Hybrid Randomized/Real-World Data Designs:
A Case Study of Semaglutide and Cardiovascular Outcomes

JICI Working Group on Integration of Observational and RCT Data

Lauren Eyler Dang, MD, MPH

Edwin Fong, Kim Clemmensen, Jens Tarp, Kajsa Kvist, John Buse, Mark van der Laan, Maya Petersen

The Forum for Collaborative Research: Liver Forum 14
April 1, 2023

Disclosure: Work funded by a philanthropic gift from Novo Nordisk to UC Berkeley.

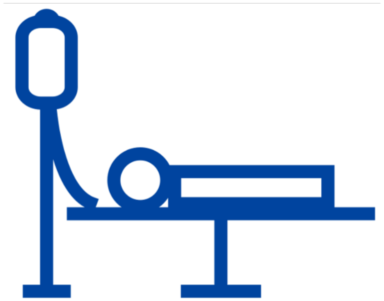
Berkeley
UNIVERSITY OF CALIFORNIA

Integrating RCTs with Real-World Data

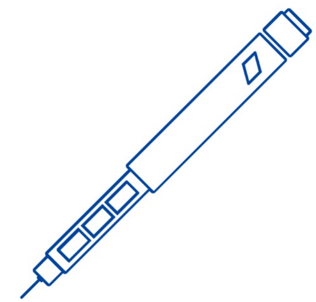
- Randomized controlled trial (RCT) considered "gold-standard"[1]
- Running an adequately powered RCT may not be feasible (e.g., rare diseases)[2]
- Unnecessary randomization to control may be considered unethical[2] (or at least undesirable to patients)

Pediatric Approvals

Severe Disease, Unmet Need



Secondary Indication, Different Route



- Hybrid randomized-external data studies
  - → Augment RCT with external data from previous trials or real-world data (RWD)
  - → Minimize number of required control arm (or total) participants
- Yet risk of introducing bias by adding non-randomized data
- **How can we incorporate external data while identifying a causal effect?**

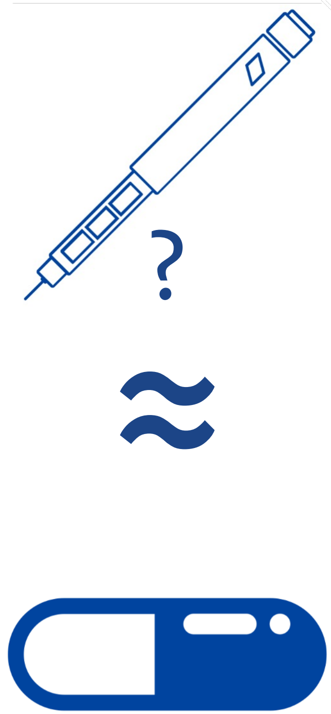
Agenda

- Use case study of Semaglutide and Cardiovascular Outcomes to:
  - Discuss hybrid randomized/external data studies
  - Discuss methods to minimize bias from considering RWD
    - Following the *Causal Roadmap*[3]
      - Step-by-step process to assist with study design and analysis
    - Statistical estimators for integration of observational and RCT data

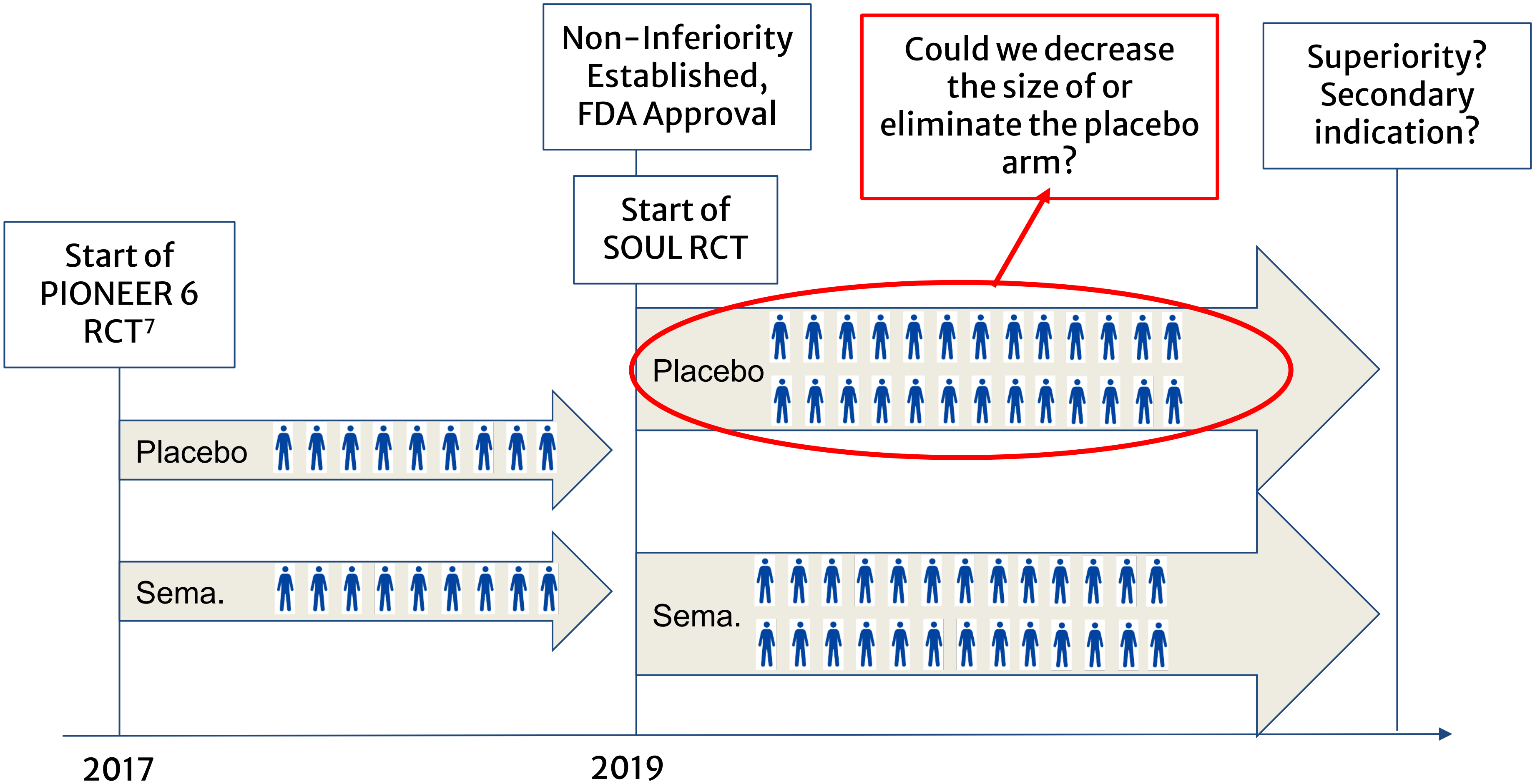
Berkeley
UNIVERSITY OF CALIFORNIA

Semaglutide and Cardiovascular Outcomes

- Semaglutide is a glucagon-like peptide-1 receptor agonist (GLP1-RA)
  - developed for the treatment of Type 2 Diabetes (T2D)
- Injectable semaglutide shown to decrease[4,5]:
  - glycated hemoglobin (HbA1c)
  - body weight
  - systolic blood pressure
  - rates of major adverse cardiovascular events (MACE)
    - death from cardiovascular causes, non-fatal stroke or MI
  - FDA approval: glycemic control, weight management, **reduce CV risk**
- Oral semaglutide shown to decrease[6]:
  - HbA1c
  - body weight
  - FDA approval: glycemic control
  - **what about MACE?**



1 Causal Question: Effect of Oral Semaglutide on MACE



Injectable sema. superior to placebo in SUSTAIN 6 trial[5], American Diabetes Association: Evidence suggests GLP1-RAs for prevention of MACE in T2D[8]

Risk-Benefit Analysis To Patients

Benefit:

- PIONEER 6[7], SUSTAIN 6[5] suggest oral sema. **likely beneficial** for CV outcomes
- Using RWD may lead to **less patient-time on inferior product (without a GLP1-RA)**

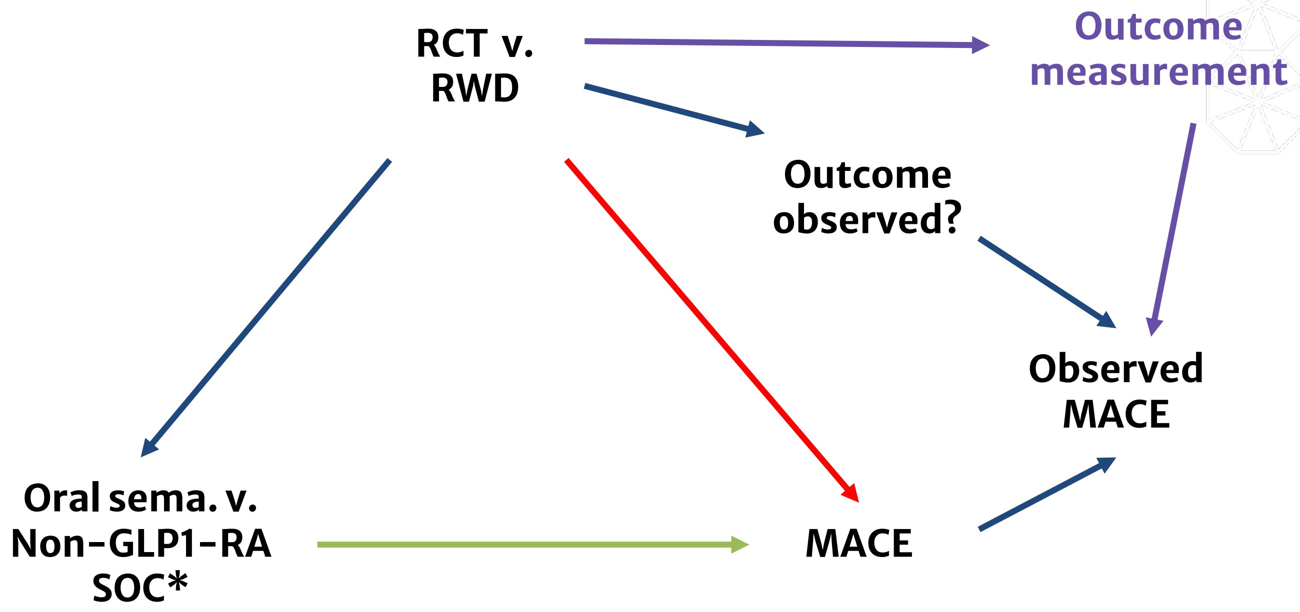
Risk:

- Causal gap[9] { Statistical parameter / Answer to causal question }
  - Minimize risk:
    - study design using causal framework
    - estimator selection to minimize risk

Simulations to help weigh risks and benefits

Berkeley
UNIVERSITY OF CALIFORNIA

2 Causal Model: Understanding (and Minimizing) the Causal Gap

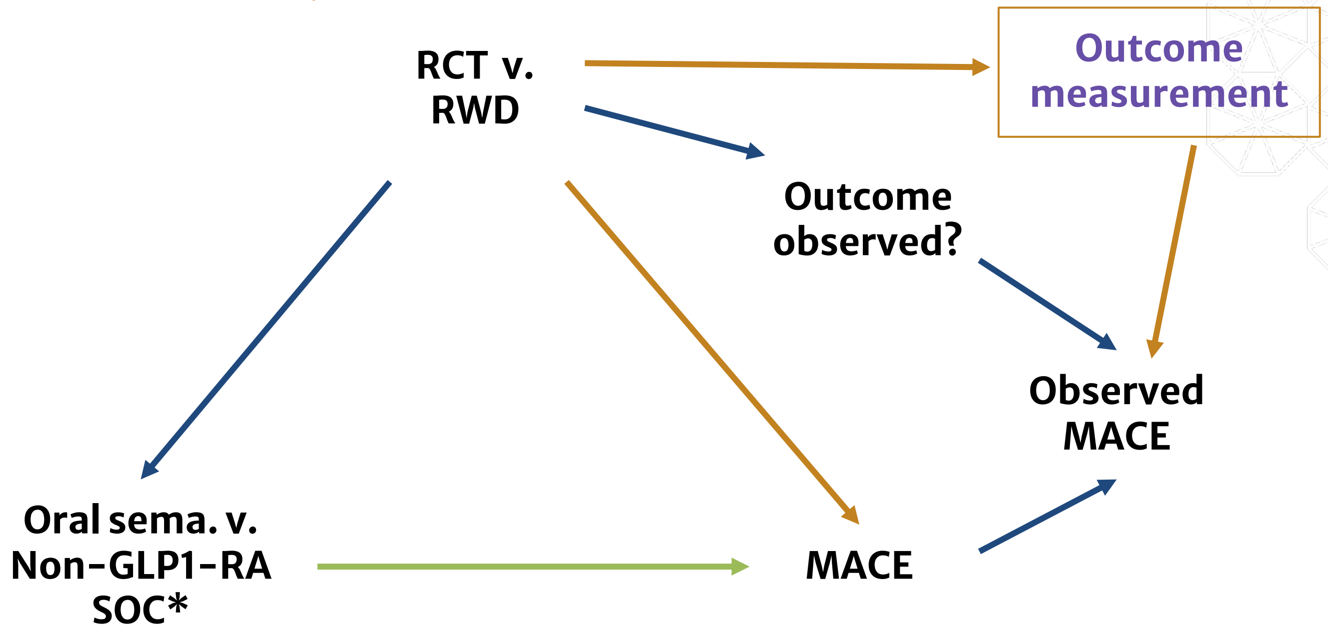


Factors contributing to causal gap: Effect of RCT on outcomes

- Placebo effect?
- Closer monitoring? Better care?
- Outcome measurement different?

*SOC: Standard of care

Ideal Changes to Study
(Not Possible in this Case)

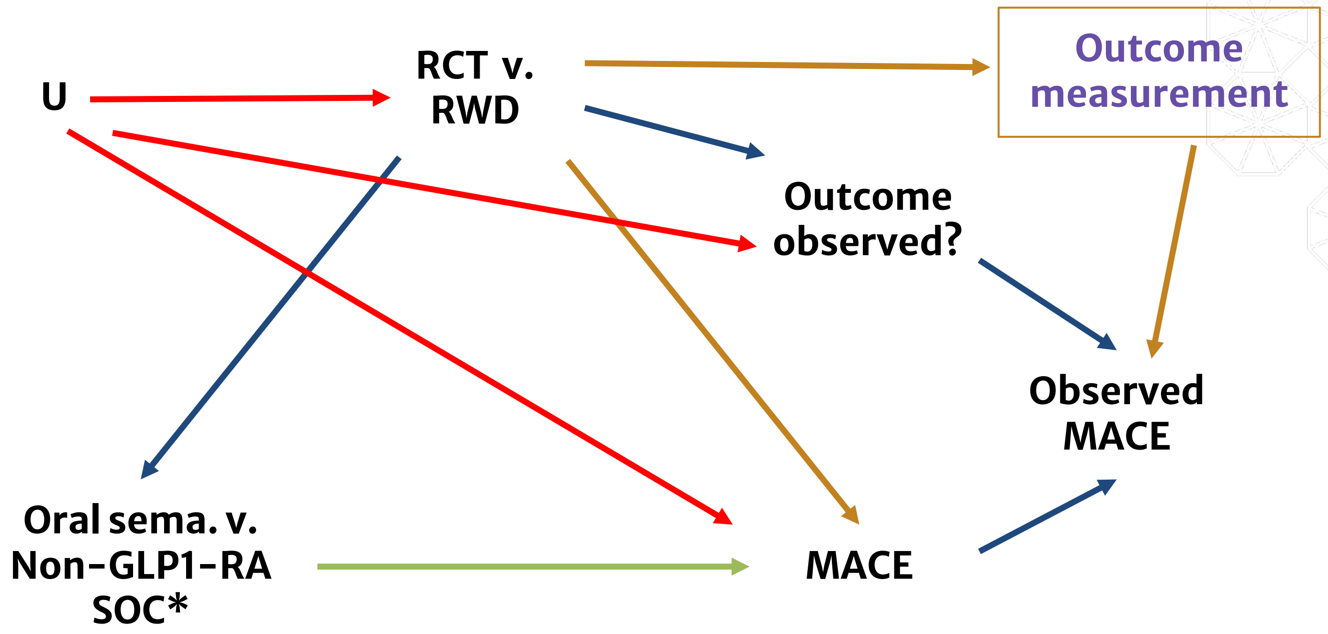


Effect of RCT on outcomes?

- Less likely with pragmatic trial[10] (if acceptable)
  - Same high-quality outcome measurement (registry?)[11]

*SOC: Standard of care

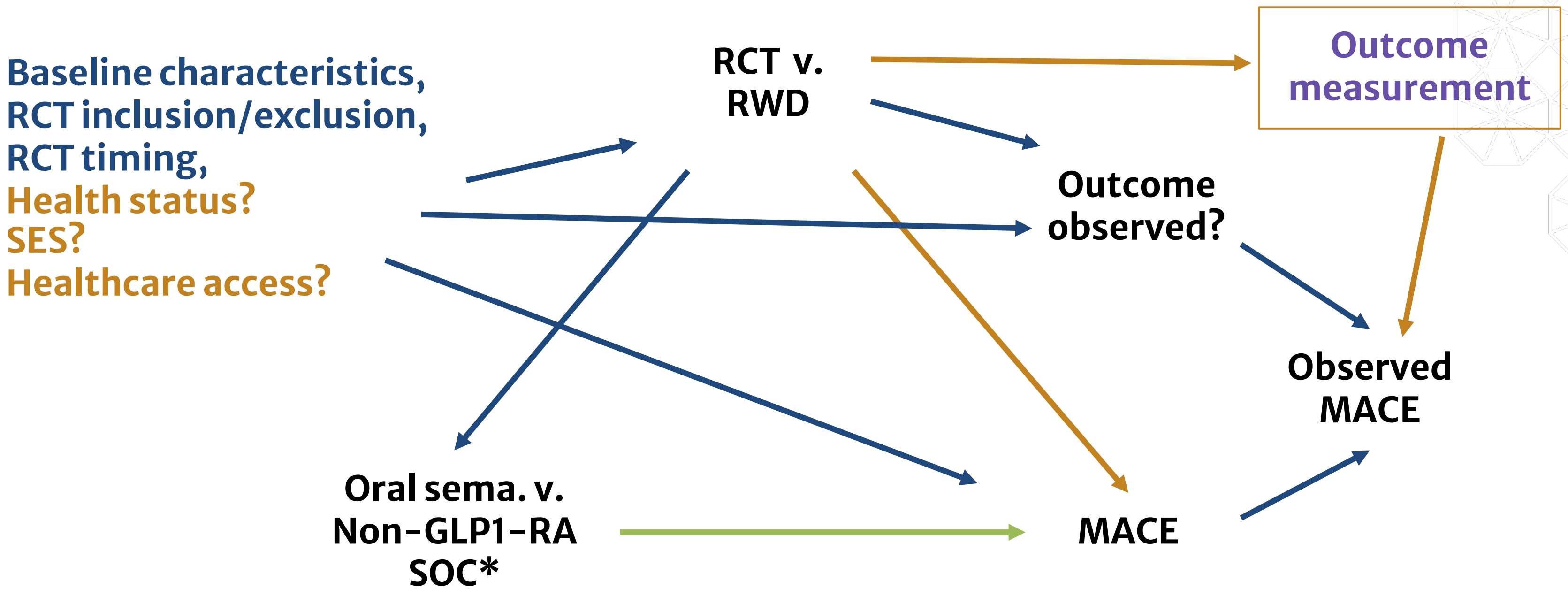
Understanding (and Minimizing) the Causal Gap



Unmeasured common causes (confounders) of trial participation or censoring and outcomes:

- Trial inclusion/exclusion criteria
- Other: Health status? Socioeconomic status (SES)? Better healthcare access? Changes in care with time?

