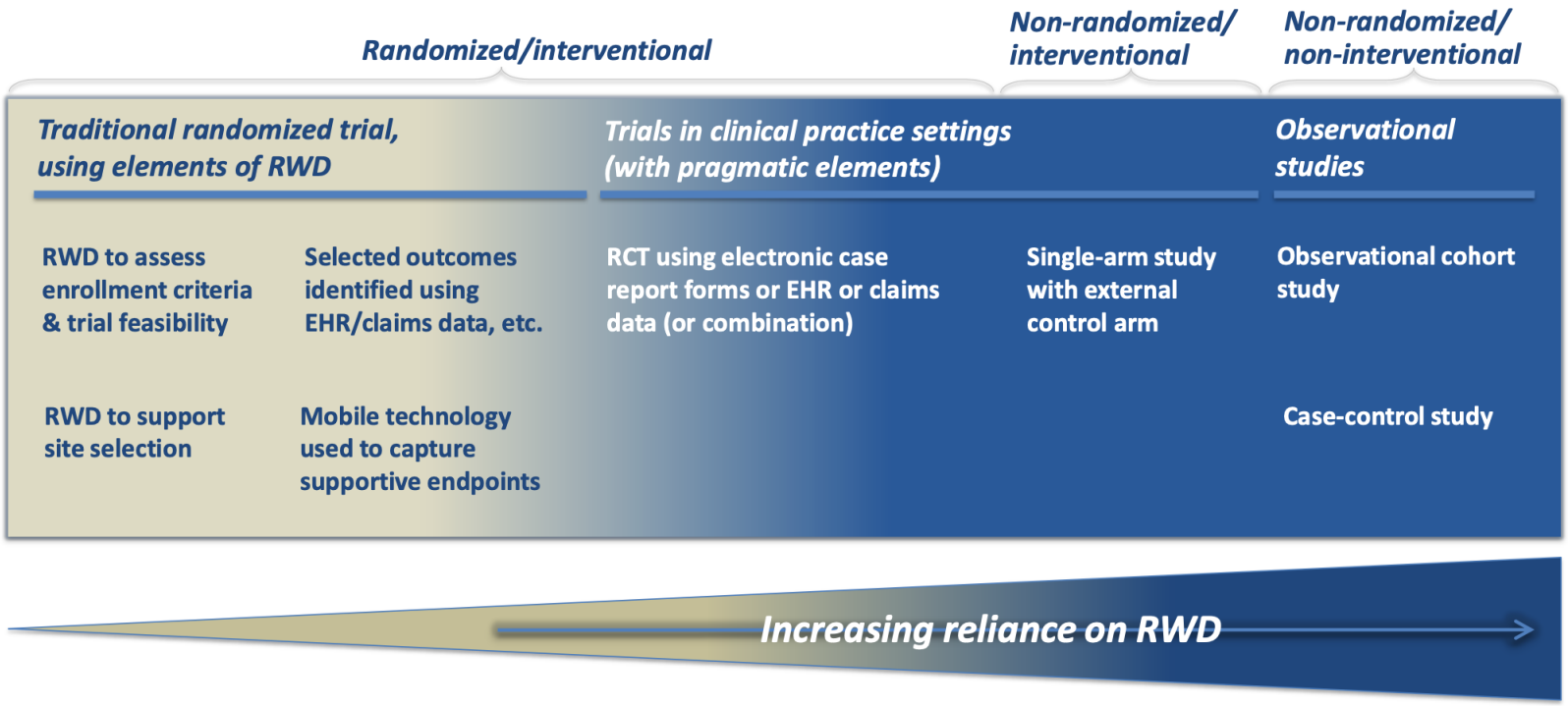
Towards Integration of Targeted Learning in Safety Analysis

Mark van der Laan

RWD/RWE Webinar 2: Statistical Analysis and Data Quality
November 8, 2021

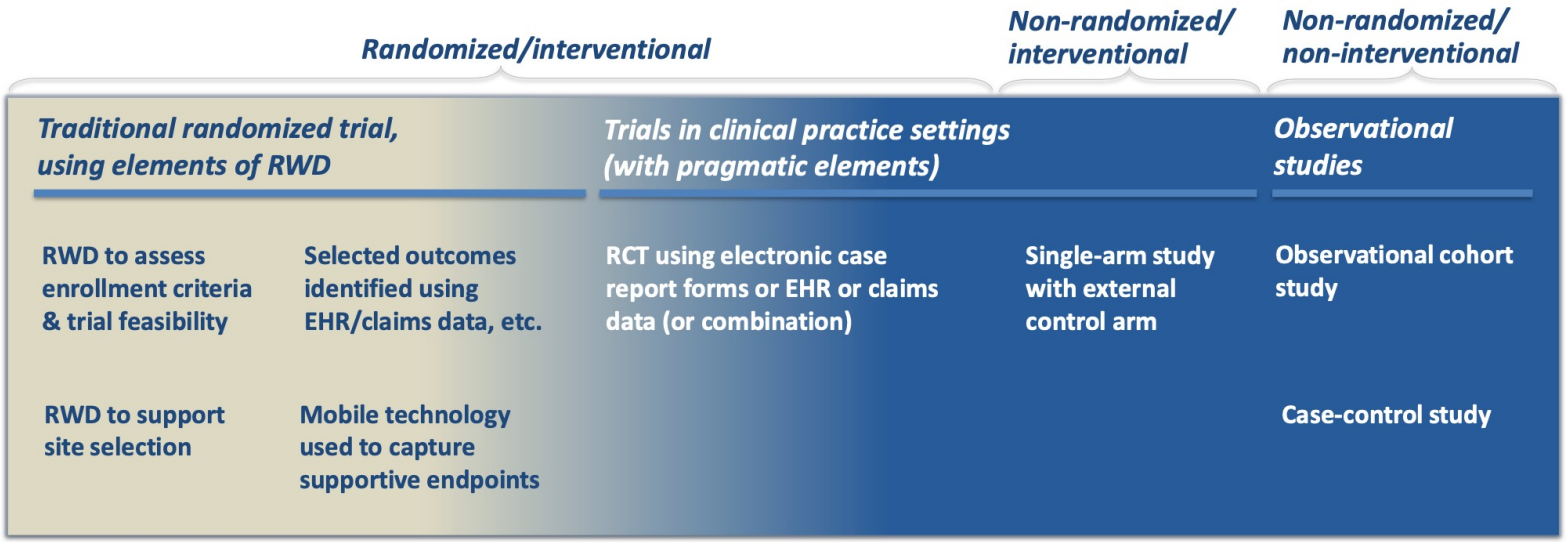
Acknowledgements: Susan Gruber, Ivana Malenica and Rachael Phillips

