Disclosure

- Jinma Ren and co-authors are employees and stockholders of Pfizer Inc.
- This presentation reflects their personal views and not the views of Pfizer or any other organization.

Comparing G-Computation, Propensity Score-Based Weighting, and Targeted Maximum Likelihood Estimation for Analyzing Externally Controlled Trials with an Unmeasured Confounder: A Simulation Study

Jinma Ren[1], Paul Cislo[2], Joseph C. Cappelleri[3], Patrick Hlavacek[2], Marco DiBonaventura[2]

ISPOR 2022
May 15-18, 2022
Washington, DC, USA Area and Virtual

[1]Pfizer Inc, Collegeville, PA, USA
[2]Pfizer Inc, New York, NY, USA
[3]Pfizer Inc, Groton, CT, USA

Introduction and objective

- In orphan and rare disease, single-arm trials (SATs) are common
- To provide context for SATs, an external control arm (ECA) can be employed, although there is a risk of bias due to confounding factors
- Per guidelines, when using ECA, adjustments for baseline characteristics is one of key step toward addressing confounders

FDA Guidelines, 2019 & 2021; ICH E10, 2021; Gray, 2020

Introduction and objective

- The highlighted methods with blue are the common approaches that use all available individual-level data, which is important in orphan and rare disease.

Common methods for minimizing the effects of potential confounders

	Data type	Methods	
Frequentist	**Individual data**	*G-computation*	One of Robins' g methods
		Propensity score (PS)-based weighting	• Inverse probability of treatment weighting (IPTW) • Overlap weighting (OW) • Standardized mortality/morbidity ratio (SMR)
		Doubly robust estimator	• Targeted Maximum Likelihood Estimation (TMLE)
		Matching/stratification	• Variable matching • Propensity score matching/stratification
	Summary data	*Network analysis, simulated treatment outcome (STC), or matching-adjusted indirect comparison (MAIC)*	
Mostly Bayesian	**Individual / summary data**	*Meta-analytic combined (MAC), meta-analytic predictive (MAP)*	

Ghadessi, 2020; Naimi, 2017; Schuler 2017

Introduction and objective

- When real-world data (RWD) is used as an ECA, randomization is not possible, so adjustments are needed
- The accuracy of each method for minimizing the effects of potential confounders is unclear
- We aimed to compare different methods (g-computation, PS-based weighting, TMLE) for adjusting for confounding while estimating average treatment effects (ATEs) in a simulated external control study with individual-level data, which included both measured and unmeasured confounders

Study design (a simulation study)

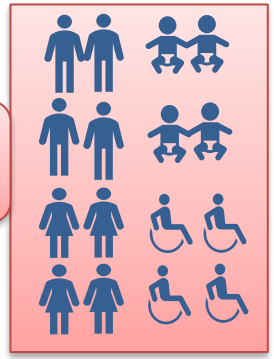
Simulate population and treatment assignment

Simulate outcomes

Estimate treatment effects

Single-arm trial

External control



- **Continuous endpoint**
- **Binary endpoint**
- **Time-to-event endpoint**

Minimizing potential confounders

Study design (simulated data set)

- A simulation study to compare g-computation, PS-based weighting, and TMLE methods was conducted with the following characteristics

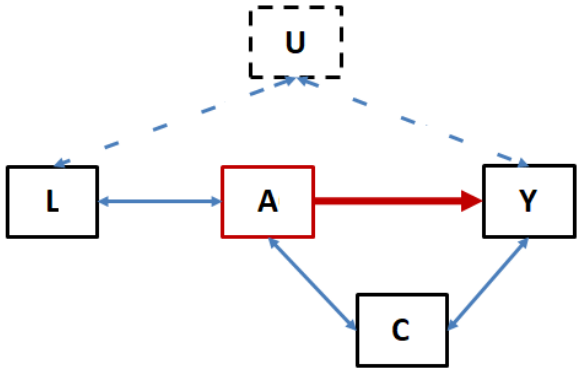
Items	Description
Population size (patients)	20000 (e.g., 10000 each group)
Sample size (patients)	100, 200
Number of samples	3000
Outcome types	Continuous, binary, and time to event
Unmeasured confounder	One (with small, medium, or strong effect)
Measured confounders	Two (one is associated with the unmeasured confounder)

