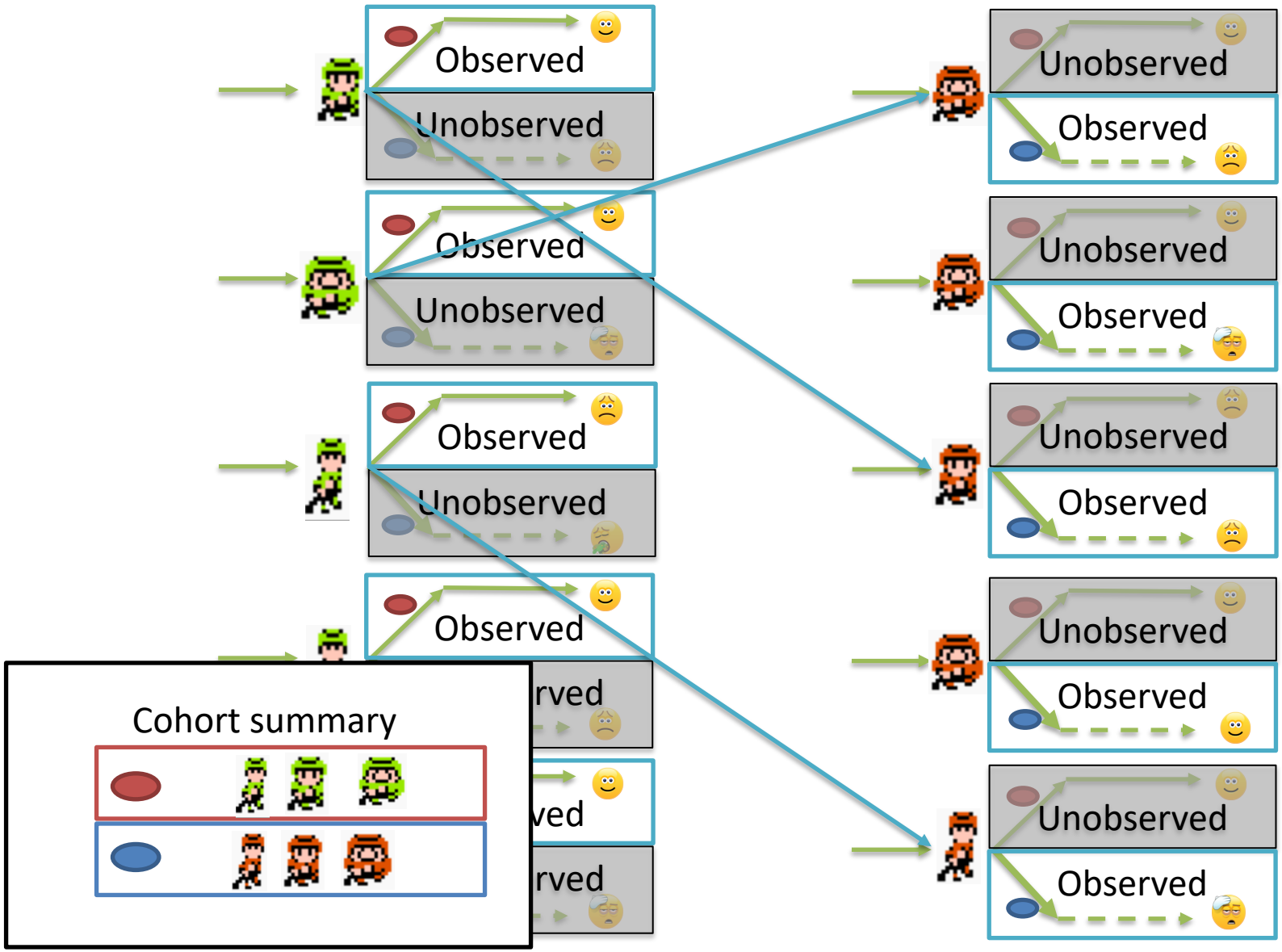
* E: exposure

Not generally recommended

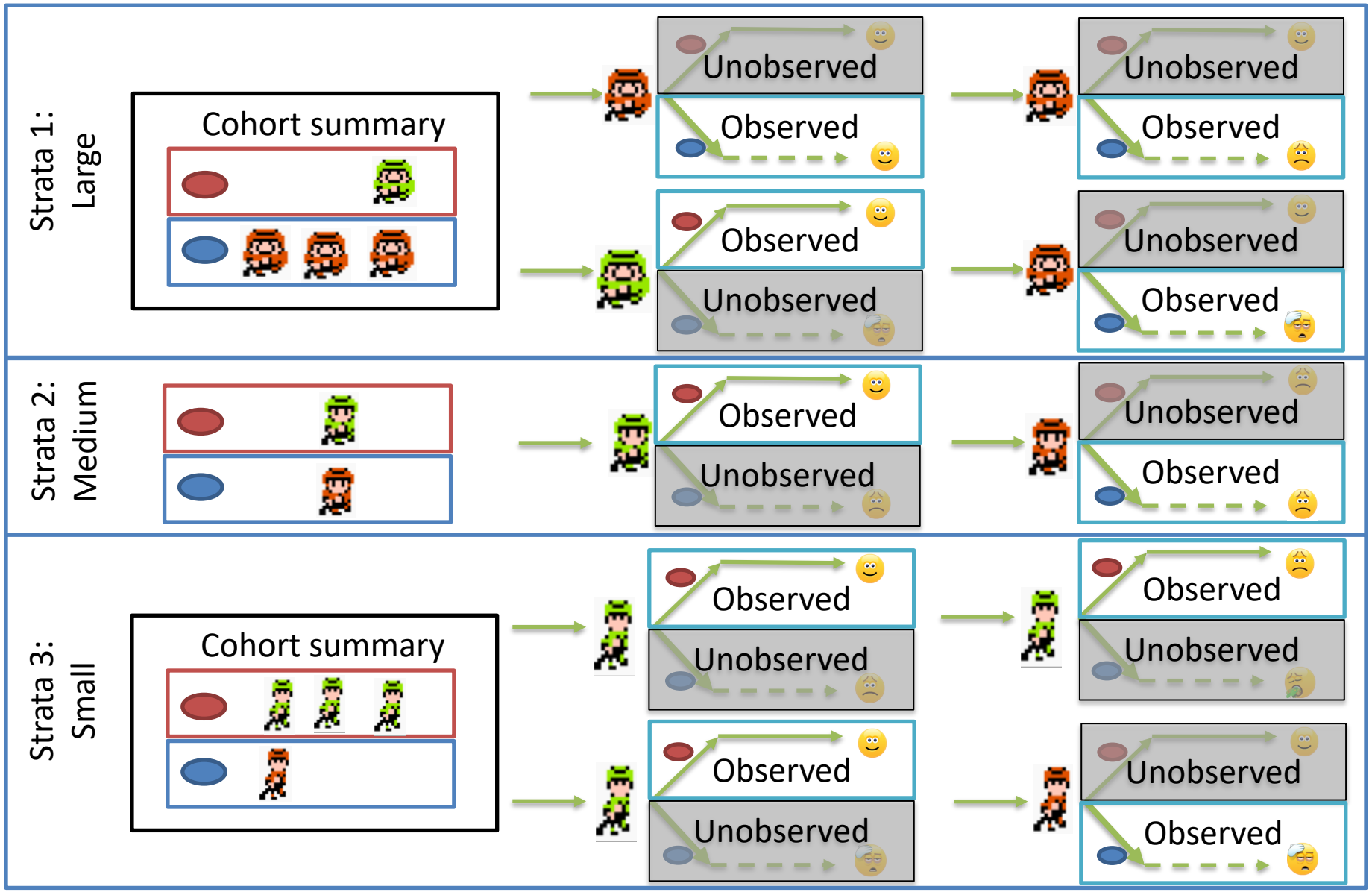
Fully implemented in OHDSI CohortMethod R package

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

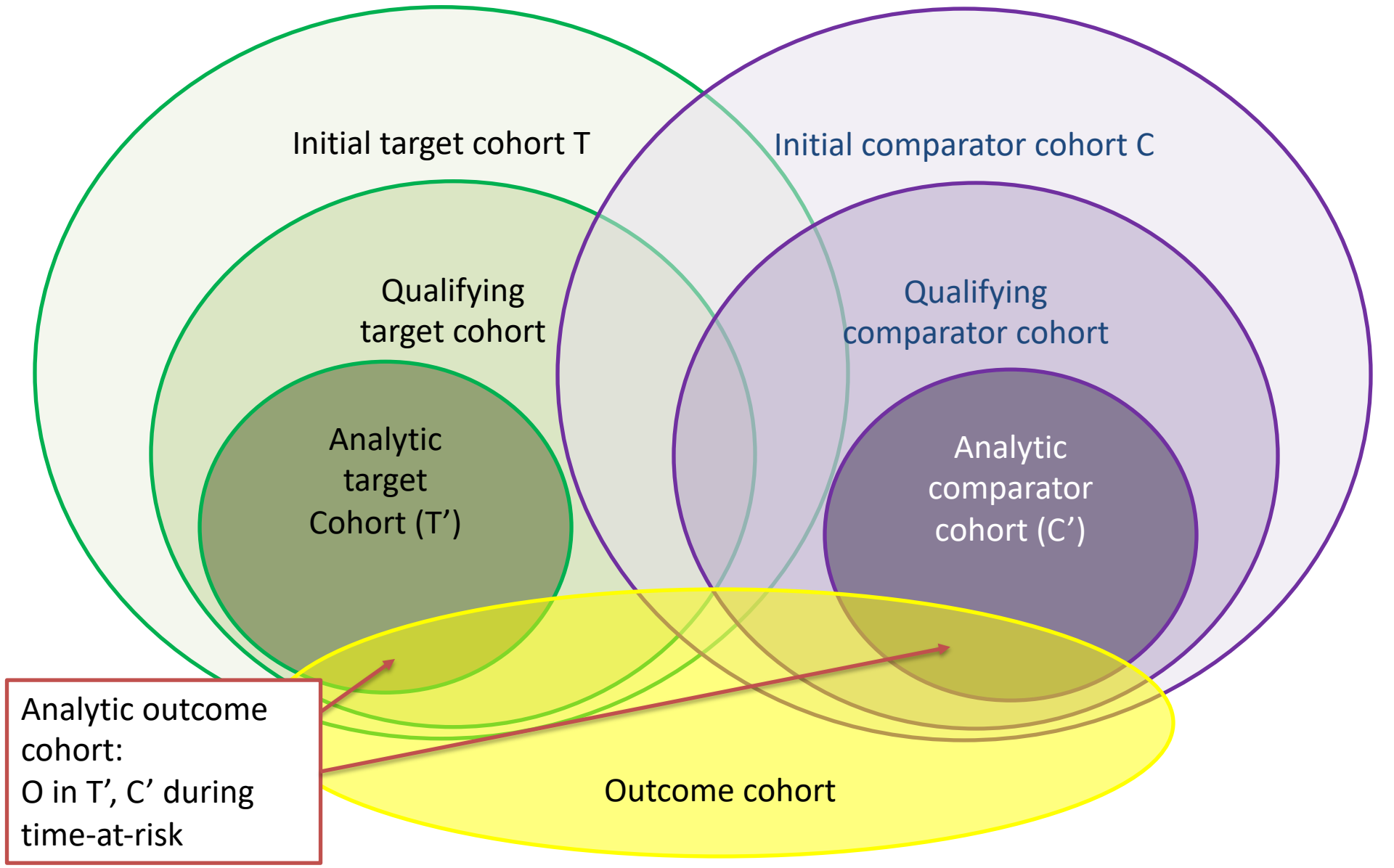
Matching as a strategy to adjust for baseline covariate imbalance



Stratification 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

When designing or reviewing a study, ask yourself:

Input parameter	Design choice
Target cohort (T)	
Comparator cohort (C)	
Outcome cohort (O)	
Time-at-risk	
Model specification	

Exercise 1

- Define your own problem

Break

Exercise 2

- Apply the framework to a published paper

Observational study design
Part #2

Patrick Ryan, PhD

Columbia University

Janssen Research and Development

Design an observational study like you would a randomized trial

American Journal of Epidemiology

Vol. 183, No. 8
DOI: 10.1093/aje/kwv254
Advance Access publication:
March 18, 2016

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

Initially submitted December 9, 2014; accepted for publication September 8, 2015.

Ideally, questions
and conducted rand
tional data. Causal i
a randomized expe
the goal is to guide
with respect to how
research using big
paring the effects o
for the criticism of

big data; causal inf

an appropriately designed
ment, we analyze observa-
ed as an attempt to emulate
question of interest. When
l data need to be evaluated
comparative effectiveness
nterfactual theory for com-
vides a structured process
s.

Protocol components to emulate:

- Eligibility criteria
- Treatment strategies
- Assignment procedures
- Follow-up period
- Outcome
- Causal contrasts of interest
- Analysis plan

- Bias = expected value of the error distribution

$$\text{BIAS}[\hat{\theta}] = E[\hat{\theta} - \theta] = E\left[\hat{\theta}\right] - \theta$$

where θ = true value, $\hat{\theta}$ = estimate of θ

- Mean squared error = metric to evaluate the quality of an estimator, accounting for both random and systematic error

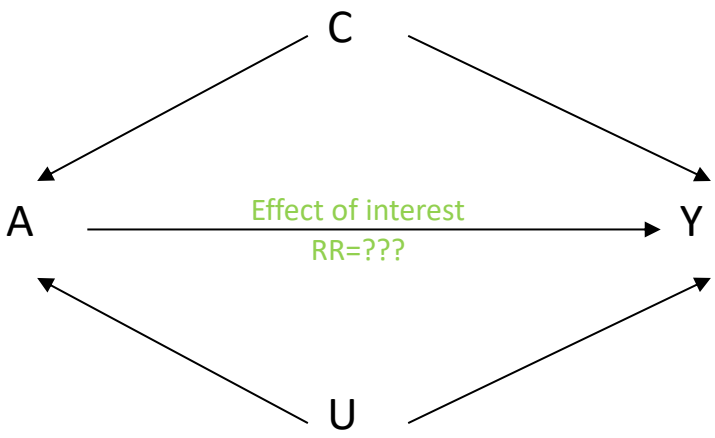
$$\text{MSE}[\hat{\theta}] = E[(\hat{\theta} - \theta)^2] = (\text{BIAS}\left[\hat{\theta}\right])^2 + Var[\hat{\theta}]$$

As studies increase in sample size, random error converges to 0 but systematic error still persists!

Types of systematic error

- Confounding
- Misclassification (Measurement error)
- Selection bias (generalizability)

Confounding



Challenge:
Producing an 'unconfounded' estimate relies on (empirically untestable) assumption that

1) all confounders were observable, and properly modeled in the design or analysis, and
2) no unobserved factors are associated with both exposure and outcome

A=exposure
Y=outcome

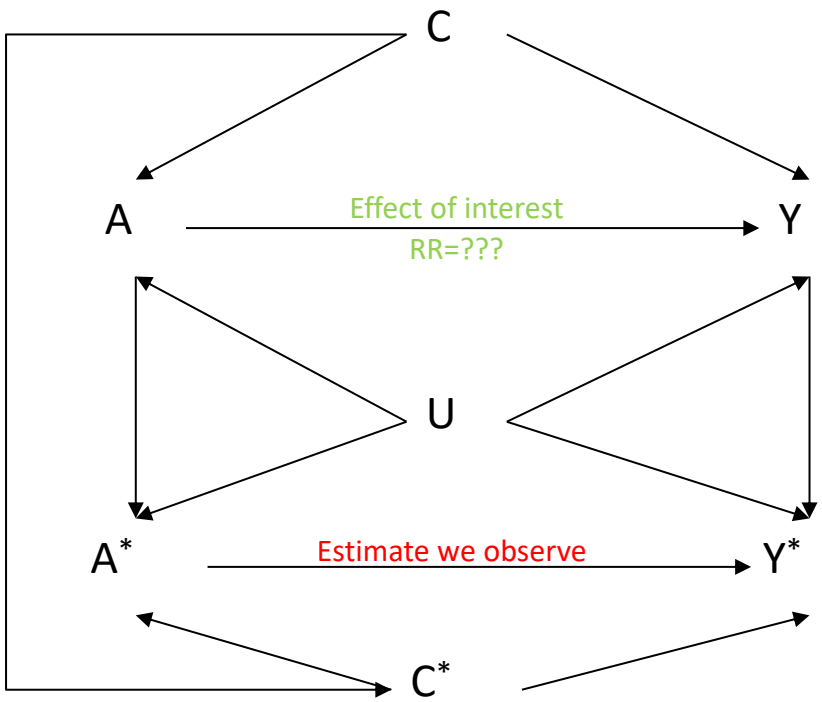
C = observed and modeled confounder

U = unobserved or mismodeled confounder

How do you assess confounding?