Statistical challenges with RWD



Courtesy of "FDA Real-World Evidence Program" Webinar by John Concato on 4 August 2021

Statistical challenges with RWD



RWD Challenges

- ❑ Selection bias
- ❑ Intercurrent events
- ❑ Informative missingness
- ❑ Treatment by indication
- ❑ High dimensional covariates
- ❑ Outcome measurement error
- ❑ Statistical model misspecification
- ❑ Differences between external controls and single trial arm RCT

Targeted Learning path supports regulatory decision making

The roadmap for learning from data

STEP 1:
DESCRIBE
EXPERIMENT

STEP 2:
SPECIFY STATISTICAL
MODEL

STEP 3:
DEFINE STATISTICAL
QUERY

STEP 4:
CONSTRUCT
ESTIMATOR

STEP 5:
OBTAIN
INFERENCE

STEP 6:
MAKE SUBSTANTIVE
CONCLUSION

What is the experiment that generated the data?

STEP 1: DESCRIBE EXPERIMENT

STEP 2: SPECIFY STATISTICAL MODEL

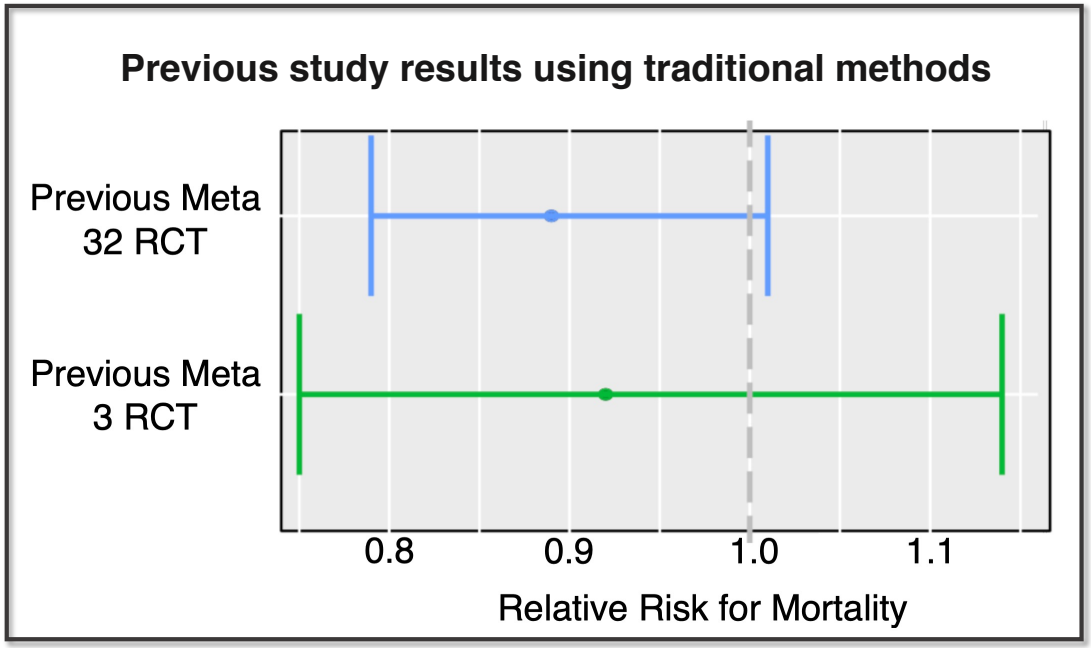
STEP 3: DEFINE STATISTICAL QUERY

STEP 4: CONSTRUCT ESTIMATOR

STEP 5: OBTAIN INFERENCE

STEP 6: MAKE SUBSTANTIVE CONCLUSION

Three multi-national RCTs assessing impact of corticosteroids on mortality among septic shock patients



What is the experiment that generated the data?

