Real World Evidence Community Forum

The Causal Roadmap, Targeted Learning and TMLE: What Is That All About?

phuse.global

Hosts & Presenters

connect·share·advance

Hosts:

- Elizabeth Merrall, *Johnson & Johnson;* Mary Anne Rutkowski, *Merck;* Nicky Newton, *PHUSE*

Presenters:

Mark van der Laan

Ela Talpes

Disclaimer: The contents, opinions and view expressed in this presentation are of authors only and not of the organisations they represent.

Content

connect·share·advance

- Introduction – 5’
- Mark– 20’
- Q&A – 5’
- Ela – 20’
- Q&A – 10’

Please enter any questions you have for the presenters in the Q+A box within Zoom

Disclaimer: The contents, opinions and view expressed in this presentation are of authors only and not of the organisations they represent.

Content

connect·share·advance

Working Group Scope

The Real-World Evidence Working Group aims to support, address and answer pertinent questions around real-world evidence. The Working Group is dedicated to sharing across the PHUSE Community (through Community Forums) and aligning on the best industry practices. Some of the questions we intend to address are:

- What are the requirements, technologies and processes/best practices needed to
- use real-world data as a source for data analysis?
- incorporate real-world data into clinical trials?
- to support the use of real-world evidence as part of regulatory submissions?

Several Projects

- Applying Advanced Data Privacy Methods to Real World Data (RWD)
- Using OMOP and Other Real World Data Standards to Support Regulatory Submissions
- Estimands For RWD/RWE
- Quality & Reusability of RWD
- RWD Guideline for Programming & Analysis Processes
- Submitting RWD
- Missing Data Imputation in RWD Exploration of Multiple Techniques on Open-Source Data

Blog Articles & White Papers
Community forums
Conference activities

https://phuse.global/Working_Groups

If you would like to get involved, email workinggroups@phuse.global

ESTIMANDS FOR RWD/RWE

Estimands for RWD/RWE

connect·share·advance

- **White Paper development (subteam 1)**
  - Co-leads: Ksenia Titorenko (ICON) and Paramita Chakraborty (IQVIA / Gilead)
  - *Establishing Robust Estimands in Real-World Evidence*
    - Literature review
    - Gap analysis
    - Combining frameworks
    - Example use cases
    - Regulatory submission aspects
    - etc
  - Preparing for internal PHUSE project team review
  - Future public review before published

Estimands for RWD/RWE

connect·share·advance

- **Webinar series (subteam 2)**
  - Lead: Matt Baldwin (Amgen)
  - *At the Intersection of Estimands and Target Trial Emulation (TTE) for RWE*
  - 22Apr2025 – Webinar 1 recording / slides:
    - *An Introduction to the Estimands and Target Trial Emulation Frameworks*
  - 05Jun2025 – Webinar 2 recording / slides:
    - *Estimands in Real-World Evidence Studies*
  - 21Jul2025 – PHUSE Blog: *What's the Hype About? PHUSE's Record-Breaking Webinars Explore Estimands and Target Trial Emulation for RWE*
  - 24Feb2026 – Webinar 3 registration:
    - *Implementing Estimands & Target Trial Emulation (TTE): Case Studies and Perspectives*

Targeted Machine Learning, Highly Adaptive Lasso, and Causal Inference for Generating RWE in Drug Development

Mark van der Laan

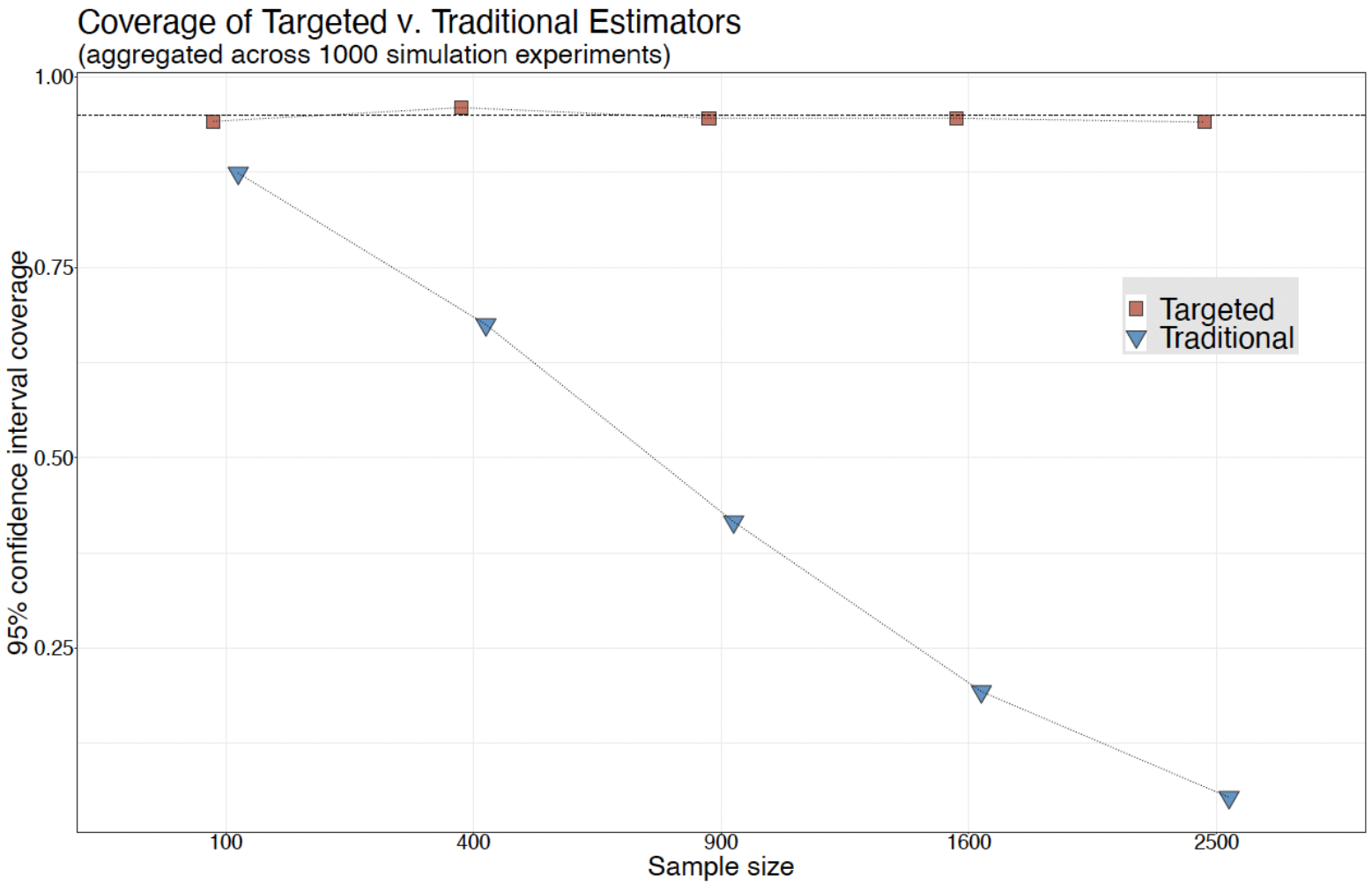
Jiann-Ping Hsu/Karl E. Peace Professor in Biostatistics & Statistics
University of California, Berkeley

January 29, 2026, RWE Forum

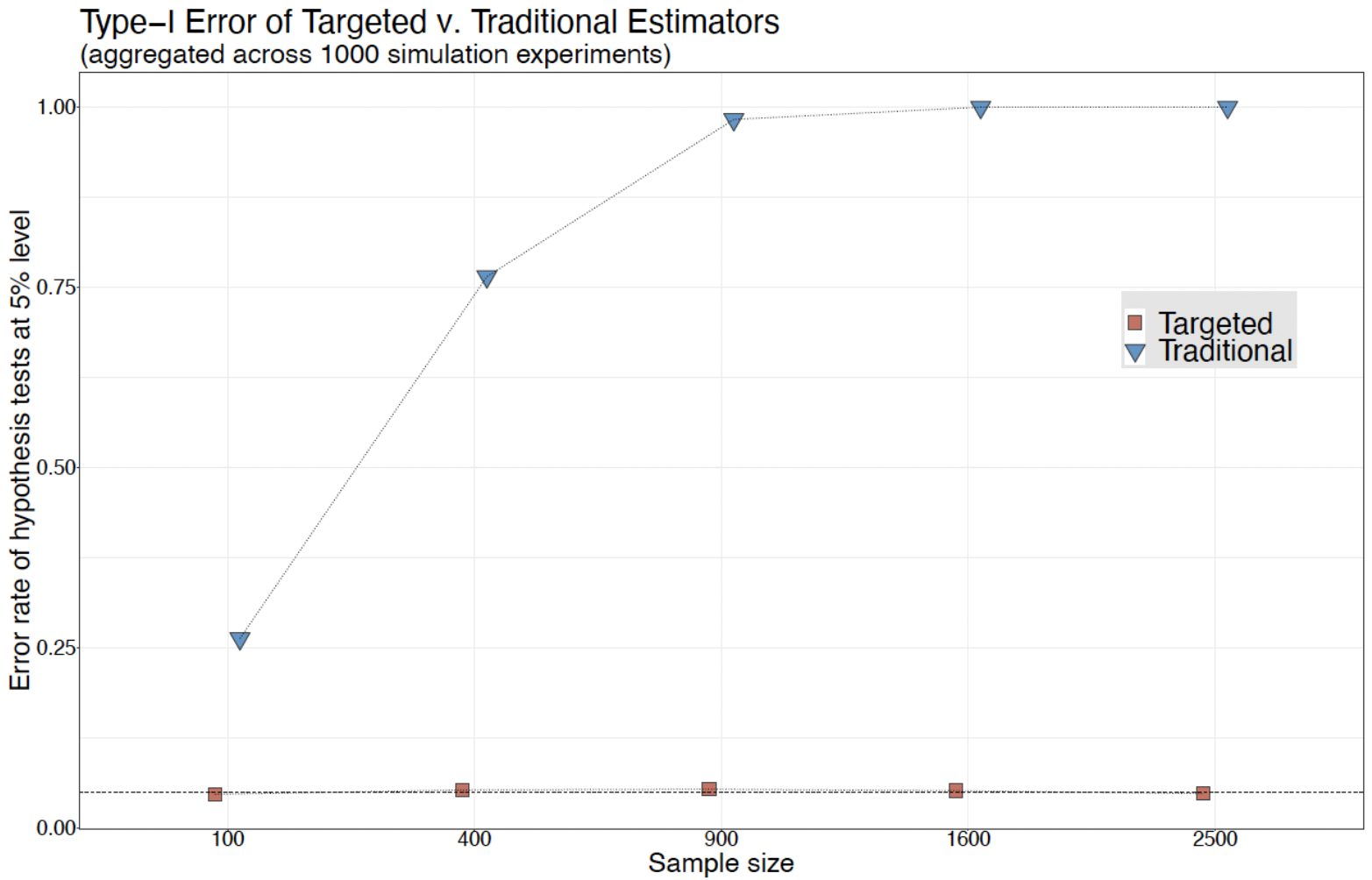
Traditional toolbox for statistics: Recipe oriented, enforces false constraints, not made for Big Data

	Type of Data			
Goal	**Measurement (from Gaussian Population)**	**Rank, Score, or Measurement (from Non- Gaussian Population)**	**Binomial (Two Possible Outcomes)**	**Survival Time**
Describe one group	Mean, SD	Median, interquartile range	Proportion	Kaplan Meier survival curve
Compare one group to a hypothetical value	One-sample ttest	Wilcoxon test	Chi-square or Binomial test **	
Compare two unpaired groups	Unpaired *t* test	Mann-Whitney test	Fisher's test (chi-square for large samples)	Log-rank test or Mantel-Haenszel*
Compare two paired groups	Paired *t* test	Wilcoxon test	McNemar's test	Conditional proportional hazards regression*
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test	Cox proportional hazard regression**
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q**	Conditional proportional hazards regression**
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients**	
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression**	Simple logistic regression*	Cox proportional hazard regression*
Predict value from several measured or binomial variables	Multiple linear regression* or Multiple nonlinear regression**		Multiple logistic regression*	Cox proportional hazard regression*

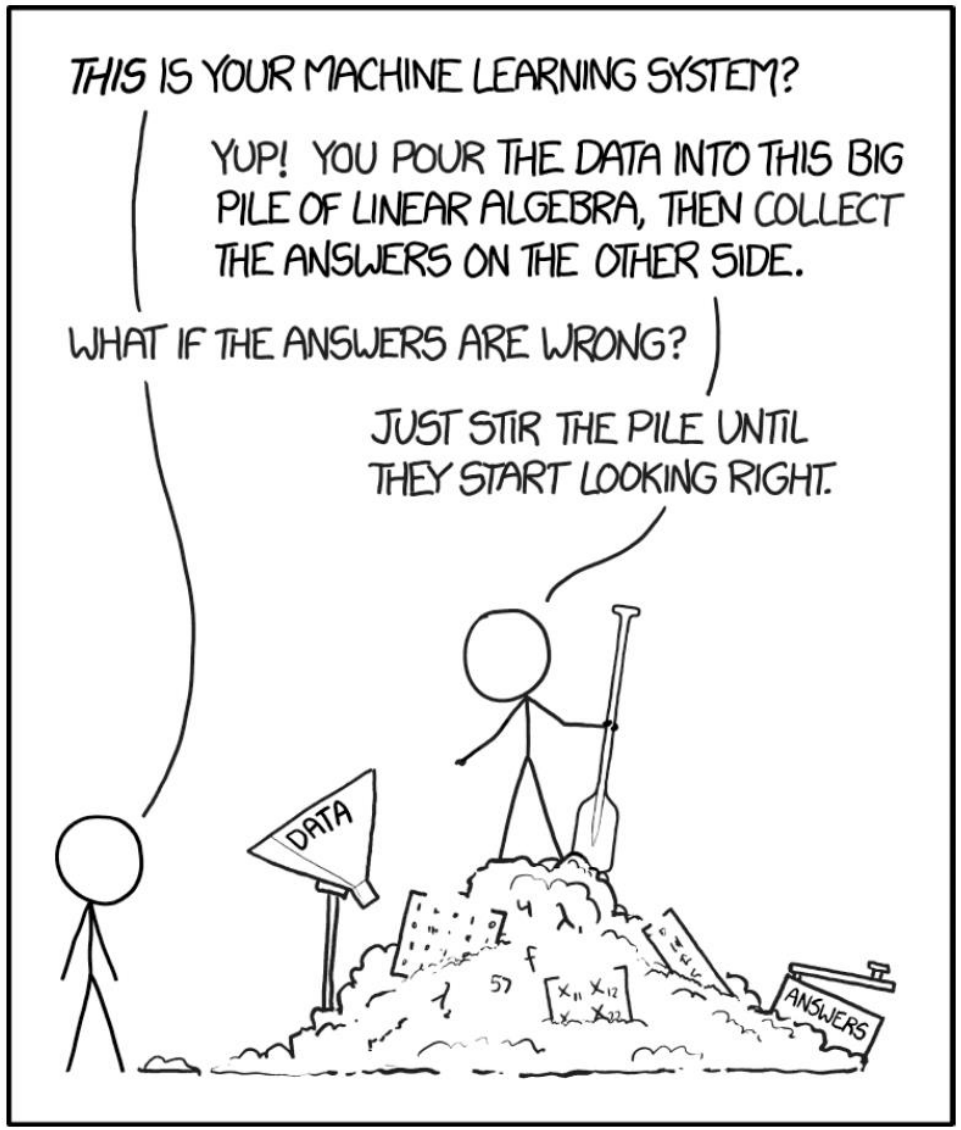
Performance of traditional tools: Coverage of Confidence Intervals deteriorates with sample size



Performance of traditional tools: Type I error deteriorates with sample size



Traditional tools invite/encourage post-hoc model manipulation



Why care about statistical inference?

