

## LETTER TO THE EDITOR: PULLING UNMEASURED CONFOUNDING OUT BY YOUR BOOTSTRAPS: TOO GOOD TO BE TRUE?

CORENTIN SÉGALAS\*

*Université de Paris, Centre of Epidemiology and Statistics (CRESS)  
Inserm, F75004, Paris, France*

*Department of Medical Statistics, London School of Hygiene and Tropical medicine, UK  
Email: corentin.segalas@inserm.fr*

CLÉMENCE LEYRAT

*Department of Medical Statistics, London School of Hygiene and Tropical medicine, UK  
Email: clemence.leyrat@lshtm.ac.uk*

ELIZABETH WILLIAMSON

*Department of Medical Statistics, London School of Hygiene and Tropical medicine, UK  
Email: elizabeth.williamson@lshtm.ac.uk*

### SUMMARY

Inverse probability of treatment weighting can account for confounding under a number of assumptions, including that of no unmeasured confounding. A recent simulation study proposed a bootstrap bias correction, apparently demonstrating good performance in removing bias due to unmeasured confounding. We revisited the simulations, finding no evidence of bias reduction.

*Keywords and phrases:* Bias-correction; Bootstrap; Ignorability; Inverse Probability of Treatment Weighting; Propensity scores; Unconfoundedness

## 1 Kundhi and Voia (2018) Methodology

When investigating the causal effect of a binary treatment  $z$  on an outcome  $y$  in counterfactual theory, two potential outcomes are assumed:  $y_i^1$  if the patient  $i$  is treated and  $y_i^0$  if not. The average treatment effect (ATE) is defined by  $E(y_i^1) - E(y_i^0)$ . However, observing both potential outcomes for the same patient is not possible, only  $y_i$  is observed. Methods based on the propensity score introduced by Rosenbaum and Rubin (1983) and defined as  $\pi(x_i) = p(z_i = 1|x_i)$ , the probability of being assigned to the treatment group  $z_i = 1$  (versus  $z_i = 0$ ), given baseline covariates  $x_i$ , have been developed to make causal inference from observational data. One of these methods is the inverse probability of treatment weighting (IPTW) which uses the inverse of the propensity score as a weight to create a pseudo-population where there is no association between confounders and treatment. Under the positivity assumption  $0 < p(z_i = 1|x_i) < 1$ , the stable unit treatment

---

\* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

value assumption which means that the two potential outcomes  $y_i^0$  and  $y_i^1$  of one patient  $i$  cannot be influenced by the treatment of another patient and the strong ignorable treatment assignment assumption,  $y^0, y^1 \perp z|x$ , IPTW provides an unbiased estimator of the ATE (Rosenbaum and Rubin, 1983; Austin, 2011). In Kundhi and Voia (2018), the IPTW estimator for the ATE is

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{y_i z_i}{\hat{\pi}(x_i)} - \frac{y_i(1-z_i)}{1-\hat{\pi}(x_i)} \right\},$$

where  $i = 1, \dots, N$  denotes the patient and  $\hat{\pi}$  an estimation of the propensity score obtained from a consistent estimator. The bootstrap bias is

$$Bias^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* - \hat{\theta},$$

where  $\hat{\theta}_b^*$  is the estimate of the ATE from the  $b$ -th nonparametric bootstrap sample. Kundhi and Voia (2018) proposed the corrected estimate:

$$\tilde{\theta}_C = \hat{\theta} - Bias^* = 2\hat{\theta} - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*.$$

An iterated version for further reducing bias is also proposed. There is no theoretical justification for a reduction in unmeasured confounding bias resulting from these corrections.