STEP 1: DESCRIBE EXPERIMENT

STEP 2: SPECIFY STATISTICAL MODEL

STEP 3: DEFINE STATISTICAL QUERY

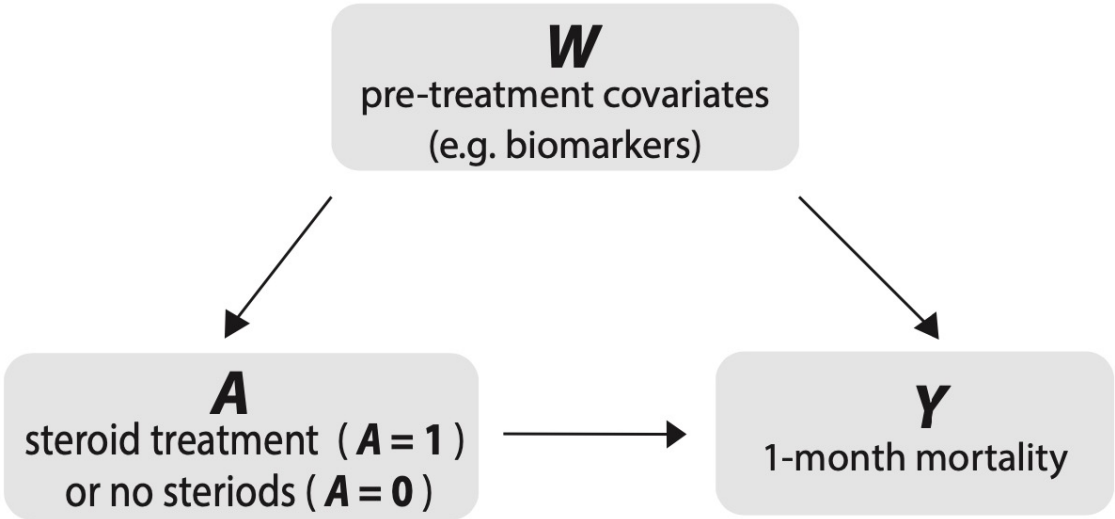
STEP 4: CONSTRUCT ESTIMATOR

STEP 5: OBTAIN INFERENCE

STEP 6: MAKE SUBSTANTIVE CONCLUSION

Three multi-national RCTs assessing impact of corticosteroids on mortality among septic shock patients

Pooled sample of $n = 1{,}300$ adults in septic shock



What is known about stochastic relations of the observed variables?