Why Most Published Research Findings Are False

John P. A. Ioannidis

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

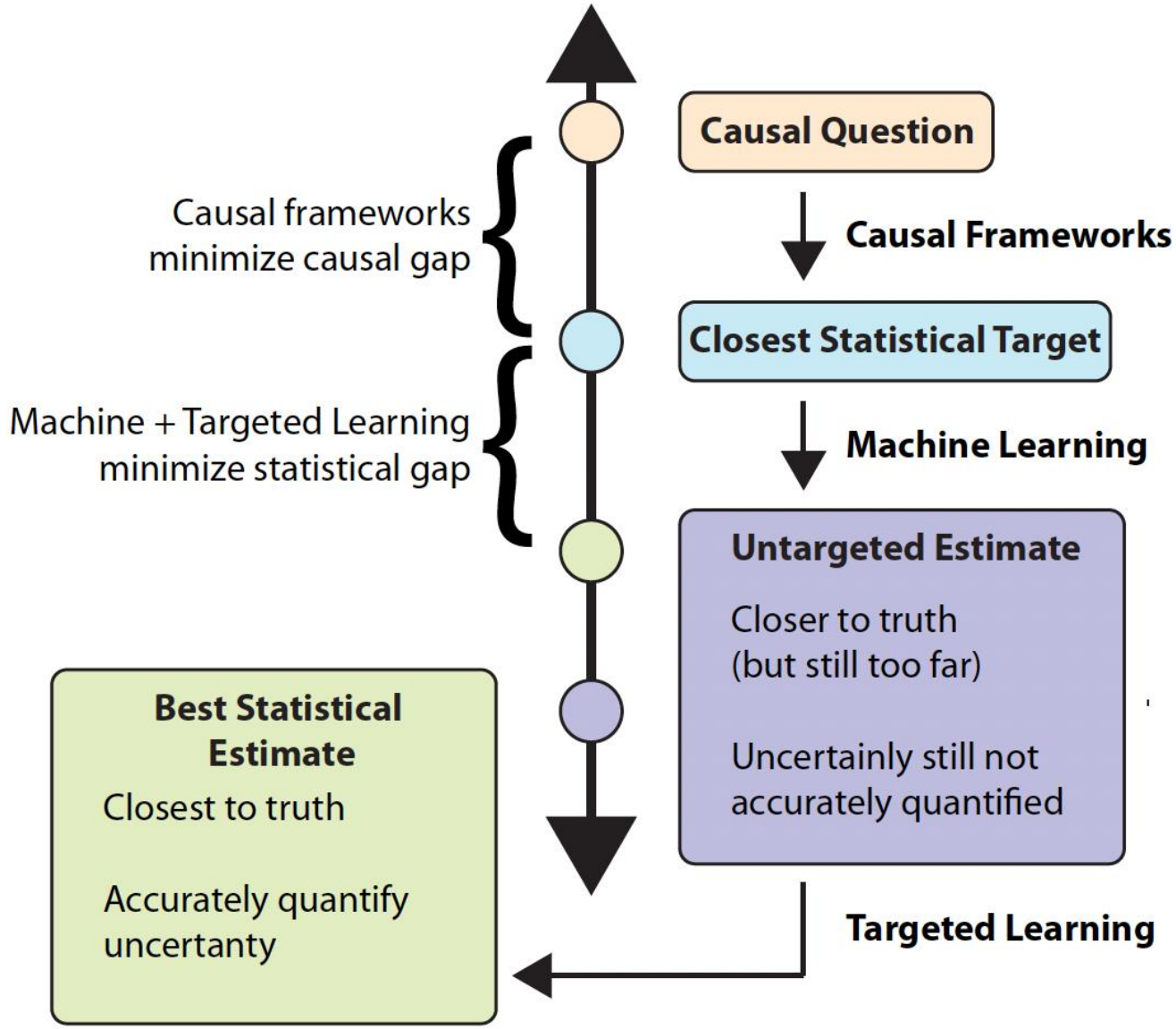
Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

The Statistical Crisis in Science

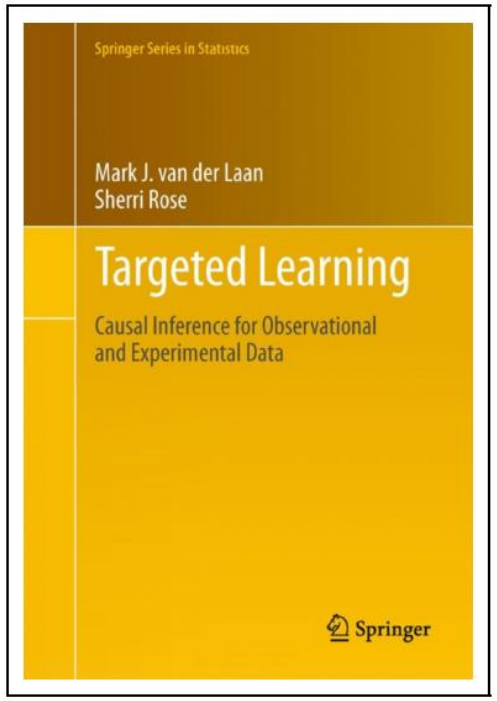
Data-dependent analysis—a "garden of forking paths"— explains why many statistically significant comparisons don't hold up.

Andrew Gelman and Eric Loken

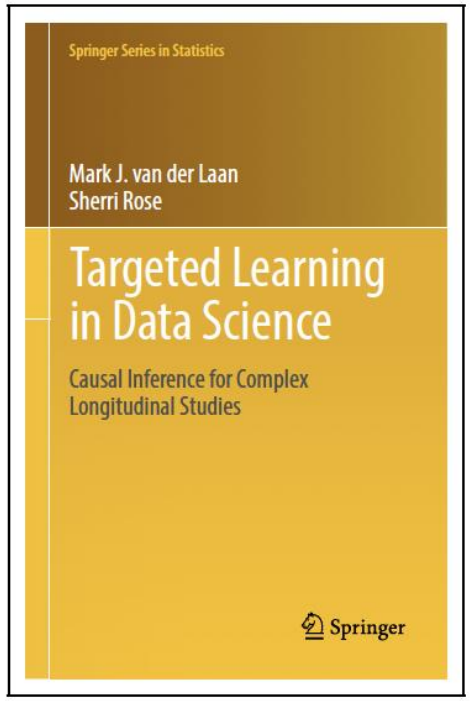
Targeted Learning for answering statistical and causal questions with confidence intervals



Targeted Learning is a subfield of statistics



van der Laan & Rose, *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer, 2011.



van der Laan & Rose, *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. New York: Springer, 2018.

The Hitchhiker's Guide to the `tlverse`

Better clinical decisions from observational data

Statistics in Medicine

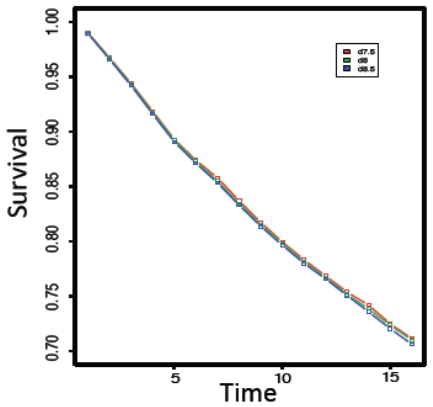
Research Article

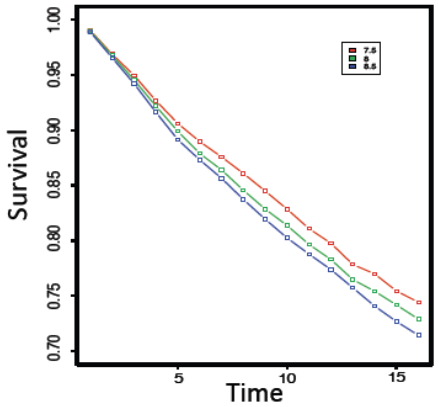
Received 24 May 2013, Accepted 5 January 2014 Published online 17 February 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6099

Targeted learning in real-world comparative effectiveness research with time-varying interventions

Romain Neugebauer,[a*†] Julie A. Schmittdiel[a] and Mark J. van der Laan[b]

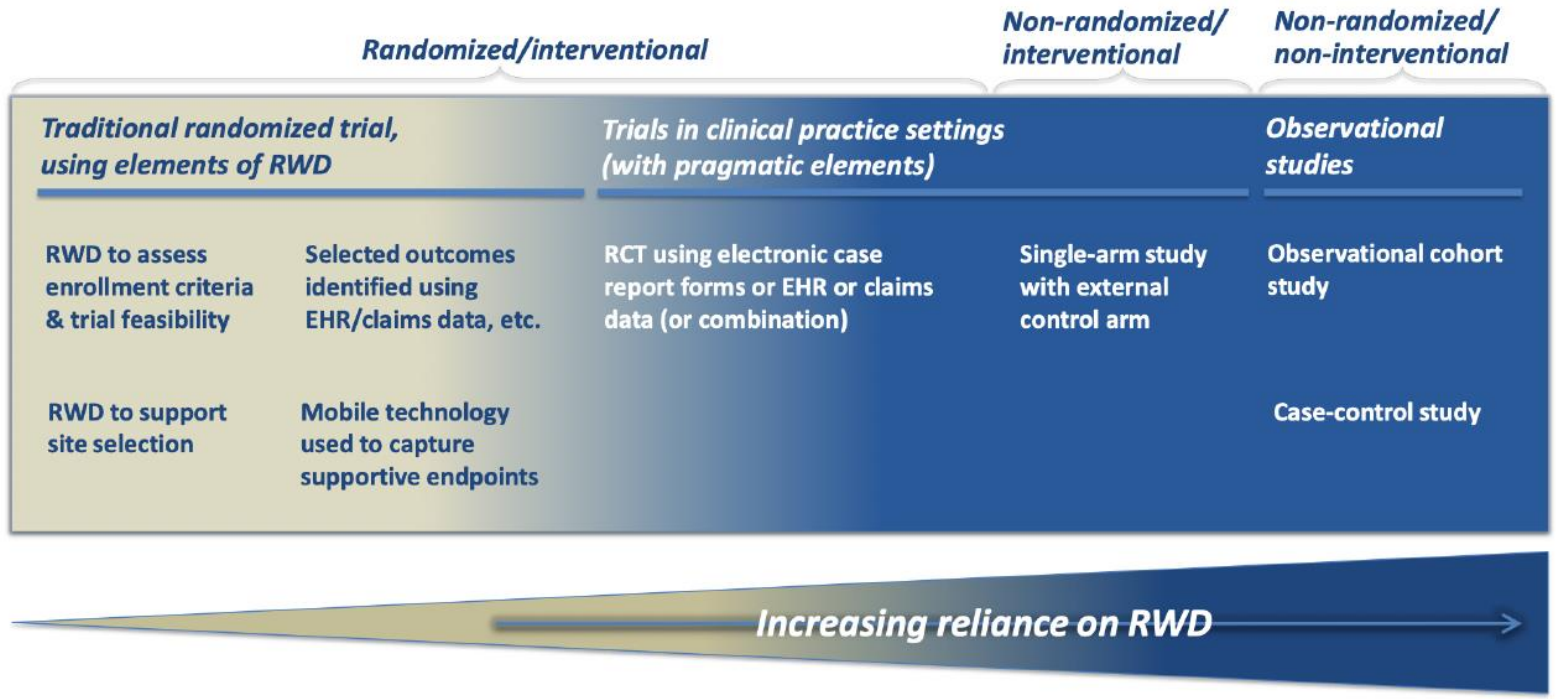




Standard methods: No benefit to more aggressive intensification strategy

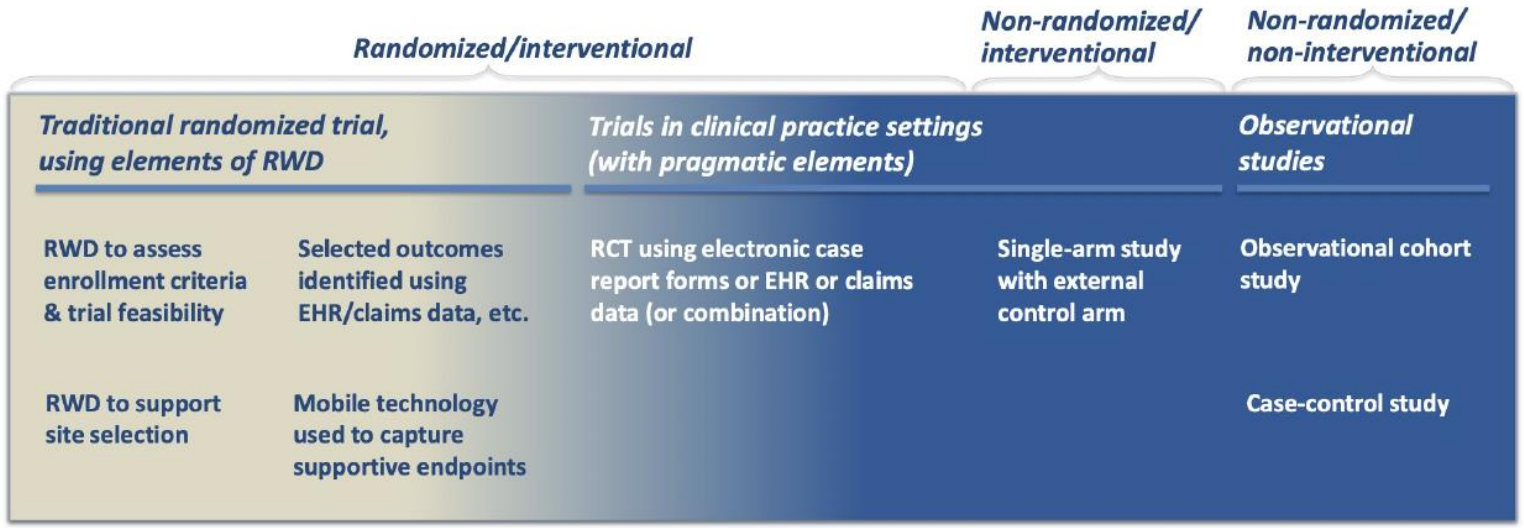
Targeted Learning: More aggressive intensification protocols result in better outcomes

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 targeted 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