- PS distribution
- Covariate balance

Misclassification (measurement error)



Challenge:
All observations are imperfect proxies for true patient status. Misclassification error can exist for all exposures, outcomes and covariates, but is generally unknown or not properly estimated (via sensitivity and specificity), and is rarely formally integrated into effect estimation.

A^*=proxy for exposure
Y^*=proxy for outcome
C^* = proxy for observed confounder

A=exposure
Y=outcome

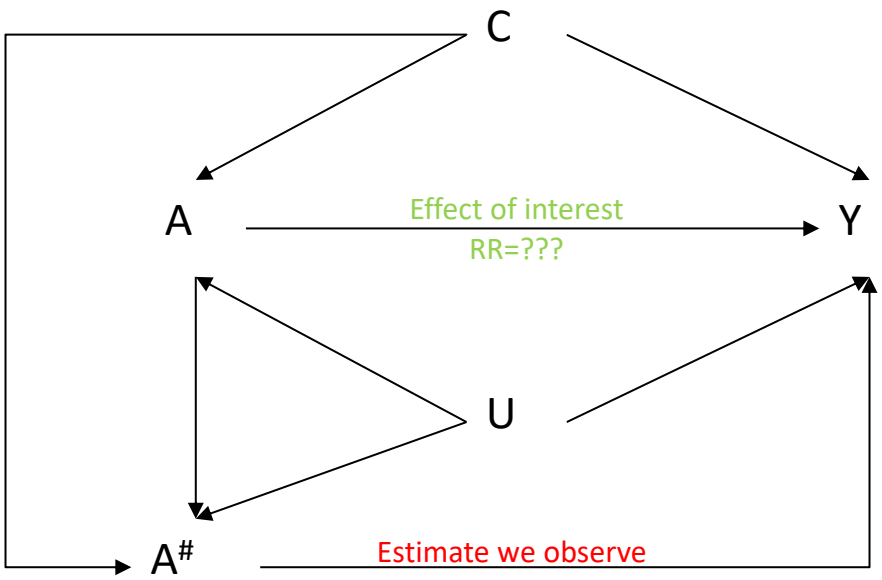
C = observed and modeled confounder

U = unobserved or mismodeled confounder

How do you assess measurement error?

- Covariate summary for exposures
- Operating characteristics for outcome phenotype
  - Sensitivity
  - Specificity
  - Positive predictive value

Selection bias and generalizability



Challenge:
A database is a non-random sample of an underlying population. A cohort is a non-random sample of the database. Study design and analysis decisions may further restrict the cohort composition. Selection bias is rarely evaluated and often empirically untestable.

$A^{\#}$=non-random sample of exposure

A=exposure
Y=outcome

C = observed and modeled confounder

U = unobserved or mismodeled confounder

How do you assess selection bias?

- Attrition table
- Covariate summary (compare before to after)

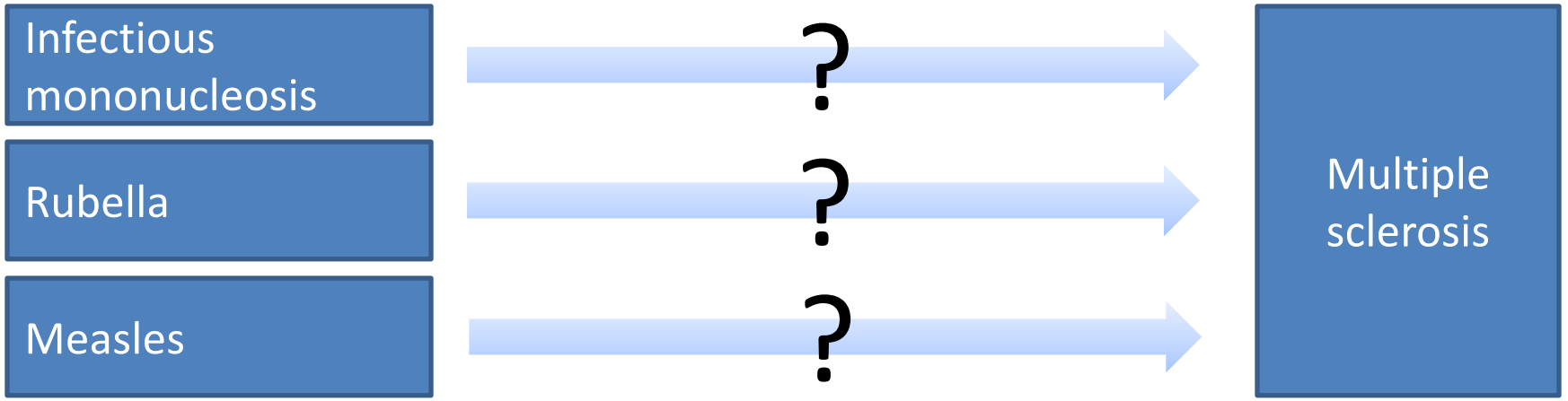
What can we do to address these challenges?

- Think really hard during study design and hope we get it right
- Equivocate in our summary of findings with a paragraph in the Discussion that reads:
  - "This study has several limitations. First, since this study relied on claims data, we had no data on <unobserved confounders>. Second, while we adjusted for <observed confounders>, residual confounding cannot be ruled out. Third, there is a potential for outcome misclassification… Fourth, there is a potential for duplicate person-years between <databases>. Lastly, as the mean follow-up was <short>, long-term effects may need to be further examined." (Kim et al., Arthritis & Rheumatology, 2017)
-
-

What can we do to address these challenges?

- Think really hard during study design and hope we get it right
- Equivocate in our summary of findings with a paragraph in the Discussion that reads:
  - "This study has several limitations. First, since this study relied on claims data, we had no data on <unobserved confounders>. Second, while we adjusted for <observed confounders>, residual confounding cannot be ruled out. Third, there is a potential for outcome misclassification… Fourth, there is a potential for duplicate person-years between <databases>. Lastly, as the mean follow-up was <short>, long-term effects may need to be further examined." (Kim et al., Arthritis & Rheumatology, 2017)
- Perform diagnostic analyses that attempt to detect if residual error may still be present
- Quantify magnitude of residual error and calibrate statistics

Examples of negative controls



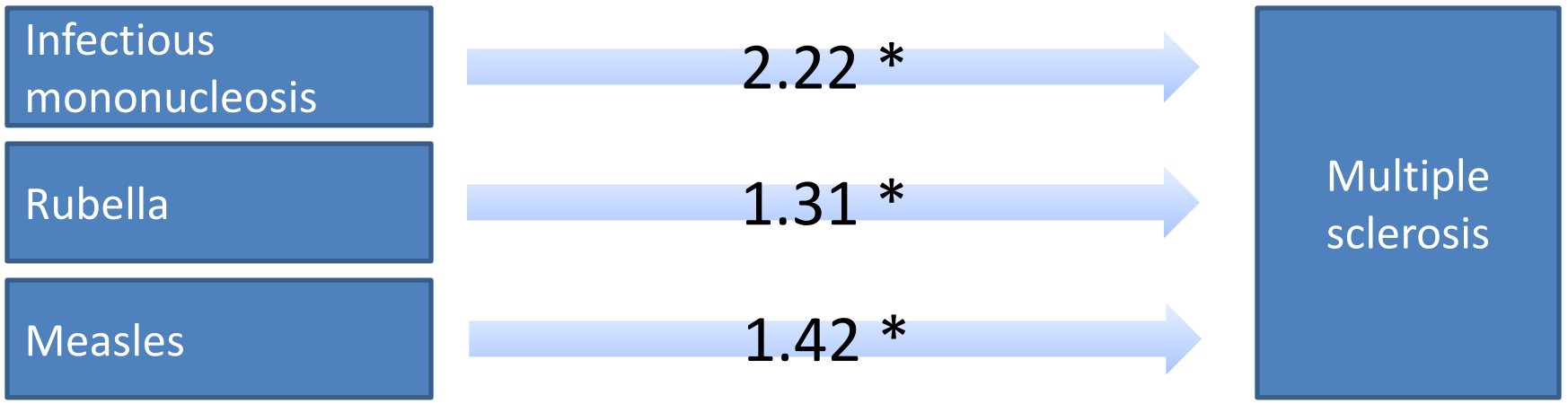
RESEARCH PAPER

Multiple Sclerosis 2008; **14**: 307–313

Selective association of multiple sclerosis with infectious mononucleosis

BM Zaadstra[1,2], AMJ Chorus[1], S van Buuren[1,3], H Kalsbeek[1] and JM van Noort[4]

Example of a negative control



* P < .05

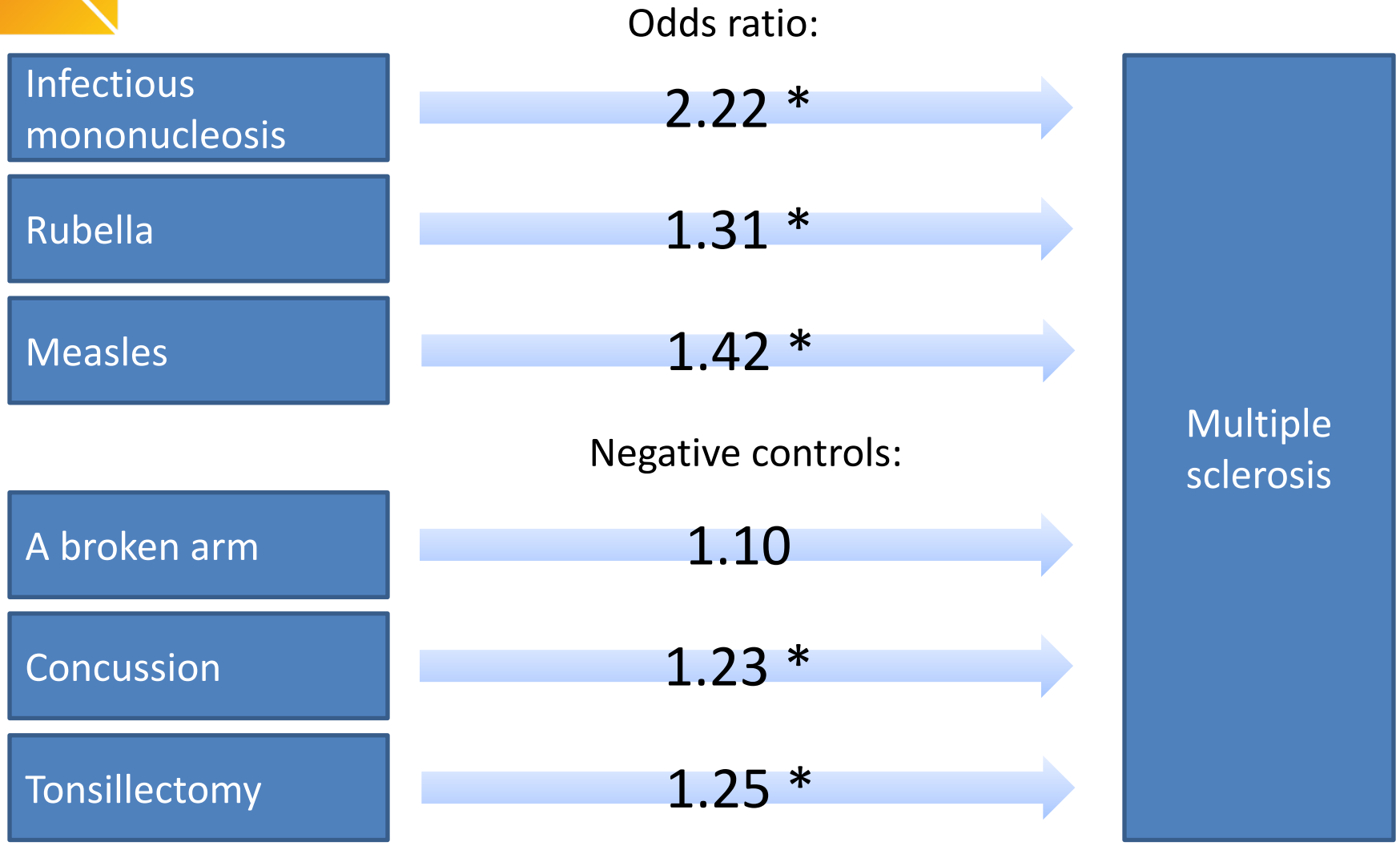
RESEARCH PAPER

Multiple Sclerosis 2008; **14**: 307–313

Selective association of multiple sclerosis with infectious mononucleosis

BM Zaadstra[1,2], AMJ Chorus[1], S van Buuren[1,3], H Kalsbeek[1] and JM van Noort[4]

Example of a negative control



* P < .05

(*Epidemiology* 2010;21: 383–388)

Negative Controls

A Tool for Detecting Confounding and Bias in Observational Studies

Marc Lipsitch,[a,b,c] *Eric Tchetgen Tchetgen,*[a,c,d] *and Ted Cohen*[a,c,e]

Key points:

- 2 types of negative controls:
  - Exposure controls
  - Outcome controls
- "In principle, the measured confounders L of the A-Y relationship need not be causes of N as well, because a properly specified model that accounted for the confounding by L of A-Y would not be misled if such confounding were absent for A-N."
- "In practice, the ideal negative control outcome should be one with incoming arrows as similar as possible to those of Y, including arrows from L"
- "In observational settings, the comparability between exposure A and negative control exposure B will be only approximate"
- "Subject matter knowledge is required for the choice of negative controls"

VIEWPOINT

JAMA®

Prespecified Falsification End Points

Can They Validate True Observational Associations?

Vinay Prasad, MD

Anupam B. Jena, MD, PhD

mur fractures and 716 atypical fractures.[5] This analysis demonstrated an increased risk of atypical fractures associated with bisphosphonate use and was validated by another large

A so
be
ing
ses
have failures
solutions to
have been su
ord not only
ducted.[2]

Key points:

- "A falsification hypothesis is a claim, distinct from the one being tested, that researchers believe is highly unlikely to be causally related to the intervention in question."
- "Falsification analysis can be operationalized by asking investigators to specify implausible hypotheses up front and then testing those claims using statistical methods similar to those used in the primary analysis."
- "Although no published recommendations exist, standardized falsification analyses with 3 or 4 prespecified or highly prevalent disease outcomes may help strengthen the validity of observational studies"

American Journal of Epidemiology

Vol. 179, No. 5
DOI: 10.1093/aje/kwt303
Advance Access publication:
December 20, 2013

Practice of Epidemiology

The Control Outcome Calibration Approach for Causal Inference With Unobserved Confounding

Eric Tchetgen Tchetgen*

* Correspondence to Dr. Eric Tchetgen Tchetgen, Department of Biostatistics, Harvard University, 677 Huntington Avenue, Kresge, Room 822, Boston, MA 02115 (e-mail: etchetge@hsph.harvard.edu).