STEP 1:
DESCRIBE
EXPERIMENT

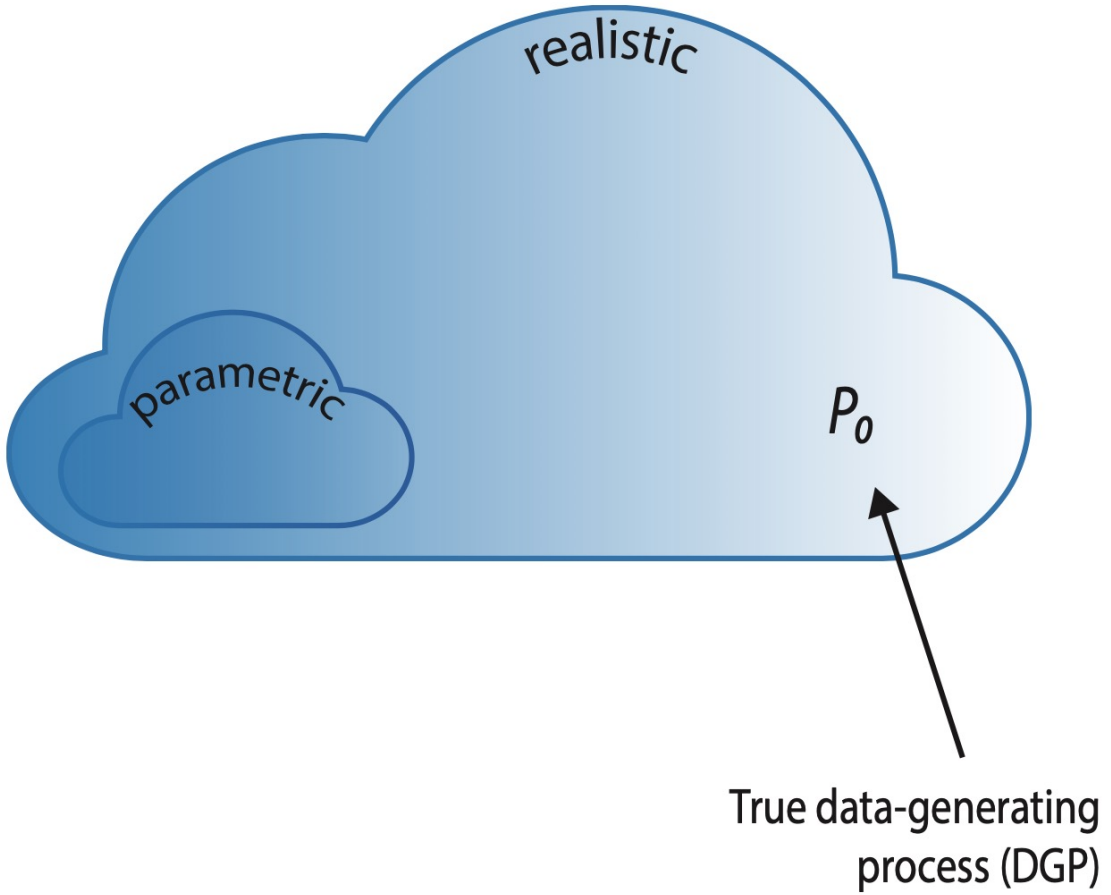
STEP 2:
SPECIFY
STATISTICAL MODEL

STEP 3:
DEFINE STATISTICAL
QUERY

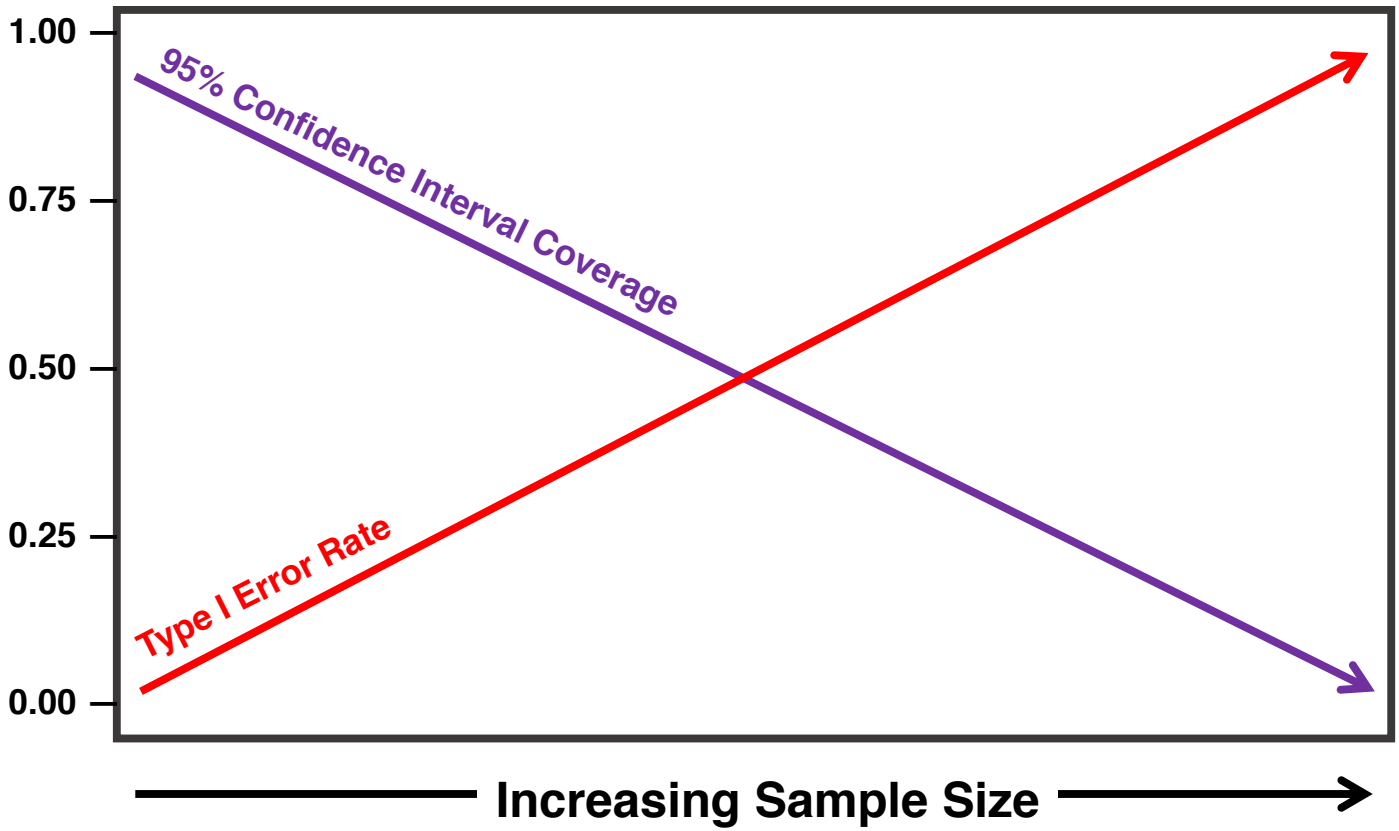
STEP 4:
CONSTRUCT
ESTIMATOR

STEP 5:
OBTAIN
INFERENCE

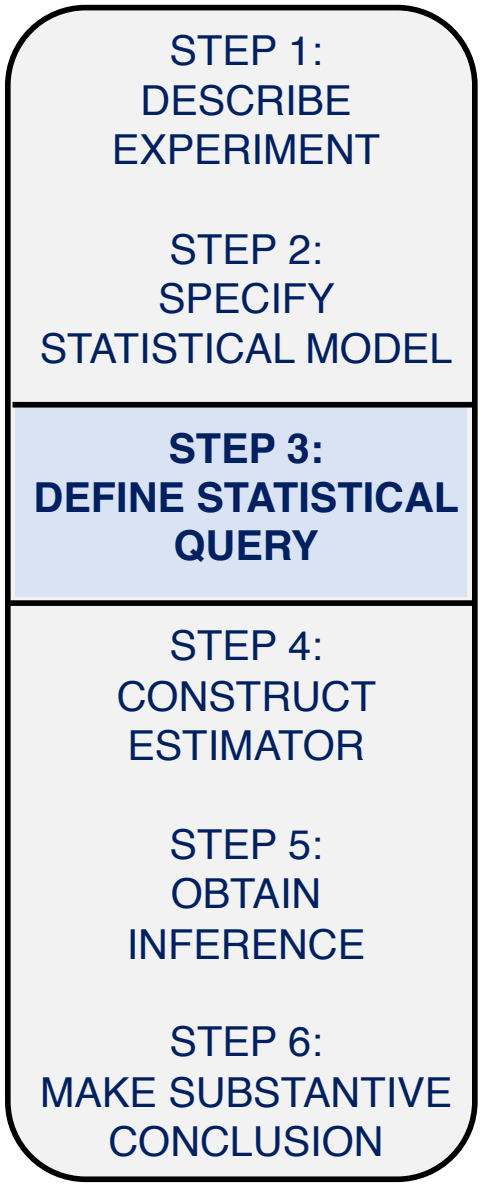
STEP 6:
MAKE SUBSTANTIVE
CONCLUSION



What happens when the statistical model is misspecified and does not contain the DGP?



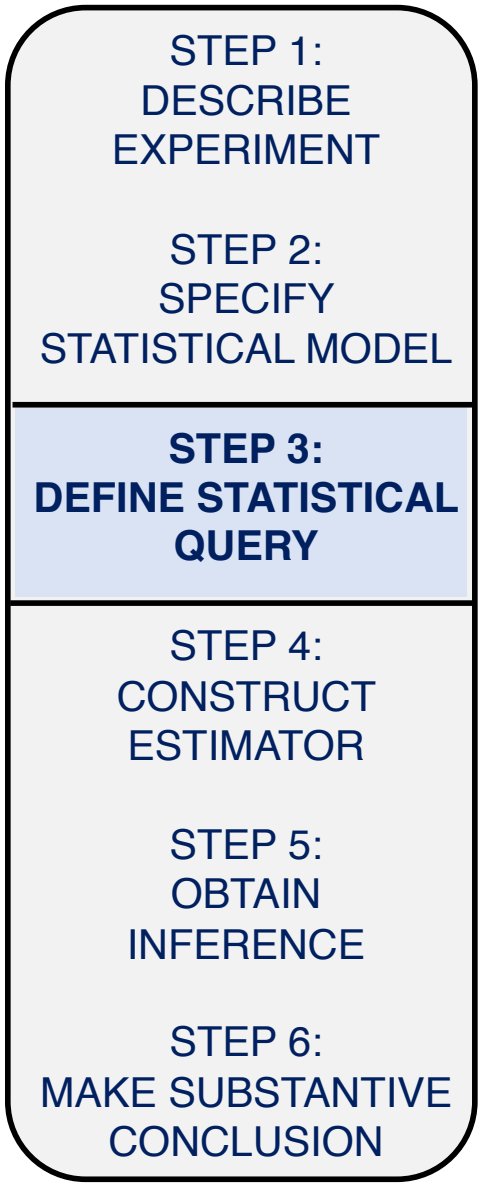
Step 3a: What is the target causal estimand that we aim to identify from the data?