Actual Changes to Study

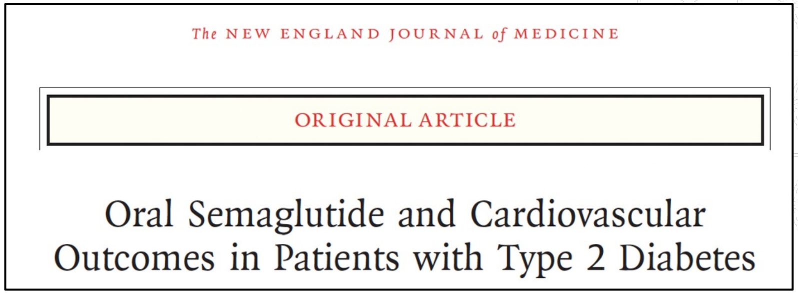


Factors affecting RCT v. RWD, Censoring, Outcomes:

- Measure relevant baseline covariates
- No RWD participants with baseline characteristics not represented in RCT
- Time period of RCT recruitment
- Active comparator in RWD
- RWD participants with relevant labs measured

3 Define the Observed data

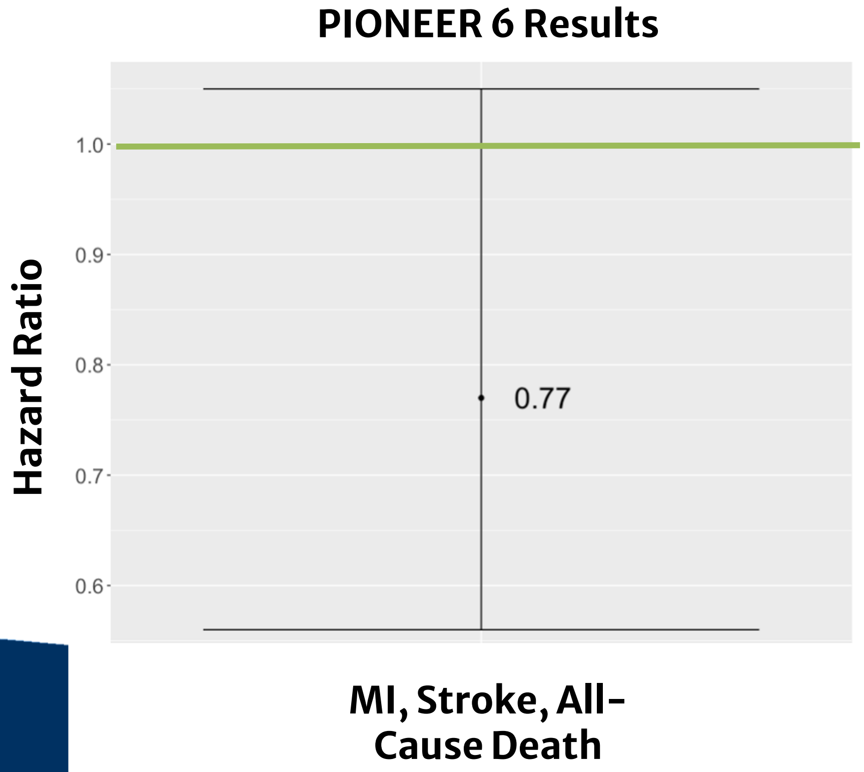
- **PIONEER 6 RCT**
- **Intervention:** Oral sema. v. placebo
- **Outcome:** First MI, stroke, all cause death (MACE)
- **Patient population:** Patients with Type 2 diabetes and high CV risk
- Powered for **non-inferiority** (N=3183)


The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes

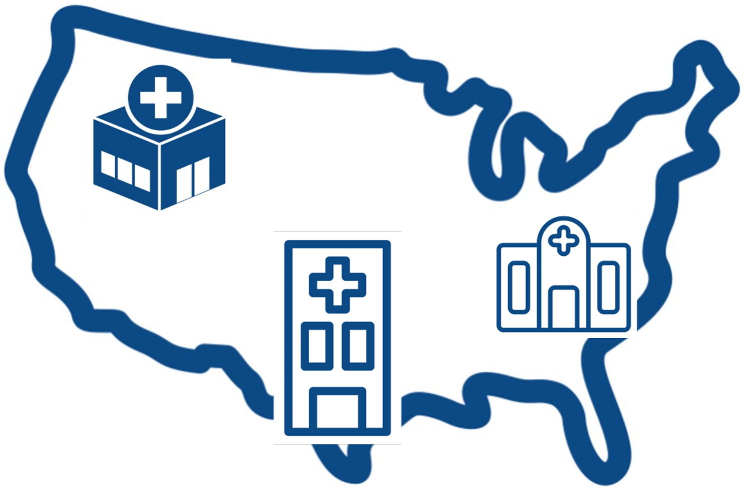
Husain et al. (2019)[7]



Berkeley
UNIVERSITY OF CALIFORNIA

Observational Data: Optum

- **Optum® Clinformatics® Data**
  - **Observational data** from inpatient and outpatient visits in the US
  - **Intervention:** No oral sema.
    - DPP4i (active comparator)
    - Index date (new prescription)
  - **Outcome:** MACE (claims data)

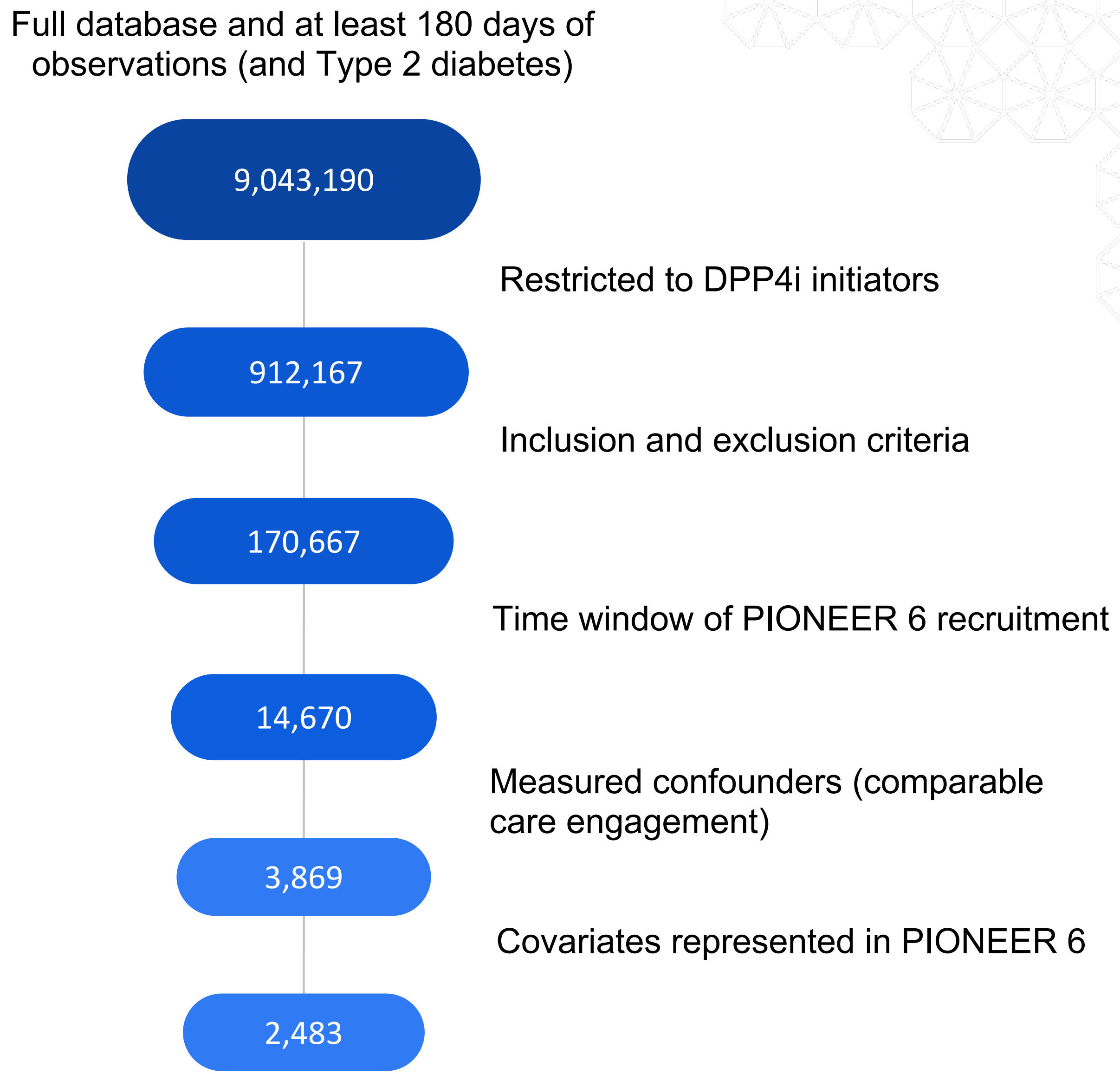


- **Possible baseline confounders:**
  - Age, sex, race, HbA1c, HDL, LDL, eGFR, prior MI, prior stroke or TIA, prior heart failure, morbid obesity, baseline glucose-lowering medications, insulin, and CV medications