Initially sub

Key points:

- "The extent to which an analysis may reveal unobserved confounding bias relies on the non-empirically verifiable assumption that the negative control outcome is carefully chosen so that it is solely influenced by observed and unobserved confounders of the exposure-outcome relationship in view"
- "We propose to use a negative control outcome not only to detect, but also to correct for unmeasured confounding bias"

BRIEF REPORT

OPEN

(*Epidemiology* 2016;27:637–641)

Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies

Benjamin F. Arnold, Ayse Ercumen, Jade Benjamin-Chung, and John M. Colford, Jr

Selection B
Structure

TABLE. Examples of Studies that Have Used Negative Controls to Detect Selection or Measurement Bias Following Bias Structures in Figures 1 and 2

Example	Bias Structure	Design	Exposure (*A*)	Outcome (*Y*)	Potential Source of Bias	Negative Control*
Selection bias						

rol
)

Negative Control
Outcomes (N_Y)

A

B

C

D

FIGURE 1. Simplified causal dia
and outcomes (N_Y). In all four s
and outcome *Y*, (B) cause of ex
and cause of outcome U_Y.

outcome itself

Key points:

- Negative controls demonstrated to detect 3 primary sources of systematic error:
  - Confounding
  - Selection bias
  - Measurement bias
- Negative controls shown to have utility across many different study types: observational vs. RCT; prospective vs. retrospective; case control vs. cohort
- "The ability of a negative control to adequately detect bias ultimately relies on the plausibility of (often untestable) assumptions encoded in its causal diagram"

Special Issue Paper

Statistics in Medicine

Received 4 November 2011, Accepted 28 August 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5620

Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership‡

Table III. Drug-adverse event outcome pairs used as reference set for methods evaluation, with overall drug and outcome counts and expected counts for each pair.

		Angioedema	Aplastic Anemia	Acute Liver Injury	Bleeding	Myocardial Infarction	Hip Fracture	Mortality after MI	Renal Failure	GI Ulcer Hospitalization	
		293,342	236,684	14,779,994	21,611,646	3,170,978	766,402	161,098	1,718,789	2,486,439	Persons with outcome
ACE Inhibitors	20,788,283	**34,249**	33,664				117,631			319,731	
Amphotericin B	11,874	23	29	987			62	82	**149**		
Antibiotics: erythromycins, sulfonamides, tetracyclines	16,089,290		21,306	**1,216,227**	1,783,940	303,832	74,798		163,165		
Antiepileptics: cabamazepine, phenytoin	1,431,777	2,282	**2,222**					2,193	17,606	20,560	
Benzodiazepines	19,619,014	29,600	27,552	1,489,451	2,258,372	400,602	**98,014**		216,380		
Beta blockers	17,380,612	28,653	28,381	1,351,351			98,914		240,375	265,769	
Bisphosphonates: alendronate	3,606,131		6,258	274,928		90,835			49,033	**61,589**	
Tricyclic antidepressants	4,977,104		7,223	385,064	581,348	**104,574**			57,875		
Typical antipsychotics	2,347,603					**53,092**			29,115	35,576	
Warfarin	4,743,694	8,179	9,266		**636,010**		34,066	9,191	74,286		
	Persons exposed										

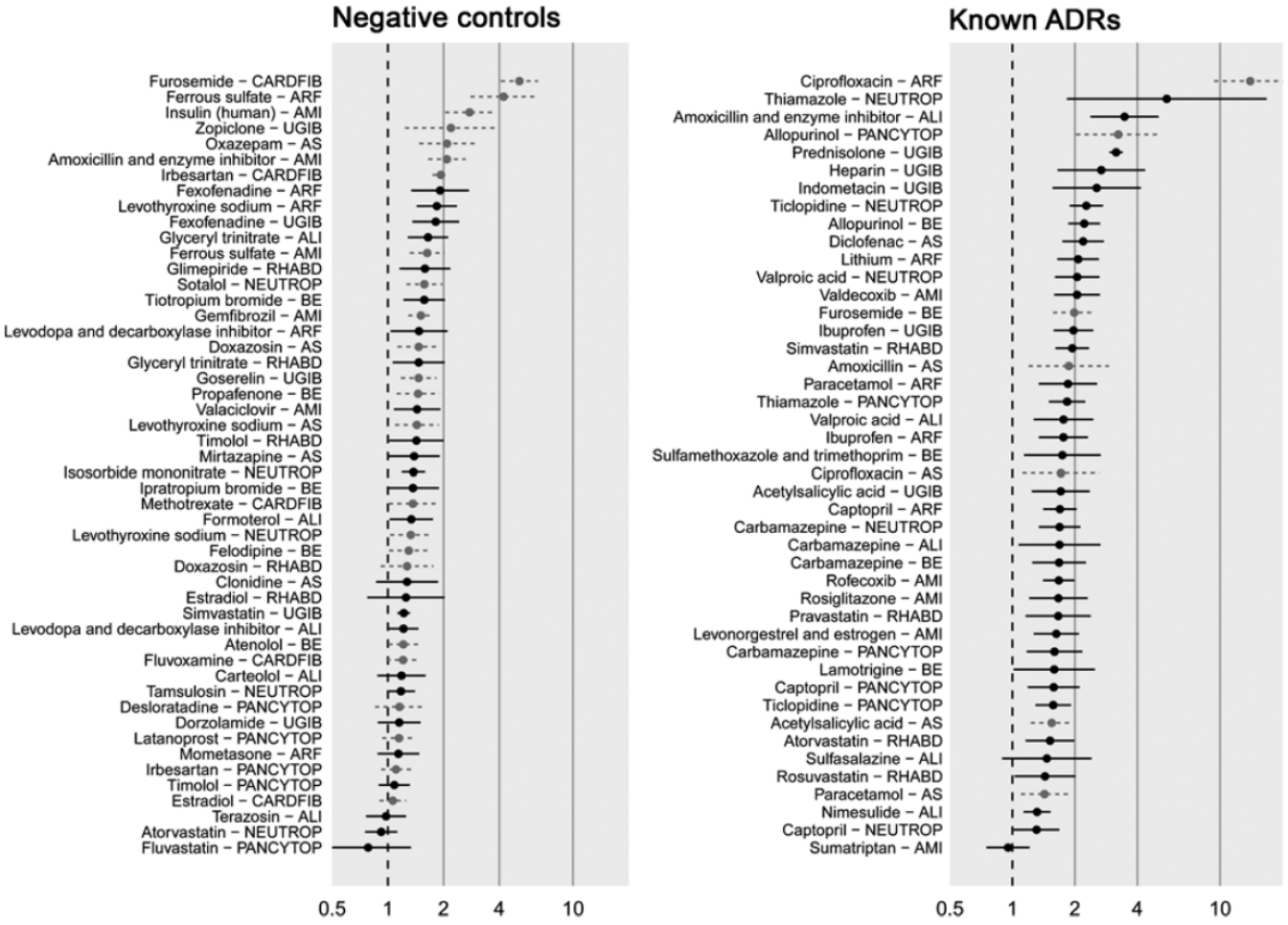
Positive controls (n=9)
Negative controls (n=44)

Medical Care • Volume 50, Number 10, October 2012

Using Electronic Health Care Records for Drug Safety Signal Detection

A Comparative Evaluation of Statistical Methods

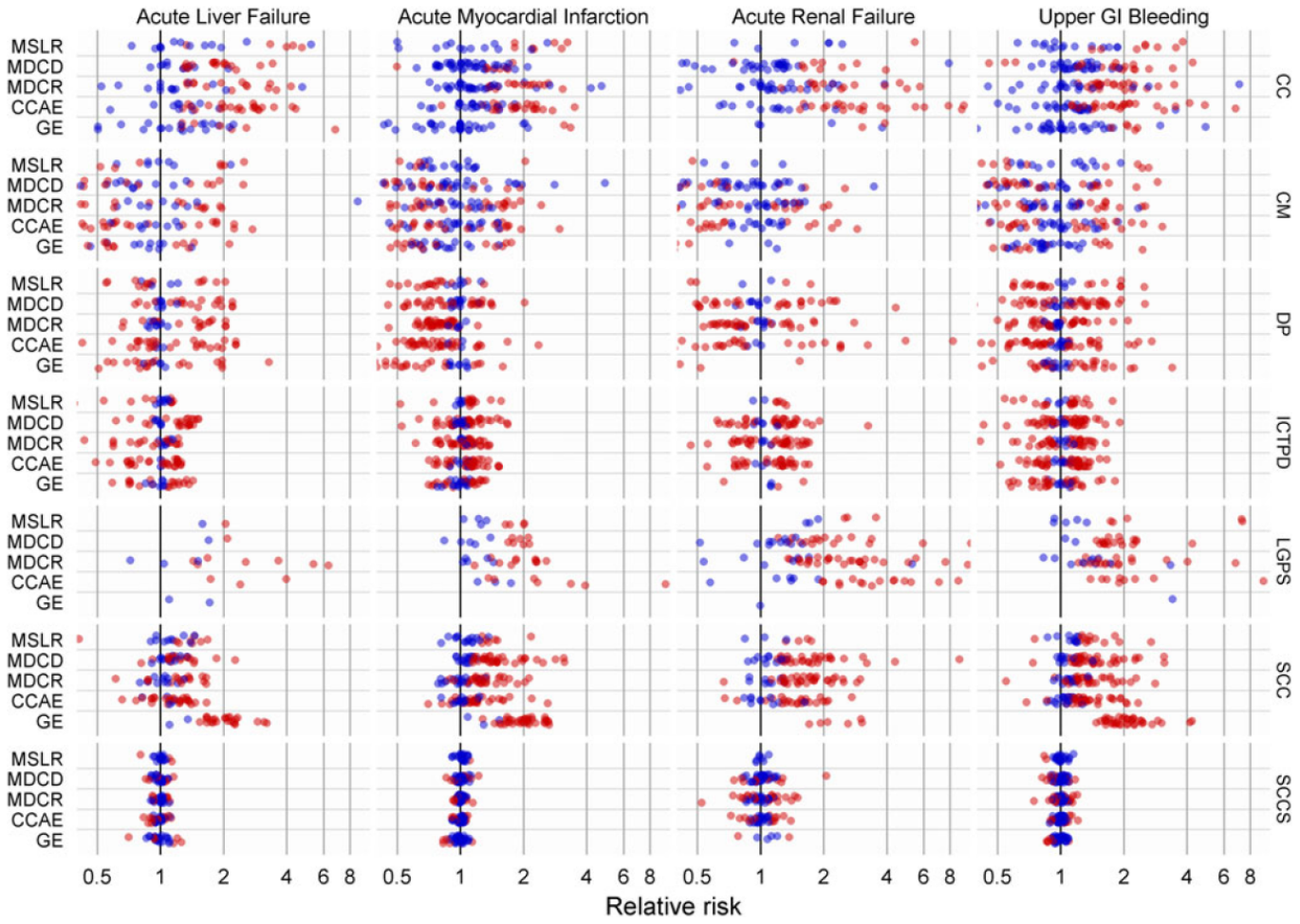
K
David
Loren
#
)*



Drug Saf (2013) 36 (Suppl 1):S143–S158
DOI 10.1007/s40264-013-0108-9

ORIGINAL RESEARCH ARTICLE

A Comparison of the Empirical Performance of Methods for a Risk Identification System



Research Article

Statistics in Medicine

Received 12 November 2012, Accepted 3 July 2013 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5925

Interpreting observational studies: why empirical calibration is needed to correct *p*-values

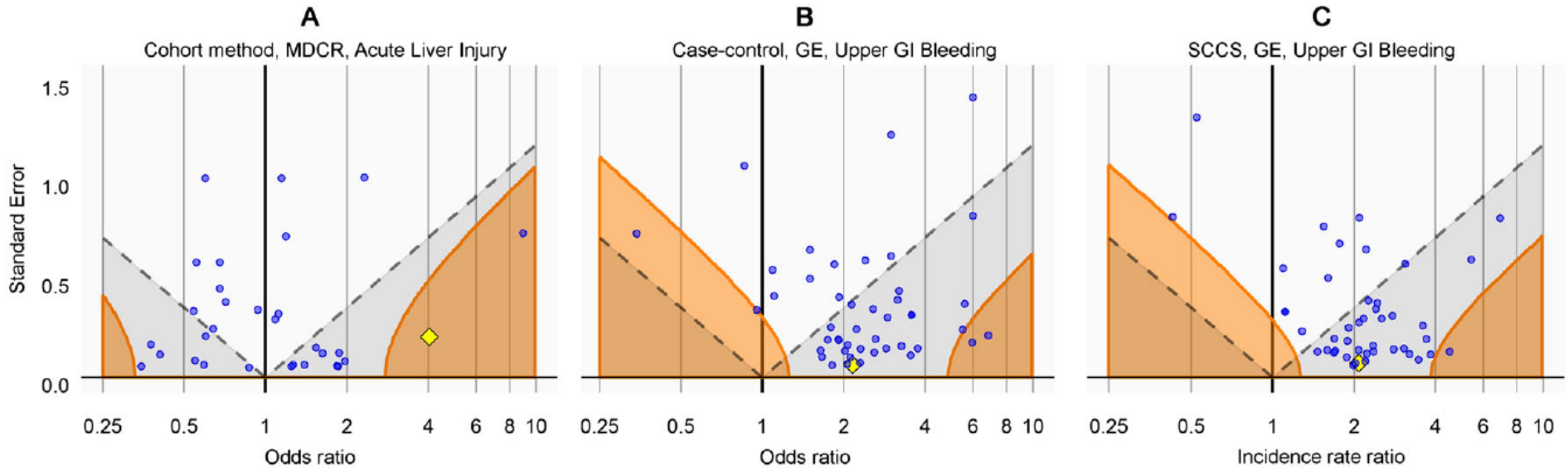


Figure 3. Traditional and calibrated significance testing. Estimates below the dashed line (gray area) have $p < 0.05$ using traditional p-value calculation. Estimates in the orange areas have $p < 0.05$ using the calibrated p-value calculation. Blue dots indicate negative controls, and the yellow diamond indicates the drugs of interest: isoniazid (A) and sertraline (B and C).

PNAS

COLLOQUIUM PAPER

Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data

Martijn J. Schuem

[a]Observational Health
[c]Department of Biome
York, NY 10032; [e]Depa
CA 90095; [g]Departmer
Los Angeles, CA 90095

Edited by Victoria Stod
2017 (received for revi

a,f,g,h

, Titusville, NJ 08560;
ian Hospital, New
lifornia, Los Angeles,
ity of California,

n T. Fiske October 26,

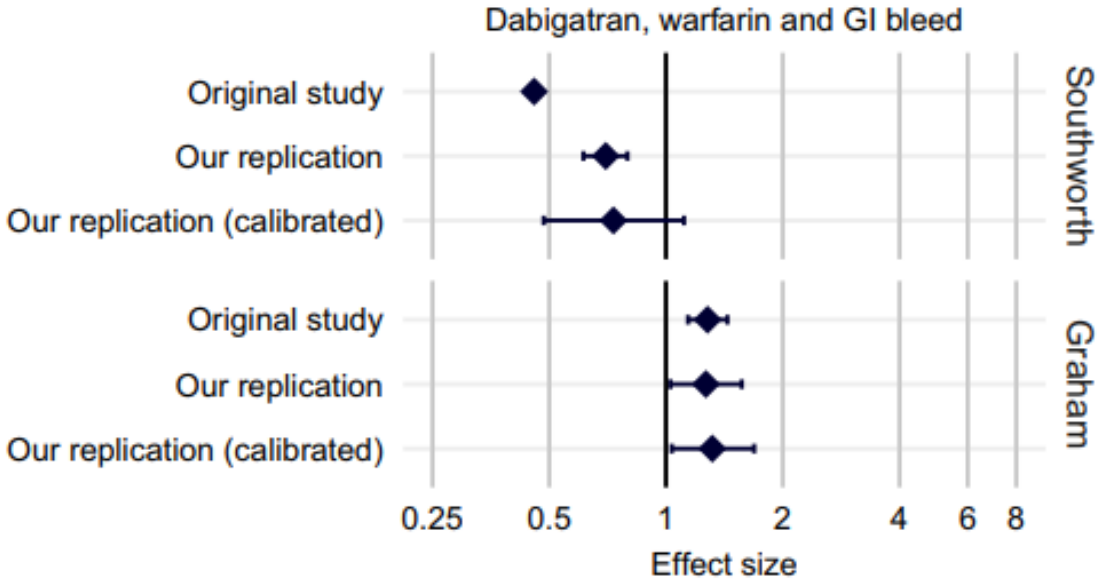


Fig. 5. Estimates from the original studies and our reproduction of the studies by Southworth et al. (12) and Graham et al. (13) both before and after calibration.