## 2 Simulations

We followed the simulations of Kundhi and Voia (2018) as closely as possible. The total number of patients was  $N$ . The  $x_{i,j}$  were assumed to be normally distributed independent centered covariates with  $Var[x_{i,j}] = 0.4, 0.4, 1.5$  for  $j = 1, 2, 3$  respectively. The treatment model used in Kundhi and Voia (2018) was not explicitly provided. Therefore we have run a Scenario A where treatment assignment was deterministic based on sample sizes and a Scenario B that included confounding. For this latter case a treatment model was included where  $z_i$  is a Bernoulli random variable whose parameter  $\pi(x_i)$  is generated as  $logit(\pi(x_i)) = -0.4054651 + 0.8x_{i,1} + 0.6x_{i,2} + 0.9x_{i,3}$ . The intercept was chosen to give a ratio of treated:untreated of 0.4 in order to match the original sample sizes  $(N_t, N_c) = (30, 50); (60, 80); (100, 150); (200, 300); (500, 750); (1000, 1500)$  where  $N_t$  and  $N_c$  are respectively the number of treated patients and the number of controls. The relationship between outcome and treatment is given by  $y_i = \theta z_i + x_{i,1} + x_{i,2} + x_{i,3} + u_i$  where  $\theta$  is the treatment effect, taking values  $\theta = 0.5, 1, 2$ , and  $u_i$  is the residual standard normal variance. Three misspecification cases were considered by the authors of the discussed paper: (1) *omitted variable*:  $x_{i,3}$  was dropped from the propensity score model; (2) *endogeneity*:  $corr(x_{i,1}, u_i) = \rho$  with  $\rho = 0.25, 0.5, 0.9$ ; (3) *lack of overlap*:  $Var[x_{i,1}] = S_t = 0.8, 1.2$  for the treated and  $Var[x_{i,1}] = S_c = 0.5$  for the controls. 1,000 datasets have been generated. For each of these datasets, the bootstrap bias correction was computed using 999 bootstrap samples.

All misspecification cases were considered in Scenario A. However, misspecification 3 was not considered in Scenario B because it is incompatible with specifying a treatment model. Relative biases for Scenario A are presented in Figure 1 and for Scenario B in Figure 2. Unlike Kundhi and Voia (2018) who found substantial bias which decreased with the bootstrap correction, we do not find any bias except for Scenario B with misspecification 1, i.e. when there is unmeasured confounding, in which case the bootstrap correction does not reduce bias. If these relative biases look equals for each value of  $\theta$  in Figure 2, this is a misleading impression created by the figures all being drawn on a common y-axis.

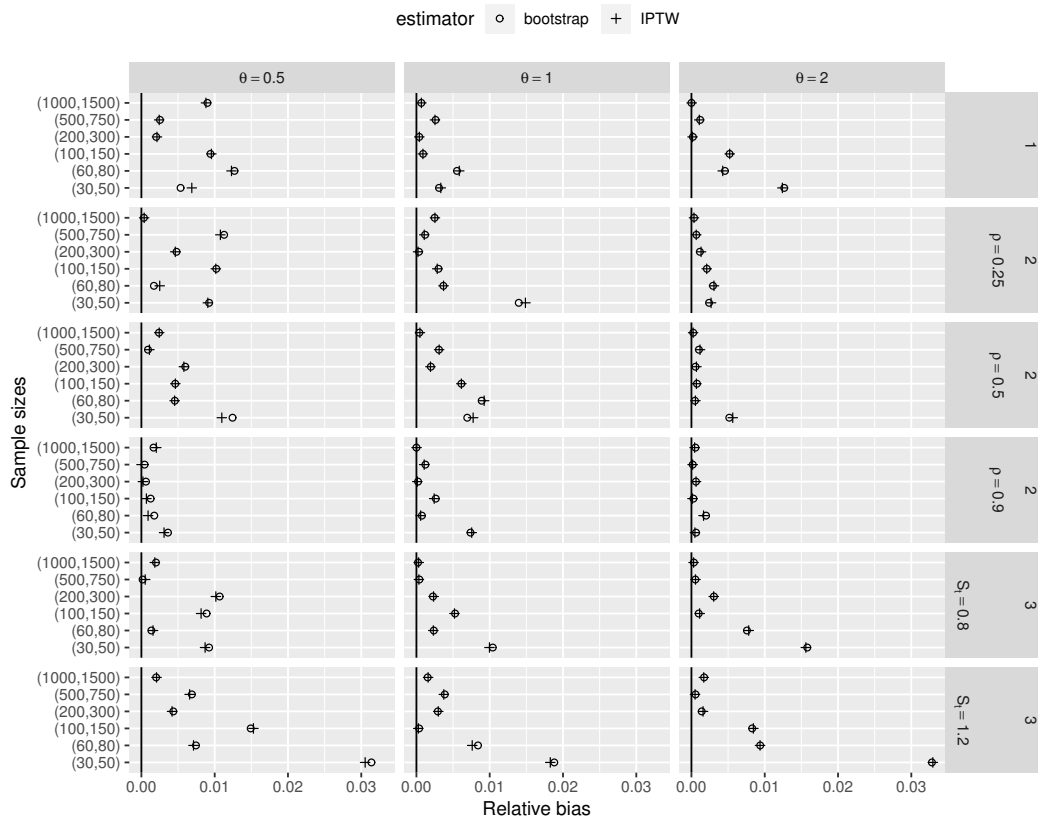


Figure 1: Relative biases according to the sample sizes  $(N_t, N_c)$  for Scenario A. Results are presented for both the bootstrap and the IPTW estimators and according to the true treatment effect  $\theta$ , the misspecification case (1, 2 or 3) and for misspecification 2 according to the values of  $\rho$  and for misspecification 3 according to the values of  $S_t$ .