What is the causal risk difference in mortality between treatment groups?

$$\psi_{causal} = E[Y_1 - Y_0]$$

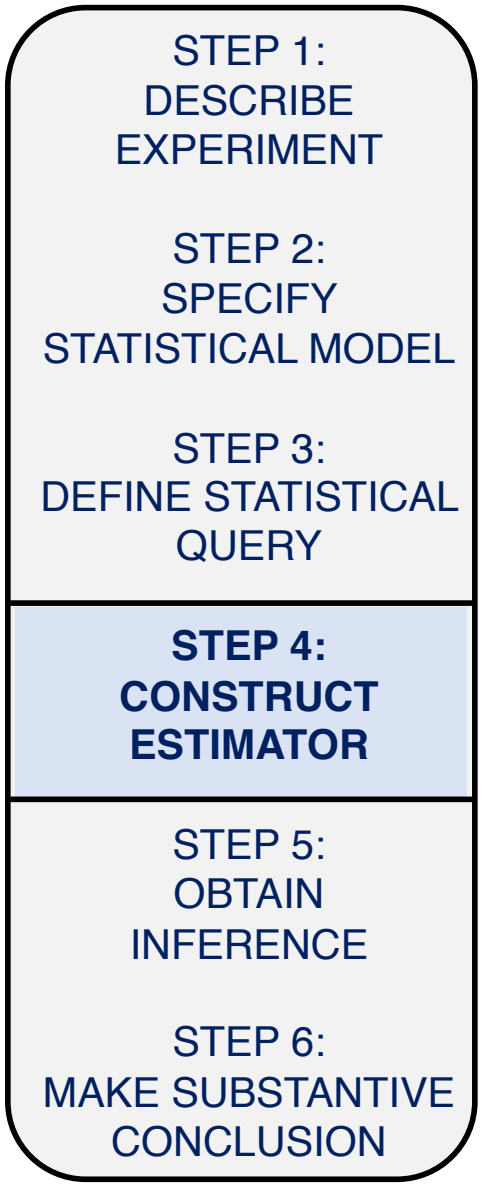
Step 3b: What is the target statistical estimand that we will learn from the data?



What is the average difference in mortality between treatment groups when adjusting for covariates?

$$\psi_{stat} = E(E[Y|A=1,W] - E[Y|A=0,W])$$

How should we estimate the target estimand?



Statistical properties to consider

- Substitution / plug-in
- Valid inference
- Efficiency
- Ability to optimize finite sample performance

Targeted Maximum Likelihood Estimation (TMLE)

STEP 1: DESCRIBE EXPERIMENT

STEP 2: SPECIFY STATISTICAL MODEL

STEP 3: DEFINE STATISTICAL QUERY

STEP 4: CONSTRUCT ESTIMATOR

STEP 5: OBTAIN INFERENCE

STEP 6: MAKE SUBSTANTIVE CONCLUSION

TMLE

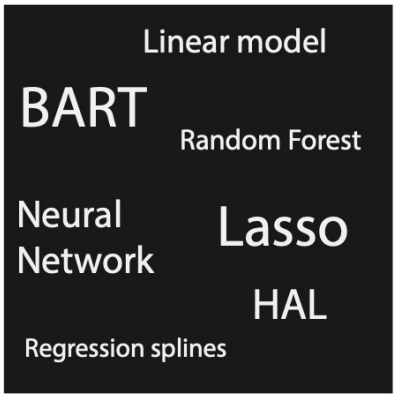
1. Initial estimation of $E[Y|A,W]$ with super (machine) learning
2. Updating initial estimate to acheive optimal bias-variance trade-off for ψ_{stat}

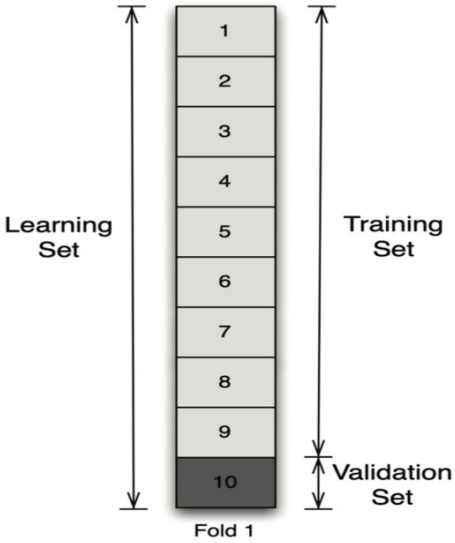
TMLE estimates are optimal:

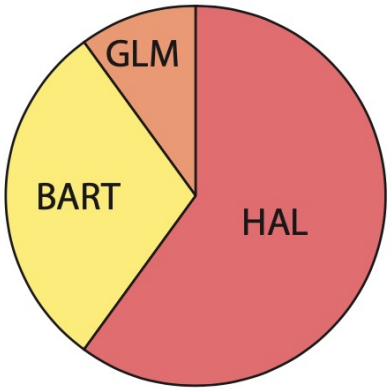
plug-in, efficient, unbiased, finite sample robust

TMLE Step 1: Super learner

Cross-validated performance of learners + ensembles







Hugely advantageous when coupled with NLP-derived covariates with EHR

How should we approximate the sampling distribution of our estimator?