Exercise 3

- Evaluate Graham, what did they do to mitigate the threat of systematic error? How do you know they were successful?

Dissecting a cohort study

Marc Suchard

Martijn Schuemie

Patrick Ryan


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Part 1

Exercise #1

Graham *et al.* (2015) Circulation

"Cardiovascular, bleeding and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation"

- Team up into groups of 4
- Identify
  - Target
  - Comparator
  - Outcome
  - Time at risk
  - Model

T and C cohorts

- Elderly (>= 65) Medicare beneficiaries (A, B and D) with nonvalvular atrial fibrillation who initiated therapy with dabigatran (T) or warfarin (C)

Is this correct?

Inclusion criteria

All patients who:

- Have any inpatient or outpatient AF or atrial flutter ICD9 codes
- Filled at least 1 prescription for either drug between Oct 19, 2010 - Dec 31, 2012

Index date: first prescription date

Exclusion criteria

All patients who:

- Have < 6 months of Medicare enrollment before index date
- Were < 65
- Received prior treatment (when?) with NOAC or warfarin
- Were in a skilled nursing facility on index date (why?)
- Were in hospice on index date (why?)
- Had a hospitalization “that extended beyond the index dispensing date”
- Undergoing dialysis (when?)
- Were kidney transplant recipients
- Had diagnoses of valvular disease, DVT, PE, joint

Outcomes

- Stroke
- Major gastrointestinal and intracranial bleeding
- Acute myocardial infarction
- Mortality

Time at risk

- Follow-up starts on index date + 1 and censored at:
  - Medicare disenrollment
  - > 3 day gap in anticoagulant supply
    RX fill for a different anticoagulant
  - Start of hospice
  - Initation of dialysis or kidney transplant
  - Admission to nursing facility
  - End of study


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Part 2

Exercise #2

Go back to Graham's paper

Discuss:

- What threats are there to the validity of the study results?
- How do Graham *et al*. address these threads?

Confounding

- Using propensity score model: Logistic regression with “initiated dabigatran” as outcome and predictors:
  - Sociodemographics
  - Prescriber characteristics
  - Baseline comorbidities
  - “Other potentially relevant variables”
- 1:1 ratio, greedy matching
- Balance assessment via:
  - Standardized mean difference (target: <= 0.1)

Measurement error

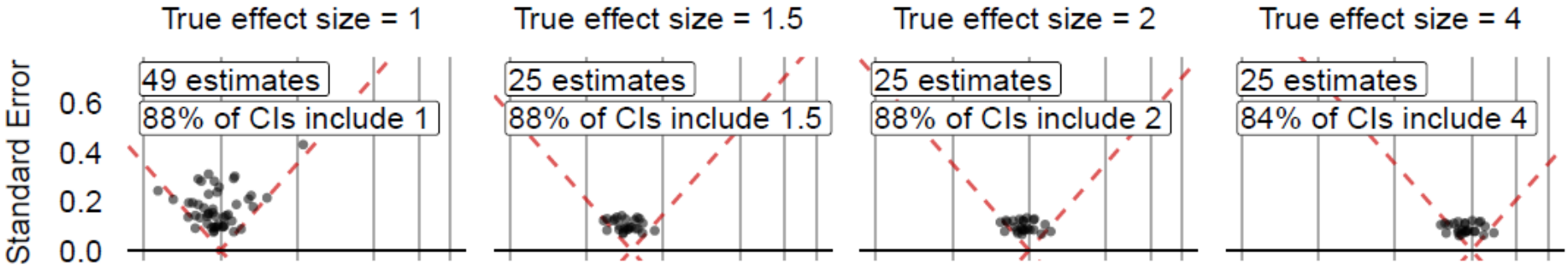
- Sensitivity and specificity of the outcome measures?

Study Outcomes

The primary outcomes were ischemic stroke, major bleeding with specific focus on intracranial and gastrointestinal bleeding, and AMI. Secondary outcomes were all hospitalized bleeding events and mortality. The *International Classification of Diseases, Ninth Revision, Clinical Modification* codes used to define these outcomes are listed in Table II in the online-only Data Supplement. The codes defining ischemic stroke have a positive predictive value (PPV) of 88% to 95%.[18–20] Major bleeding was defined as

Overall systematic error

- Negative controls could show amount of residual bias



Walkthrough of implementing a cohort study using OHDSI tools

Cardiovascular, Bleeding, and Mortality Risks in Elderly Medicare Patients Treated With Dabigatran or Warfarin for Nonvalvular Atrial Fibrillation

David J. Graham, MD, MPH; Marsha E. Reichman, PhD; Michael Wernecke, BA; Rongmei Zhang, PhD; Mary Ross Southworth, PharmD; Mark Levenson, PhD; Ting-Chang Sheu, MPH; Katrina Mott, MHS; Margie R. Goulding, PhD; Monika Houstoun, PharmD, MPH; Thomas E. MaCurdy, PhD; Chris Worrall, BS; Jeffrey A. Kelman, MD, MMSc

Background—The comparative safety of dabigatran versus warfarin for treatment of nonvalvular atrial fibrillation in general practice settings has not been established.

Methods and Results—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134414 patients with 37587 person-years of follow-up, there were 2715 primary outcome events. The hazard ratios (95% confidence intervals) comparing dabigatran with warfarin (reference) were as follows: ischemic stroke, 0.80 (0.67–0.96); intracranial hemorrhage, 0.34 (0.26–0.46); major gastrointestinal bleeding, 1.28 (1.14–1.44); acute myocardial infarction, 0.92 (0.78–1.08); and death, 0.86 (0.77–0.96). In the subgroup treated with dabigatran 75 mg twice daily, there was no difference in risk compared with warfarin for any outcome except intracranial hemorrhage, in which case dabigatran risk was reduced. Most patients treated with dabigatran 75 mg twice daily appeared not to have severe renal impairment, the intended population for this dose. In the dabigatran 150-mg twice daily subgroup, the magnitude of effect for each outcome was greater than in the combined-dose analysis.

Conclusions—In general practice settings, dabigatran was associated with reduced risk of ischemic stroke, intracranial hemorrhage, and death and increased risk of major gastrointestinal hemorrhage compared with warfarin in elderly patients with nonvalvular atrial fibrillation. These associations were most pronounced in patients treated with dabigatran 150 mg twice daily, whereas the association of 75 mg twice daily with study outcomes was indistinguishable from warfarin except for a lower risk of intracranial hemorrhage with dabigatran. (***Circulation***. **2015;131:157-164. DOI: 10.1161/CIRCULATIONAHA.114.012061.**)

Key Words: anticoagulant ■ pharmacoepidemiology ■ safety ■ thrombin inhibitor ■ warfarin

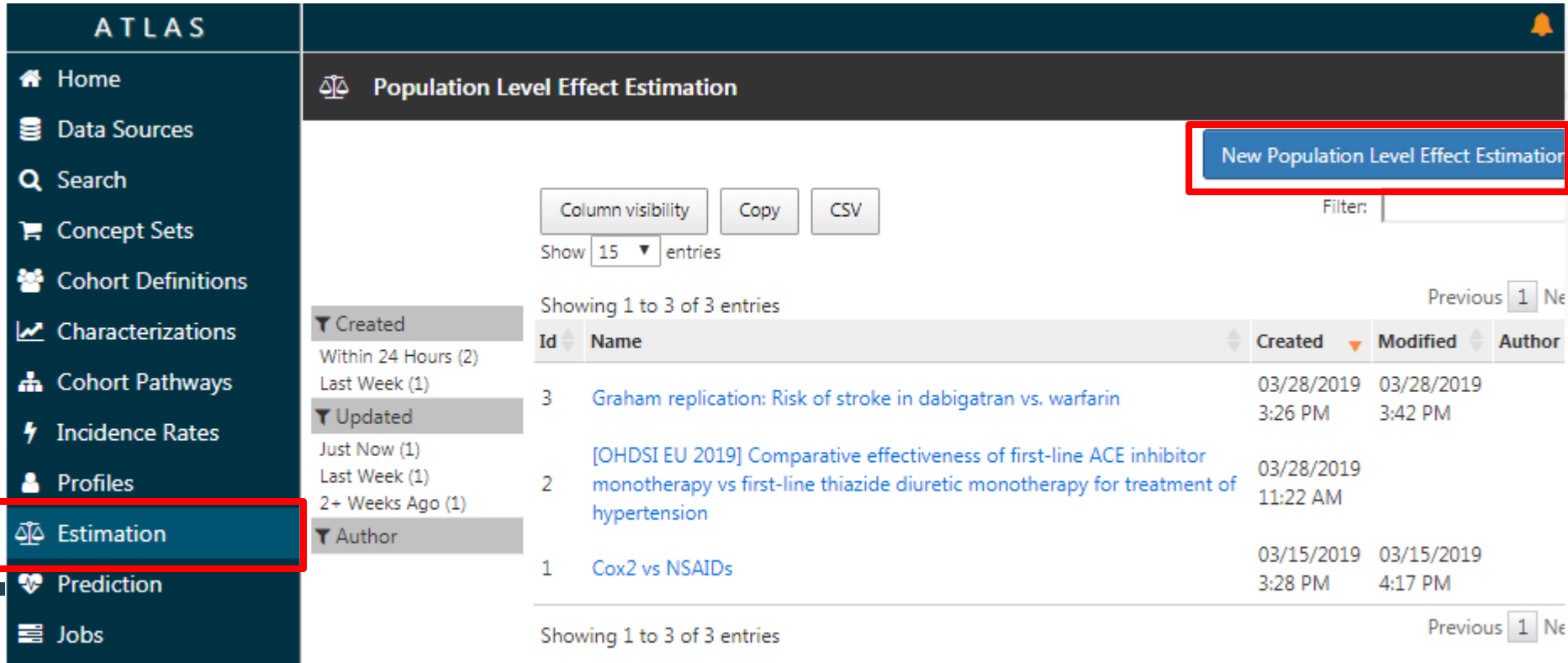
What is the design used by Graham et al?

Input parameter	Design choice
Target cohort (T)	dabigatran new users with prior atrial fibrillation
Comparator cohort (C)	warfarin new users with prior atrial fibrillation
Outcome cohort (O)	Ischemic stroke
Time-at-risk	1 day after cohort start → cohort end
Model specification	1:1 propensity score-matched univariable conditional Cox proportional hazards

Click-a-long

Go to http://ohdsitutorialtest.eu-central-1.elasticbeanstalk.com

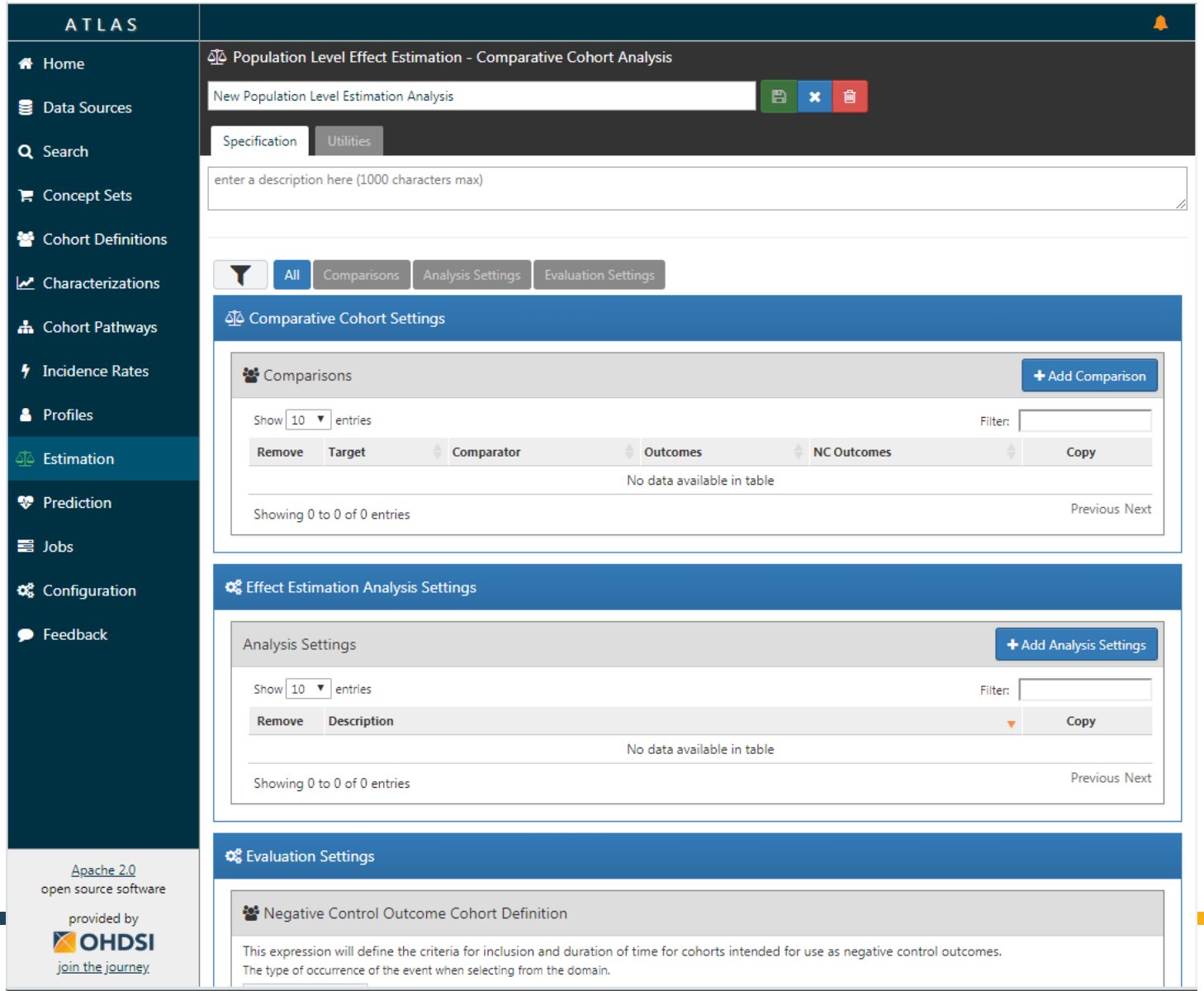
Create your own estimation study!