*Sampling with replacement

Study design (3 scenarios for unmeasured confounder)

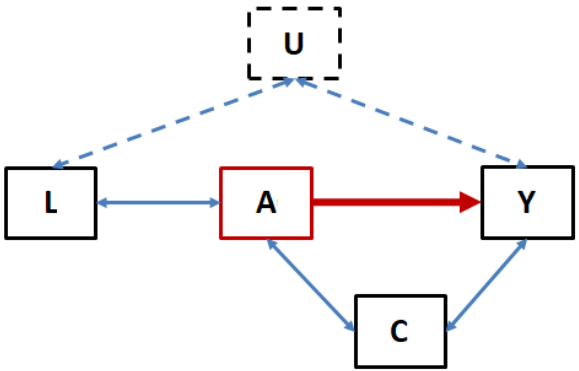
Small confounding

a) Unmeasured confounder *U* associated with *L* and *Y* (*small confounding effect*)



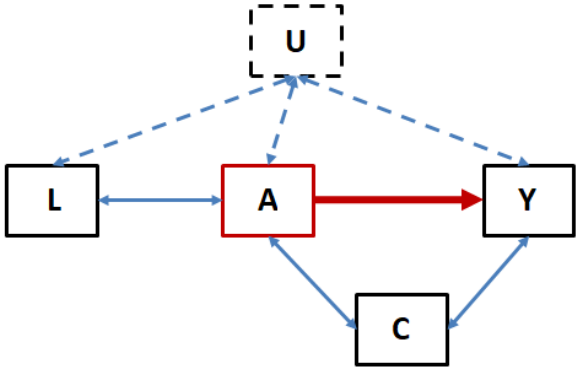
Medium confounding

b) Unmeasured confounder *U* associated with *L* and *Y* (*medium confounding effect*)



Large confounding

c) Unmeasured confounder *U* associated with *L*, *A*, and *Y* (*large confounding effect*)



Y : Outcome
A: Treatment at baseline
L : Confounder at baseline
C: Other baseline confounders
U: Unmeasured confounder

Measured causation
Measured association
Unmeasured association

Study design (model comparison)

- Selected methods
  - G-computation
  - Propensity score (PS)-based weighting: IPTW, OW, SMR
  - Targeted Maximum Likelihood Estimation (TMLE)
- Reference method: raw model without adjusting any confounders
- True effects: adjustment for both measured and unmeasured confounders using the target population
- Comparing the true effect with the results (point estimate and 95%CI) from each model, and calculating root mean squared error (RMSE), coverage and width of 95% confidence interval

Results (summary of confounders in the simulated population)

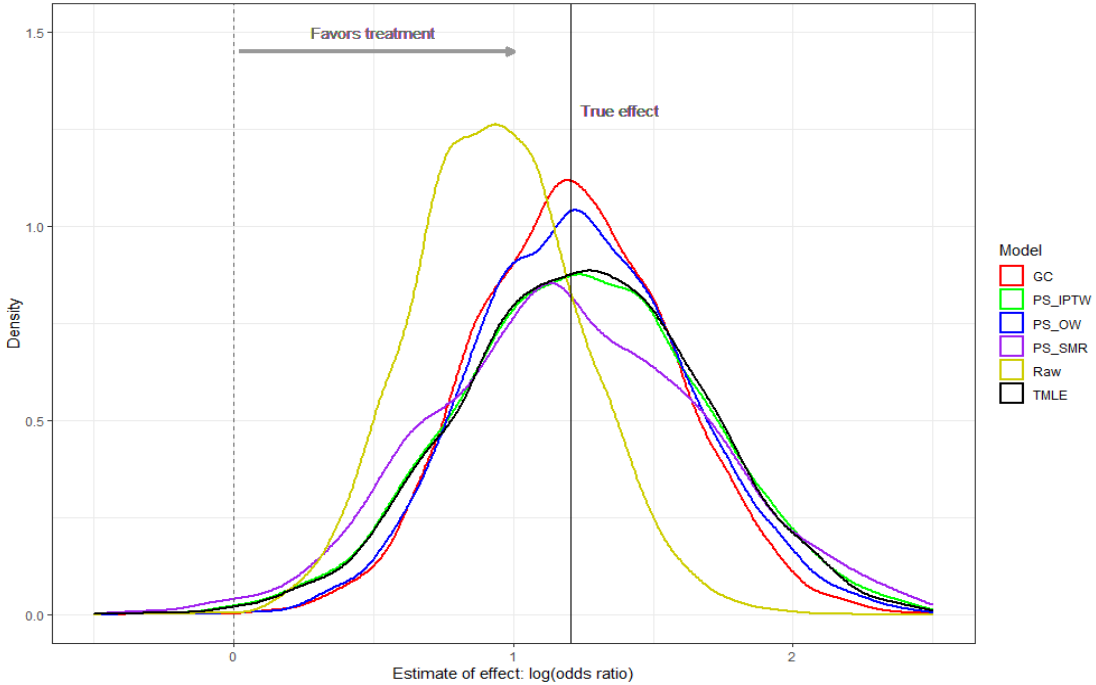
- Unmeasured confounder (U) had different confounding in the three scenarios