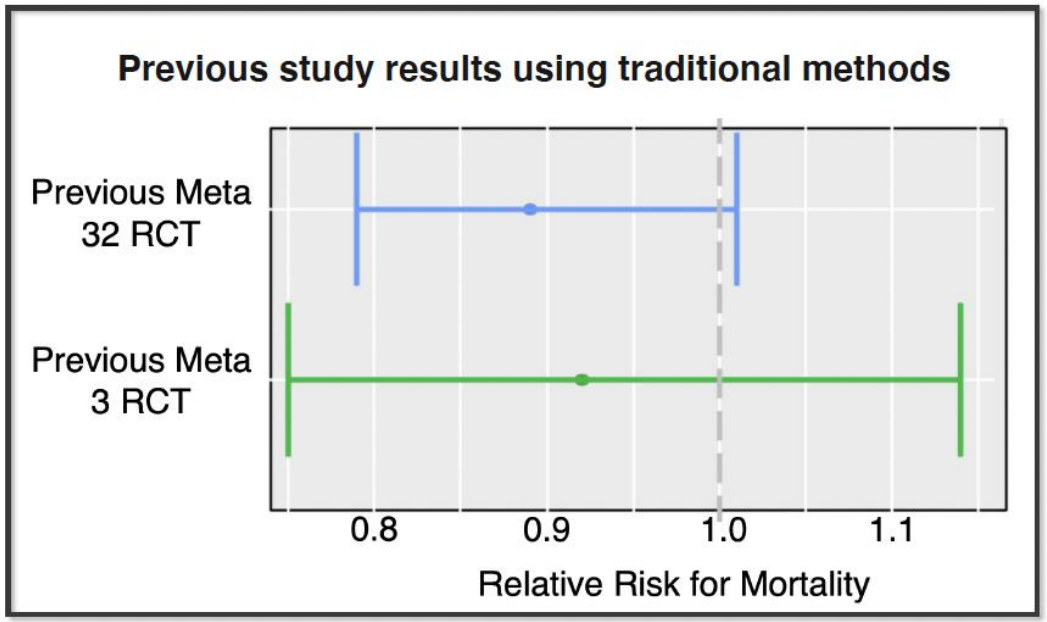
STEP 3: DEFINE STATISTICAL QUERY

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

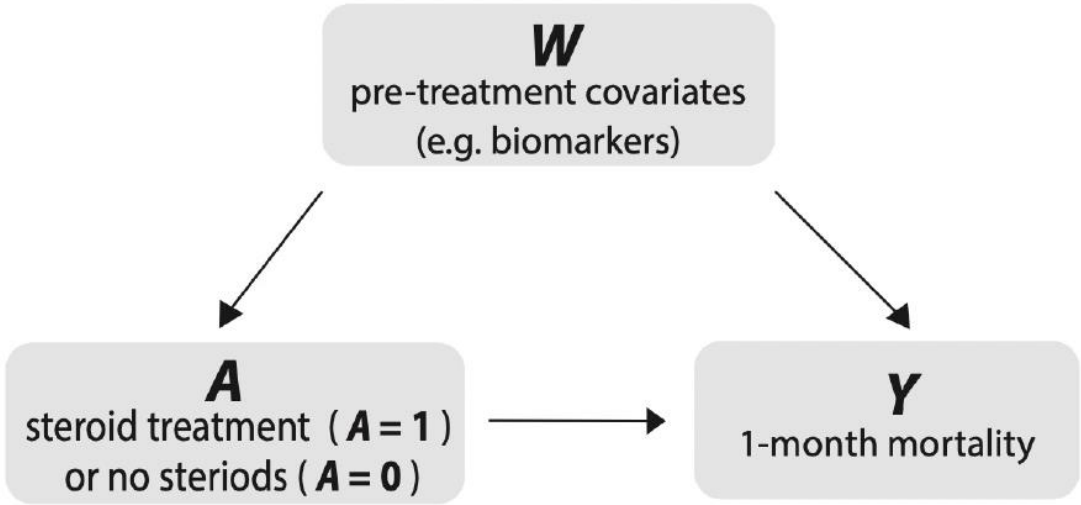
STEP 4: CONSTRUCT ESTIMATOR

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

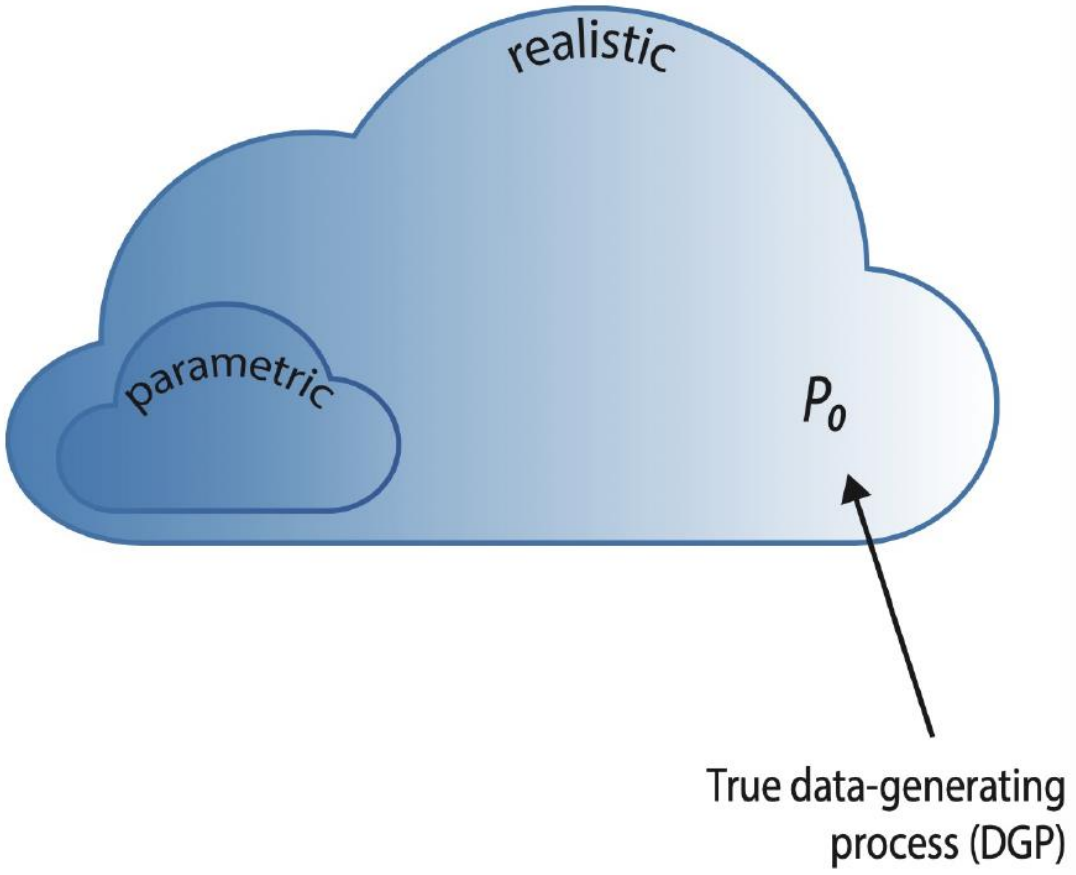
STEP 2:
SPECIFY
STATISTICAL MODEL

STEP 3:
DEFINE STATISTICAL
QUERY

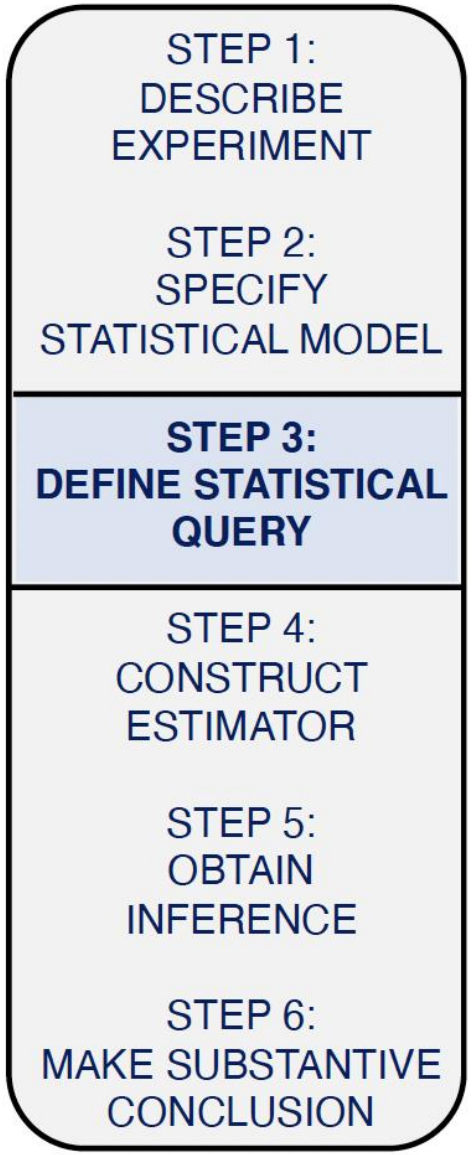
STEP 4:
CONSTRUCT
ESTIMATOR

STEP 5:
OBTAIN
INFERENCE

STEP 6:
MAKE SUBSTANTIVE
CONCLUSION



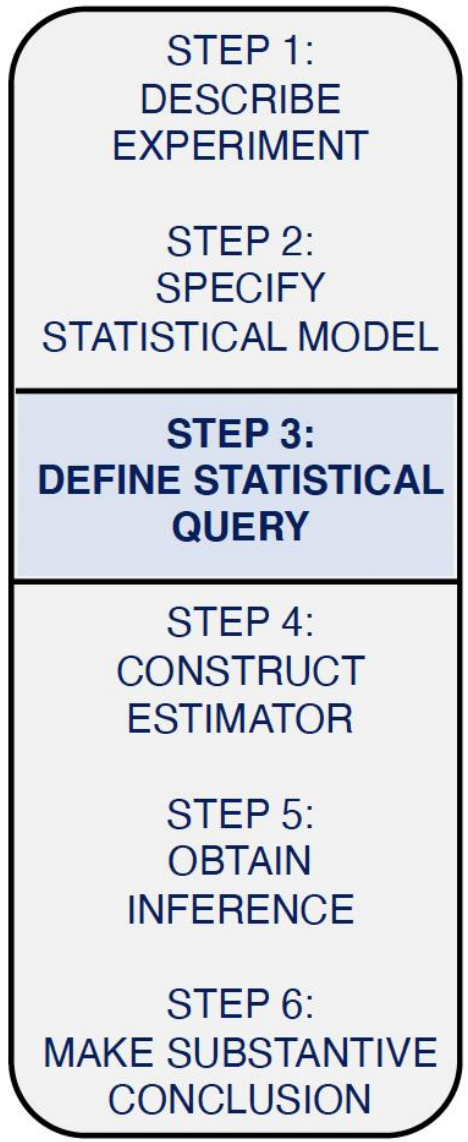
Step 3a: What is the target causal estimand that we aim to identify from the data?



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

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

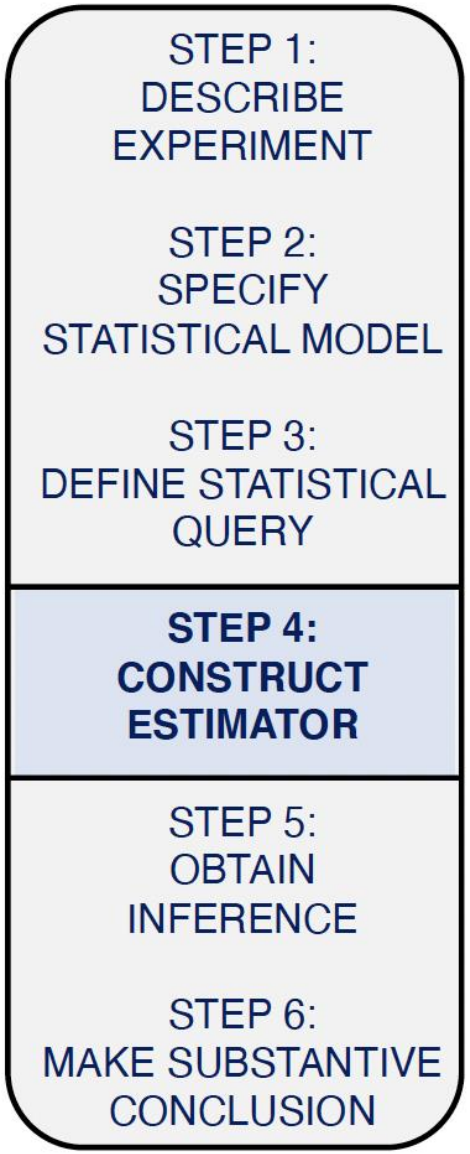
Step 3b: What is the target statistical estimand that we will learn from the data?



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

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

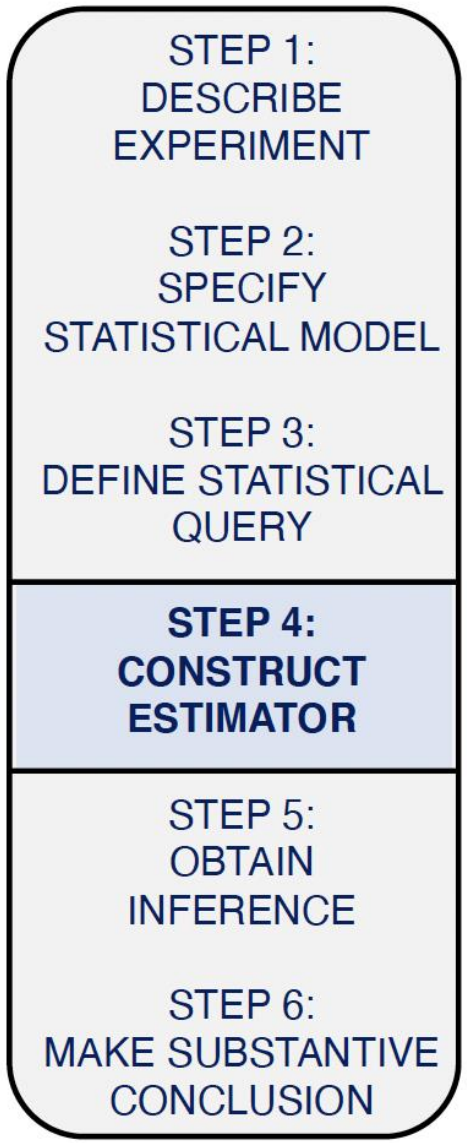
How should we estimate the target estimand?



Statistical properties to consider

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

Targeted Maximum Likelihood Estimation (TMLE)



TMLE

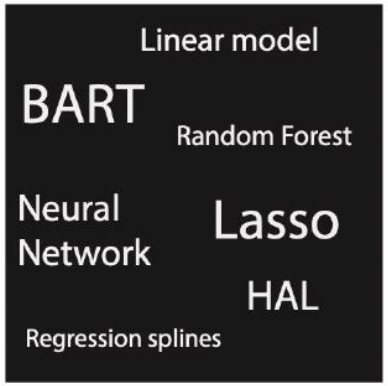
1. Initial estimation of $E[Y|A,W]$ with super (machine) learning
2. Updating initial estimate to acheive optimal bias-variance trade-off for ψ_{stat}

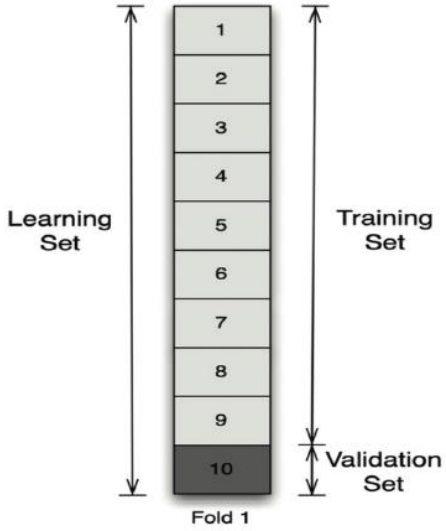
TMLE estimates are optimal:

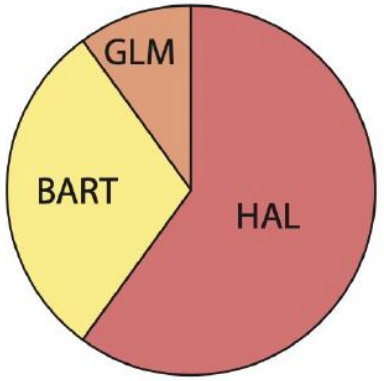
plug-in, efficient, unbiased, finite sample robust