*Translation: ICD9/10 codes, AHFS drug codes, LOINC lab codes

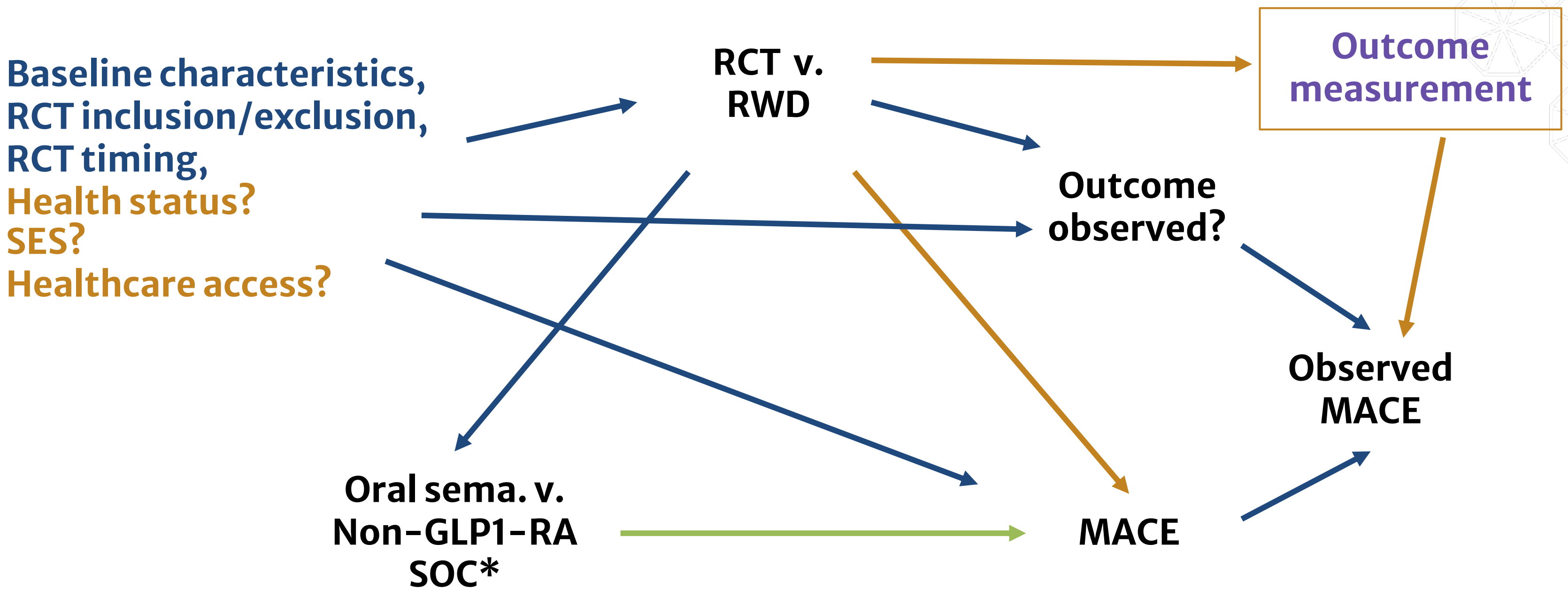
Berkeley
UNIVERSITY OF CALIFORNIA

Flowchart



Berkeley
UNIVERSITY OF CALIFORNIA

5 Assess Identifiability: Can we estimate a causal effect?

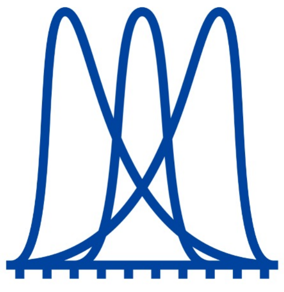


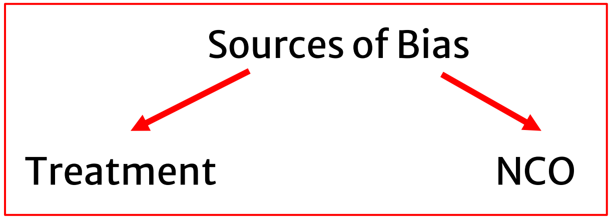
After modifications to RWD control group:

- Plausible that causal gap is small
- Is it small enough for nominal type 1 error control?

6 Choose a Statistical Estimator
7 Causal Sensitivity Analysis

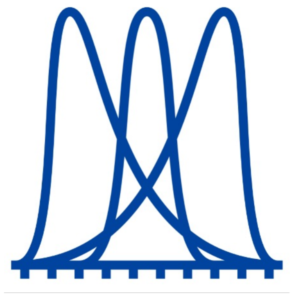
- We tried to design a compatible study.
  - **Sensitivity Analysis:** How large could the causal gap be?
- Choose a statistical estimator that
  - uses **evidence about the causal gap** to decide
  - whether to include RWD or analyze RCT alone
- **Commonly used evidence of bias:**
  - Difference in outcomes between RCT and RWD controls
  - Effect of treatment on a negative control outcome (NCO)

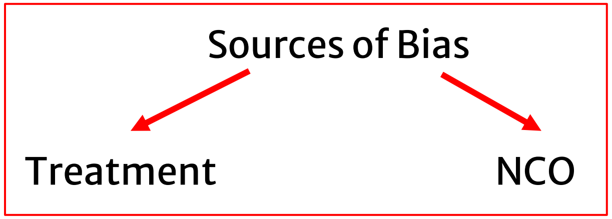




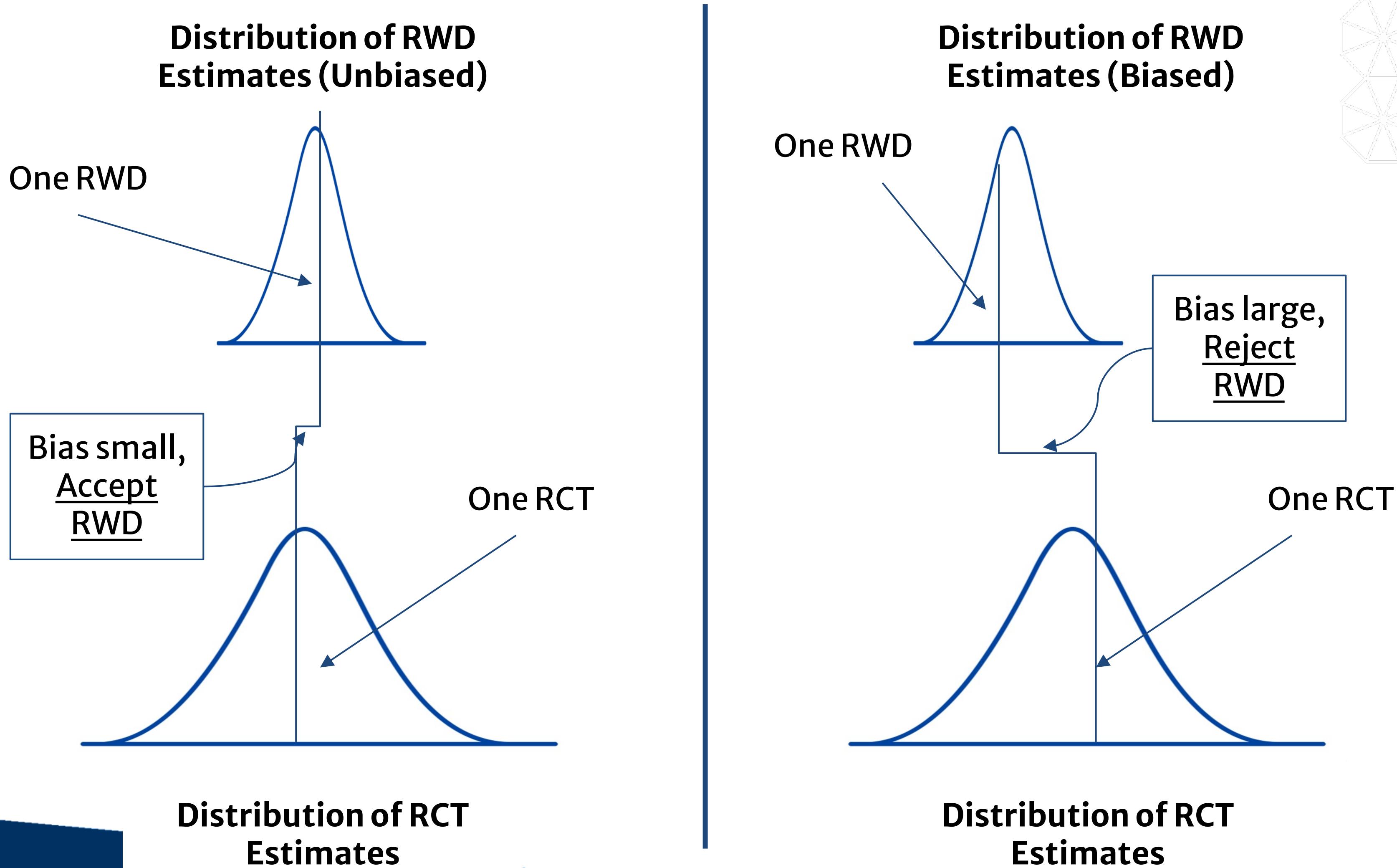
6 Choose a Statistical Estimator
7 Causal Sensitivity Analysis

- We tried to design a compatible study.
  - **Sensitivity Analysis:** How large could the causal gap be?
- Choose a statistical estimator that
  - uses **evidence about the causal gap** to decide
  - whether to include RWD (or how to weight RWD)
- **Commonly used evidence of bias:**
  - Difference in outcomes between RCT and RWD controls
  - Effect of treatment on a negative control outcome (NCO)