Scenario	Confounders	Control (A=0)	Treatment (A=1)	Standardized mean difference (SMD)
Small confounding	U (mean ± SD)	0.45 ± 0.99	0.56 ± 1.01	0.08
	C (mean ± SD)	47.0 ± 6.12	54.4 ± 6.06	0.86
	L (n, %)	2080 (18.0)	2860 (34.0)	0.42
Medium confounding	U (mean ± SD)	0.40 ± 0.97	0.62 ± 1.02	0.16
	C (mean ± SD)	46.7 ± 6.08	54.2 ± 5.93	0.88
	L (n, %)	2355 (20.6)	3466 (40.4)	0.44
Large confounding	U (mean ± SD)	0.27 ± 0.95	0.76 ± 1.00	0.36
	C (mean ± SD)	46.7 ± 6.14	53.8 ± 6.05	0.82
	L (n, %)	1880 (17.4)	3941 (42.9)	0.55

*U, unmeasured confounder; C & L, measured confounders

Results (N = 200, Endpoint = binary outcome, Scenario = small confounding)

- ❑ **The estimated treatment effects by selected methods were <u>approximately unbiased</u> when the unmeasured confounding was <u>small</u>** *(SMD=0.08)*
- ❑ **GC had the smallest RMSE followed by PS_OW**