TMLE Step 1: Super learner

Cross-validated performance of learners + ensembles







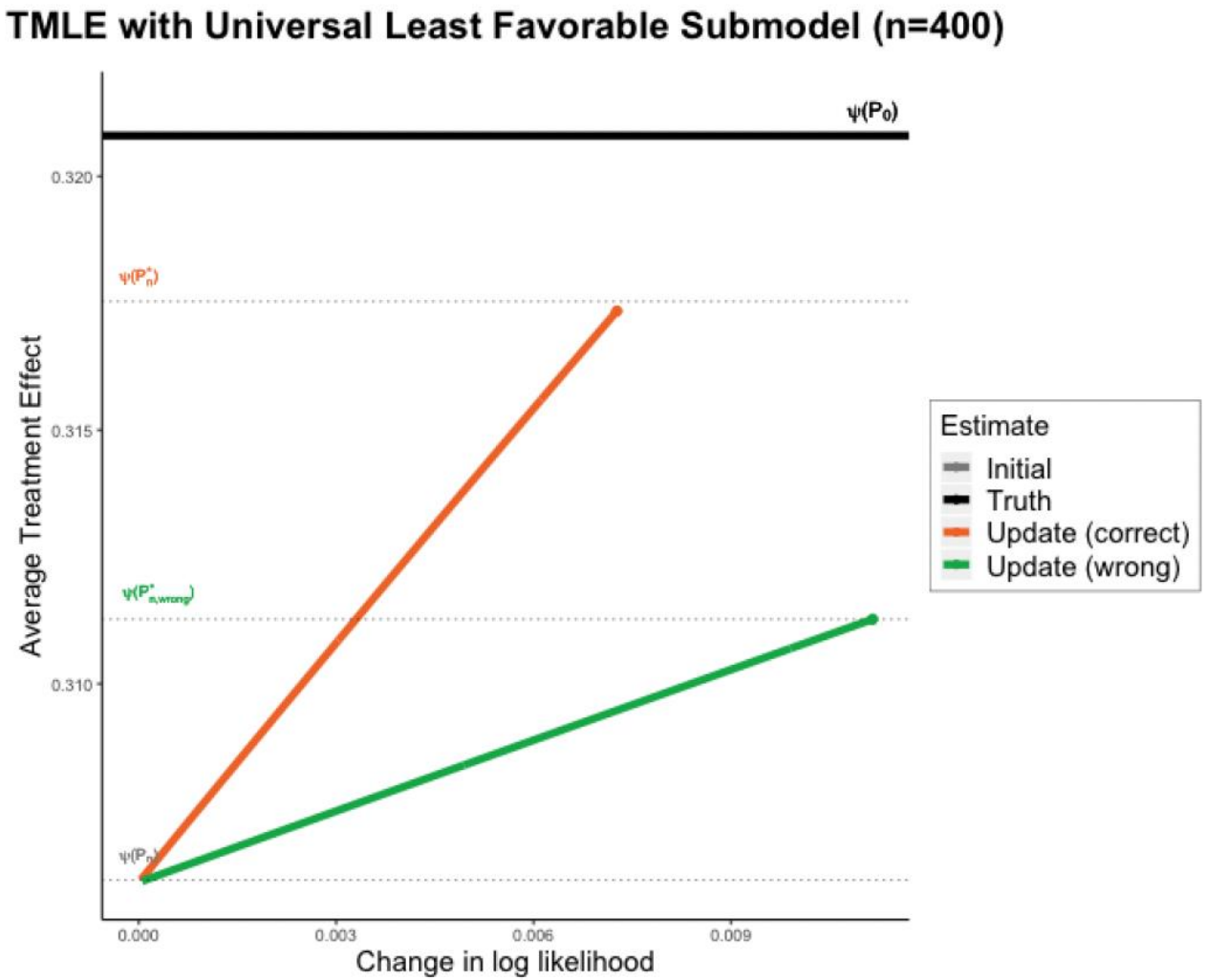
Hugely advantageous when coupled with NLP-derived covariates with EHR

Highly Adaptive Lasso (HAL)

- Any d-dimensional cadlag function (i.e. right-continuous) can be represented as an nfinite linear combination of spline basis functions.
- The variation norm / complexity of a function is the L_1-norm of the vector of coefficients in this representation.
- HAL minimizes the empirical risk, such as MSE for least squares regression, over all cadlag functions with variation norm bounded by a constant.
- This L_1-norm is chosen with cross-validation.

Zero-order HAL converges to true function at rate $n^{-1/3}(\log n)^{d/2}$ (only assuming finite variation norm!)

TMLE Step 2: Targeting follows a path of maximal change in target estimand per unit likelihood



Arriving at the substantive conclusion

STEP 1: DESCRIBE EXPERIMENT

STEP 2: SPECIFY STATISTICAL MODEL

STEP 3: DEFINE STATISTICAL QUERY

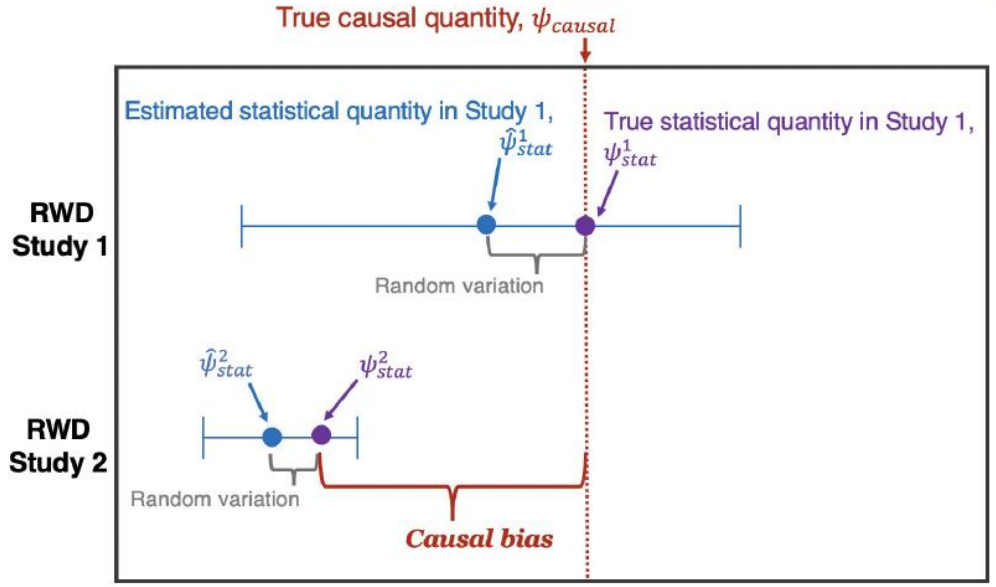
STEP 4: CONSTRUCT ESTIMATOR

STEP 5: OBTAIN INFERENCE

STEP 6: MAKE SUBSTANTIVE CONCLUSION

Investigate causal bias with sensitivity analysis

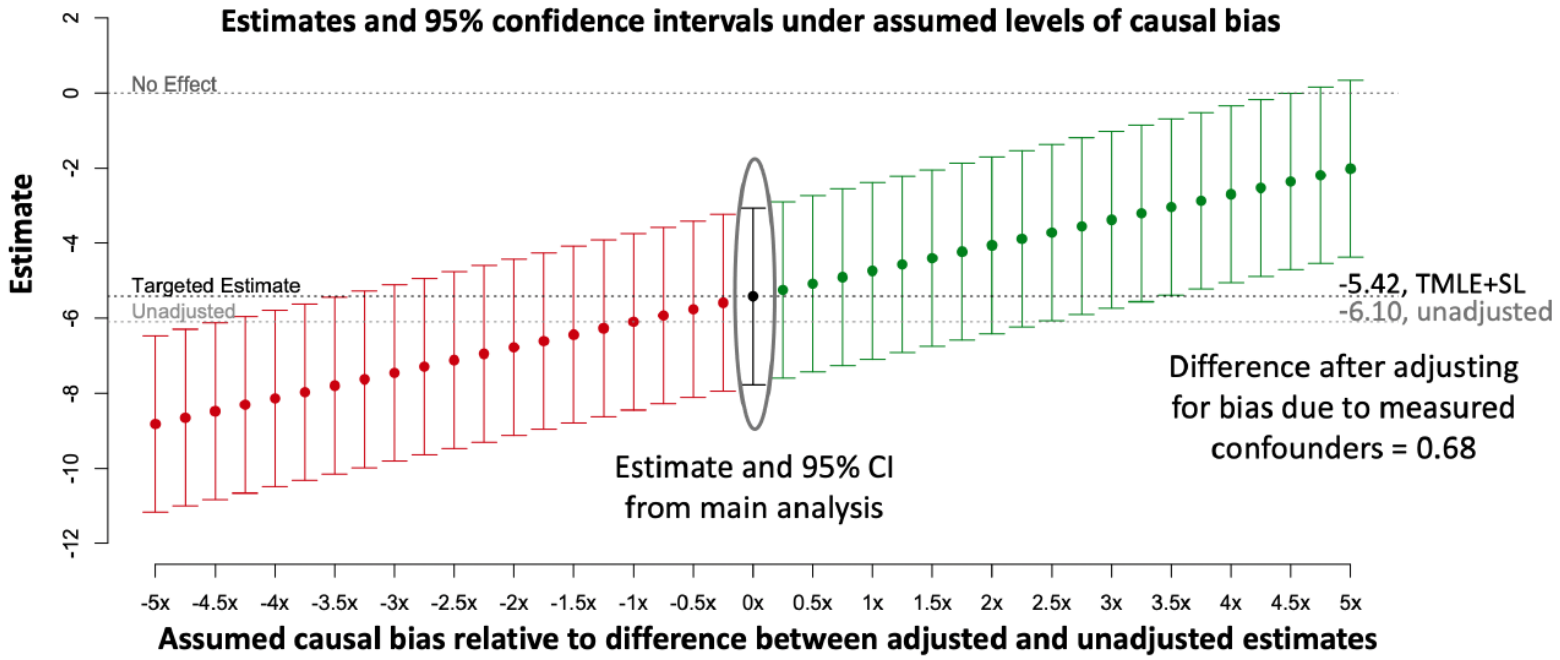
Causal bias: Gap between estimate and truth due to violations of any of the causal assumptions (e.g., unmeasured confounding)*



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

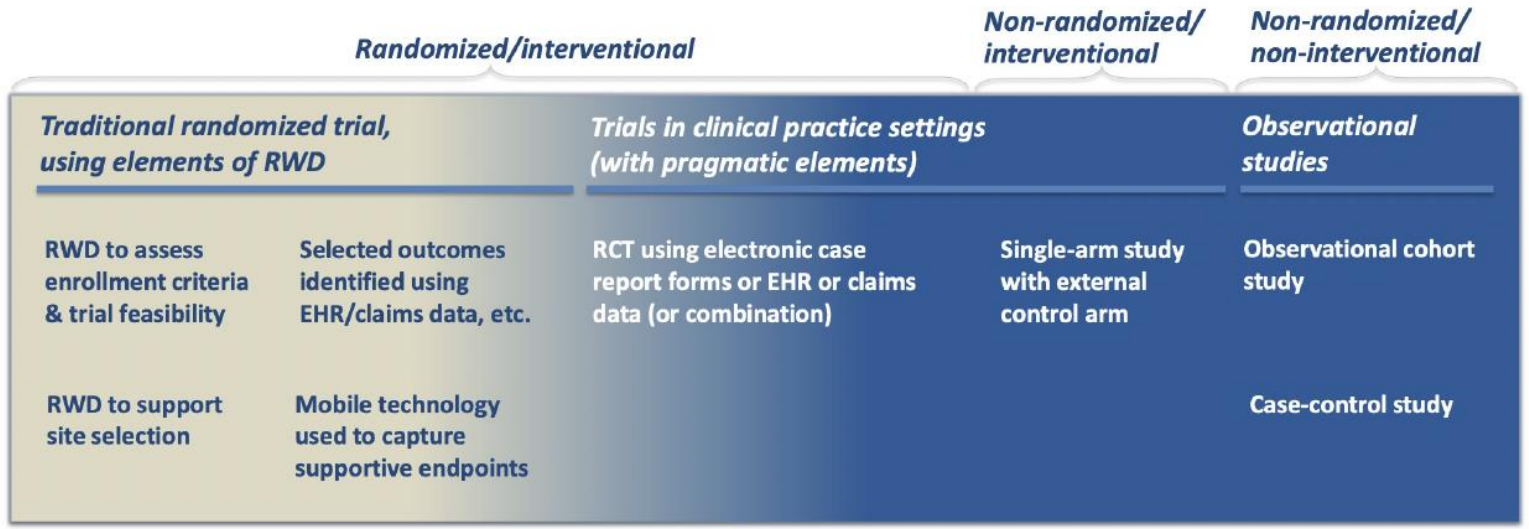
* Sensitivity analysis can be extended to incorporate statistical bias

TL-based non-parametric sensitivity analysis RCT with 25% LTFU example



Courtesy of "Targeted-Learning Based Statistical Analysis Plan" Webinar by Susan Gruber on 28 April 2021

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

Outcome blind simulations to define specification of TMLE: Wyss et al., 2023

Step 4. Estimation and Inference using TMLE + SL

TMLE software default settings are robust across many situations, ***but not optimized for your analysis***

Recommendation

Compare **coverage, type I error, power** for different specifications using outcome blind simulations, or plasmode simulations based on external data (pilot, Phase 1) sharing key features with your data

R. Wyss et al. (2023). Targeted Learning with the Collaborative Controlled Lasso for Large-Scale Covariate Adjustment in Healthcare

Step 4. Estimation and Inference using TMLE + SL

Case study 1: Evaluating safety and efficacy with **high dimensional RWD**

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

Challenges

- Lack of baseline randomization
- How to model propensity score and outcome regression?
  - Thousands of covariates in linked EHR + claims
  - Traditional parametric model too high dimensional, unknown functional form

Step 4. Estimation and Inference using TMLE + SL

Wyss investigated multiple specificiations

(only 4 discussed today)

- Outcome regression used Lasso
  - 91 expert-selected covariates forced into the model, all others were main term candidates
- PS estimation: considered four variants of lasso
  - Investigate to find settings that minimize mean squared error (MSE) in plasmode simulation studies

synthetically generate datasets to combine study data with simulated causal effects
- bootstrap covariates from dataset
- simulate treatment & outcome with desired prevalence

Lasso for estimating PS (four variants)

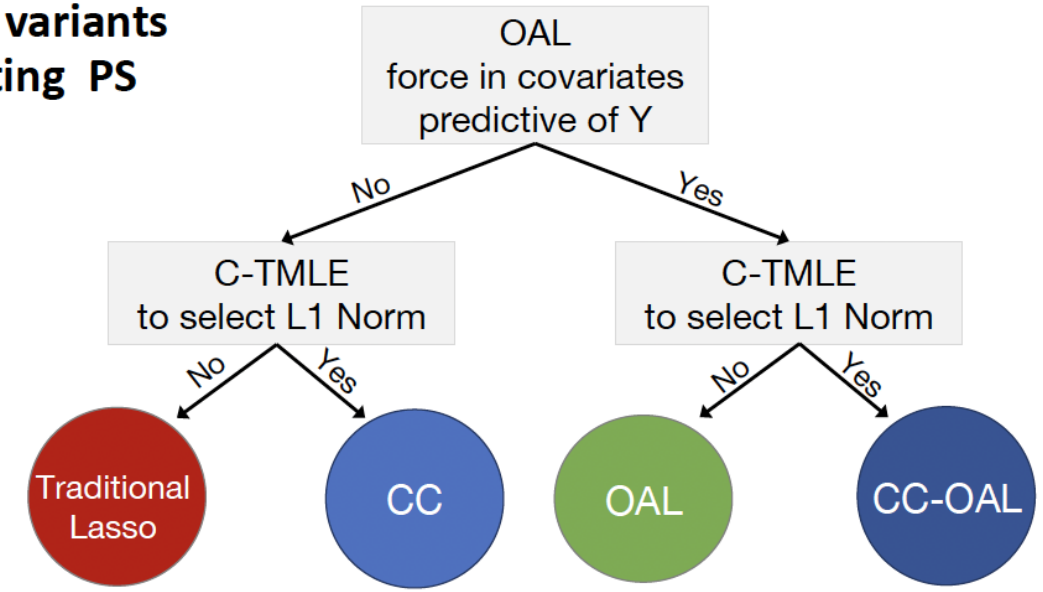
Considerations

- Lasso or Outcome-adaptive Lasso (OAL)*
  - OAL favors variables associated with outcome by assigning covariate-specific lasso penalties.
    - covariates strongly associated with outcome are less penalized
  - Wyss forced covariates previously selected for outcome regression into PS model, and all others were candidates for inclusion
- Choose L1 norm based on goodness-of-fit for treatment or goodness-of-fit for outcome (*"collaborative control"* using C-TMLE)
  - "Collaborative control" (CC) chooses L1-norm based on increase in likelihood for Q after TMLE update

*Shortreed SM, Ertefaie A. Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*. 2017 Dec;73(4):1111-1122.