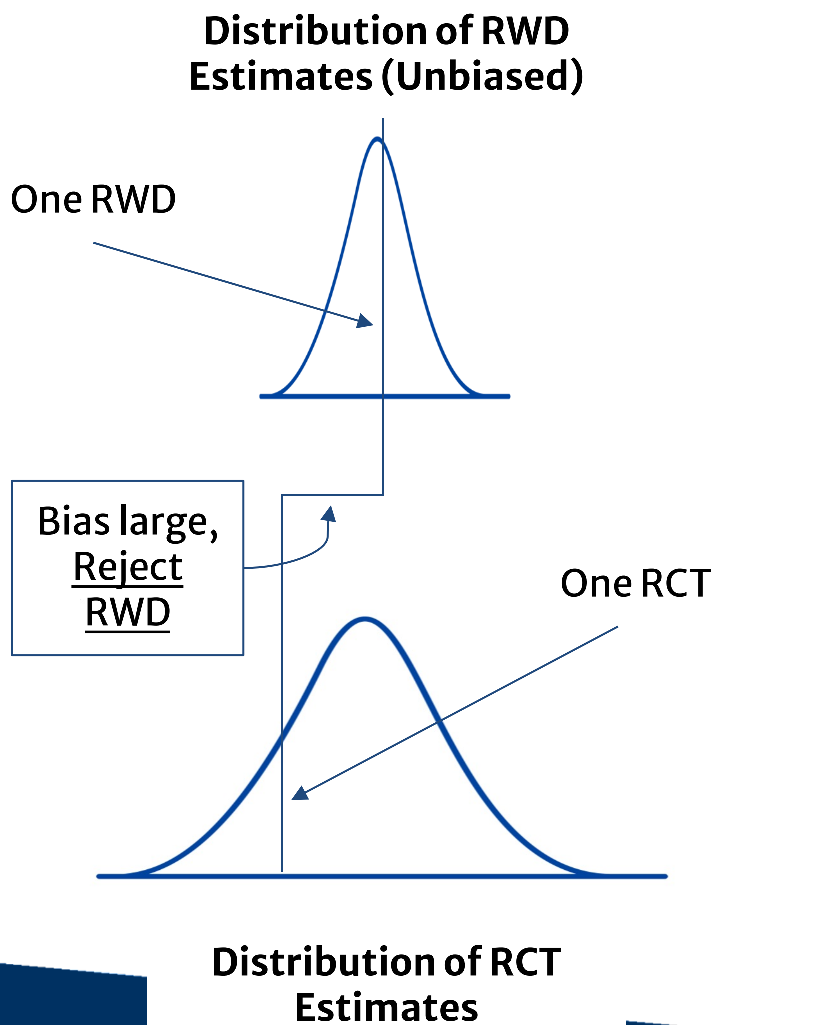


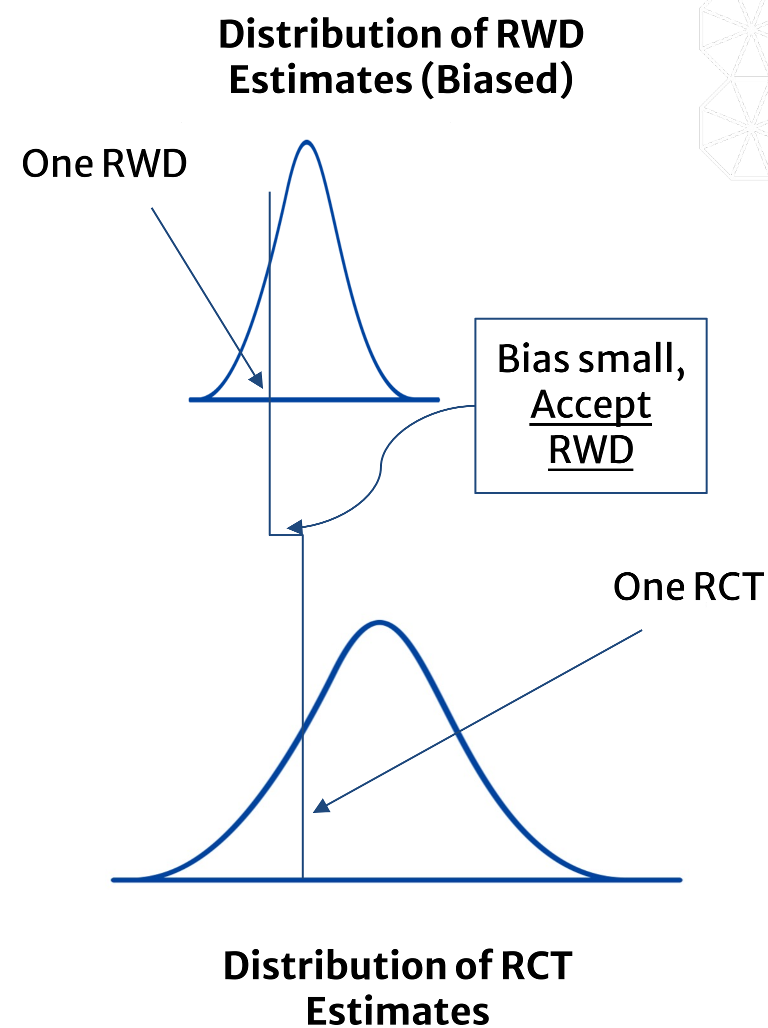
Difference in RCT/RWD Outcomes



Berkeley
UNIVERSITY OF CALIFORNIA

Challenge: Bias Estimated

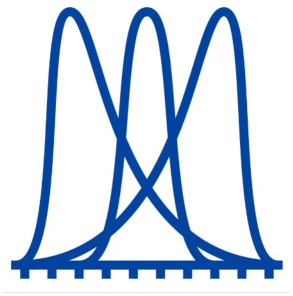




Berkeley
UNIVERSITY OF CALIFORNIA

6 Choose a Statistical Estimator
7 Causal Sensitivity Analysis

- We tried to design a compatible study.
  - **Sensitivity Analysis:** How large could the causal gap be?
- Choose a statistical estimator that
  - uses **evidence about the causal gap** to decide
  - whether to include RWD (or how to weight RWD)
- **Commonly used evidence of bias:**
  - Difference in outcomes between RCT and RWD controls

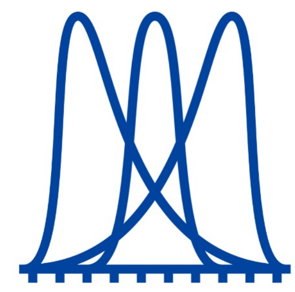


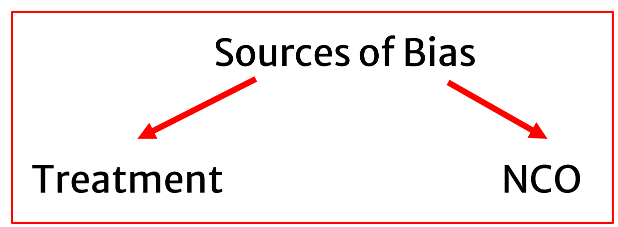
  - Better type 1 error control than simple pooled estimate
  - Tradeoff between ability to
    - include unbiased RWD (increase power)
    - exclude RWD with non-negligible bias (maintain nominal type 1 error)

Berkeley
UNIVERSITY OF CALIFORNIA

6 Choose a Statistical Estimator
7 Causal Sensitivity Analysis

- We tried to design a compatible study.
  - **Sensitivity Analysis:** How large could the causal gap be?
- Choose a statistical estimator that
  - uses **evidence about the causal gap** to decide
  - whether to include RWD (or how to weight RWD)
- **Commonly used evidence of bias:**
  - Difference in outcomes between RCT and RWD controls
  - Effect of treatment on a negative control outcome (NCO)

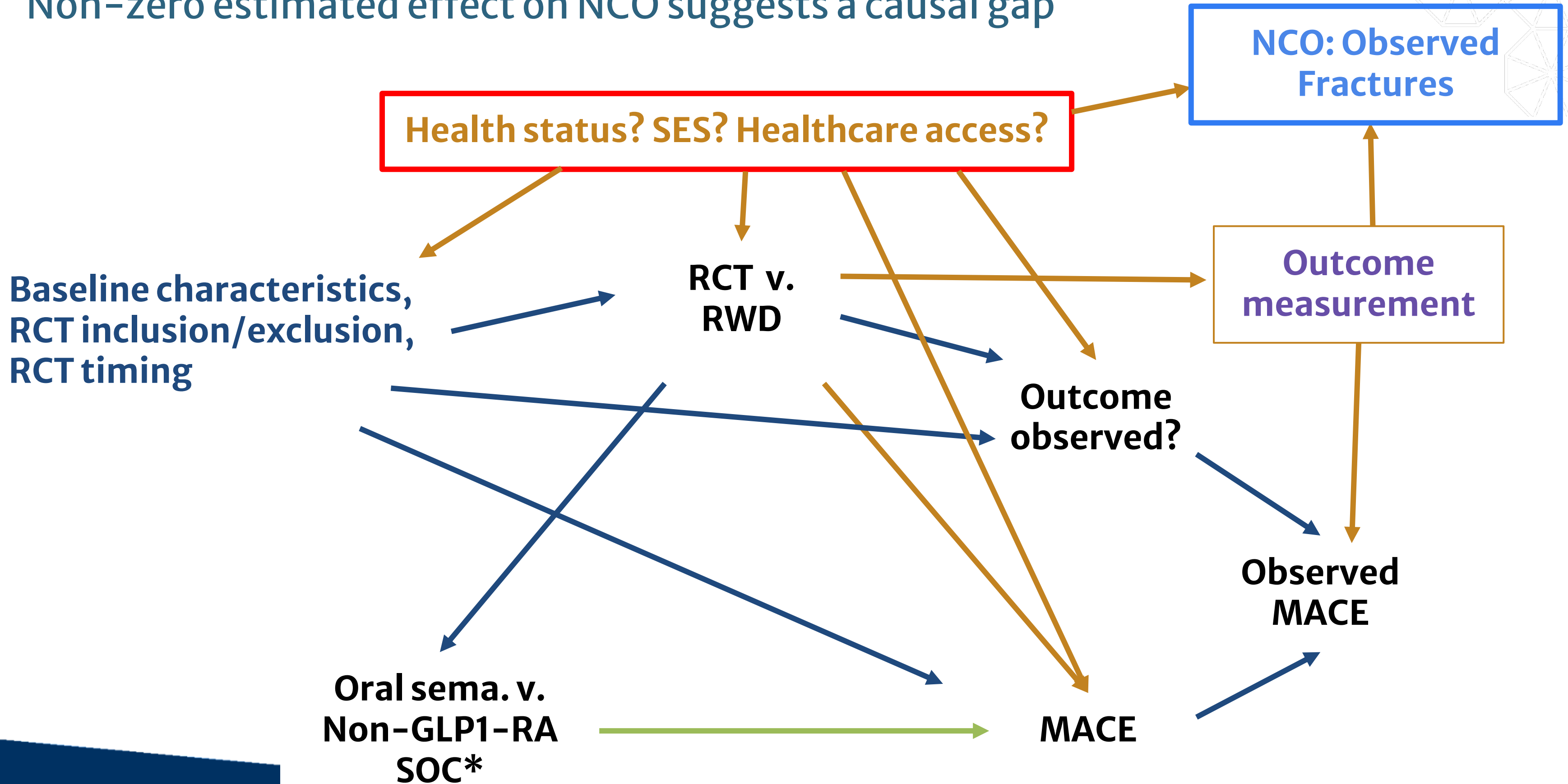




Berkeley
UNIVERSITY OF CALIFORNIA

Negative Control Outcome

- NCO[12–14]:
  - Not affected by treatment
  - Affected by unmeasured factors causing bias
- Non-zero estimated effect on NCO suggests a causal gap



Berkeley
UNIVERSITY OF CALIFORNIA

Estimators for Integration of RCT & RWD

Class of Estimator	Examples
Comparison of RCT/RWD Outcomes	
Bayesian Dynamic Borrowing	Ibrahim et al. (2000)[15], Hobbs et al. (2011)[16], Schmidli et al. (2014)[17],...
Test-then-pool/ Equivalence test	Viele et al. (2014)[18], Hartman & Hidalgo (2018)[19], Li et al. (2020)[20], …
Shrinkage estimators	Green and Strawderman (1991)[21], Rosenman et al. (2020)[22], …
Optimize bias-variance tradeoff	Yang et al. (2020)[23], Chen et al. (2021)[24], Cheng et al. (2021)[25], Oberst et al. (2022)[26], Dang et al. (2022)[27], …
Effect of Treatment on a Negative Control Outcome (NCO)	
Test or adjust for bias using NCO	Sofer et al. (2016)[12], Miao et al. (2020)[13], Shi et al. (2020)[14], Dang et al. (2022)[27], …

Data fusion estimators!

