Improving Causal Inference from Observational Data

A Comparative Analysis of Confidence Interval Methods in Sequential Target Trial Emulation

26 May 2025

Florian Metwaly

Utrecht University

Causal Inference

- **Randomized control trials** are the **gold standard** of causal inference
- But, not always feasible
  - Ethical limitations
  - Practical limitations

→ More and more common to use **observational data**

- Observational prone to biases

Sequential Target Trial Emulation

- Copying of observations
- **Confidence Intervals** can’t be estimated directly
  - One individual can be in multiple trials
    → Violates independence assumption
- Literature recommends **non-parametric bootstrap**
- In practice: **sandwich-type estimators**

How do non-parametric bootstrap confidence intervals compare to the sandwich-type confidence intervals?

Methods

- Simulation Study
  - Sandwich-type estimator
  - Non-parametric bootstrap
    - Empirical bootstrap
    - Percentile bootstrap
  - 81 scenarios
- Development of `TargetTrialEmulation.jl`
  - A Julia package for computational efficient estimation of bootstrap confidence intervals

Results

- **Bootstrap CIs** are **narrower** in some cases in small and medium sample sizes
- Bootstrap **coverage** is **more often closest to 95%**
- **Performance degrades** at high event rates or large sample sizes

Method	Closest to 95%
Sandwich	24.7%
Empirical	36.3%
Percentile	39.0%

Table: Proportion of simulation scenarios where the method's coverage was closest to the nominal 95% target.

Results

- **Point estimates are biased**, especially with:
  - Small sample sizes
  - High outcome event rates
  - Later follow-up times
- **Bootstrap distributions are skewed**
  - Empirical bootstrap assumes symmetry
  - Percentile more robust to skew, but **sensitive to bias**
- **Undercoverage is mainly due to bias**

Conclusion & Future Outlook

- Non-parametric bootstrap shows potential, but **more research** needed
- Alternative: **Bias-corrected accelerated bootstrap**
- Computational efficiency
  - **ABC** interval

Utrecht University

Sources

Austin, P. C. (2016). *Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis*. *Statistics in Medicine, 35*(30), 5642–5655. https://doi.org/10.1002/sim.7084

Fu, E. L. (2023). *Target trial emulation to improve causal inference from observational data: What, why, and how? Journal of the American Society of Nephrology, 34*(8), 1305. https://doi.org/10.1681/ASN.0000000000000152

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*.

Hernán, M. A. (2018). *How to estimate the effect of treatment duration on survival outcomes using observational data*. *BMJ, 360*, k182. https://doi.org/10.1136/bmj.k182

Su, L., Rezvani, R., Seaman, S. R., Starr, C., & Gravestock, I. (2024). *TrialEmulation: An R package to emulate target trials for causal analysis of observational time-to-event data*. *arXiv*. https://arxiv.org/abs/2401.12345