STEP 1: DESCRIBE EXPERIMENT

STEP 2: SPECIFY STATISTICAL MODEL

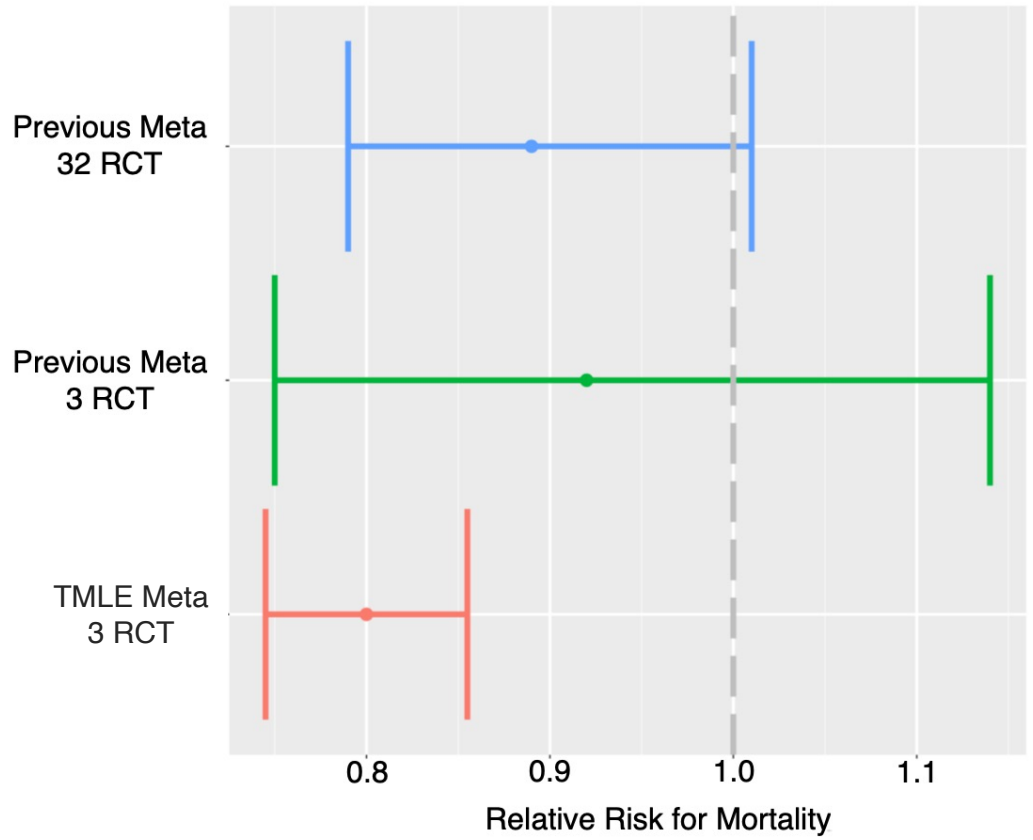
STEP 3: DEFINE STATISTICAL QUERY

STEP 4: CONSTRUCT ESTIMATOR

STEP 5: OBTAIN INFERENCE

STEP 6: MAKE SUBSTANTIVE CONCLUSION

Due to targeting (step 2), the TMLE behaves as the ***sample mean*** of efficient influence function