Berkeley
UNIVERSITY OF CALIFORNIA

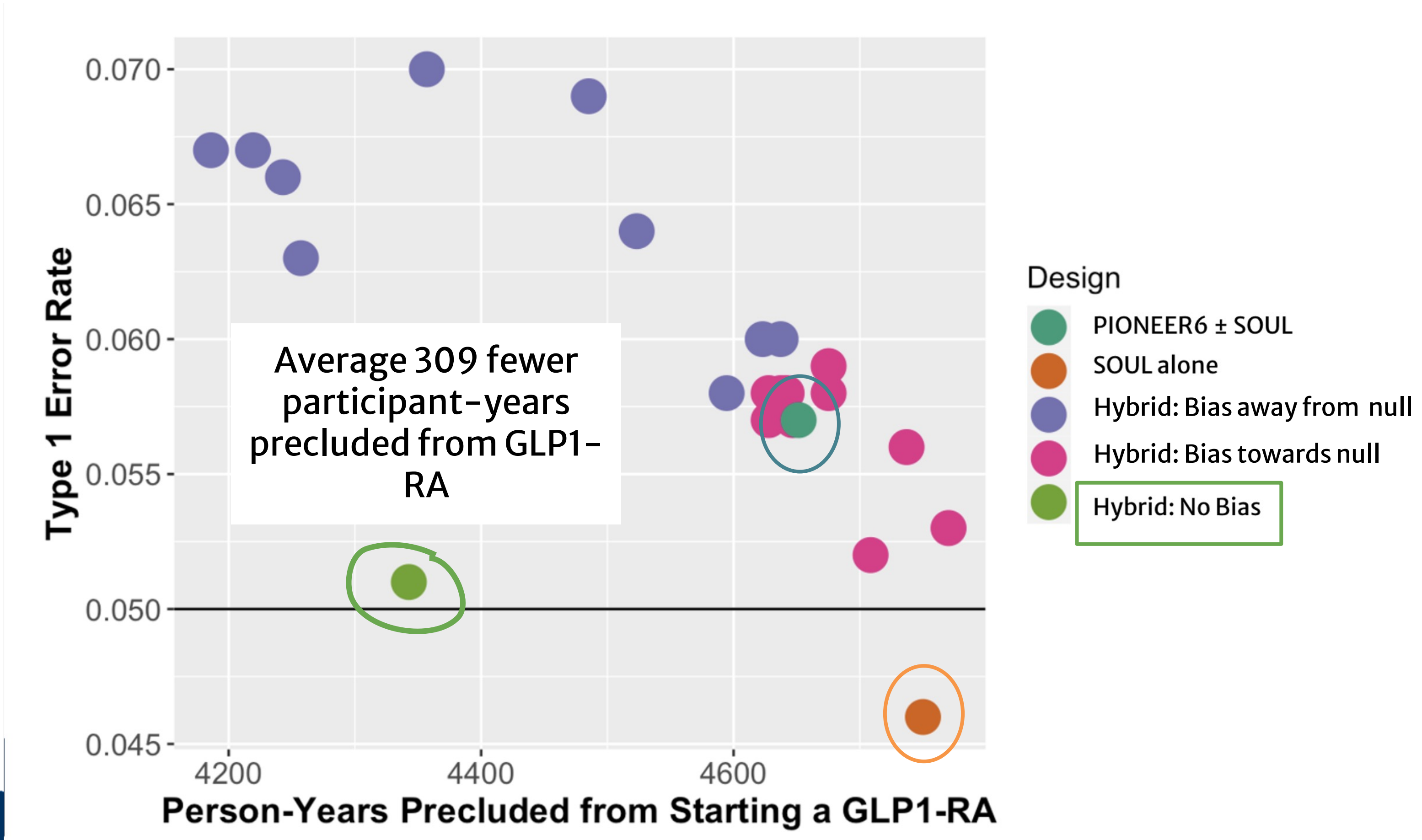
8 Compare Possible Study Designs

- Specified one possible study design: **Hybrid RCT-RWD**
- Other possible designs?
  - **PIONEER 6 then SOUL if non-significant for superiority**
  - **Single superiority RCT**
  - Others (e.g., adaptive designs[28])
- How should we compare them?
  - Standard metrics:
    - **Type 1 error**
    - Power
    - Bias, variance, 95% CI coverage, mean squared error…
  - Why would we consider the hybrid design instead of RCTs?
    - **Person-time precluded from receiving a GLP1-RA**

Berkeley
UNIVERSITY OF CALIFORNIA

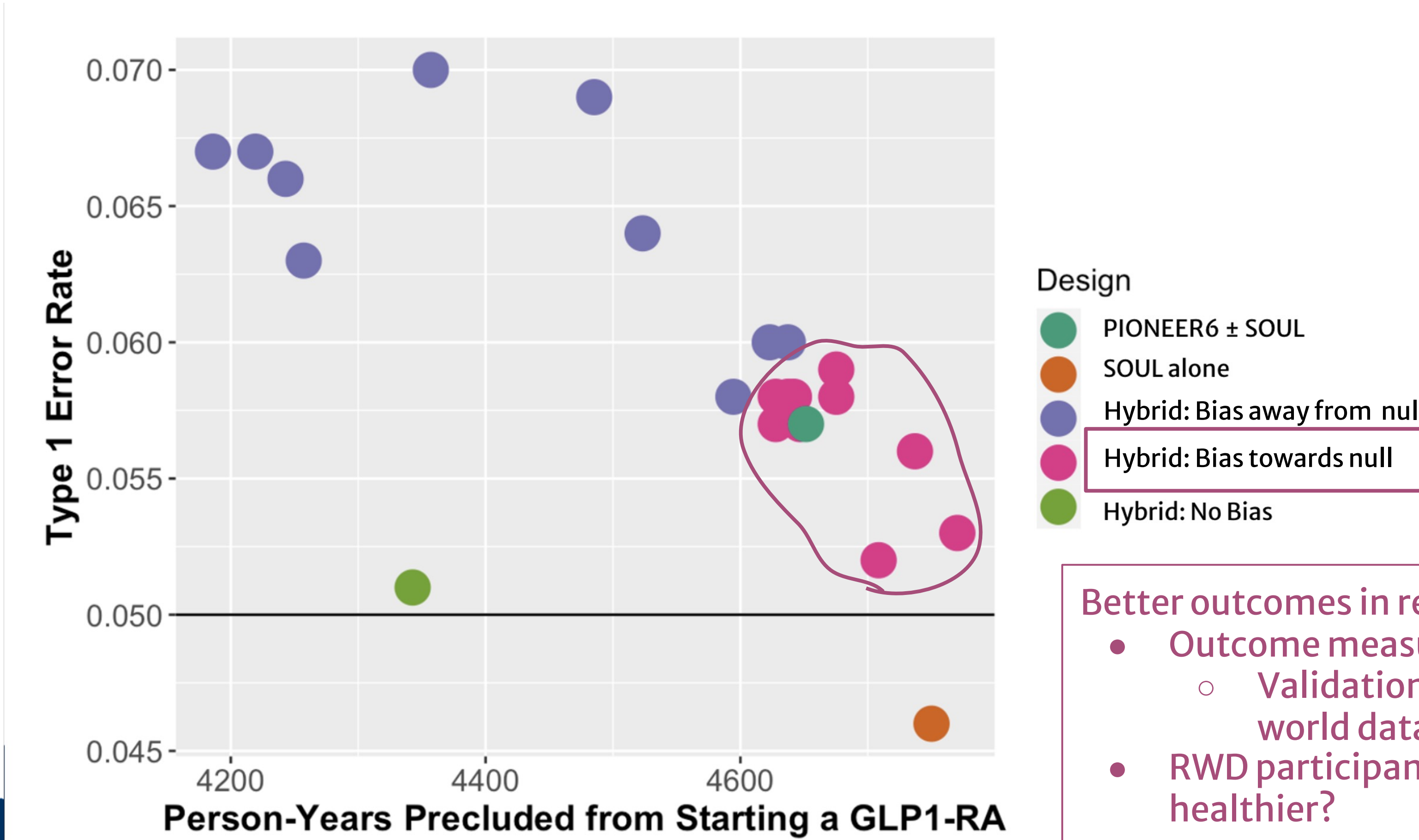
Simulation: 1000 Iterations

- Mimic real data (sample size, event rates, censoring ...)
- 10 magnitudes of RWD bias in positive and negative directions up to ±2.1%



Simulation: 1000 Iterations

- Mimic real data (sample size, event rates, censoring …)
- 10 magnitudes of RWD bias in positive and negative directions up to ±2.1%



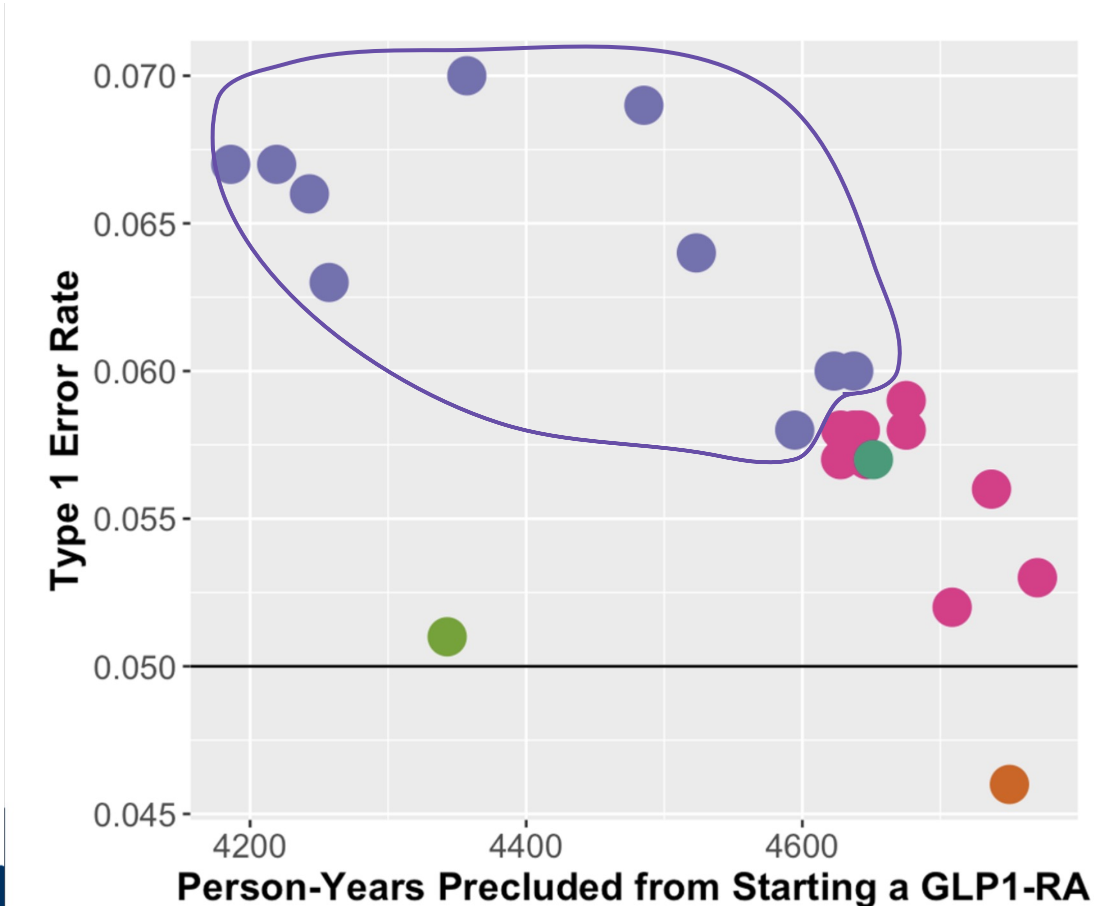
Better outcomes in real world?