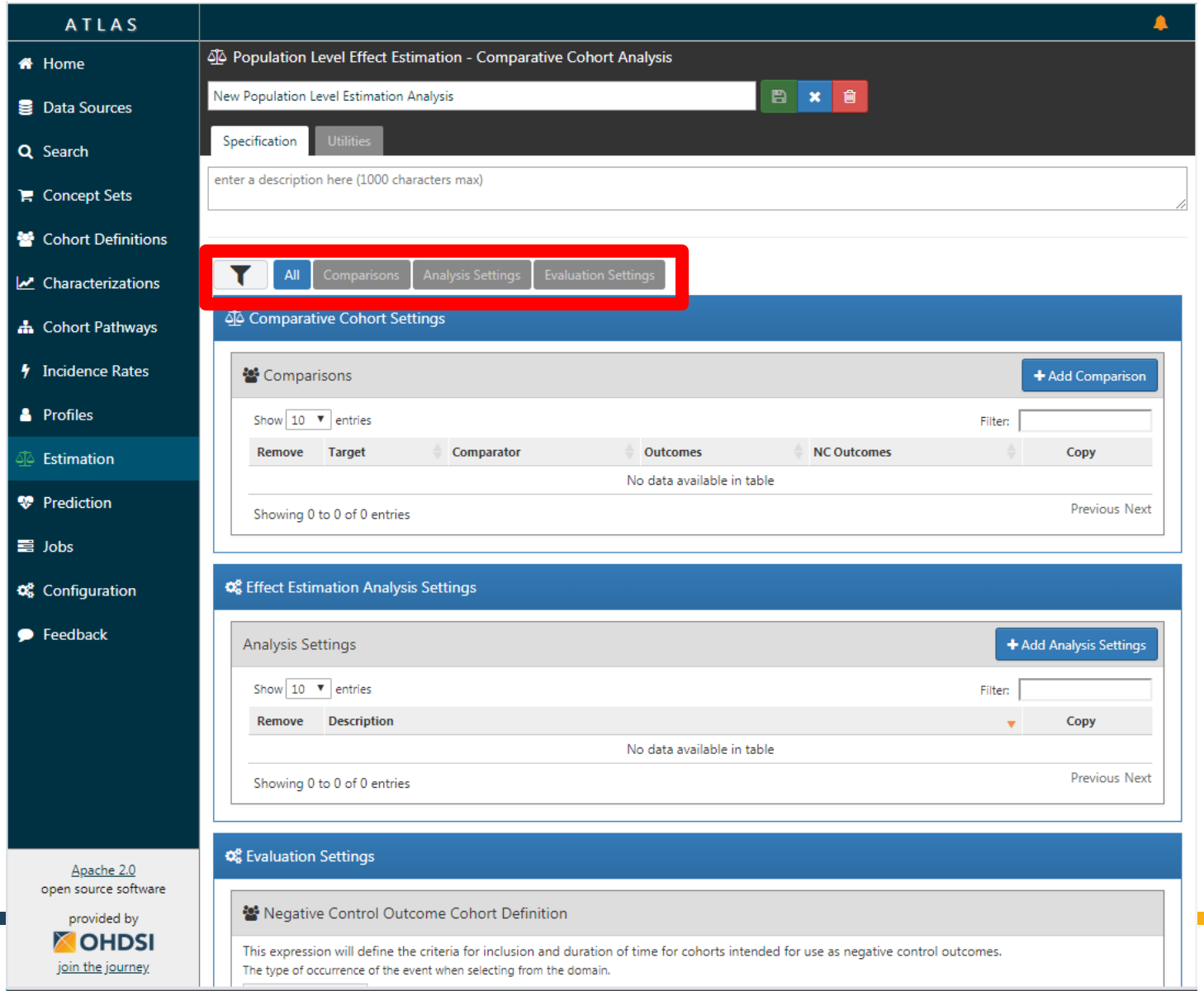
Graham et al. description of the cohort(s)

A new-user retrospective cohort design was used to compare patients initiating dabigatran or warfarin for the treatment of nonvalvular AF.[10] We identified all patients with any inpatient or outpatient diagnoses of AF or atrial flutter based on *International Classification of Diseases, Ninth Revision* coding who also filled at least 1 prescription for either drug from October 19, 2010 (US dabigatran approval date) through December 31, 2012, the study end date. Patients were excluded if they had <6 months of enrollment in Medicare before their index dispensing, were aged <65 years, received prior treatment with a study medication or rivaroxaban or apixaban (anticoagulants approved during the study), were in a skilled nursing facility or nursing home, or were receiving hospice care on the date of their cohort-qualifying prescription. Patients were also excluded if they had a hospitalization that extended beyond the index dispensing date. Patients discharged from the hospital on the same day as their index dispensing were included. Patients undergoing dialysis and kidney transplant recipients were also excluded. Additionally, because warfarin is approved for indications other than AF, we excluded patients with diagnoses indicating the presence of mitral valve disease, heart valve repair or replacement, deep vein thrombosis, pulmonary embolism, or joint replacement surgery in the preceding 6 months.

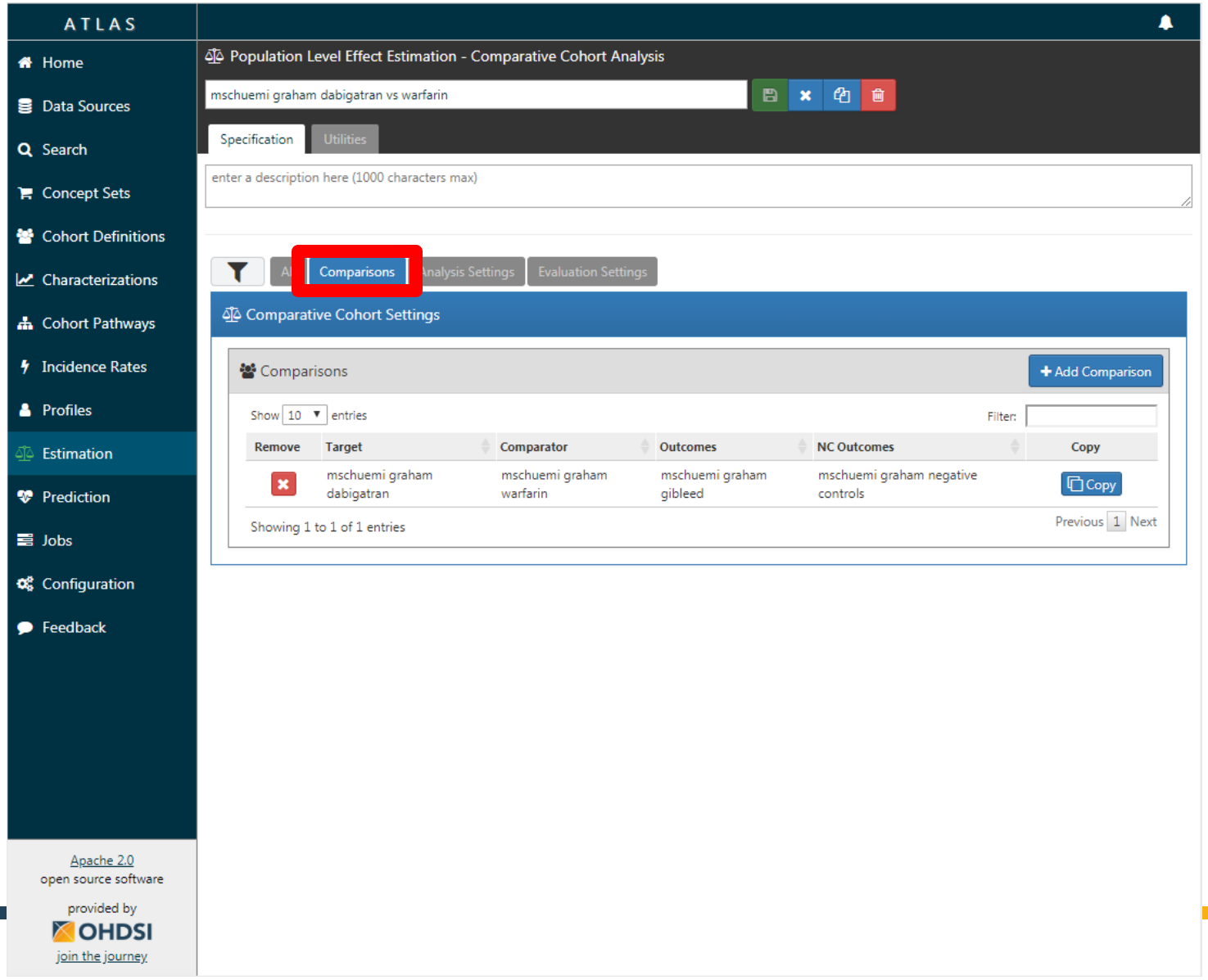
Graham et al. replication: Designing the study in ATLAS



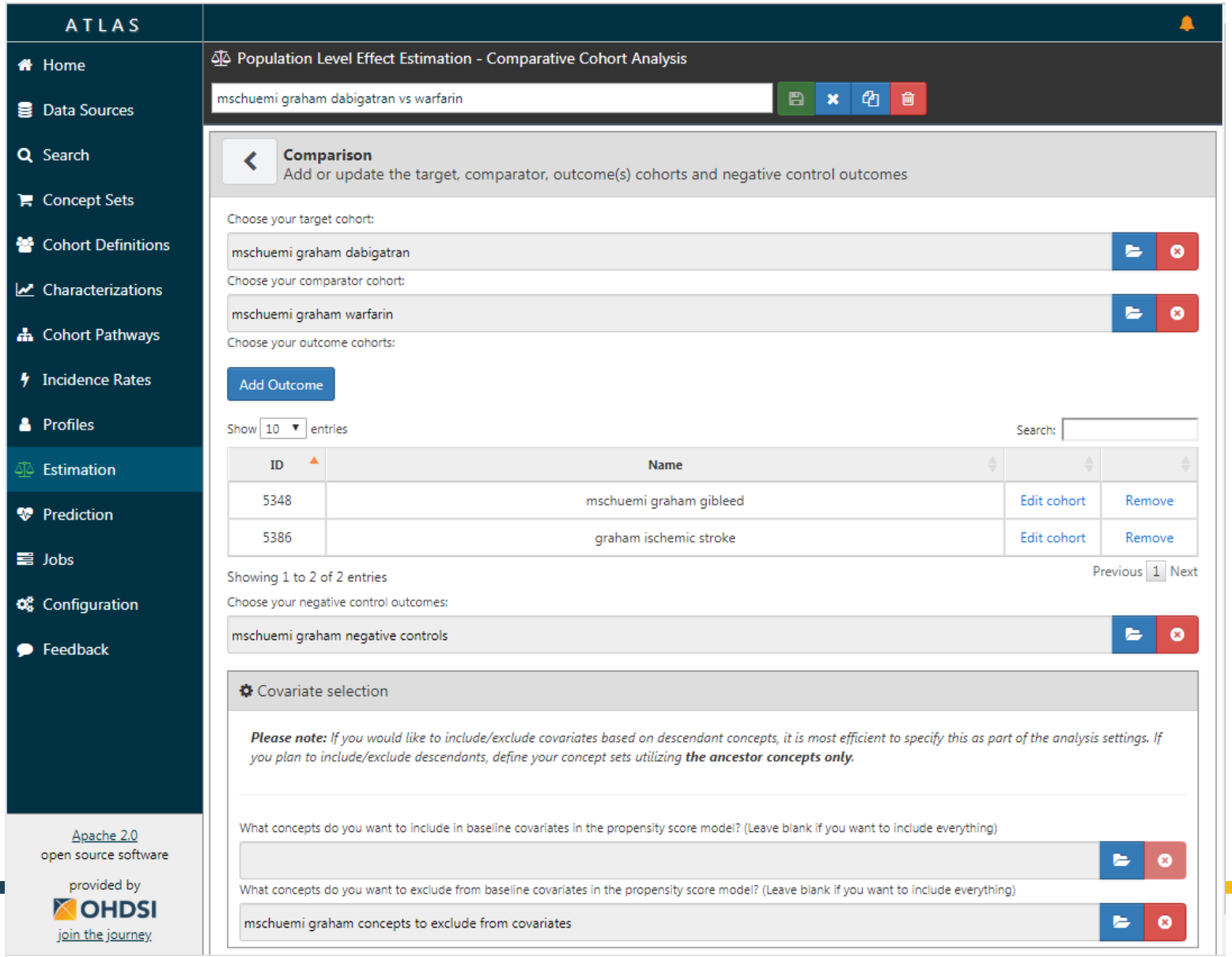
Graham et al. replication: Designing the study in ATLAS



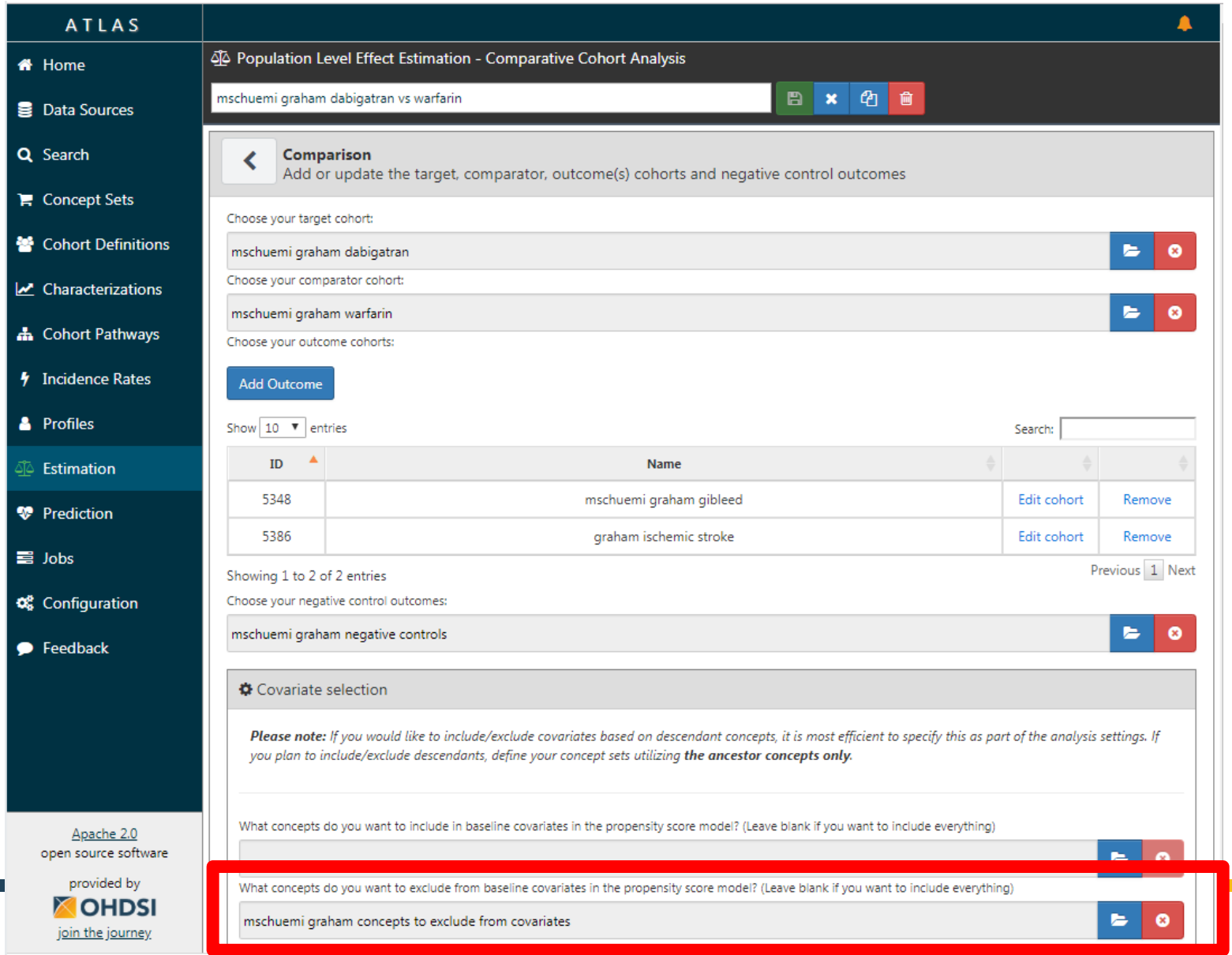
Graham et al. replication: Specifying the comparison