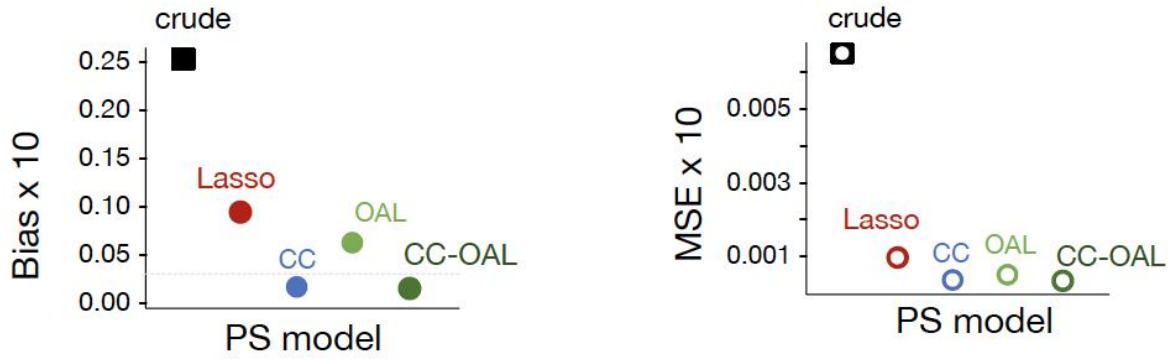
Step 4. Estimation and Inference using TMLE + SL

Four Lasso variants for estimating PS



Step 4. Estimation and Inference using TMLE + SL

Plasmode study results



Collaborative control greatly reduced bias and improved MSE

- Less regularization captured more relevant confounder information in PS

Step 4. Estimation and Inference using TMLE + SL

Case study 1: Data analysis

- **Analyze study data**
  - Outcome regression: Lasso with expert-selected covariates forced in
  - PS: Collaborative control Outcome-adaptive lasso (CC-OAL)
    - force in covariates selected in the outcome, others as candidates
    - C-TMLE to choose L1 norm
- **Result: RD = 0.005 (95% CI: -0.027, 0.038)**

  crude RD = 0.024 (95% CI: 0.018, 0.030)

Crude estimate indicates NSAIDS increase risk for AKI

Adjustment moves point estimate towards the null

The Roadmap in Action: Case Studies & Ongoing Projects

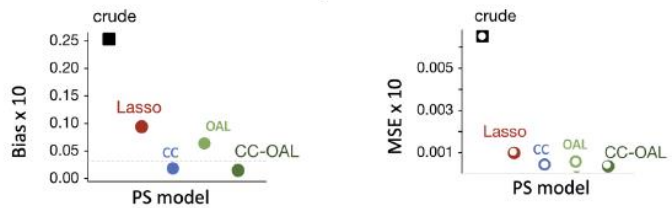
Berkeley
UNIVERSITY OF CALIFORNIA

FDA Sentinel Innovation Center

- **Safety evaluation with high dimensional data:**
  Wyss et al. (2024), AJE, Targeted Learning with an Undersmoothed Lasso Propensity Score Model for Large Scale Covariate Adjustment i Healthcare Database Studies.

- **Subset calibration/two-stage designs:**
  -Ongoing project evaluating methods such as the two-stage design TMLE for study designs that involve a subset of subjects with carefully curated confounders and or outcomes, and a remaining set of subjects.
  -This is a common type of design to obtain desired causal identification from RWD while still gaining efficiency from the less curated data set.

Plasmode study results



Collaborative control greatly reduced bias and improved MSE
- Less regularization captured more relevant confounder information in PS

These projects involve multi-author working groups with FDA/Pharma/Academics/Kaiser Permanente.

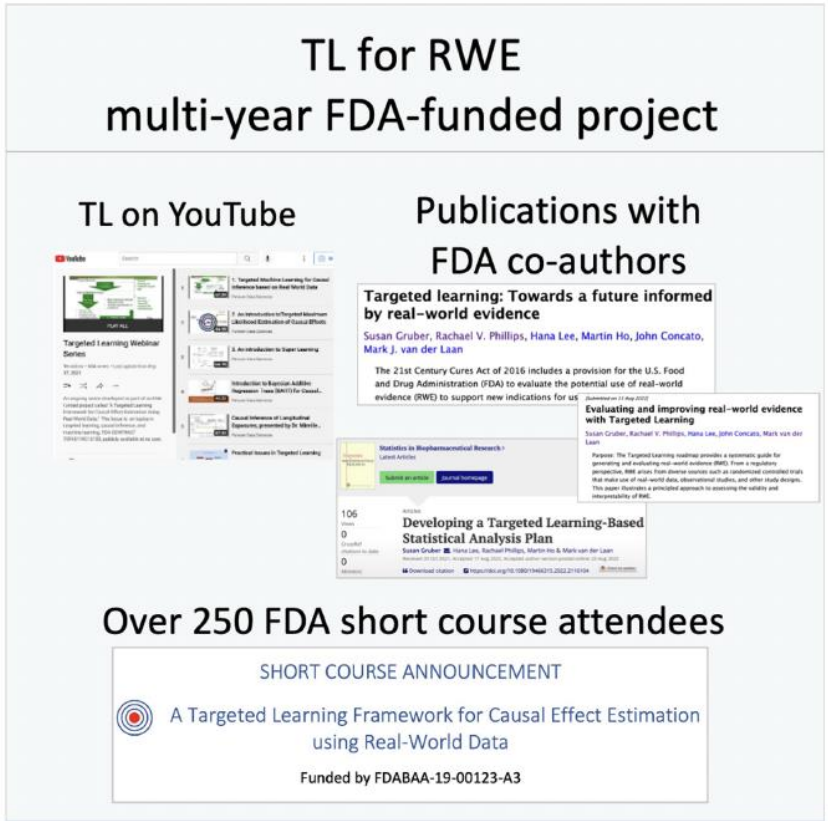
The Sentinel Innovation Center is funded by the FDA through the Department of Health and Human Services (HHS) Task order 75F40119D10037.

Deep LTMLE: Toru Shirakawa, Yi Li, Sky Qiu, vdL

- LTMLE (Bang, Robins, 2005, Petersen et al., Gruber, vdL, etc, relies on **sequential regression**.
- Targeted **maximum likelihood** requires estimation of all conditional densities (vdL 2010a,b; Stitelman etal (2011)).
- Rijtgaard, Gerds, vdL (2023) **combines** TM-likelihood based estimation of **intensities** with estimation of a **conditional mean function integrating over time dependent covariates**.
- Toru et al. Deep LTMLE utilize **transformer architecture for temporal difference learning**, simultaneously modeling and estimating the sequential regressions: 1) *Computationally superior*; 2) continuous time monitoring; 3) large histories; 4) TMLE update step uses transformer gradient descent algorithm.

Comparative Effectiveness: Targeted Learning FDA Demonstration Project

Resulted in various collaborative relations with FDA statisticians



Berkeley UNIVERSITY OF CALIFORNIA

Comparative Effectiveness: GLP1-RA and Dementia in the Danish National Registry

Population: Insulin-naive T2-DM initating 2nd-line therapy (N=104,928)

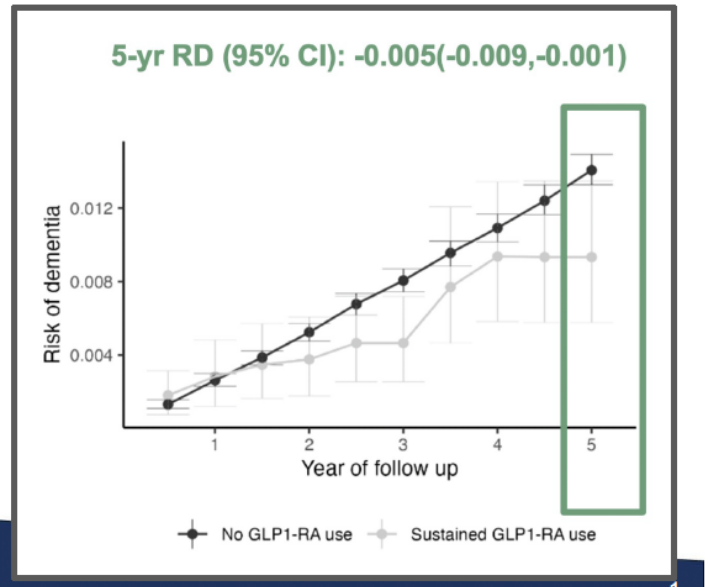
Causal Estimand: Difference in 5 year incident dementia under hypothetical protocols:

- continuous GLP1-RA vs. no- GLP1-RA use
- no censoring

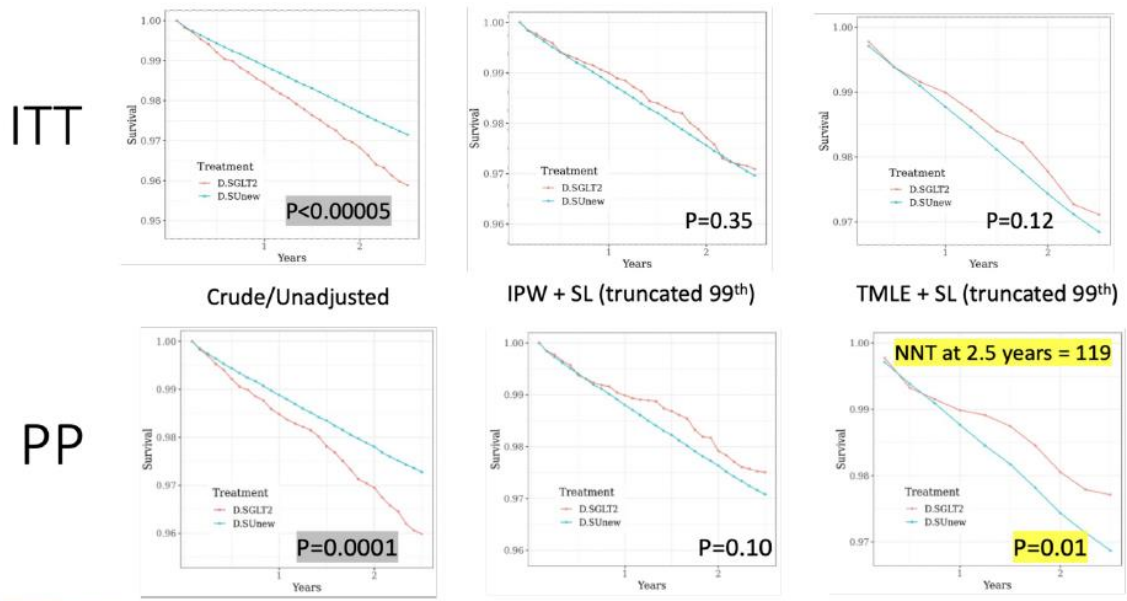
Identification: Baseline & post-baseline confoudner adjustment

Estimation & Inference: Longitudinal TMLE w/ machine learning; simulation-based specification

Nance et al, 2023, arXiv:2310.03235



Causal Effect of SLGT2/SU on MACE, controlling for >100 baseline confounders, and time-dependent confounders of drop-out



Berkeley UNIVERSITY OF CALIFORNIA

Preliminary results from the ON TARGET DM Study (PCORI #DB-2020C2-20318, Neugebauer & O'Connor)

RCT analysis: Improved Precision

SEARCH Study (NCT:0186460

- Pragmatic Cluster RCT
  - N=320,000, 16 communities
- TMLE & "Adaptive pre-specification"
  - Flexible pre-specified adjustment
- Primary pre-specified analysis in trial
- Efficiency gains
  - Substantial for some outcomes
  - Never harmful

Similar findings across many trials

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa

D.V. Havlir, L.B. Balzer, E.D. Charlebois, T.D. Clark, D. Kwarisiima, J. Ayieko,

Focusing Stage 2 estimation:	Effect	95%CI	Efficiency*
Three-year cumulative HIV Incidence			
Unadjusted	0.98	(0.66, 1.45)	-
TMLE with Adaptive Prespec.	0.96	(0.80, 1.17)	4.6
Incidence of HIV-associated TB or death			
Unadjusted	0.79	(0.64, 0.98)	-
TMLE with Adaptive Prespec.	0.80	(0.69, 0.91)	2.6
Population-level Viral Suppression			
Unadjusted	1.15	(1.11, 1.20)	-
TMLE with Adaptive Prespec.	1.15	(1.11, 1.20)	1.0

Berkeley UNIVERSITY OF CALIFORNIA

e.g., Balzer, et al Stat Med 2016; Biometrics 2024

*relative to unadjusted

RCT analysis: Intercurrent Events/Drop-in

1. LEADER Cardiovascular outcomes trial: RCT comparing Liraglutide vs. SoC for treating diabetes patients at high cardiovascular risk
2. The study showed a significant Cox-PH RH with CI [0.78-0.97] on MACE time to event outcome
3. TMLE reproduced similar effect, with more precision; robust across many subgroups (Chen et al., 2023)
4. There was **significant additional drop-in insulin use in the control arm.**
5. An L-TMLE **stochastic intervention direct effect** analysis controlling for the differential post-treatment use of insulin was carried out

Results:

-*Significant* ***stochastically controlled*** additive direct treatment effect

-*Non-significant* ***statically controlled*** additive direct treatment effect

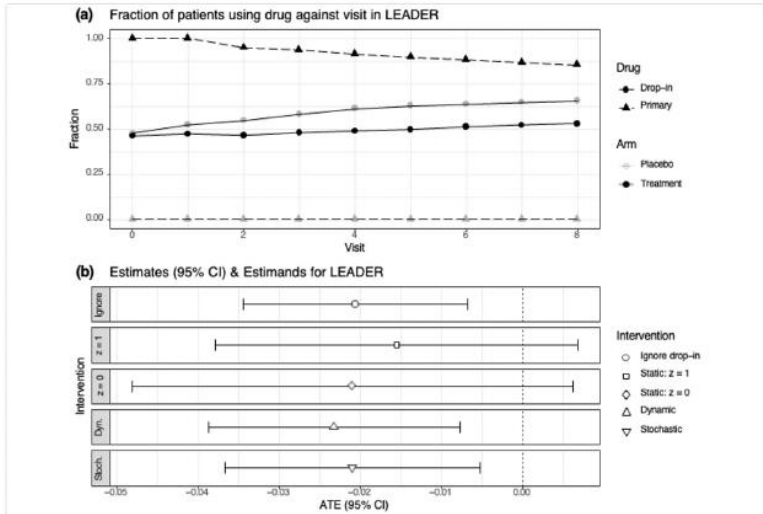


FIGURE 4 (a) Fraction of those at risk using main/drop-in treatment in each arm for LEADER and (b) LTMLE estimates with 95% confidence intervals.