- Outcome measurement?
  - Validation of real-world dataset?
- RWD participants healthier?

Berkeley
UNIVERSITY OF CALIFORNIA

Simulation: 1000 Iterations

- Mimic real data (sample size, event rates, censoring …)
- 10 magnitudes of RWD bias in positive and negative directions up to ±2.1%

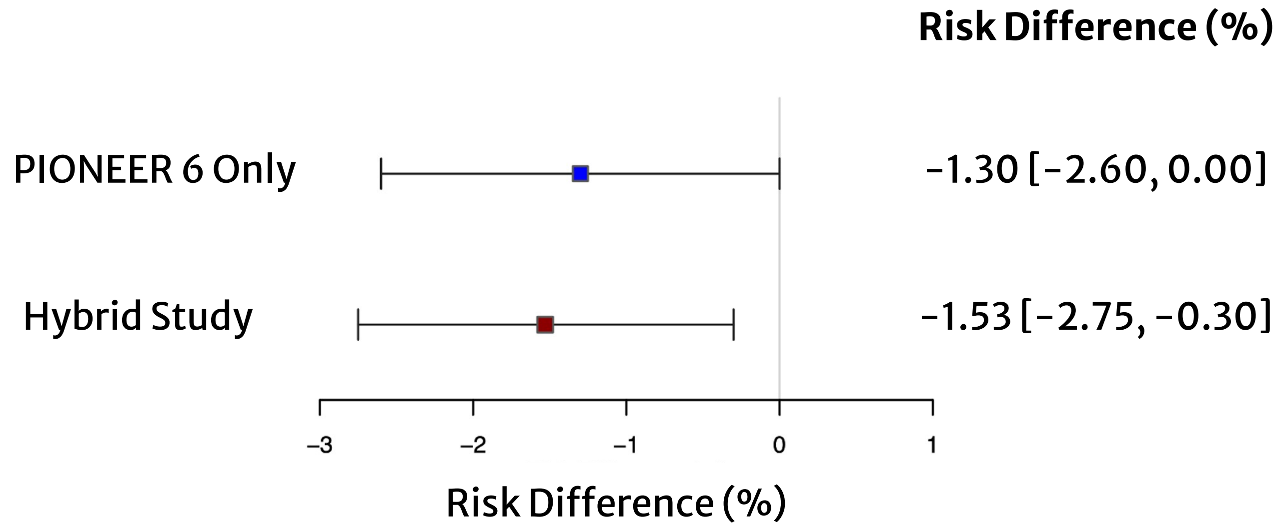


Worse outcomes in real world?
- Monitoring, care
- Access, engagement
- etc.

Plausible range of bias helps to:
- Interpret final results
- Choose between designs

PIONEER 6 + Optum Results

Estimated Causal Risk Difference (%) of MACE within 1 Year



- Confidence intervals narrower
- Point estimate shifted by -0.23%
  - Normal sample variability
  - Modified target population
  - Causal gap: worst plausible type 1 error control from simulations

Berkeley
UNIVERSITY OF CALIFORNIA

Summary: Lessons Learned

- **Hybrid RCT–RWD Studies:**
  - Potential to reduce size of RCT control arm or avoid large RCT entirely
  - With protection against bias
- Optimizing CI coverage:
  - Careful consideration of controls and covariates (causal framework)
  - Data fusion estimator: capable of providing **nominal or close to nominal coverage** across a range of potential bias (bad controls)
- Optimizing power:
  - More inclusive RCTs
  - RWD that also includes treatment
  - Adaptive designs
- Simulations can clarify motivations and facilitate stakeholder discussions
- Case study:
  - SOUL trial to report results in 2024
  - Pioneer 6 + RWD supports superiority
  - **Role of hybrid trials for secondary indications?**

berkeley
UNIVERSITY OF CALIFORNIA

Thank you!

Acknowledgements:

- JICI IORD Working Group:
- UC Berkeley: Mark van der Laan, Maya Petersen, Andrew Mertens, David Chen, Ahmed Alaa, Rachael Phillips
- Novo Nordisk: Edwin Fong, Kim Clemmensen, Jens Tarp, Kajsa Kvist, Trine Abrahamsen, Henrik Ravn, Lars Damgaard
- University of Copenhagen: Theis Lange, Thomas Gerds, Torben Martinussen, Zehao Su, Frank Eriksson
- Clinician Partners: John Buse, Richard Pratley, Steve Marso

Questions?

Comments?