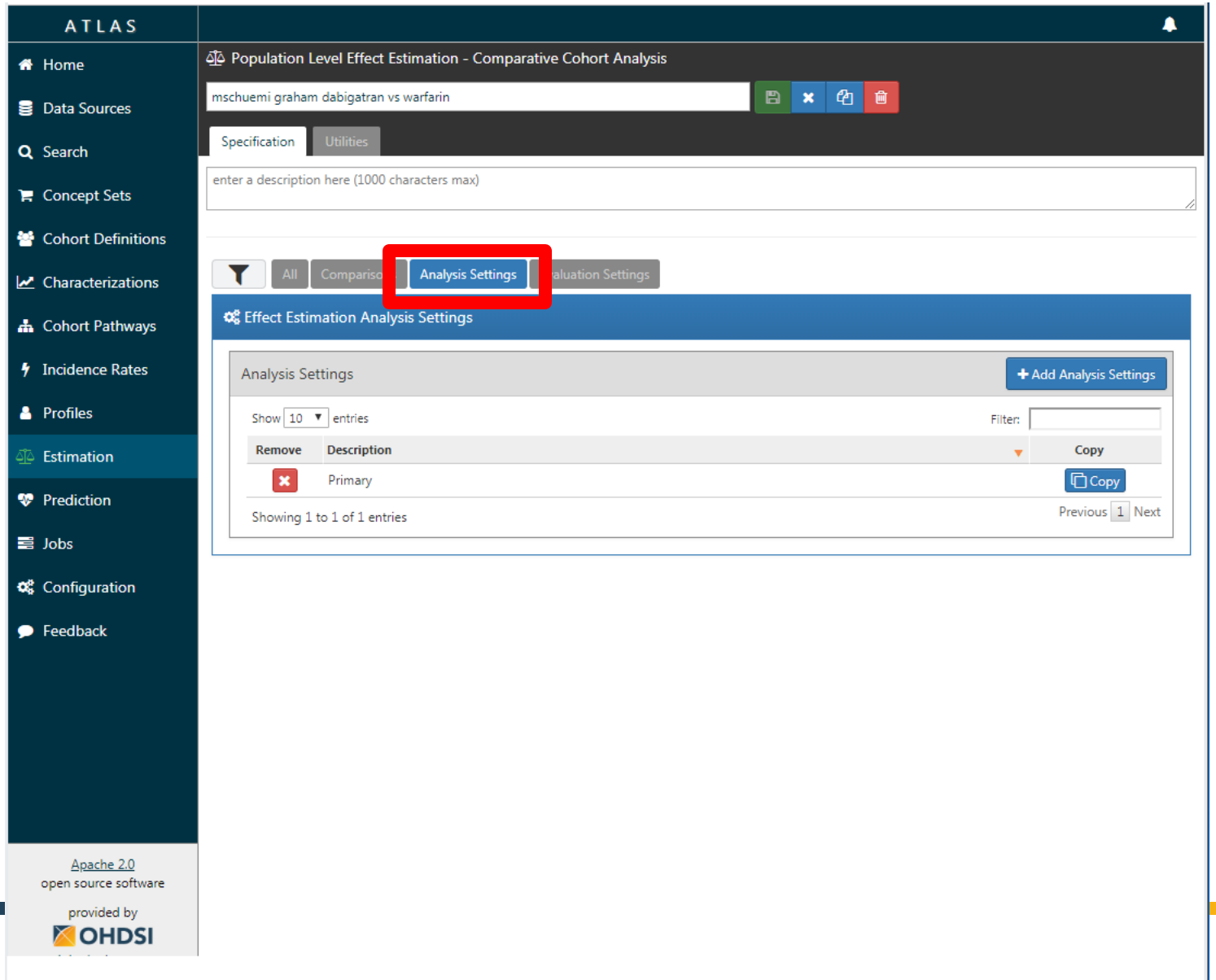
Graham et al. replication: Specifying the comparison



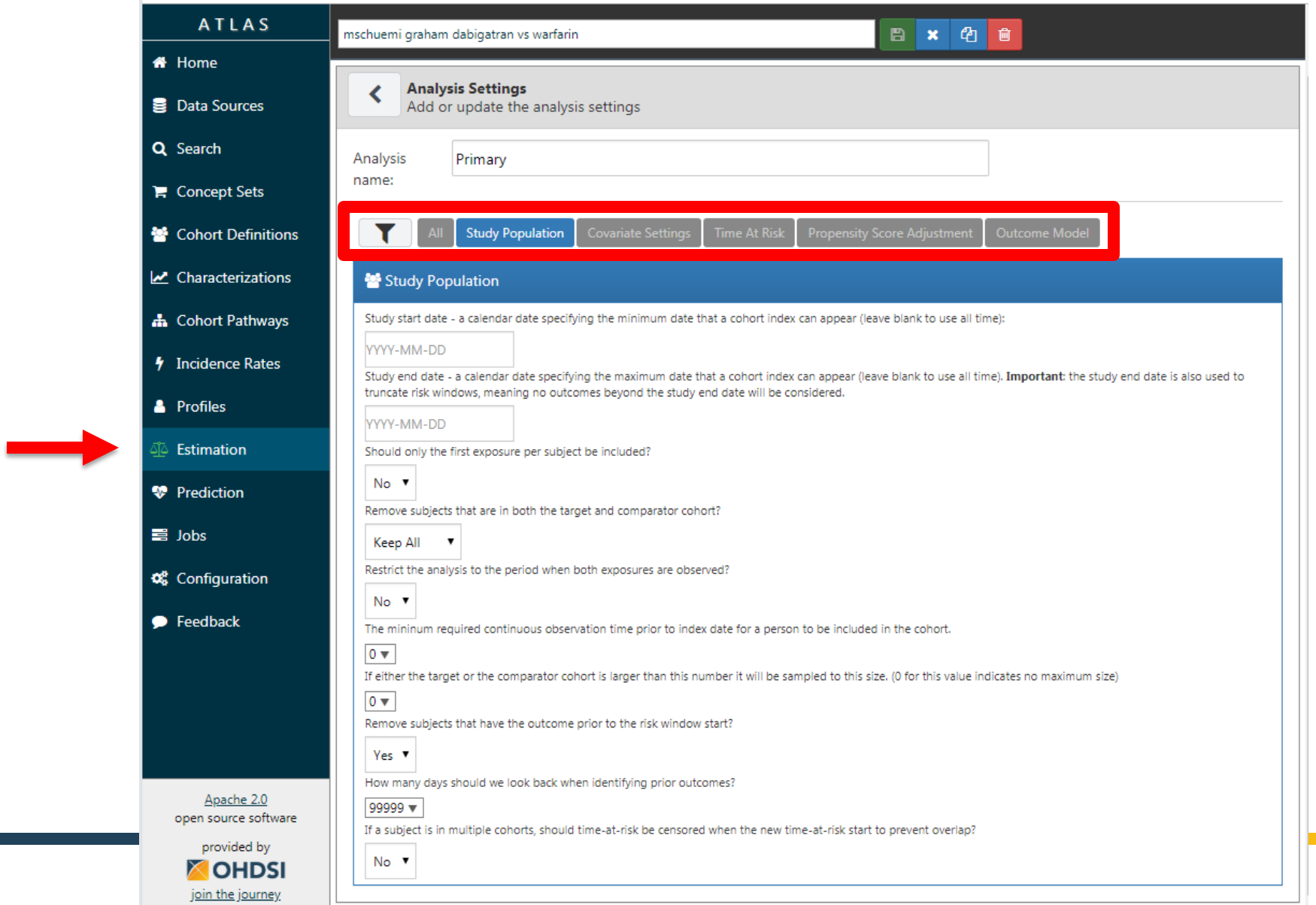
Graham et al. replication: Specifying the comparison



Graham et al. replication: Specifying the analysis settings



Graham et al. replication: Specifying the analysis settings

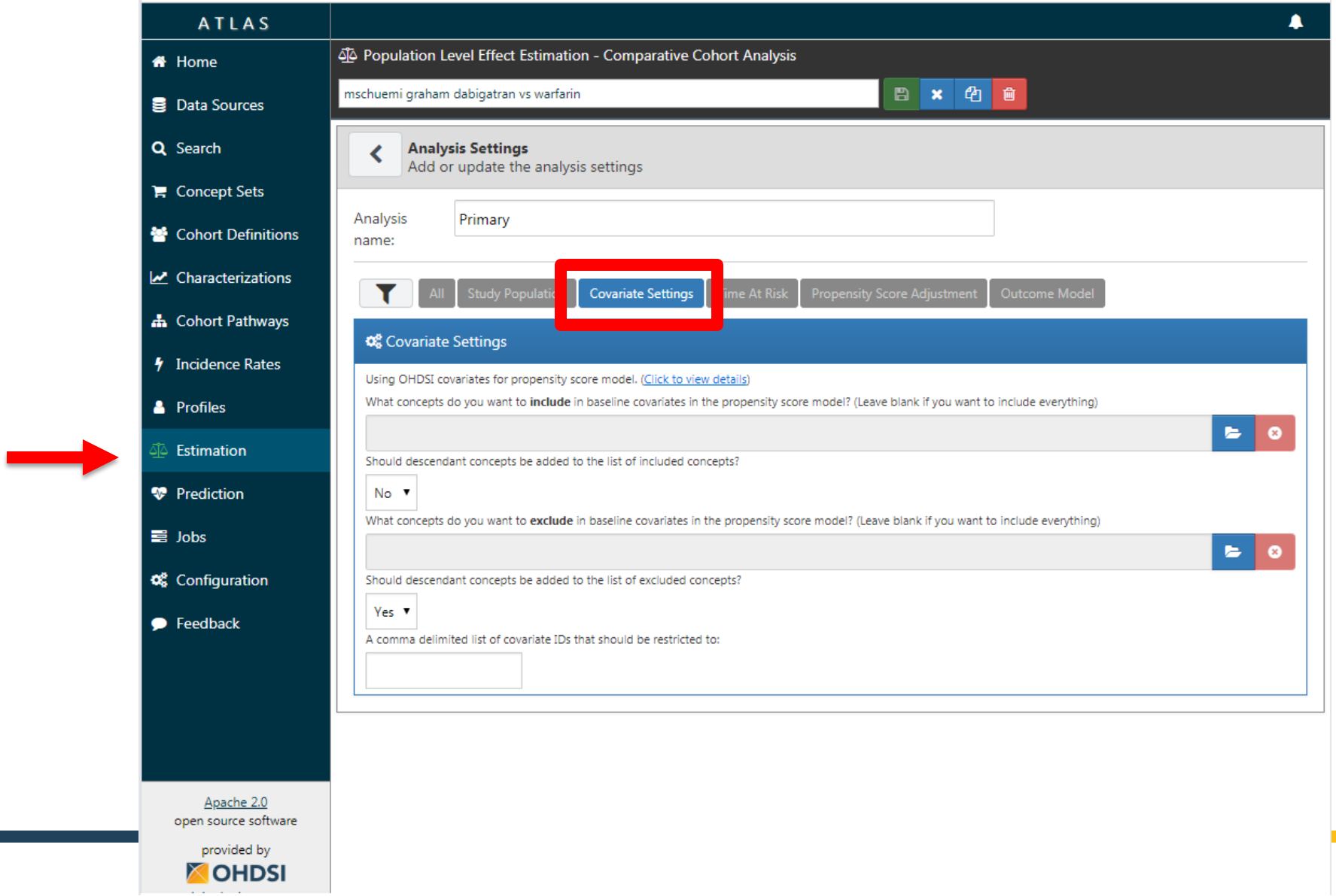


Graham et al. replication: covariates for confounding adjustment

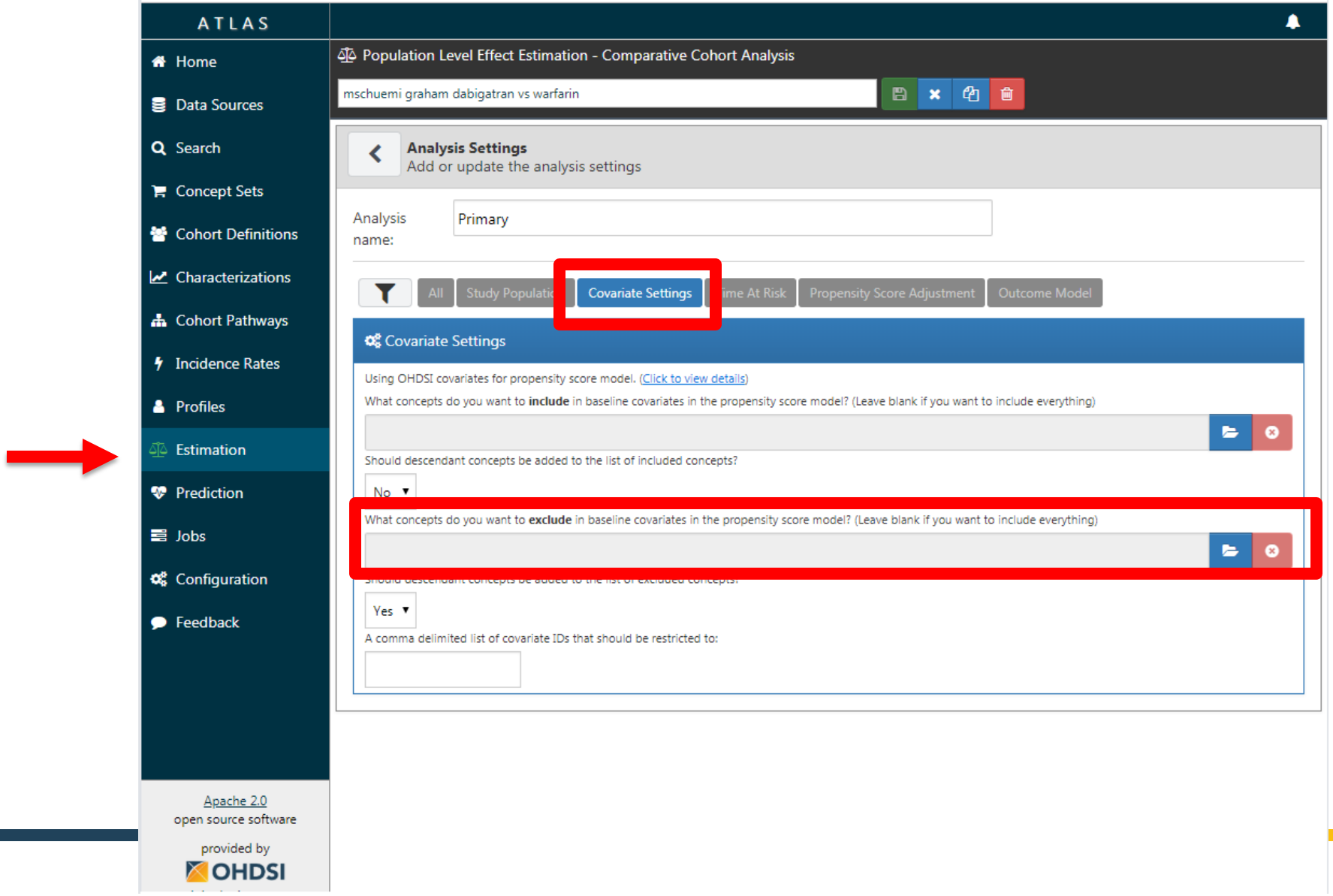
Claims data on chronic medical conditions, cardiovascular risk factors, risk factors for bleeding events, and healthcare utilization were collected for each patient during the 6 months preceding their cohort-qualifying prescription fill. We also collected data on prescriptions for medications used for treatment of cardiovascular disease and other chronic medical conditions, as well as potentially interacting medications that might alter warfarin or dabigatran pharmacokinetics. Finally, to the extent possible using claims data, we calculated the $CHADS_2$ score,[11] which predicts the risk of stroke in patients with AF, and the HAS-BLED score,[12,13] which predicts the risk of bleeding in patients with AF treated with warfarin.

To reduce confounding due to imbalance in study covariates, propensity score matching was used.[14–16] Unconditional logistic regression was used to estimate the predicted probability of patients initiating dabigatran therapy given their sociodemographic characteristics, baseline medical comorbidities, medications used during the preceding 6 months, prescriber characteristics, and other potentially relevant variables (Table 1 and

Graham et al. replication: covariates for confounding adjustment



Graham et al. replication: covariates for confounding adjustment



Graham et al. replication: covariates for confounding adjustment

- Leverages FeatureExtraction package
- Default settings create covariates for all drug and condition group concepts, and procedure, measurement, observation, and device exposure concepts during 2 lookback window
- Defaults also include demographics and 4 risk indexes

Select Covariates...