Arriving at the substantive conclusion

STEP 1: DESCRIBE EXPERIMENT

STEP 2: SPECIFY STATISTICAL MODEL

STEP 3: DEFINE STATISTICAL QUERY

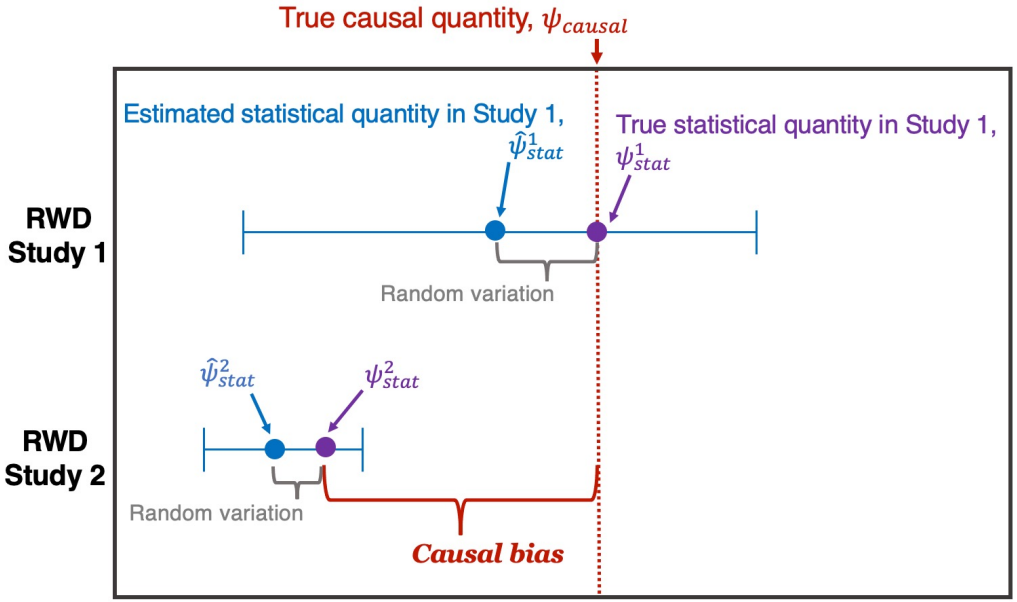
STEP 4: CONSTRUCT ESTIMATOR

STEP 5: OBTAIN INFERENCE

STEP 6: MAKE SUBSTANTIVE CONCLUSION

Investigate causal bias with sensitivity analysis

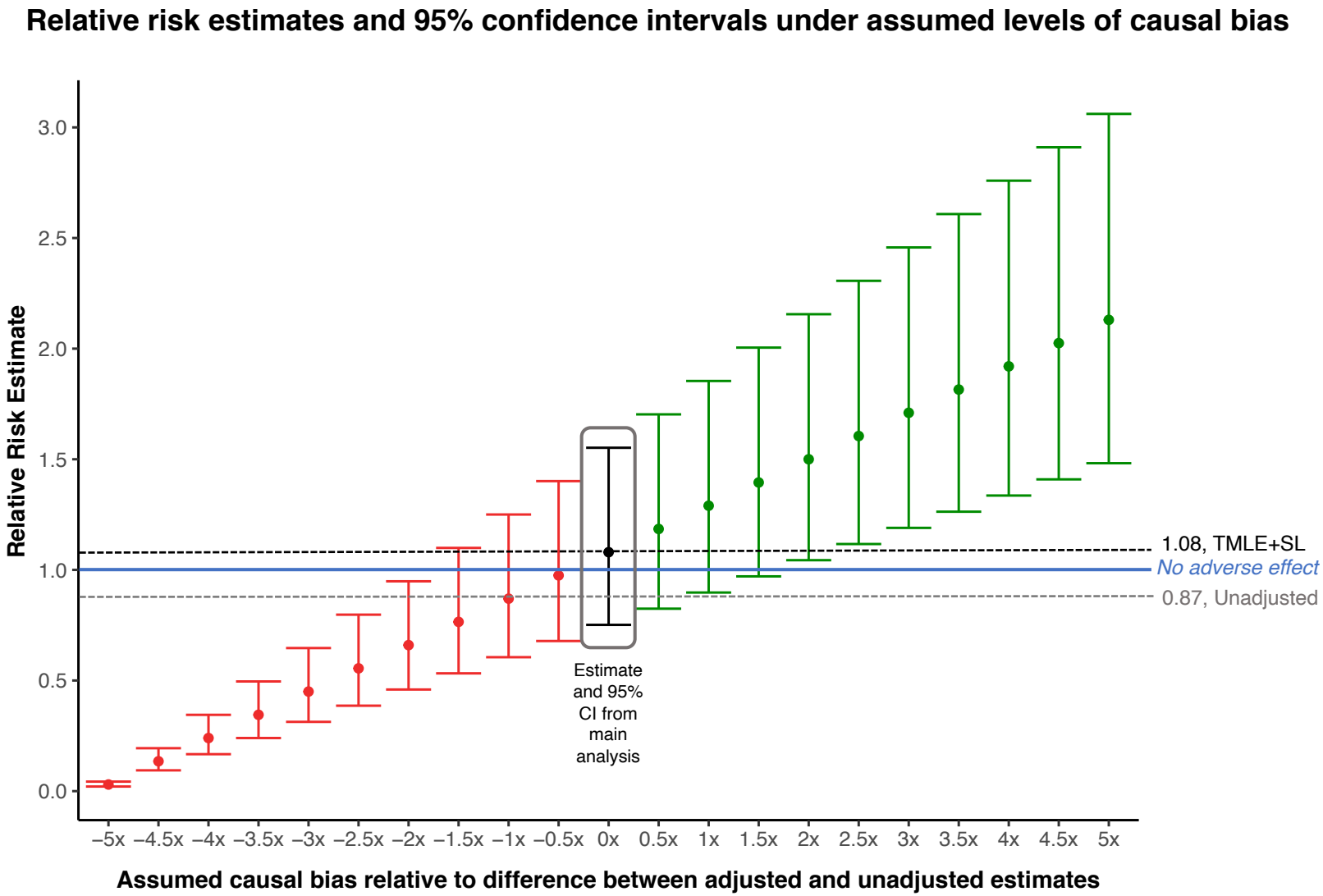
Causal bias: Gap between estimate and truth due to violations of any of the causal assumptions (e.g., unmeasured confounding)*



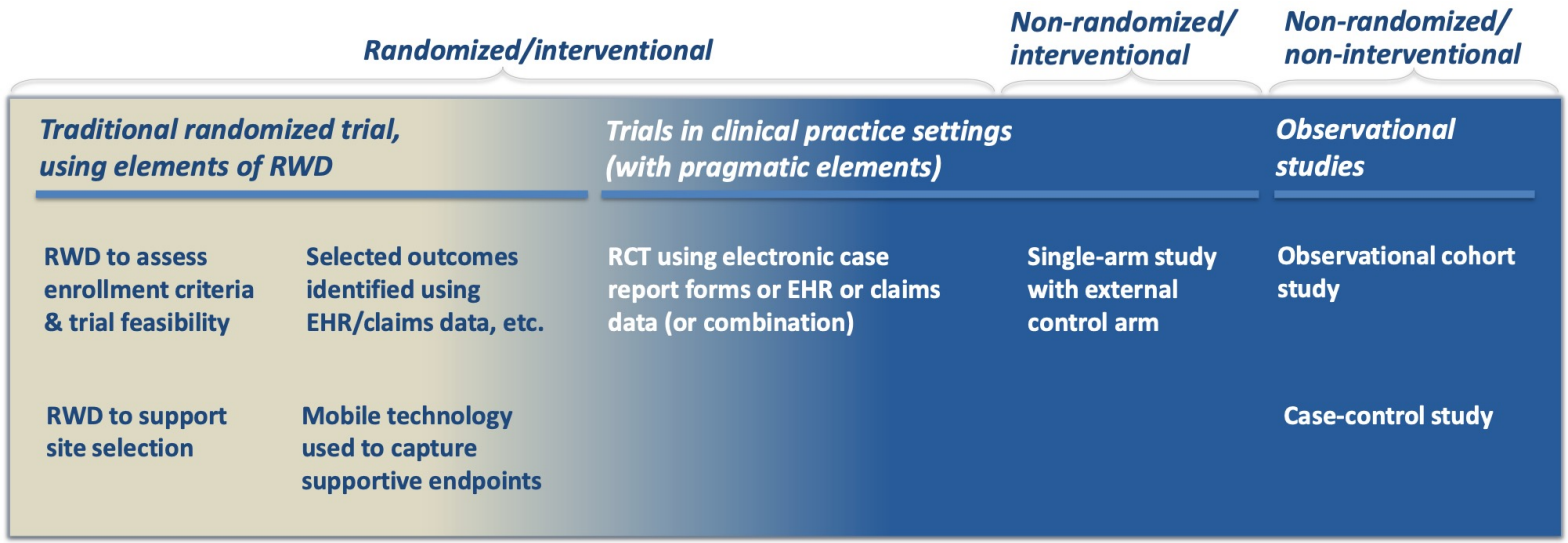
Sensitivity Analysis: Model-free assessment of how reasonable departures from causal assumptions would impact study findings