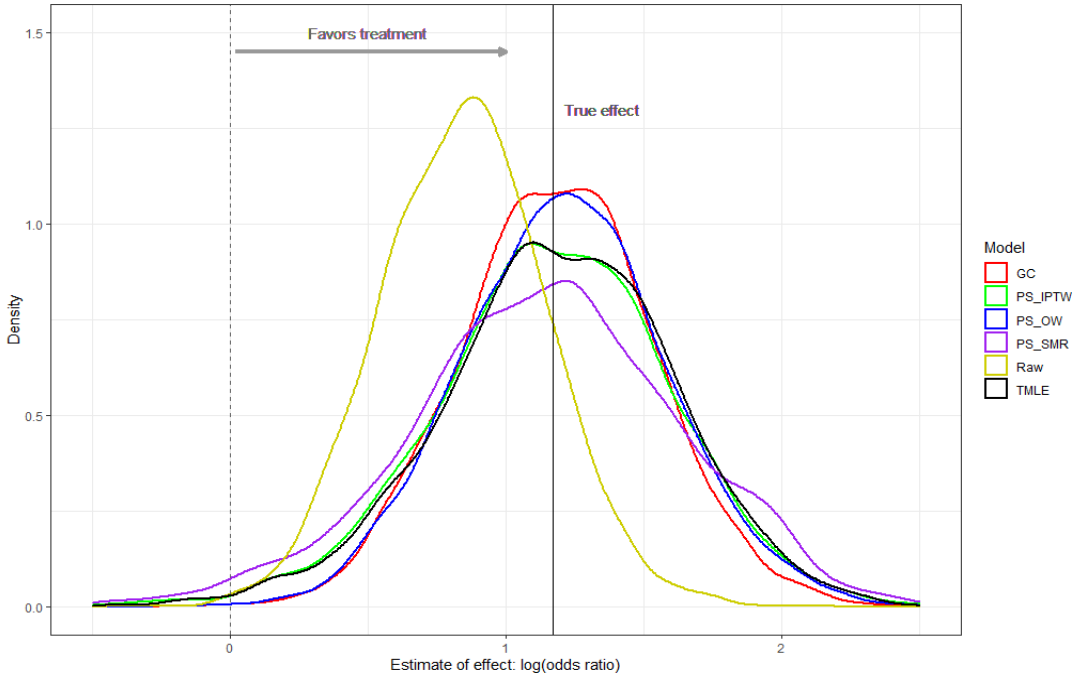


Method	RMSE	Coverage
GC	0.355	0.938
PS_IPTW	0.443	0.836
PS_OW	0.381	0.894
PS_SMR	0.500	0.777
Raw	0.405	0.854
TMLE	0.432	0.974

GC, g-computation; IPTW, inverse probability of treatment weighting; PS-, propensity score-based; RMSE, root mean squared error; SMD, standardized mean difference; SMR, standardized mortality or morbidity ratio; OW, overlap weighting; TMLE, targeted maximum likelihood estimation

Results (N = 200, Endpoint = binary outcome, Scenario = medium confounding)

- ❑ **The estimated treatment effects by selected methods were <u>relatively unbiased</u> when the unmeasured confounding was <u>medium</u>** *(SMD=0.16)*
- ❑ **GC had the smallest RMSE followed by PS_OW**

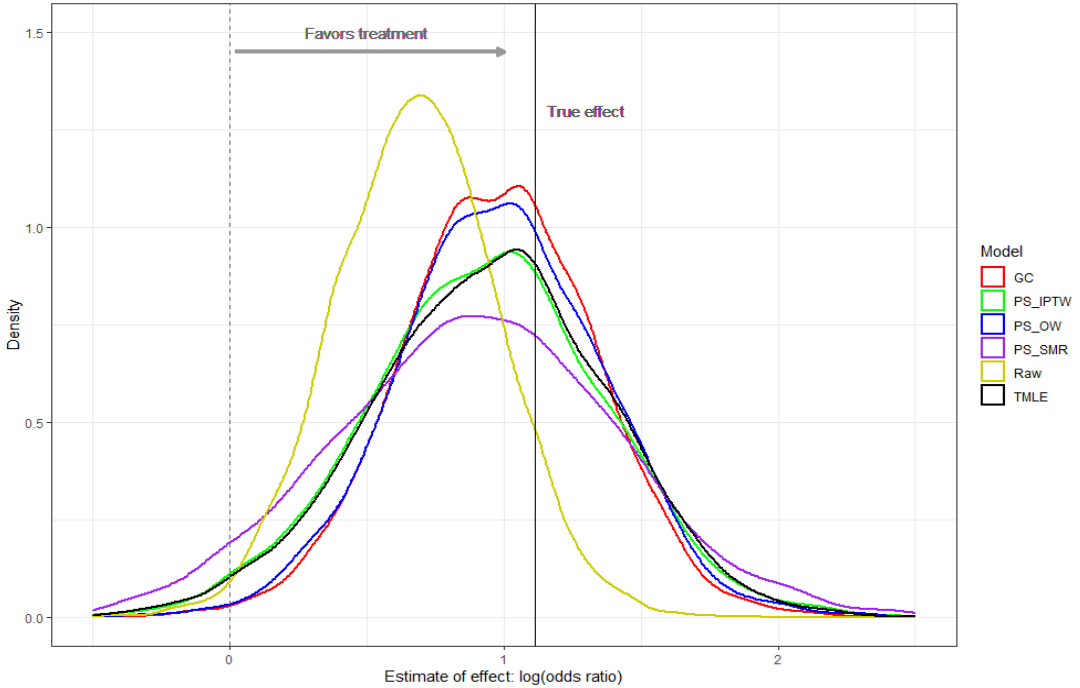


Method	RMSE	Coverage
GC	0.349	0.934
PS_IPTW	0.430	0.848
PS_OW	0.372	0.891
PS_SMR	0.497	0.773
Raw	0.443	0.796
TMLE	0.423	0.974