Berkeley UNIVERSITY OF CALIFORNIA

Hypothetical treatment interventions to handle treatment drop-in in randomized controlled trials

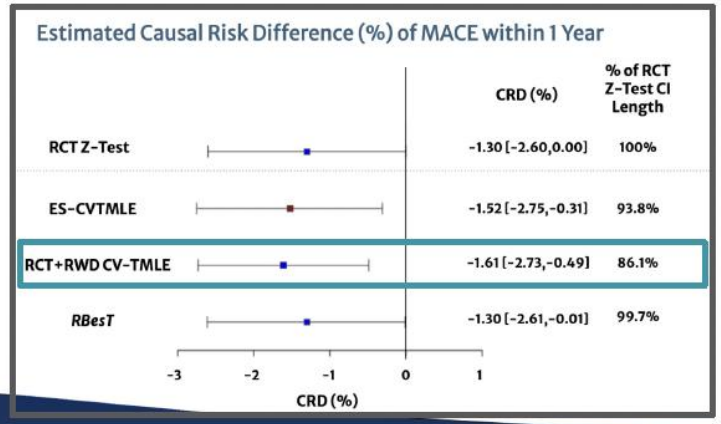
HELENE C. W. RYTGAARD*, EDWIN FONG, JENS M. TARP,

THOMAS A. GERDS, SØREN RASMUSSEN, MARK J. VAN DER LAAN, HENRIK RAVN

RCTs Augmented with External Controls
Ex. Semaglutide and Major Adverse Cardiovascular Event (MACE)

- **Objective**: Potentially augment control arm while controlling Type 1 error
  - Pioneer 6 RCT
  - Optum observational data
- **Experiment-Selector CV-TMLE**
  - Select experiment (RCT only or RCT with RWD) that optimizes bias-variance tradeoff
  - Separate experiment-selection from effect estimation using cross-validation
  - Integrates negative control outcomes

Husain et al. (2019)

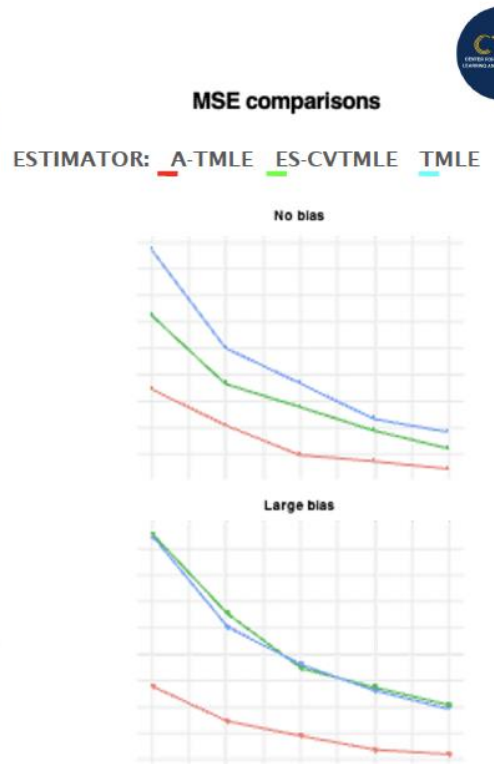


Berkeley UNIVERSITY OF CALIFORNIA

Dang, et al, 2022; arXiv:2210.05802
Dang, et al, 2023, J. Clin & Trans. Science

External Controls: Adaptive-TMLE

- Dang et al (2023): ES-CV-TMLE that data adaptively decides to **accept or reject** external data based on evaluating bias relative to gain in variance.
- Adaptive-TMLE is an advance in TMLE.
- It yields a new method for integrating external controls that **estimates the bias** from external data and **corrects** the pooled TMLE of ATE.
- The method is shown to be asymptotically **valid coverage** and type-I error **without any assumptions beyond RCT**.
- Advantage is that it yields a way forward with biased external control data while preserving valid inference.



Adaptive debiased machine learning
using data-driven model selection techniques

Lars van der Laan, Marco Carone, Alex Luedtke, Mark van der Laan

July 25, 2023

Berkeley UNIVERSITY OF CALIFORNIA

arXiv:2307.12544

Concluding Remarks

- Roadmap for causal inference and Targeted Learning provides scientific approach for generating RWE.
- Integrates all advances in machine learning, statistical theory and causal identification.
- SL and TMLE can be tailored towards particular estimation problem in pre-specified manner using outcome blind simulations.

Q&A

AMGEN®
Pioneering science delivers vital medicines™

Enhancing Validity of Externally Controlled Trials: A TMLE Application in Pediatric ALL

Ela Talpes, Junyi Zhou, Spencer Woody

Center for Design & Analysis, Amgen

January 29, 2026, RWE Forum

Acknowledgements:

- Amy Kimball
- Belle Fang
- Chris Kim
- Li Chen
- May Mo
- Qing Liu
- Wenqing Jiang

AMGEN

Overview

The Problem

Background

- Causal inference in clinical development
- Risks and mitigation strategies

Solution: novel study design

- Sources of data
- Implementation

Results

Remarks

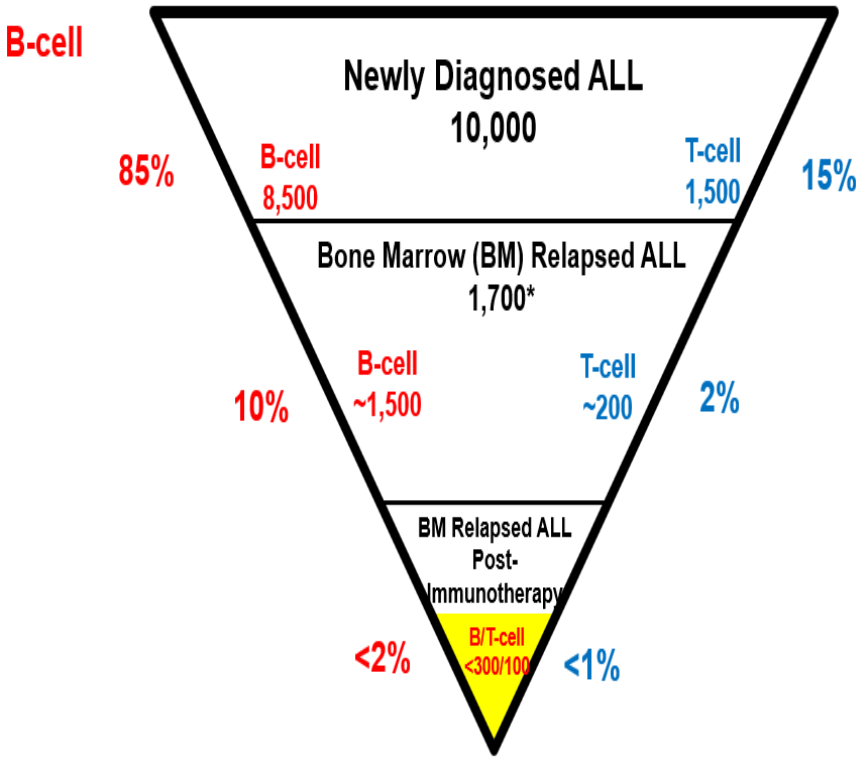
Outline

The Problem

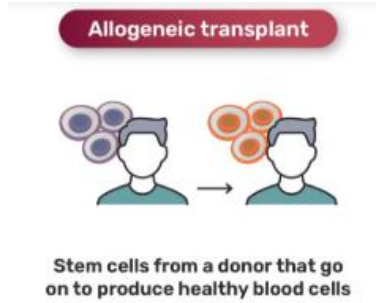
Explorations in Pediatric ALL

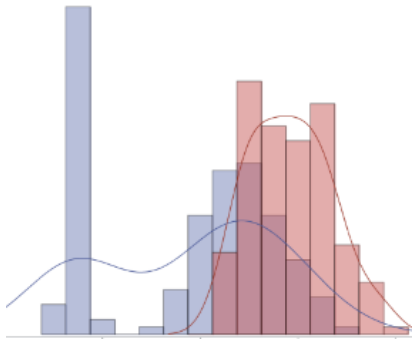
Objective: Achieving superior complete remission for experimental arm versus control arm

Rare Disease:
Pediatric Relapsed/ Refractory B-cell/ T-cell ALL



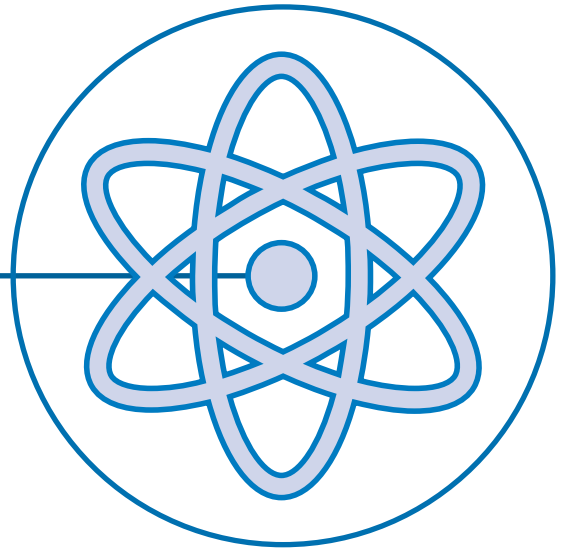
- Original study design (2014) included BLRM phase 1 followed by controlled **randomized** phase 2
- **Phase 1 Results (2019)**: Induction with CFZ-VXLD, increased the chances to reach Complete Remission
- By 2019, **planned phase 2 no longer feasible** due to lack of appropriate active control
- Limited **Real World Data (RWD) for evidence** in unmet need due to scarce data





AMGEN

Outline



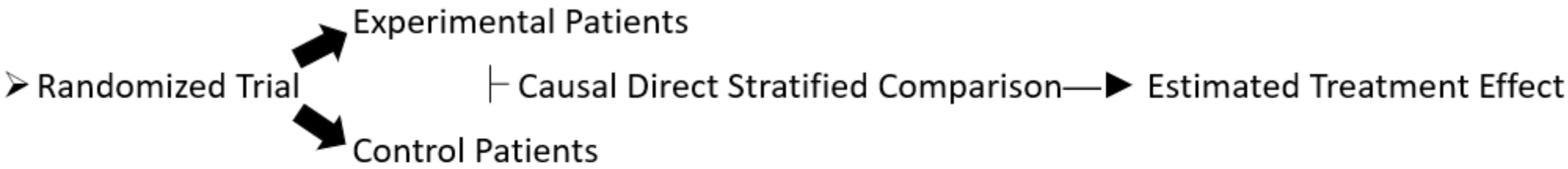
Background

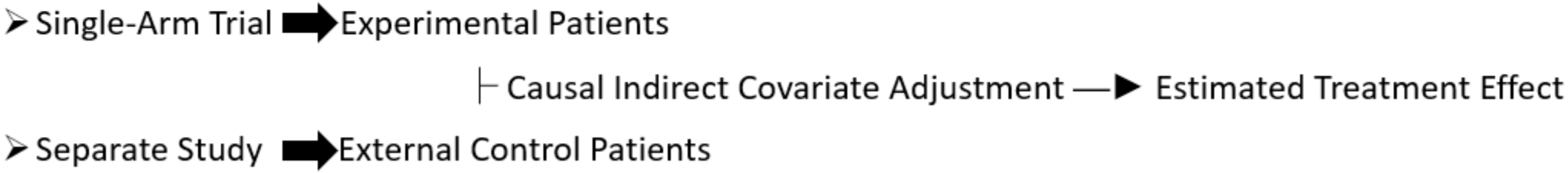
AMGEN

Causal Inference:
Estimate Clinically Interpretable Treatment Effects

Core assumptions across treatment arms for validity

- ✓ **Exchangeability: n**o unmeasured confounding
- ✓ **Positivity:** overlap in inclusion/ exclusion criteria
- ✓ **Consistency:** similar endpoint definitions





AMGEN

Externally Controlled Trial:
Addressing Bias in a Pre-specified and Transparent Way

Risk	Mitigation
Non-randomized populations	Eligibility alignment + causal adjustment
Unmeasured confounding	Sensitivity analyses
Poor overlap	Overlap diagnostics for matching or weighting
Endpoint differences	Harmonized collection and definition
Subsequent therapy	Estimand strategies
RWE data limitations	Quality checks + robustness checks
Regulatory uncertainty	Supportive evidence within the totality of the data

AMGEN

Outline

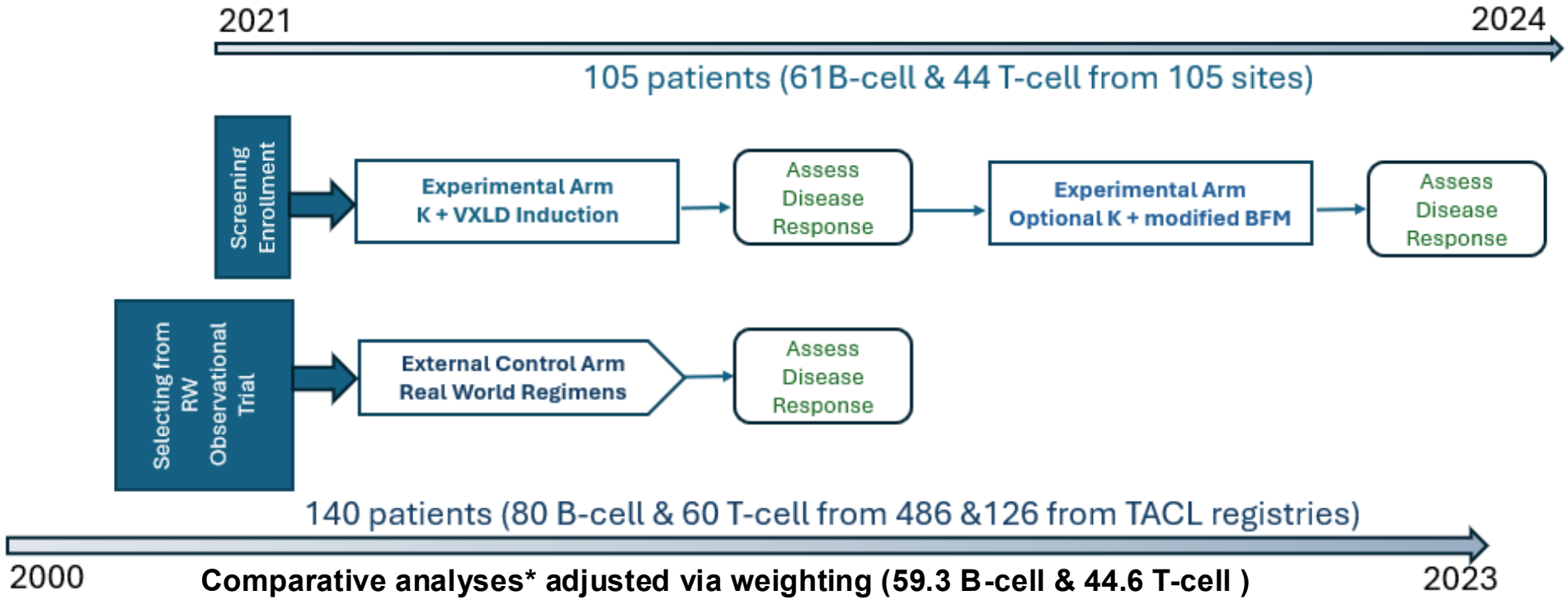
The problem

Background

Solution

AMGEN

Novel Study Design
Phase 2: Externally Controlled Single Arm Trial



Main Endpoints

- **CR**
- CRi+
- EFS
- OS

Sample size considerations- analyses done separately per ALL phenotype (B-cell/ T-cell):

At least 100 subjects: (min of 50 B-cell ALL / 30 T-cell ALL) for more than 70% power for at least one phenotype at 5% alpha

AMGEN

Efficient Trial Execution
Pre-specified Statistical Analysis Plan

FDA's expectations (guidelines/ RTQs)

- ***SAP will be submitted to FDA for review and agreement prior to study analysis***
  - *Comparability of enrolled pts to the external control*
  - *Sensitivity of the results to key model assumptions, including unmeasured confounding*

- **Validation of key study design elements:**
  - Population
  - Treatment
  - Outcomes