Coverage

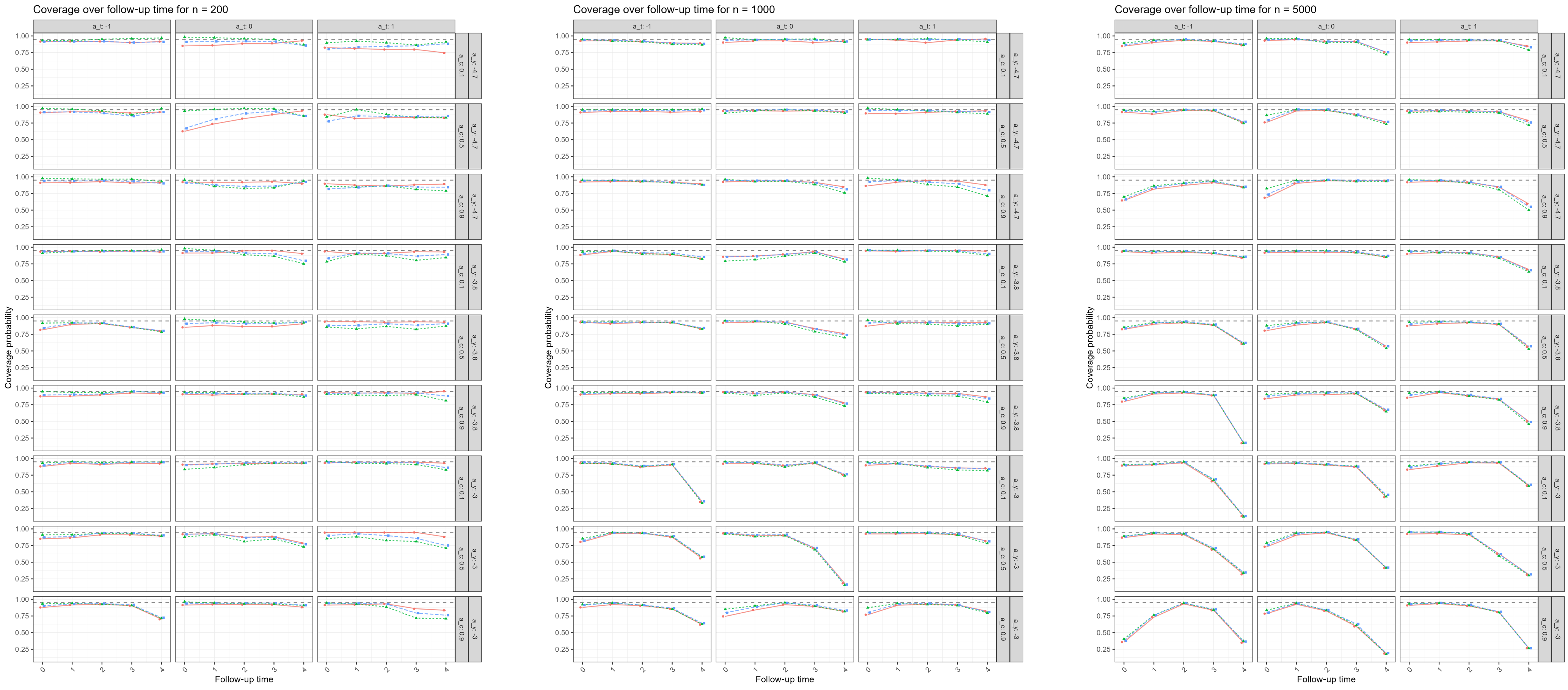

Figure: Coverage of 95% intervals results for sample sizes 200, 1000, and 5000. The green line denotes the empirical bootstrap CI, the blue line denotes percentile bootstrap CI, and the red line denotes the sandwich-type CI.

Utrecht University

Width

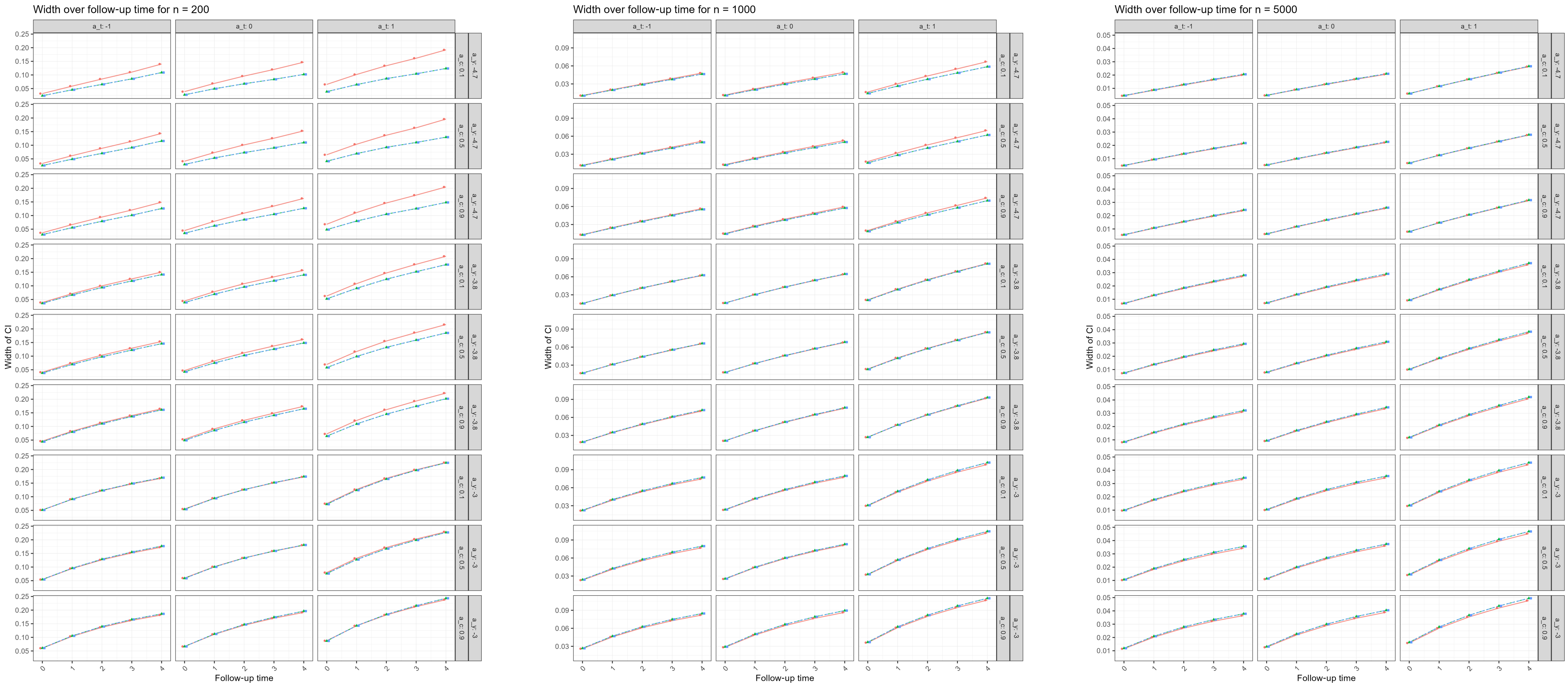

Figure: Width of 95% intervals results for sample sizes 200, 1000, and 5000. The red line denotes the bootstrap CIs, the blue line denotes the sandwich-type CIs.