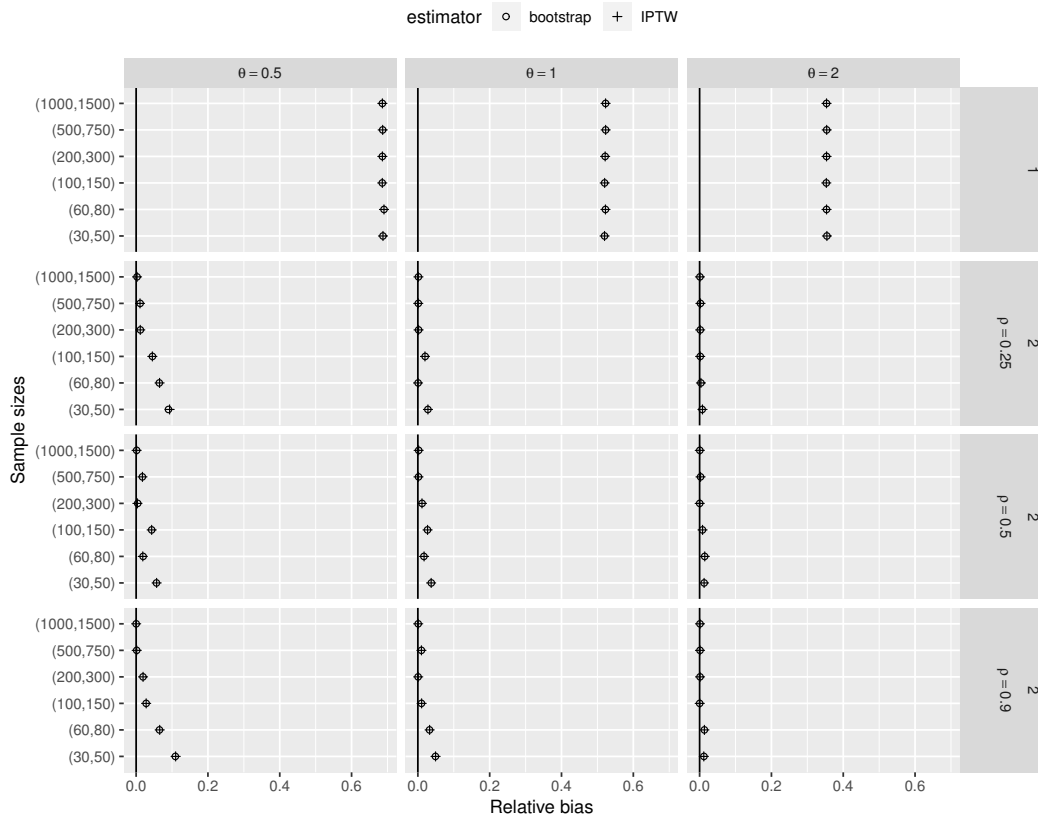


Figure 2: Relative biases according to the sample sizes  $(N_t, N_c)$  for Scenario B. Results are presented for both the bootstrap and the IPTW estimators and according to the true treatment effect  $\theta$ , the misspecification case (1 or 2) and for misspecification 2 according to the values of  $\rho$ .

### 3 Conclusion

Suggesting that being able to remove bias due to unmeasured confounding with no additional information using only the bootstrap appears, indeed, too good to be true.

### Acknowledgement

This work was supported by a grant from the UK Medical Research Council MRS01442X1. CL is supported by the UK Medical Research Council (Skills Development Fellowship MR/T032448/1)

## References

- Austin, P. C. (2011), “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies,” *Multivariate Behavioral Research*, 46, 399–424.
- Kundhi, G. and Voia, M. (2018), “Bootstrap bias correction for average treatment effects with inverse propensity weights,” *Journal of Statistical Research*, 52, 187–200.
- Rosenbaum, P. R. and Rubin, D. B. (1983), “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, 70, 41–55.

Received: August 16, 2021

Accepted: October 19, 2021