	Gender	Age	Age Groups	Race	Ethnicity	Index Year	Index Month	Prior Observation Time	Post Observation Time	Time In Cohort	Index Year & Month
Demographics	☑	☐	☑	☑	☑	☑	☑	☐	☐	☐	☐

Time bound covariates

Set the time windows for the time bound covariates in days relative to the cohort index

	Any Time Prior	Long Term	Medium Term	Short Term	End Days
Time Windows	All Time	-365	-180	-30	0

Set the time bound era covariates

Domain	Any Time Prior	Long Term (-365 days)	Medium Term (-180 days)	Short Term (-30 days)	Overlapping	Era Start: Long Term (-365 days)	Era Start: Medium Term (-180 days)	Era Start: Short Term (-30 days)
Condition	☐	☐	☐	☐	☐	☐	☐	☐
Condition Group	☐	☑	☐	☑	☐	☐	☐	☐
Drug	☐	☐	☐	☐	☐	☐	☐	☐
Drug Group	☐	☑	☐	☑	☑	☐	☐	☐

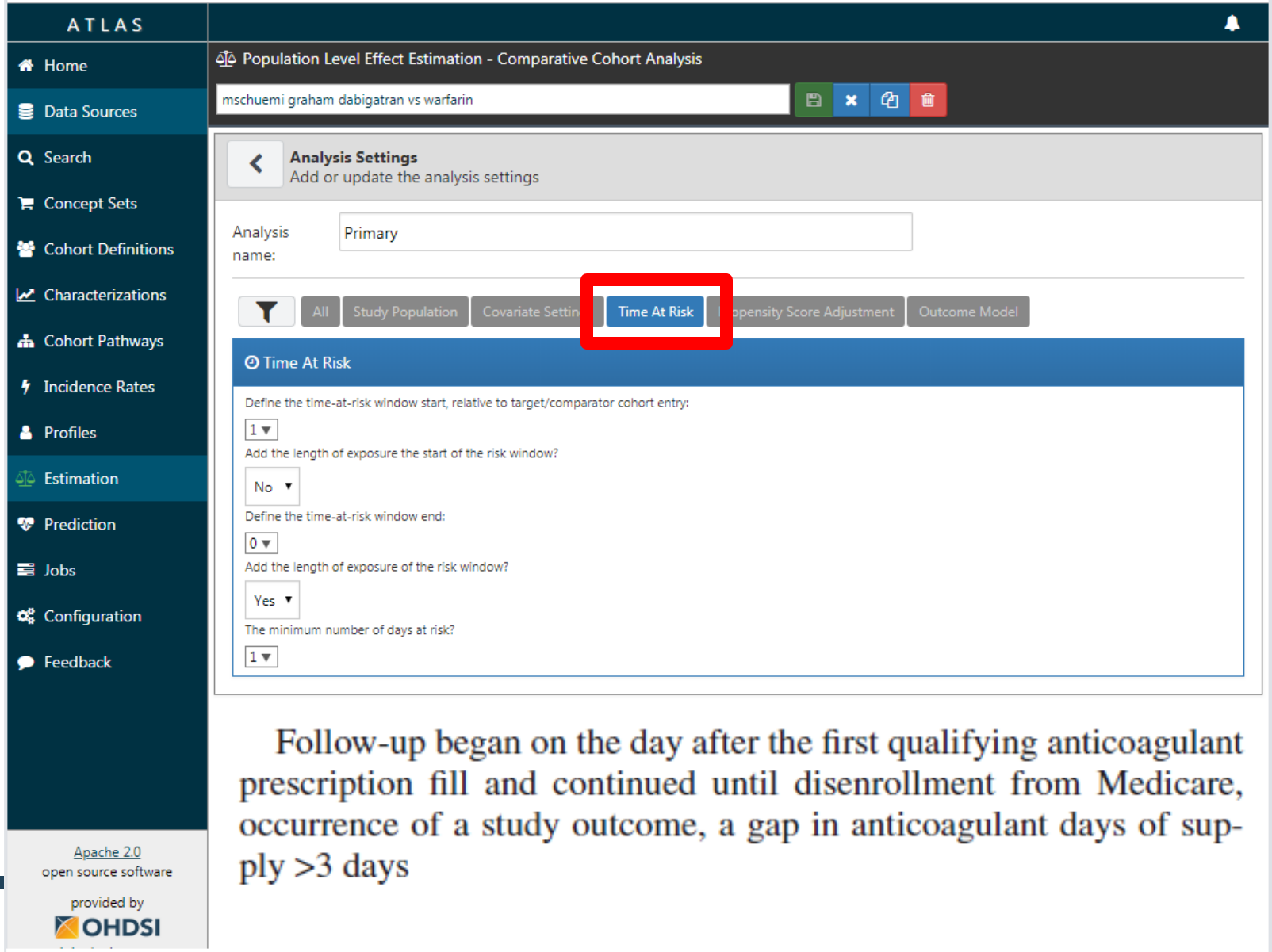
Set the time bound covariates

Domain	Any Time Prior	Long Term (-365 days)	Medium Term (-180 days)	Short Term (-30 days)	Distinct Count: Long Term (-365 days)	Distinct Count: Medium Term (-180 days)	Distinct Count: Short Term (-30 days)
Condition	☐	☐	☐	☐	☐	☐	☐
Condition - Primary Inpatient	☐	☐	☐	☐			
Drug	☐	☐	☐	☐	☐	☐	☐
Procedure	☐	☑	☐	☑	☐	☐	☐
Measurement	☐	☑	☐	☑	☐	☐	☐
Measurement - Value	☐	☐	☐	☐			
Measurement - Range Group	☐	☑	☐	☐			
Observation	☐	☑	☐	☑	☐	☐	☐
Device	☐	☑	☐	☑			
Visit - Count		☐	☐	☐			
Visit - Concept Count		☐	☐	☐			

Set the index score covariates

Index Score Type	
$CHADS_2$	☑
CHA_2DS_2VASc	☑
DCSI	☑
Charlson	☑

Graham et al. replication: specifying time-at-risk

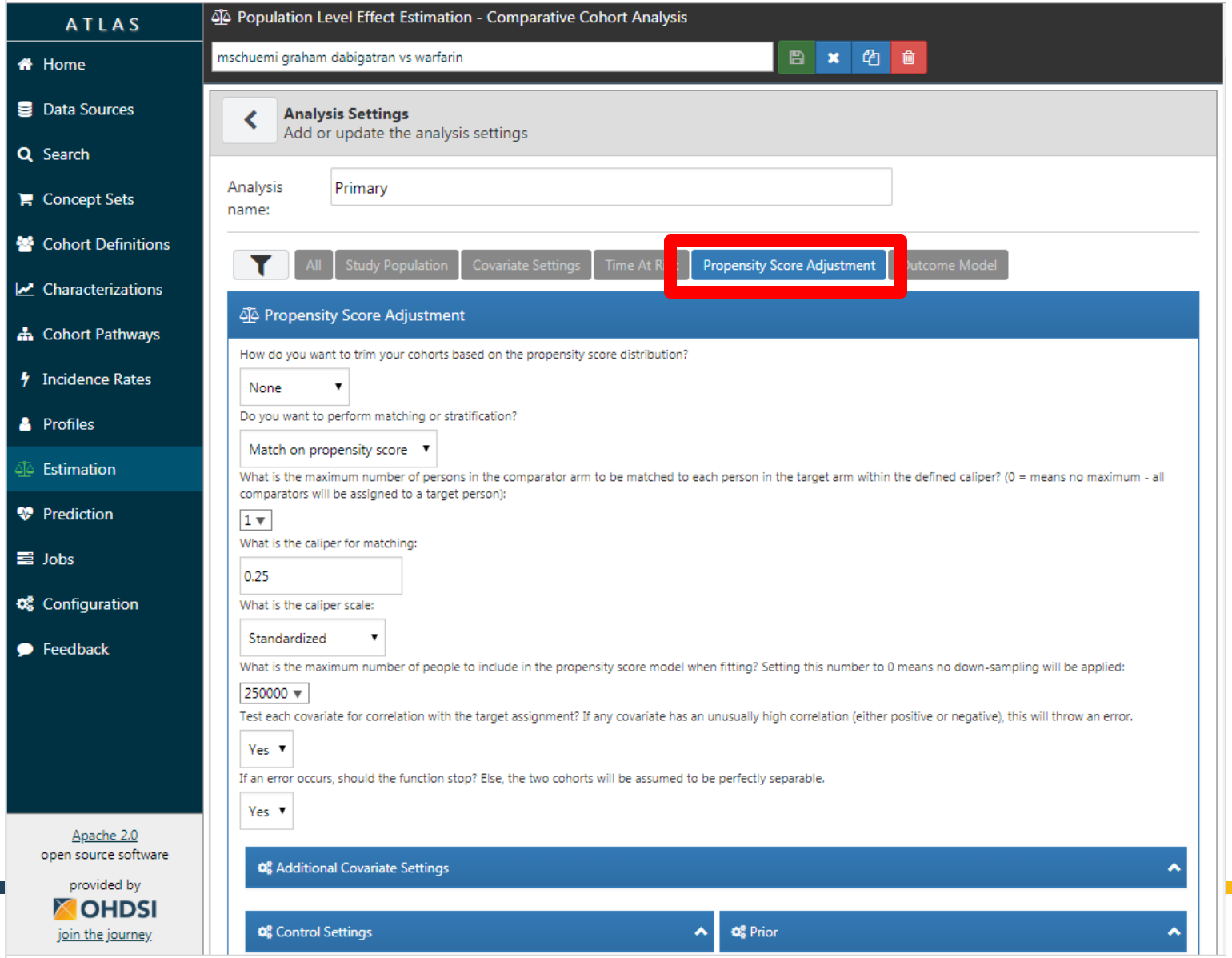


Follow-up began on the day after the first qualifying anticoagulant prescription fill and continued until disenrollment from Medicare, occurrence of a study outcome, a gap in anticoagulant days of supply >3 days

Graham et al. replication: specifying the propensity score model

Dabigatran users were propensity score matched to warfarin users in a 1:1 ratio with the use of a greedy matching algorithm. The balance of measured covariates between the matched cohorts was assessed with the standardized mean difference, a measure not influenced by sample size and thus useful for comparing cohorts in large observational studies.[17] A standardized mean difference of ≤0.1 indicates a negligible difference in the measured variables between groups.[17]

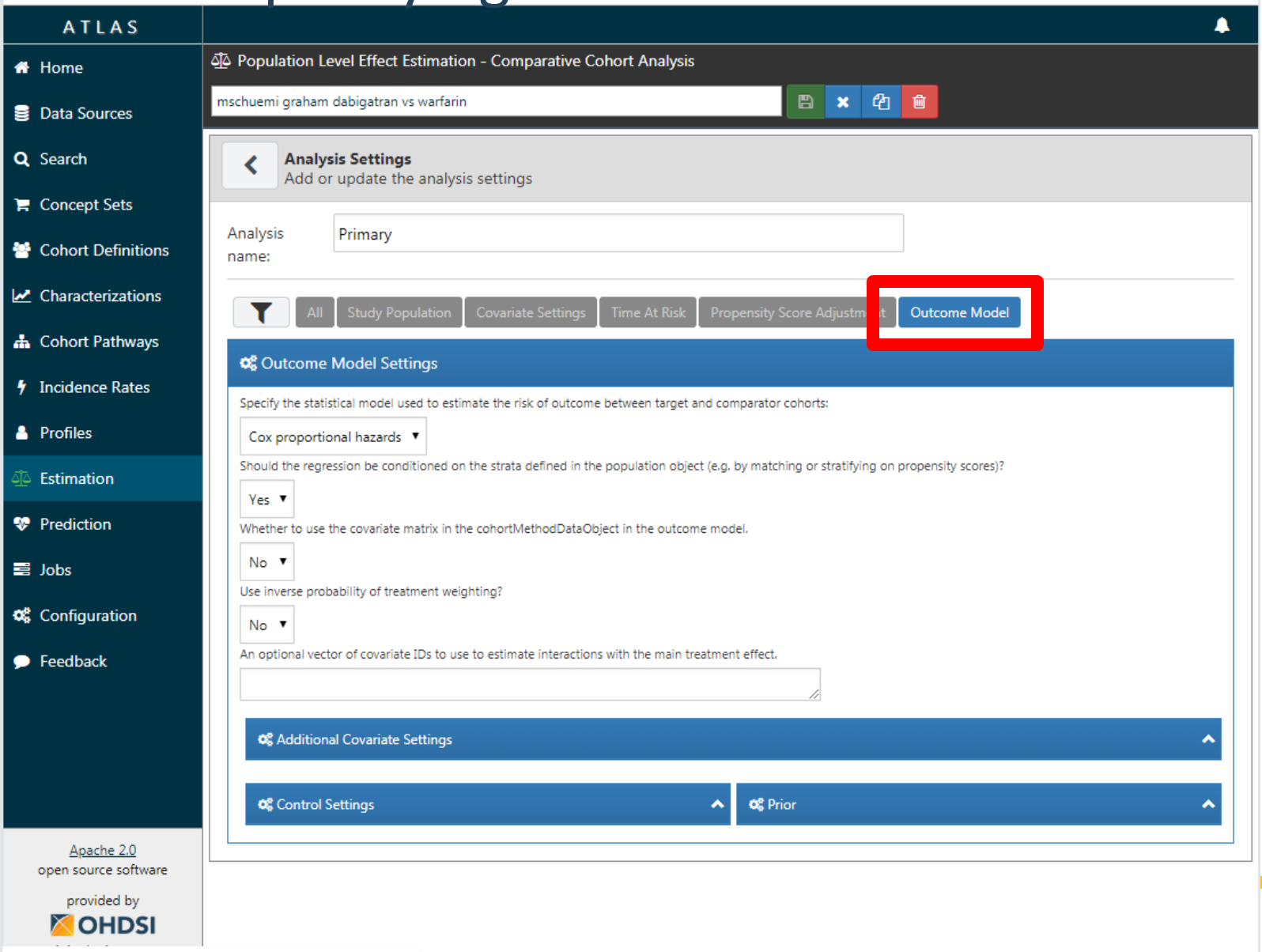
Graham et al. replication: specifying the propensity score model



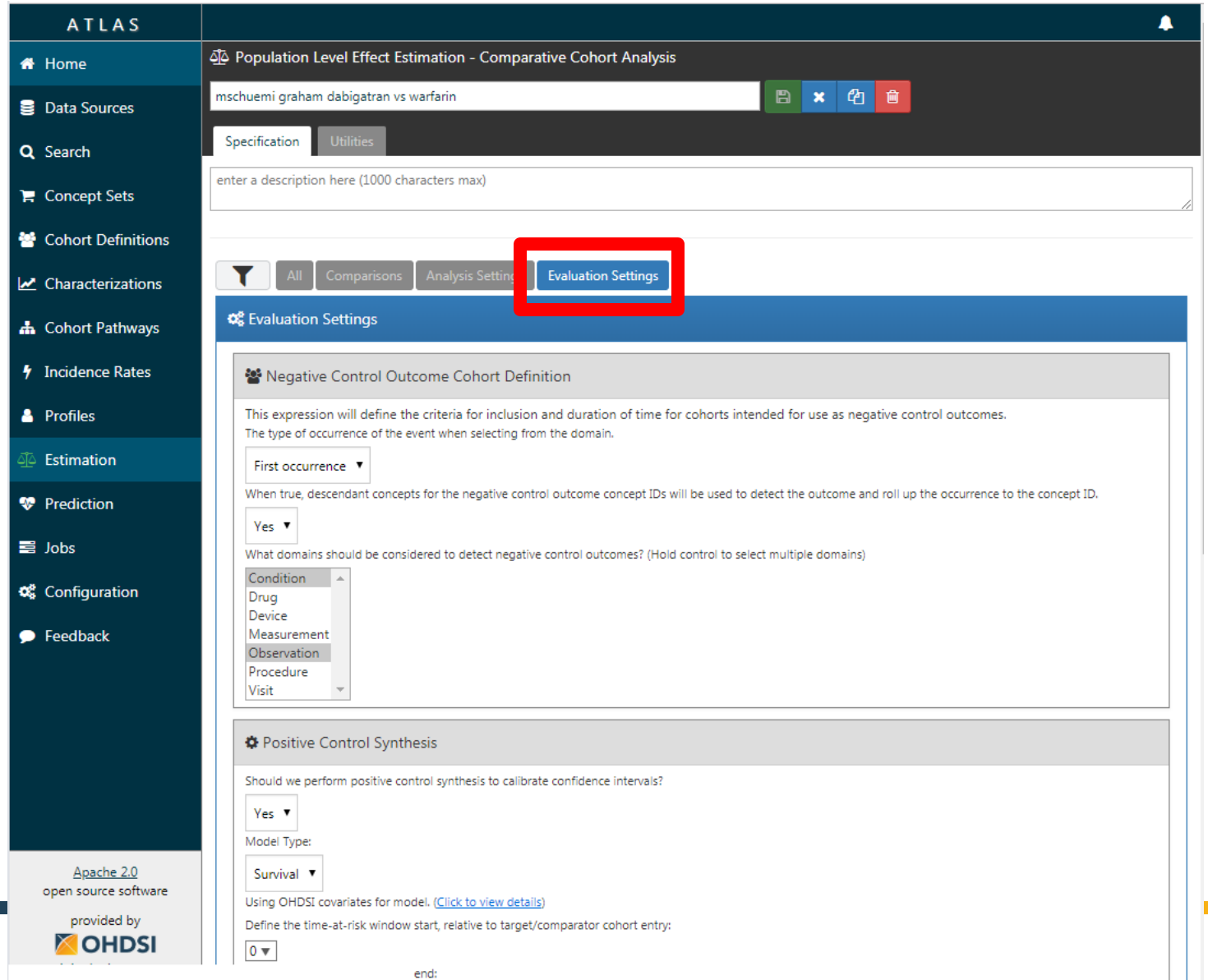
Graham et al. replication: specifying the outcome model

Analyses were performed on the propensity score–matched cohorts, thereby accounting for the potential confounding factors shown in Table 1 and in the online-only Data Supplement. Incidence rates were estimated with the use of event counts and exposure follow-up time. Kaplan–Meier plots were generated to characterize the contour of risk over time for each outcome. Cox proportional hazards regression was used to compare time to event in dabigatran compared with warfarin (reference) cohorts.