GC, g-computation; IPTW, inverse probability of treatment weighting; PS-, propensity score-based; RMSE, root mean squared error; SMD, standardized mean difference; SMR, standardized mortality or morbidity ratio; OW, overlap weighting; TMLE, targeted maximum likelihood estimation

Results (N = 200, Endpoint = binary outcome, Scenario = large confounding)

- ❑ **The estimated treatment effects by selected methods became <u>biased</u> when the unmeasured confounding was <u>large</u> *(SMD=0.36)***
- ❑ **GC had the smallest RMSE followed by PS_OW**



Method	RMSE	Coverage
GC	0.376	0.917
PS_IPTW	0.470	0.794
PS_OW	0.394	0.867
PS_SMR	0.573	0.703
Raw	0.524	0.687
TMLE	0.462	0.958

GC, g-computation; IPTW, inverse probability of treatment weighting; PS-, propensity score-based; RMSE, root mean squared error; SMD, standardized mean difference; SMR, standardized mortality or morbidity ratio; OW, overlap weighting; TMLE, targeted maximum likelihood estimation

Results for other outcomes

- The results for continuous and time-to-event outcomes were similar with the results for binary outcome
- However, the unmeasured confounder had the least impact on the estimate of treatment effect for the time-to-event outcome compared to the other two outcomes, because
  - The correlation (Pearson's) between the time-to-event outcome and the unmeasured confounder was the smallest compared to the other two outcomes. For instance:
    - Continuous outcome: -0.44
    - Binary outcome: -0.23
    - Time-to-event outcome: -0.12

Discussion

- As expected, both the unmeasured and measured factors played a role of potential confounder in the simulated data.
- G-computation had the smallest RMSEs in the simulated settings followed by PS-based OW, which is consistent with the other recent simulation study (Chatton et al. 2020). But Chatton's study investigated binary outcome only and did not emphasize on unmeasured confounders.
- Limitations:
  - Only one unmeasured confounder was investigated
  - Time-varying confounders were not explored
  - The current R package "LTMLE" is not able to provide a hazard ratio (HR) directly for time-to-event outcomes

Conclusions

- The selected methods can be used to reduce the biases due to measured and unmeasured confounders in ECA studies
- The extent of bias reduction depends on the technique used and the strength of association between the unmeasured confounder and the outcome or treatment
- The findings in this study suggest that g-computation and PS-based OW might be preferable alternative approaches we investigated, producing relatively unbiased estimates

Reference

- Chatton, A., et al., *G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study.* Sci Rep, 2020. **10**(1): p. 9219.
- FDA. *Rare Diseases: Natural History Studies for Drug Development Guidance for Industry*. 2019 [cited 2022 February 17]; Available from: https://www.fda.gov/media/122425/download.
- FDA. *Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products*. 2021 [cited 2022 January 3]; Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug.
- Gray, C.M., et al., *A Framework for Methodological Choice and Evidence Assessment for Studies Using External Comparators from Real-World Data.* Drug Saf, 2020. **43**(7): p. 623-633.
- Ghadessi, M., et al., *A roadmap to using historical controls in clinical trials - by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG).* Orphanet J Rare Dis, 2020. **15**(1): p. 69.
- ICH. *ICH E10 Choice of control group in clinical trials*. 2001 [cited 2022 January 3]; Available from: https://www.ema.europa.eu/en/ich-e10-choice-control-group-clinical-trials.
- Naimi, A.I., S.R. Cole, and E.H. Kennedy, *An introduction to g methods.* Int J Epidemiol, 2017. **46**(2): p. 756-762.
- Schuler, M.S. and S. Rose, *Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies.* Am J Epidemiol, 2017. **185**(1): p. 65-73.

Q & A

- My contact: Jinma.ren@pfizer.com