* Sensitivity analysis can be extended to incorporate statistical bias

TL-based non-parametric sensitivity analysis: Safety analysis example



Targeted Learning with RWD



RWD Challenges

- ❑ Selection bias
- ❑ Intercurrent events
- ❑ Informative missingness
- ❑ Treatment by indication
- ❑ High dimensional covariates
- ❑ Outcome measurement error
- ❑ Statistical model misspecification
- ❑ Differences between external controls and single trial arm RCT

Targeted Learning path supports regulatory decision making

Targeted Learning

- ✓ Roadmap for causal and statistical inference
- ✓ Realistic statistical model
- ✓ Statistical estimand approximates answer to causal question
- ✓ Flexible estimation and dimension reduction with Super Learner
- ✓ Model-free sensitivity analysis
- ✓ Generate RWE with confidence

FDA Funded Demonstration Project

FDA has funded a two year demonstration project of TL (led by Susan Gruber) involving

- Simulations imitating real world studies demonstrating the roadmap and showcasing that TMLE outperforms propensity score matching and other current methods of choice.
- Weekly meetings with senior FDA statisticians and us (S. Gruber, Rachael Philips, MvdL).
- Monthly meetings updating the leadership of real world analytics group at FDA.
- Workshop on TL at FDA
- Publications of various articles reporting on findings.
- Regular seminars on topics in TL, recorded and made public.
- Educational short videos on key concepts in TL.

FDA funded Sentinel Innovation Center on Causal inference with Real World Data

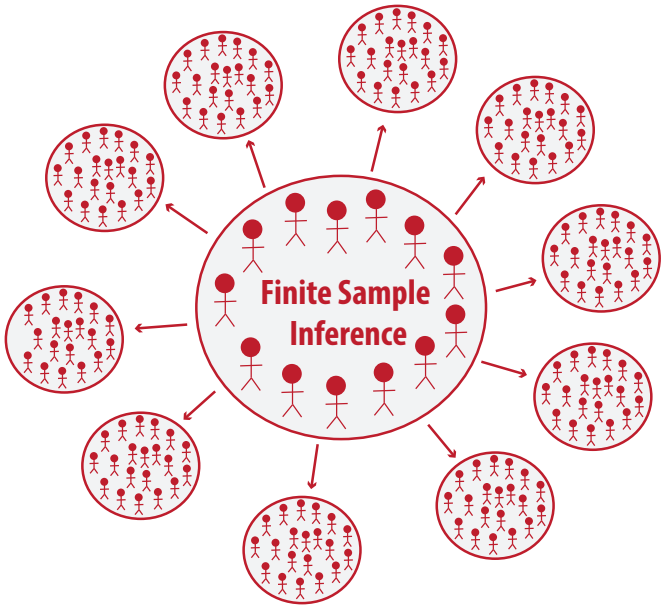
- Sentinel is the FDA national electronic system transforming the way researchers monitor the safety of FDA-regulated medical products. Launched in response to FDA Amendments Act of 2007.
- Innovation Center is led by Department of Pharmacoepidemiology of Harvard University
- Working group includes FDA, Pharma, and academic statisticians.
- Full focus on how to apply TL to real world data sets in Sentinel system, and evaluating its performance relative to other approaches.

Using Innovation Center to showcase how to set up TL Statistical Analysis Plan (SAP)

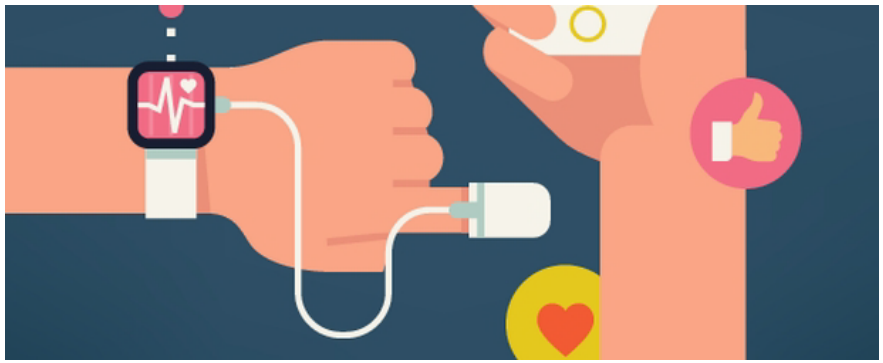
- Specification of a TMLE relies on various choices that can be tailored towards precise application in question: e.g., library of super-learner; truncation method; type of TMLE, e.g., collaborative TMLE or not.
- We use outcome blind version of data set in question to set up simulation of (similar) data sets for which we know the truth.
- We then then select a TMLE that performs best w.r.t. coverage of 0.95 confidence intervals, bias and mean squared error, optimizing power while controlling type-I error and coverage.
- Initial results demonstrate for rare outcomes C-TMLE is superior thereby providing the choice of SAP, which will then be applied to real data.

Future of Targeted Learning

FDA U.S. FOOD & DRUG ADMINISTRATION



novo nordisk®



accenture