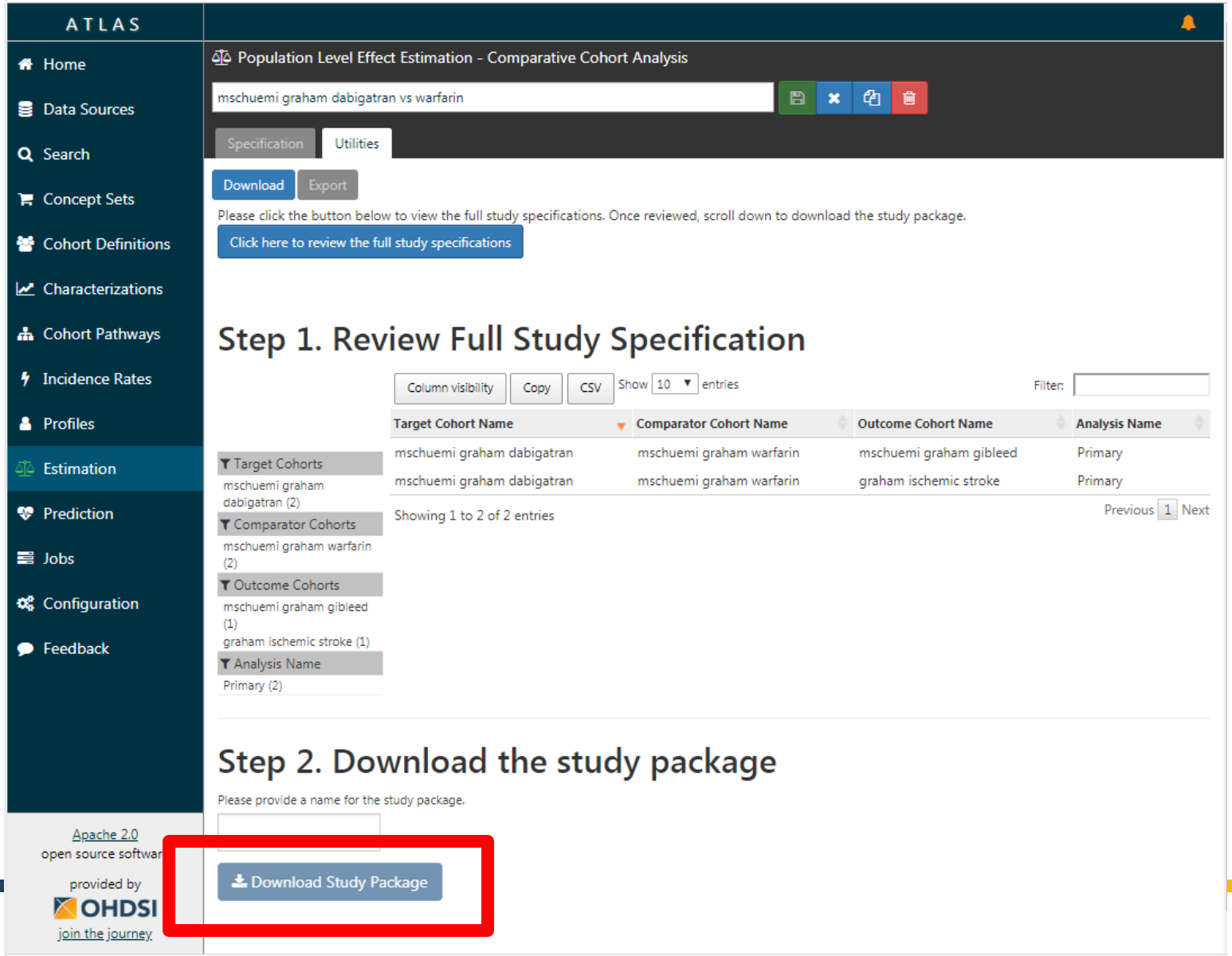
Graham et al. replication: specifying the outcome model



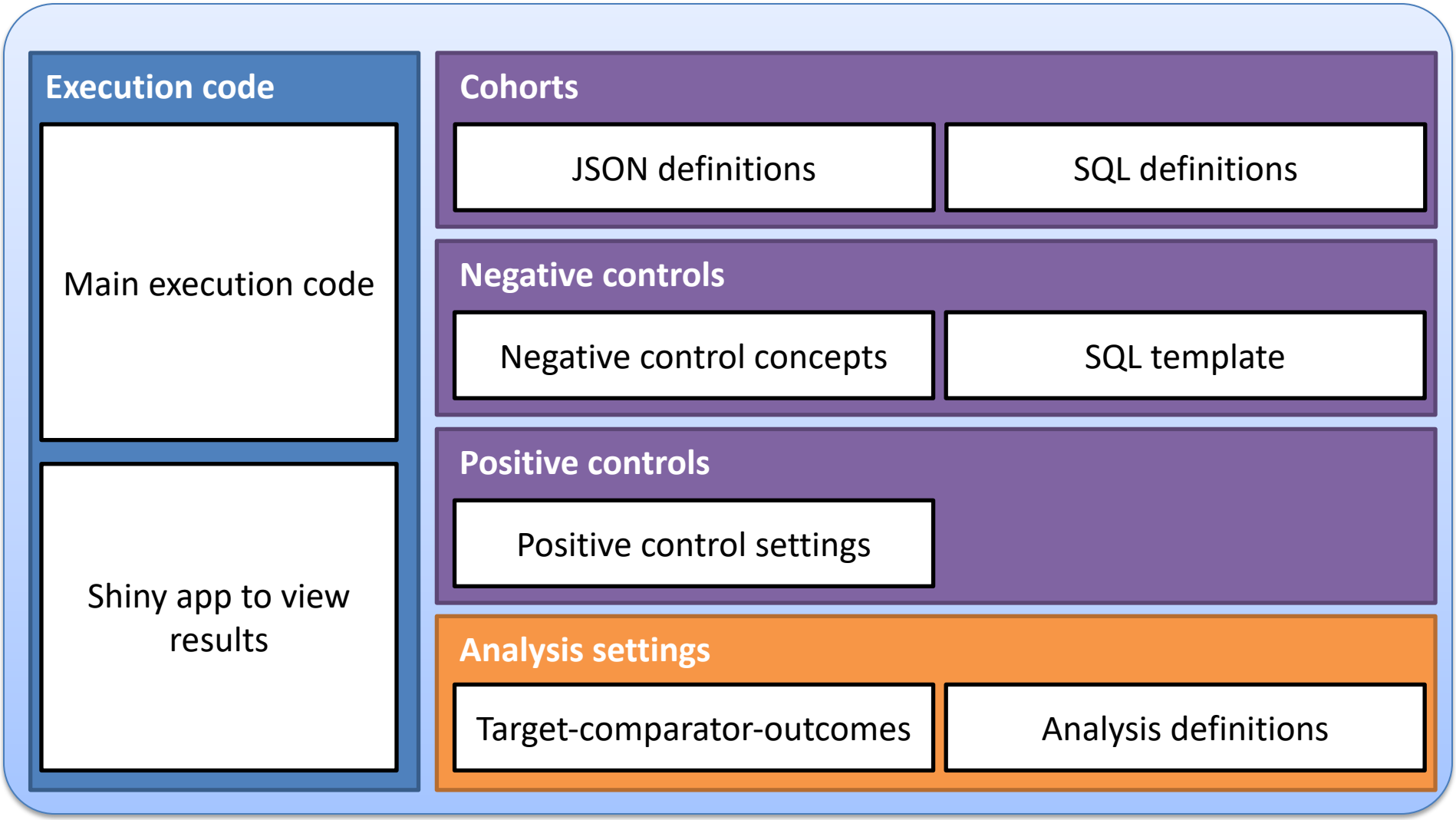
Graham et al. replication: specifying the evaluation settings



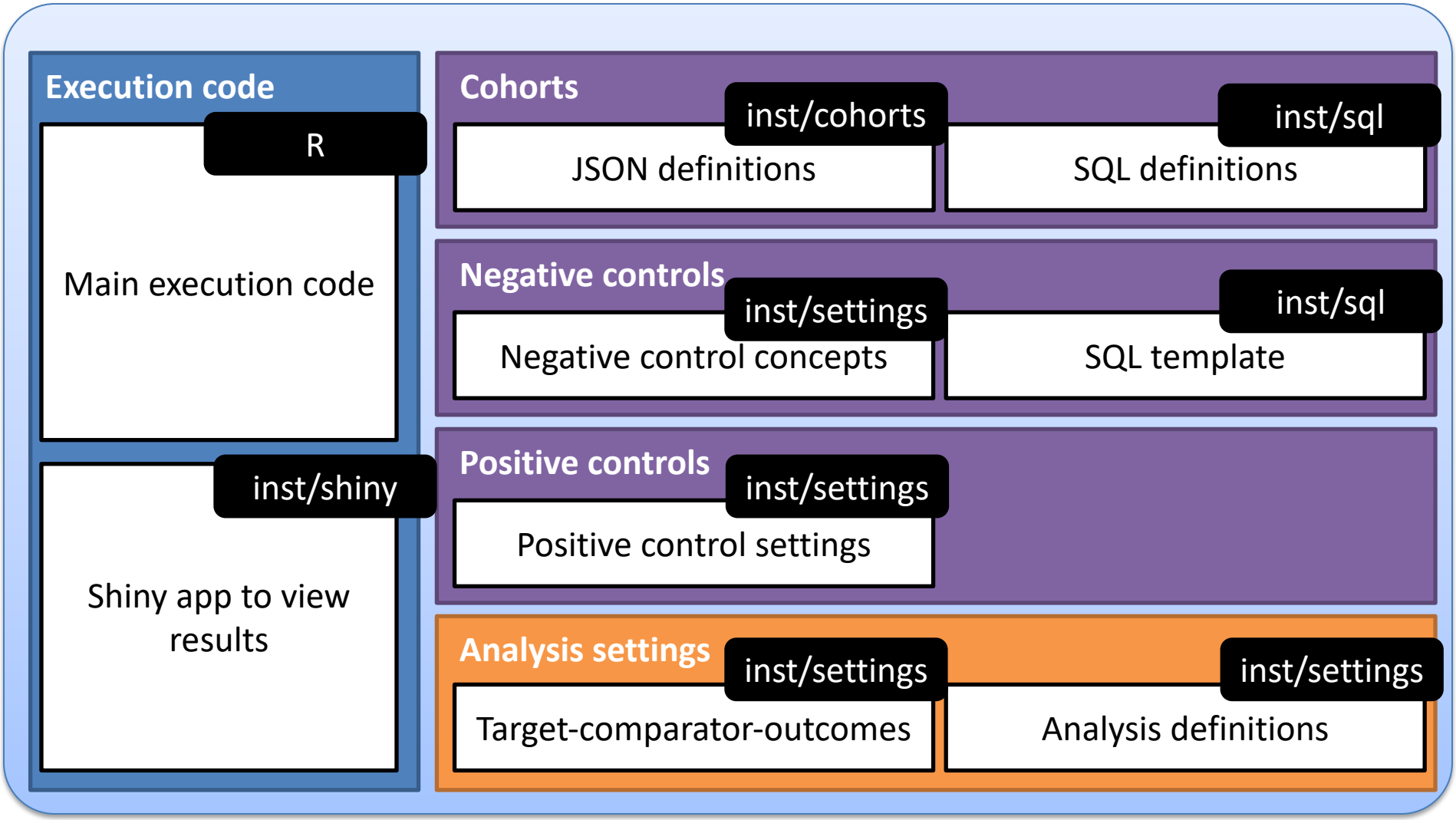
Graham et al. replication: exporting the study package for execution



Anatomy of the study package



Anatomy of the study package



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 depencies (see readme)
  - Install with devtools::install_github

Installing dependencies

How to run

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

```
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

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"
```

Export data model

Study specification

analyses

cohort_method_analysis
- analysis_id
- description
- definition

covariate_analysis
- covariate_analysis_id
- covariate_analysis_name

exposures

exposure_of_interest
- exposure_id
- exposure_name
- definition

outcomes

outcome_of_interest
- outcome_id
- outcome_name
- definition

negative_control_outcome
- outcome_id
- outcome_name

positive_control_outcome
- outcome_id
- outcome_name
- exposure_id
- negative_control_id
- effect_size

Generated results

metadata

database
- database_id
- database_name
- description
- is_meta_analysis

exposure_summary
- database_id
- exposure_id
- min_date
- max_date

comparison_summary
- database_id
- target_id
- comparator_id
- min_date
- max_date

attrition
- database_id
- exposure_id
- [target_id]
- [comparator_id]
- [outcome_id]
- [analysis_id]
- sequence_number
- description
- subjects*

cm_follow_up_dist
- database_id
- target_id
- comparator_id
- outcome_id
- analysis_id
- target_min_days
- target_p10_days
- target_p25_days
- target_median_days
- target_p75_days
- target_p90_days
- target_max_days
- comparator_min_days
- comparator_p10_days
- comparator_p25_days
- comparator_median_days
- comparator_p75_days
- comparator_p90_days
- comparator_max_days

covariate
- database_id
- covariate_id
- covariate_name
- covariate_analysis_id

main results

cohort_method_result
- database_id
- target_id
- comparator_id
- outcome_id
- analysis_id
- rr
- ci_95_lb
- ci_95_ub
- p
- [i_2]
- log_rr
- se_log_rr
- target_subjects*
- comparator_subjects*
- target_days
- comparator_days
- target_outcomes*
- comparator_outcomes*
- calibrated_p

cm_interaction_result
- database_id
- target_id
- comparator_id
- outcome_id
- analysis_id
- interaction_covariate_id
- rrr
- ci_95_lb
- ci_95_ub
- p
- [i_2]
- log_rrr
- se_log_rrr
- target_subjects*
- comparator_subjects*
- target_days
- comparator_days
- target_outcomes*
- comparator_outcomes*

diagnostics

covariate_balance
- database_id
- target_id
- comparator_id
- [outcome_id]
- [analysis_id]
- [interaction_covariate_id]
- covariate_id
- target_mean_before*
- comparator_mean_before*
- std_diff_before
- target_mean_after*
- comparator_mean_after*
- std_diff_after

preference_score_dist
- database_id
- target_id
- comparator_id
- preference_score

kaplan_meier_dist
- database_id
- target_id
- comparator_id
- outcome_id
- analysis_id
- time
- [target_at_risk*]
- [comparator_at_risk*]
- target_survival
- target_survival_lb
- target_survival_ub
- comparator_survival
- comparator_survival_lb
- comparator_survival_ub

propensity_model
- database_id
- target_id
- comparator_id

underscore indicates primary key

[ ] indicates nullable

* indicates fields with a minimum value to avoid identifiability

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:

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

Source: readme.md

Evidence Explorer

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.

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