Berkeley
UNIVERSITY OF CALIFORNIA

References

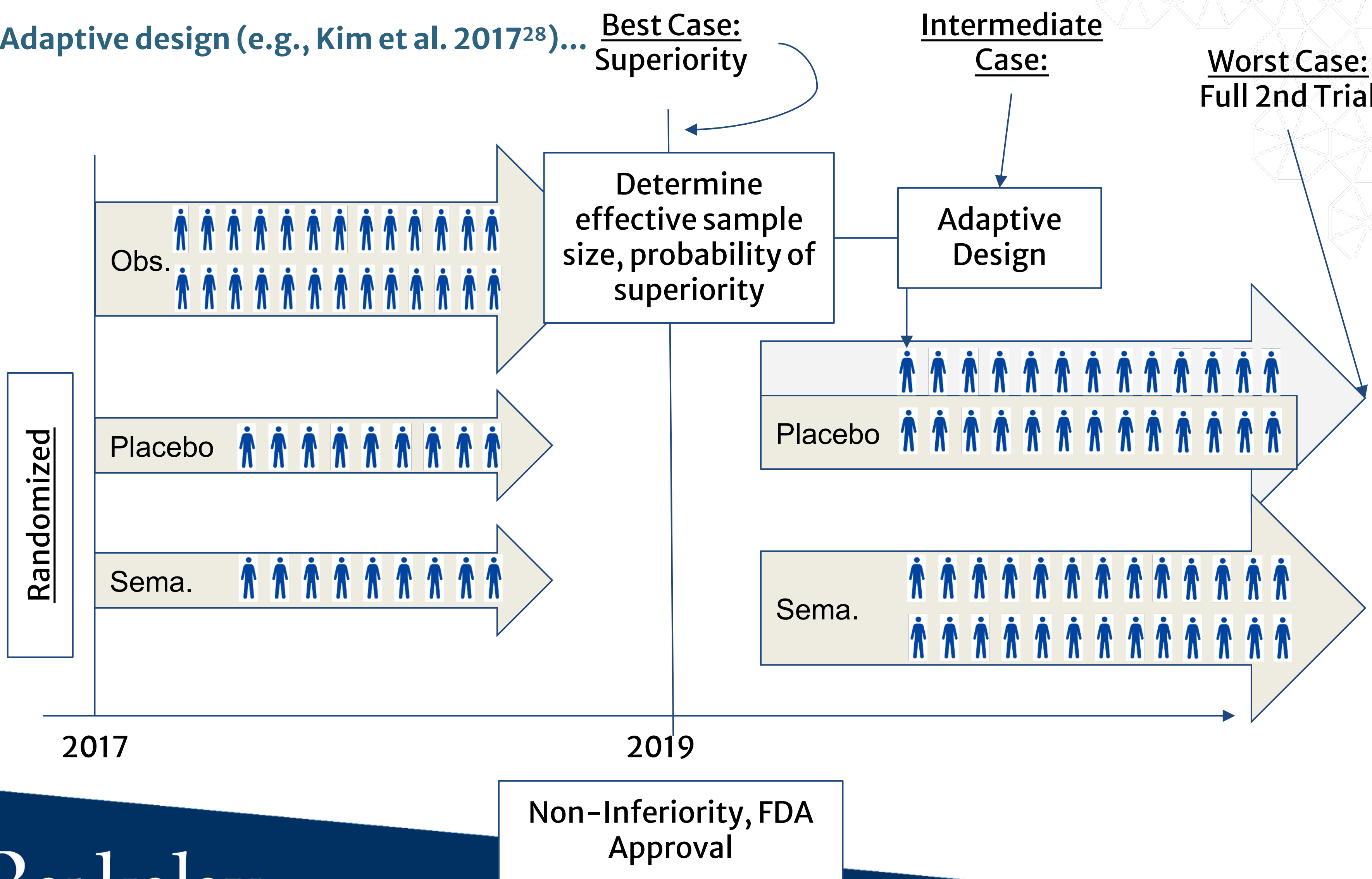
1. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG: Int J Obstet Gy*. 2018;125(13):1716-1716. doi:10.1111/1471-0528.15199
2. Rivera D. Designing External Controls using Real World Data for Pediatric Cancer Drug Development. Published online May 2021. Accessed July 19, 2021. https://www.fda.gov/media/148543/download
3. Petersen ML, van der Laan MJ. Causal Models and Learning from Data: Integrating Causal Modeling and Statistical Estimation. *Epidemiology*. 2014;25(3):418-426. doi:10.1097/EDE.0000000000000078
4. Rodbard HW, Lingvay I, Reed J, et al. Semaglutide Added to Basal Insulin in Type 2 Diabetes (SUSTAIN 5): A Randomized, Controlled Trial. *The Journal of Clinical Endocrinology & Metabolism*. 2018;103(6):2291-2301. doi:10.1210/jc.2018-00070
5. Marso, S. P., Bain, S. C., Consoli, A., Eliaschewitz, F. G., Jódar, E., Leiter, L. A., Lingvay, I., Rosenstock, J., Seufert, J., Warren, M. L., Woo, V., Hansen, O., Holst, A. G., Pettersson, J., & Vilsbøll, T. (2016). Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. *New England Journal of Medicine*, *375*(19), 1834–1844. https://doi.org/10.1056/NEJMoa1607141
6. Aroda VR, Rosenstock J, Terauchi Y, et al. PIONEER 1: Randomized Clinical Trial of the Efficacy and Safety of Oral Semaglutide Monotherapy in Comparison With Placebo in Patients With Type 2 Diabetes. *Diabetes Care*. 2019;42(9):1724-1732. doi:10.2337/dc19-0749
7. Husain, M., Birkenfeld, A. L., Donsmark, M., Dungan, K., Eliaschewitz, F. G., Franco, D. R., Jeppesen, O. K., Lingvay, I., Mosenzon, O., Pedersen, S. D., Tack, C. J., Thomsen, M., Vilsbøll, T., Warren, M. L., & Bain, S. C. (2019). Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. *New England Journal of Medicine*, *381*(9), 841–851. https://doi.org/10.1056/NEJMoa1901118
8. Buse JB, Wexler DJ, Tsapas A, et al. 2019 Update to: Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*. 2020;43(2):487-493. doi:10.2337/dci19-0066
9. Gruber S, Phillips RV, Lee H, Ho M, Concato J, van der Laan MJ. Targeted learning: Towards a future informed by real-world evidence. *Statistics in Biopharmaceutical Research*. Published online February 21, 2023:1-23. doi:10.1080/19466315.2023.2182356
10. Ghadessi, M., Tang, R., Zhou, J., Liu, R., Wang, C., Toyoizumi, K., Mei, C., Zhang, L., Deng, C. Q., & Beckman, R. A. (2020). A roadmap to using historical controls in clinical trials – by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). *Orphanet Journal of Rare Diseases*, *15*(1), 69. https://doi.org/10.1186/s13023-020-1332-x
11. Rylance RT, Wagner P, Omerovic E, et al. Assessing the external validity of the VALIDATE-SWEDEHEART trial. *Clinical Trials*. 2021;18(4):427-435. doi:10.1177/17407745211012438
12. Sofer, T., Richardson, D. B., Colicino, E., Schwartz, J., & Tchetgen Tchetgen, E. J. (2016). On Negative Outcome Control of Unobserved Confounding as a Generalization of Difference-in-Differences. *Statistical Science*, *31*(3). https://doi.org/10.1214/16-STS558
13. Miao, W., Shi, X., & Tchetgen, E. T. (2020). A Confounding Bridge Approach for Double Negative Control Inference on Causal Effects. *ArXiv:1808.04945 [Stat]*. http://arxiv.org/abs/1808.04945
14. Shi, X., Miao, W., Nelson, J. C., & Tchetgen Tchetgen, E. J. (2020). Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(2), 521–540. https://doi.org/10.1111/rssb.12361
15. Ibrahim, J., & Chen, M.-H. (2000). Power Prior Distributions for Regression Models. *Statistical Science*, *15*(1), 46–60.
16. Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., & Sargent, D. J. (2011). Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. *Biometrics*, *67*(3), 1047–1056. https://doi.org/10.1111/j.1541-0420.2011.01564.x
17. Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., & Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information: Robust Meta-Analytic-Predictive Priors. *Biometrics*, *70*(4), 1023–1032. https://doi.org/10.1111/biom.12242
18. Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., Micallef, S., Roychoudhury, S., & Thompson, L. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, *13*(1), 41–54. https://doi.org/10.1002/pst.1589
19. Hartman, E., & Hidalgo, F. D. (2018). An Equivalence Approach to Balance and Placebo Tests. *American Journal of Political Science*, *62*(4), 1000–1013. https://doi.org/10.1111/ajps.12387
20. Li, W., Liu, F., & Snavely, D. (2020). Revisit of test-then-pool methods and some practical considerations. *Pharmaceutical Statistics*, *19*(5), 498–517. https://doi.org/10.1002/pst.2009
21. Green, E., & Strawderman, W. (1991). A James-Stein Type Estimator for Combining Unbiased and Possibly Biased Estimators. *Journal of the American Statistical Association*, *86*(416), 1001–1006.
22. Rosenman, E., Basse, G., Owen, A., & Baiocchi, M. (2020). *Combining Observational and Experimental Datasets Using Shrinkage Estimators*. https://doi.org/10.48550/ARXIV.2002.06708
23. Yang, S., Zeng, D., & Wang, X. (2020). *Elastic Integrative Analysis of Randomized Trial and Real-World Data for Treatment Heterogeneity Estimation*. https://arxiv.org/abs/2005.10579
24. Chen, S., Zhang, B., & Ye, T. (2021). *Minimax Rates and Adaptivity in Combining Experimental and Observational Data*. https://doi.org/10.48550/ARXIV.2109.10522
25. Cheng, D., & Cai, T. (2021). *Adaptive Combination of Randomized and Observational Data*. https://doi.org/10.48550/ARXIV.2111.15012
26. Oberst, M., D'Amour, A., Chen, M., Wang, Y., Sontag, D., & Yadlowsky, S. (2022). *Bias-robust Integration of Observational and Experimental Estimators*. https://doi.org/10.48550/ARXIV.2205.10467
27. Dang, L. E., Tarp, J. M., Abrahamsen, T. J., Kvist, K., Buse, J. B., Petersen, M., & van der Laan, M. (2022). *A Cross-Validated Targeted Maximum Likelihood Estimator for Data-Adaptive Experiment Selection Applied to the Augmentation of RCT Control Arms with External Data*. https://doi.org/10.48550/ARXIV.2210.05802
28. Kim, M., Harun, N., Liu, C., Khoury, J. C., & Broderick, J. P. (2018). Bayesian selective response-adaptive design using the historical control. *Statistics in Medicine*, *37*(26), 3709–3722. https://doi.org/10.1002/sim.7836

Berkeley
UNIVERSITY OF CALIFORNIA

Extra Slides

Effect of Oral Semaglutide on MACE

Adaptive design (e.g., Kim et al. 2017[28])...



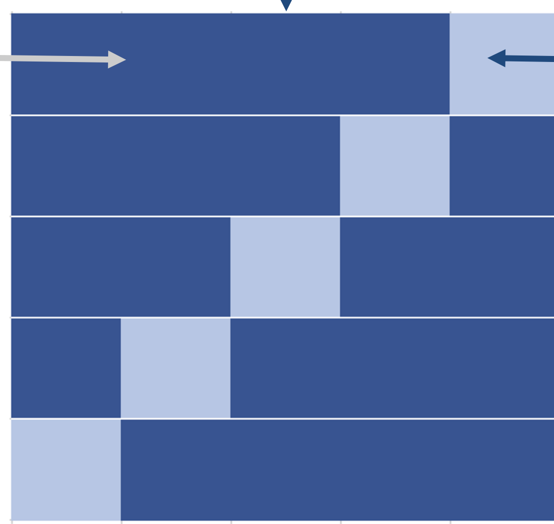
Experiment-Selector CV-TMLE[27]

- **Goal:** Select the experiment (RCT only or RCT with RWD) that optimizes the bias-variance tradeoff for the target parameter
- Separate experiment-selection from effect estimation using cross-validation

Divide the data into V folds

For each experiment-selection set:
For each experiment:
1. Estimate bias from including RWD (bias estimate may include effect on NCO)
2. Estimate variance

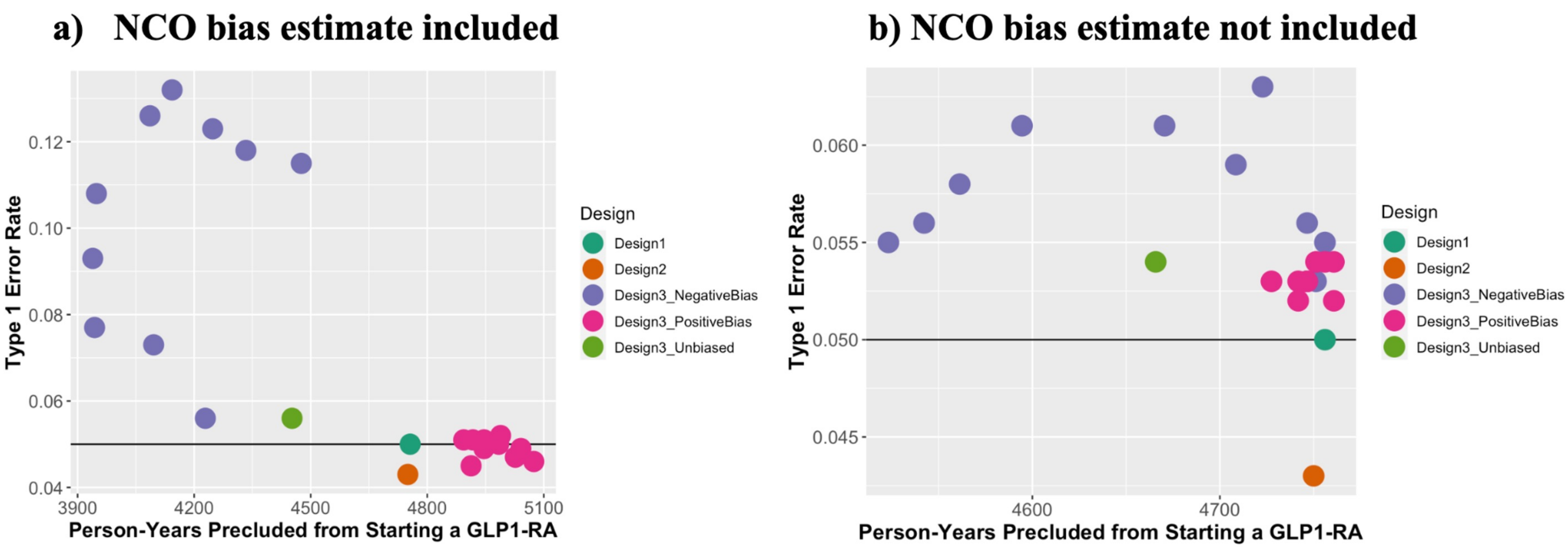
Select experiment with lowest estimated mean squared error ($Bias^2$ + Variance)



For each estimation set:
1. Select the experiment chosen in the corresponding experiment-selection set
2. Use TMLE to estimate the causal risk difference (CRD) for that experiment

Final CRD estimate is an average of fold-specific CRD estimates

Supplementary Figure 1: Simulation Results by Study Design with Different Amounts of RWD Bias when Bias has No Effect on NCO



Berkeley
UNIVERSITY OF CALIFORNIA