Amgen's steps to minimize bias

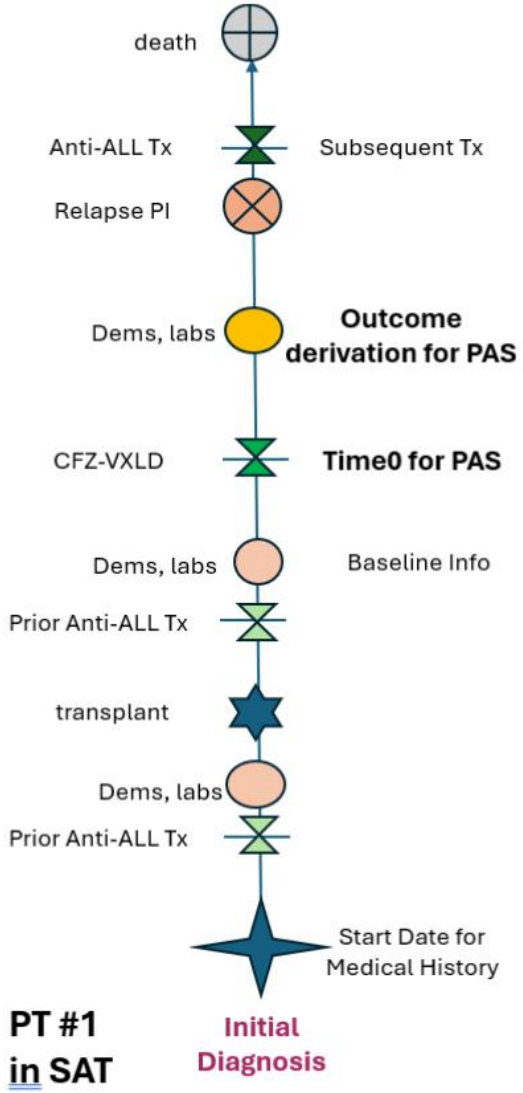
- Robust measures for Data Quality
  - Objective imputation and derivation rules

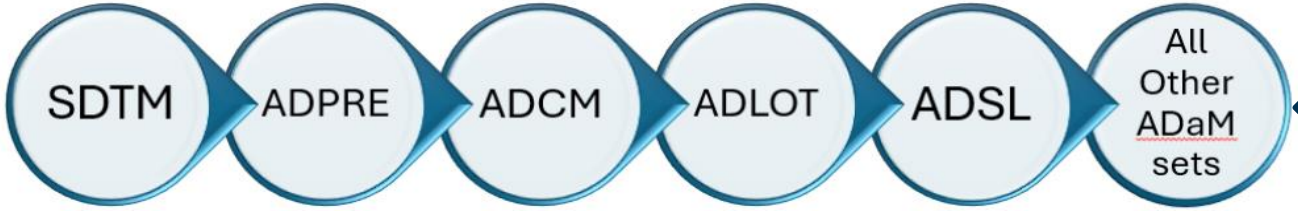
- Rigorous methodology for analyses
  - IPTW models and diagnostic checks
  - Validity of modeling assumptions
  - Concordance analyses

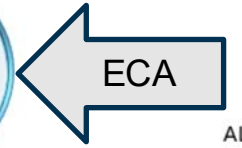
- Trial Integrity measures
  - Fire walls between experimental and control arm (outcomes)
  - Independent external DMC

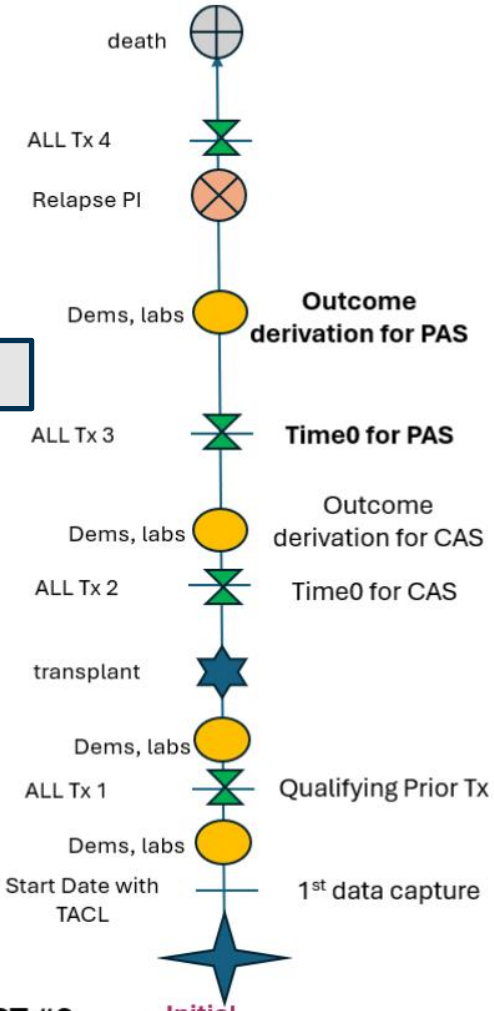
AMGEN

Mapping Data into C-DISC Standards









Prior Response (labs/procedures)

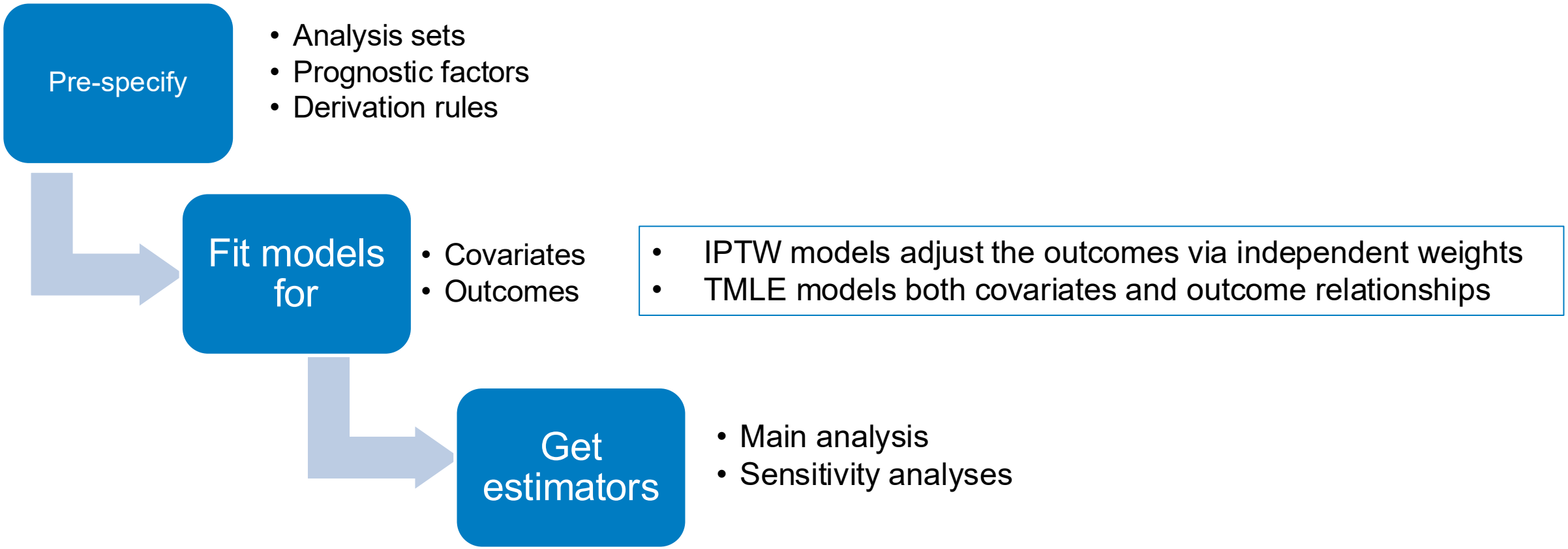
ConMeds (any medication in MedDRA)

Line of Therapy (anti-ALL therapy w. flags for I/E)

Flags for ECA used in PA and SA

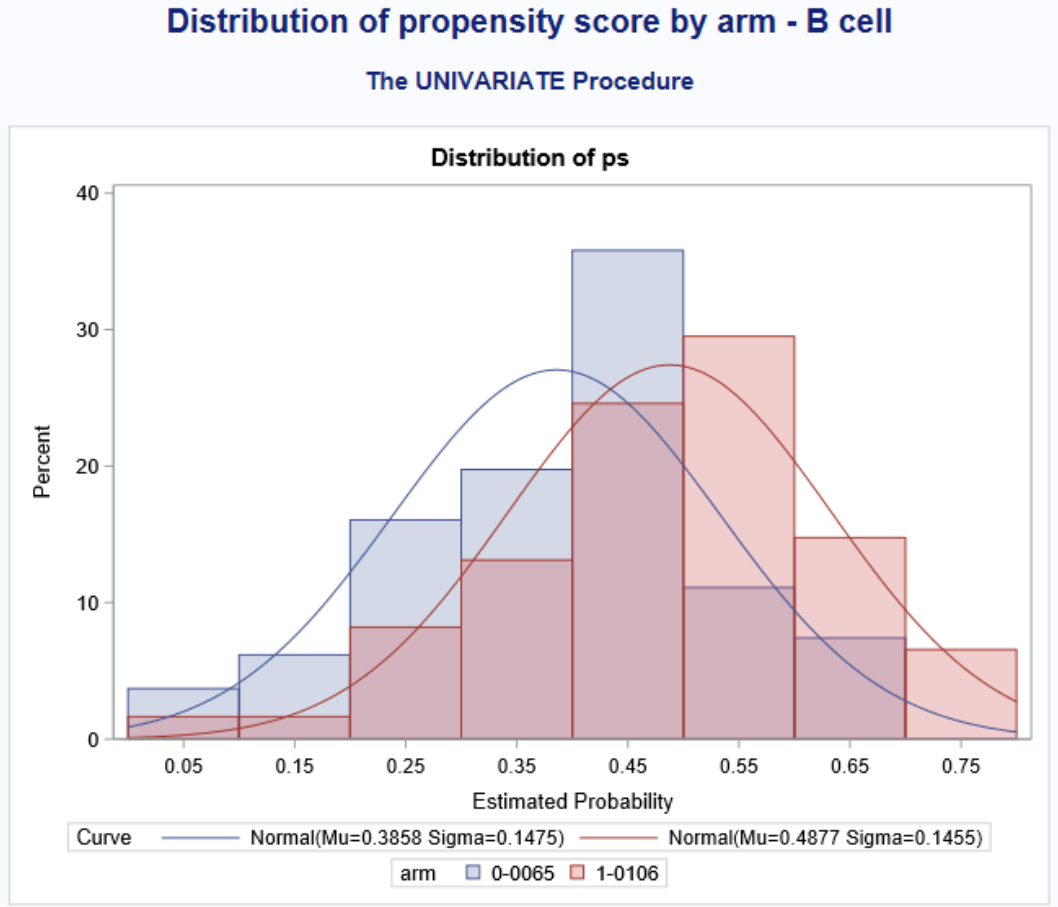
ECA=external control arm; I/E= inclusion/ exclusion; PAS=primary analysis set; PT=patient; SAT= single arm trial; Tx= treatment

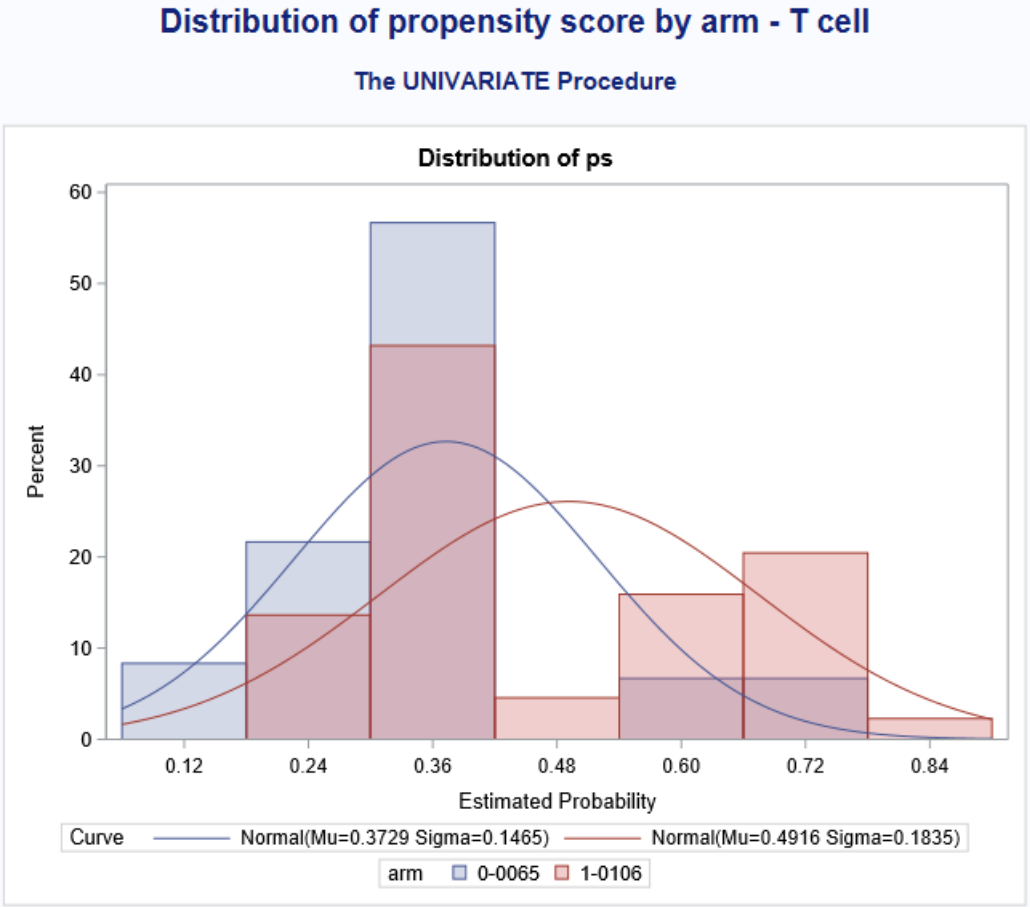
Analysis Plan for Inference



Propensity Score Balance Diagnoses

Overall balance is sufficient if ≥ 25% of control PS overlaps with the inner 95th % of experimental PS



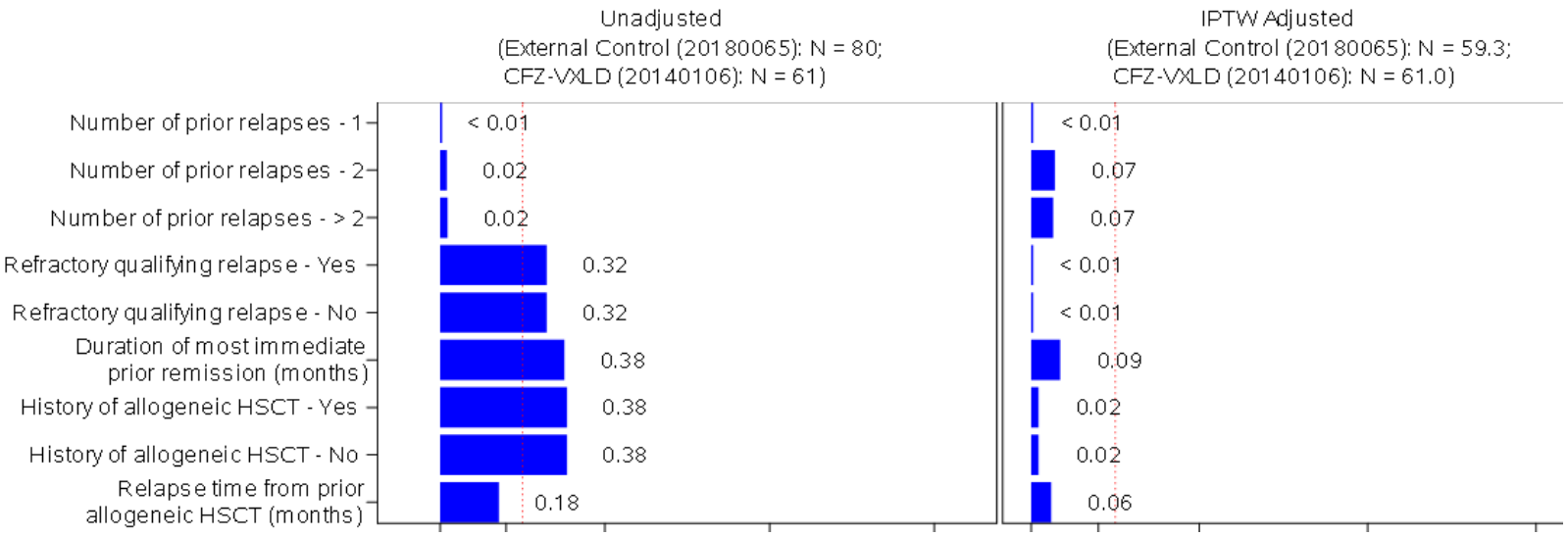


AMGEN

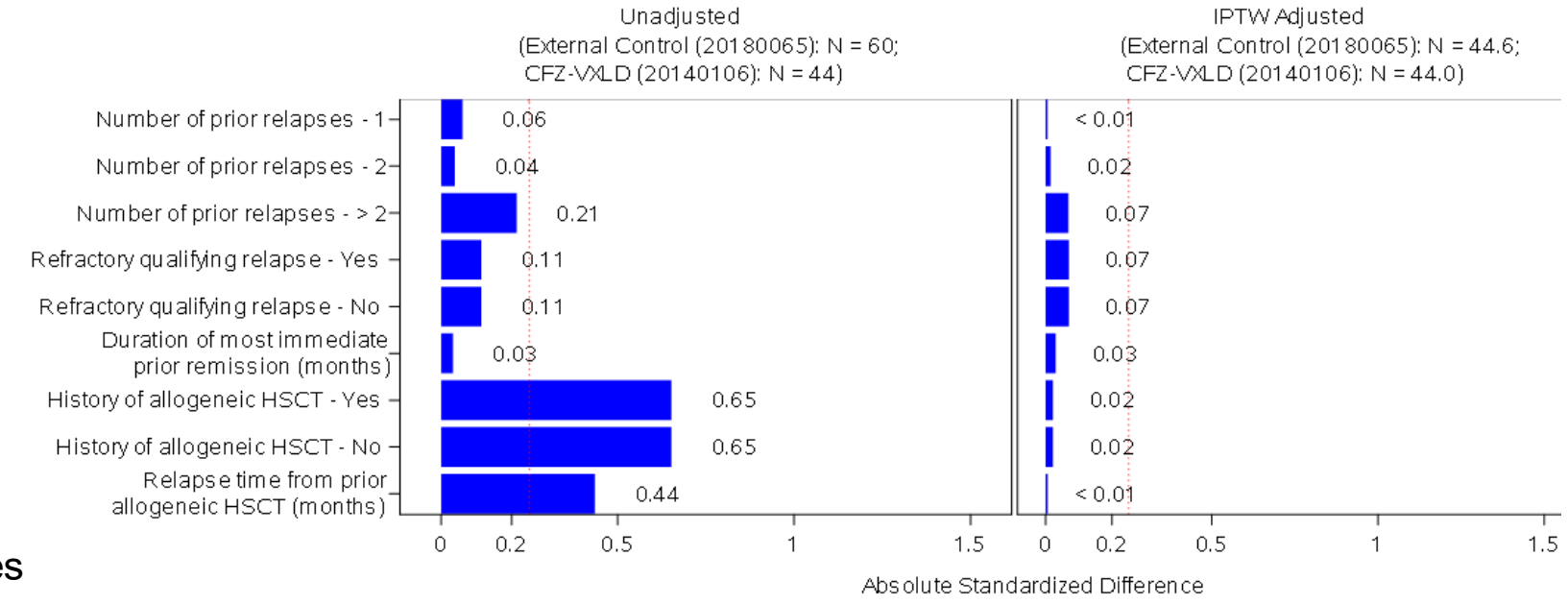
IPTW Balance of Baseline Prognostic Factors (C1)

Balance is sufficient if ASD≤0.25

B-cell (5)



T-cell (5)



C1 = set of covariates used for main analyses

Variety of Sensitivity Analyses

Outcome assessment

Bias in clinical outcomes (sponsor derived using local lab)

- Investigator assessed using local laboratory
- Sponsor derived using central laboratory

Patient population

Dependence of treatment effect on control arm selection

- Chemotherapy Analysis Set
- Targeted Salvage Therapy Analysis Set

Modeling specifications

Check modeling assumptions

- Tipping point analyses for outliers/ imputation thresholds
- Quantitative bias analysis
- Additional prognostic factors
- Selection method for prognostic factors
- G- Computation
- Target Maximum Likelihood Estimation (TMLE)

AMGEN

Outline

Background

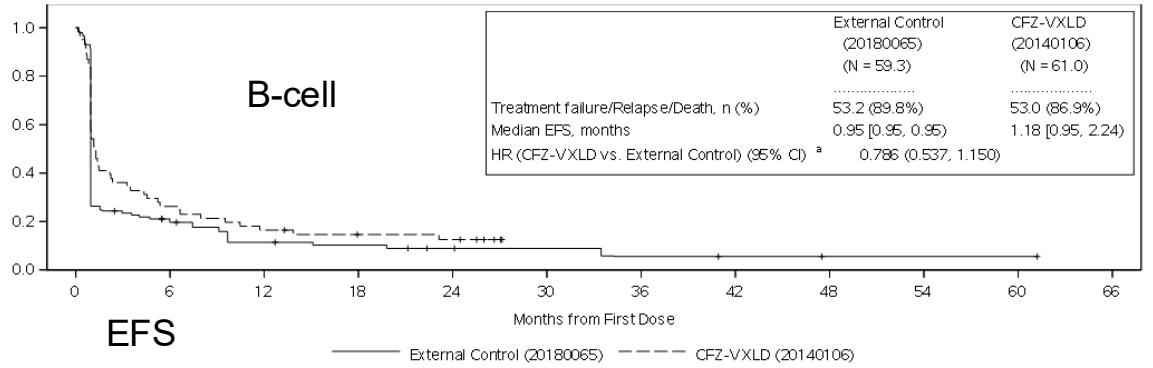
Solution

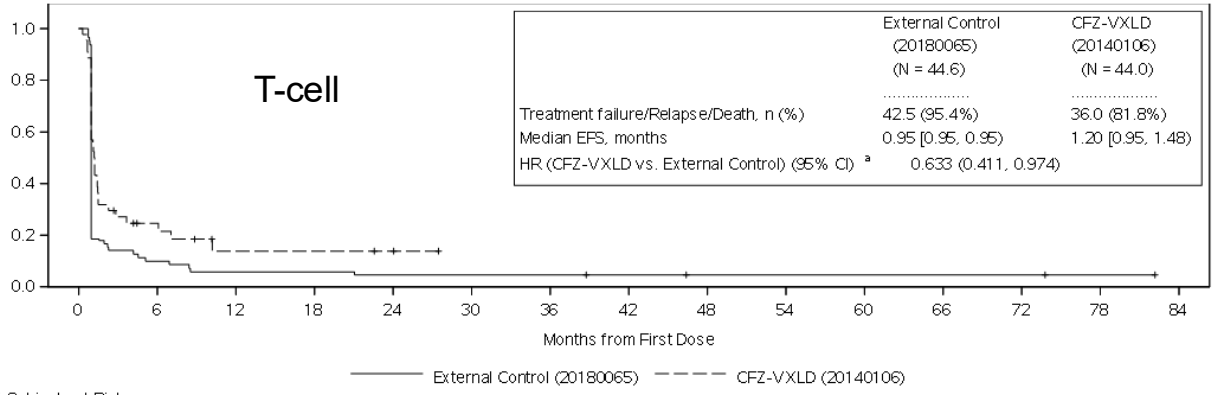
Results

AMGEN

Selected Results

	B-cell hypothesized 50% vs 25%	T-cell hypothesized 30% vs 15%
CR rates OR (95%CI)	14.8% vs 7.8% 2.04 (0.54;7.66)	13.6% vs 9.1% 1.58 (0.47;5.31)
CRi+ rates OR (95%CI)	42.6% vs 26.3% 2.08 (0.97;4.48)	27% vs 18.6% 1.65 (0.64;4.25)





Consistency of the Sensitivity Analyses supported the validity of the trial

Selected Sensitivity Results for B-cell

Analysis	Adjustment methods	External Control Remission Rate (%) (95% CI)	Experimental Remission Rate (%) (95% CI)	Odds Ratio (95% CI)
Primary	IPTW –all C1	7.8 (1.0, 14.7)	14.8 (5.9, 23.7)	2.04 (0.54, 7.66)
Sensitivity	IPTW –all C1 & C2	6.5 (0.1, 12.9)	14.8 (5.9, 23.7)	2.49 (0.53, 11.77)
	IPTW- stepwise selection	7.2 (0.7, 13.8)	14.8 (5.9, 23.7)	2.22 (0.62, 7.97)
	G-computation–all C1	6.0 (1.6, 13.0)	14.8 (6.6, 24.6)	2.70 (0.86, 10.09)
	G-comp–LASSO C1 (3)	5.9 (1.8, 13.2)	14.8 (8.2, 23.0)	2.75 (0.97, 8.96)
	G-comp–LASSO C1 & C2 (4)	5.9 (1.7, 12.7)	14.8 (6.6, 24.6)	2.78 (0.90, 9.61)
	TMLE –all C1	6.7 (1.7, 14.5)	16.4 (6.8, 26.5)	2.72 (0.89, 10.17)
	TMLE –LASSO C1 (3)	6.6 (1.9, 14.3)	16.8 (7.5, 25.7)	2.87 (0.99, 10.46)
	TMLE –LASSO C1 & C2 (4)	6.3 (1.7, 13.6)	16.2 (6.5, 26.1)	2.85 (0.93, 10.09)
	TMLE- SL C1	7.9 (2.7, 17.9)	17.8 (8.6, 28.1)	2.52 (0.79, 6.08)

Consistent results enhance the validity of the design

AMGEN

Selected Sensitivity Results for T-cell

Analysis	Adjustment methods	External Control Remission Rate (%) (95% CI)	Experimental Remission Rate (%) (95% CI)	Odds Ratio (95% CI)
Primary	IPTW –all C1	9.1 (0.7, 17.5)	13.6 (3.5, 23.8)	1.58 (0.47, 5.31)
Sensitivity	IPTW –all C1 & C2	7.1 (0.0, 15.2)	13.6 (3.5, 23.8)	2.06 (0.51, 8.24)
	IPTW- stepwise selection	8.0 (0.0, 16.0)	13.6 (3.5, 23.8)	1.83 (0.55, 6.08)
	G-comp–all C1	11.5 (4.7, 21.9)	13.6 (4.5, 25.0)	1.22 (0.35, 3.41)
	G-comp–LASSO C1 (all)	11.5 (4.7, 21.9)	13.6 (4.5, 25.0)	1.22 (0.35, 3.41)
	G-comp –LASSO C1 & C2 (6)	11.5 (4.3, 21.6)	13.6 (4.5, 25.0)	1.22 (0.34, 3.77)
	TMLE –all C1	11.6 (2.9, 21.3)	14.2 (0.0, 25.5)	1.26 (0.37, 3.50)
	TMLE –LASSO C1 (all)	11.6 (2.9, 21.3)	14.2 (0.0, 25.5)	1.26 (0.37, 3.50)
	TMLE –LASSO C1 & C2 (6)	11.6 (3.7, 21.5)	14.3 (2.3, 25.5)	1.27 (0.36, 3.82)
	TMLE- SL C1	8.7 (4.3, 19.1)	13. 4 (3.8, 26.5)	1.63 (0.36, 4.54)

Consistent results enhance the validity of the design

AMGEN

Outline

Background

Solution

Results

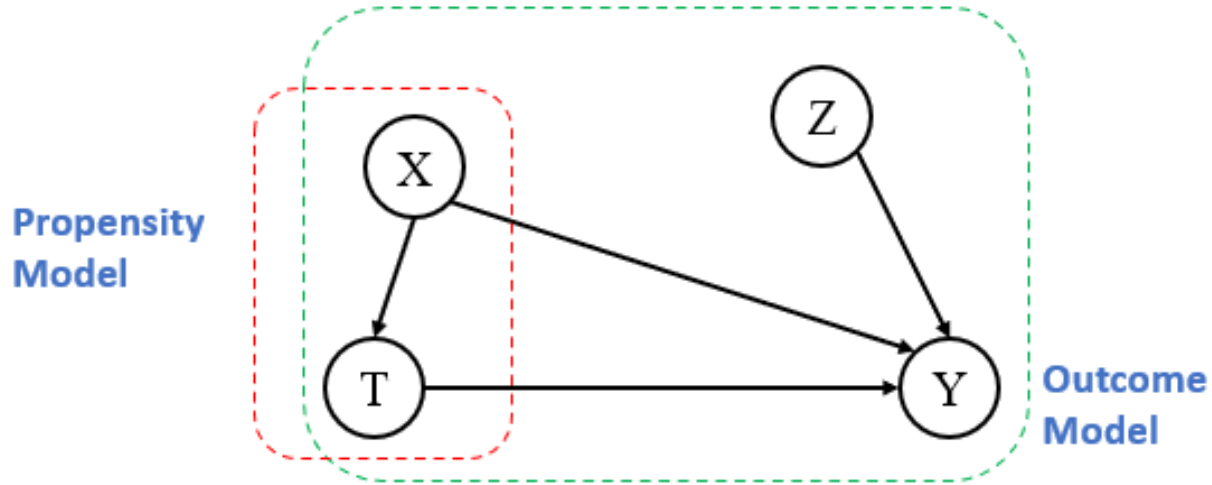
Remarks

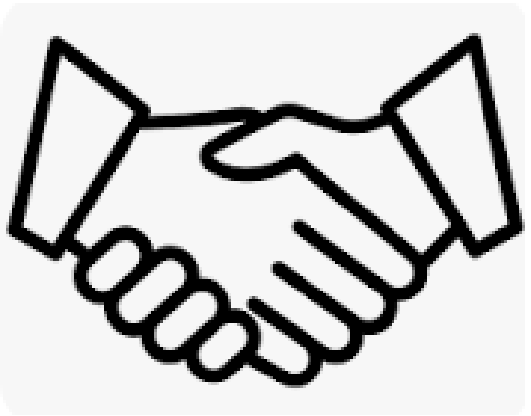
AMGEN

Remarks

- Embrace Complexity with Flexible Innovative Designs
- Doble Robust Estimators consider the Totality of Info
- Build Robust Data and Process Foundations
- Strengthen Collaboration Across Functions and Regulators

Doubly Robust Estimators





Establish Pathways for Future Success

AMGEN

Q&A

connect·share·advance

RWE Pavilion at US Connect

Click here to play!

phuse

connect·share·advance

Join us in Austin, Texas

Full Event: Sunday 22 - Thursday 26 March

RWE Pavilion: Wednesday 25 - Thursday 26 March

A Full Connect Ticket gives you access to all 5 event days...

Check Out the RWE Pavilion

Register Now

Find Out More

phuse

Keep in Touch!

connect·share·advance

Slides & recording will shortly be available on our event page

Welcome to reach out to any of the presenters directly:

- Mark: laan@berkeley.edu
- Ela: mobreja@amgen.com

If you would like to get involved, email workinggroups@phuse.global

If you would like to get involved, email: workinggroups@phuse.global

https://phuse.global/Working_Groups

connect·share·advance

THANK YOU!