

Online Supplementary Material for
“Estimating the Marginal Hazard Ratio by Simultaneously Using
A Set of Propensity Score Models: A Multiply Robust Approach”
by

Di Shu^{1*}, Peisong Han², Rui Wang^{1,3}, and Sengwee Toh¹

¹Department of Population Medicine, Harvard Medical School and
Harvard Pilgrim Health Care Institute

²Department of Biostatistics, University of Michigan

³Department of Biostatistics, Harvard T.H. Chan School of Public Health

Simulation Results Using Conventional Weights

In this online supplementary material, we report simulation results using conventional weights.

Figures S1 and S2 report the empirical relative bias using conventional weights with sample sizes $n = 500$ and $n = 5000$, respectively. Figures S3 and S4 report the empirical coverage using conventional weights with sample sizes $n = 500$ and $n = 5000$, respectively. Figures S5 and S6 report the average widths of 95% confidence intervals using conventional weights with sample sizes $n = 500$ and $n = 5000$, respectively. Figure S7 reports simulation results when all postulated propensity score models are wrong using conventional weights with $n = 5000$.

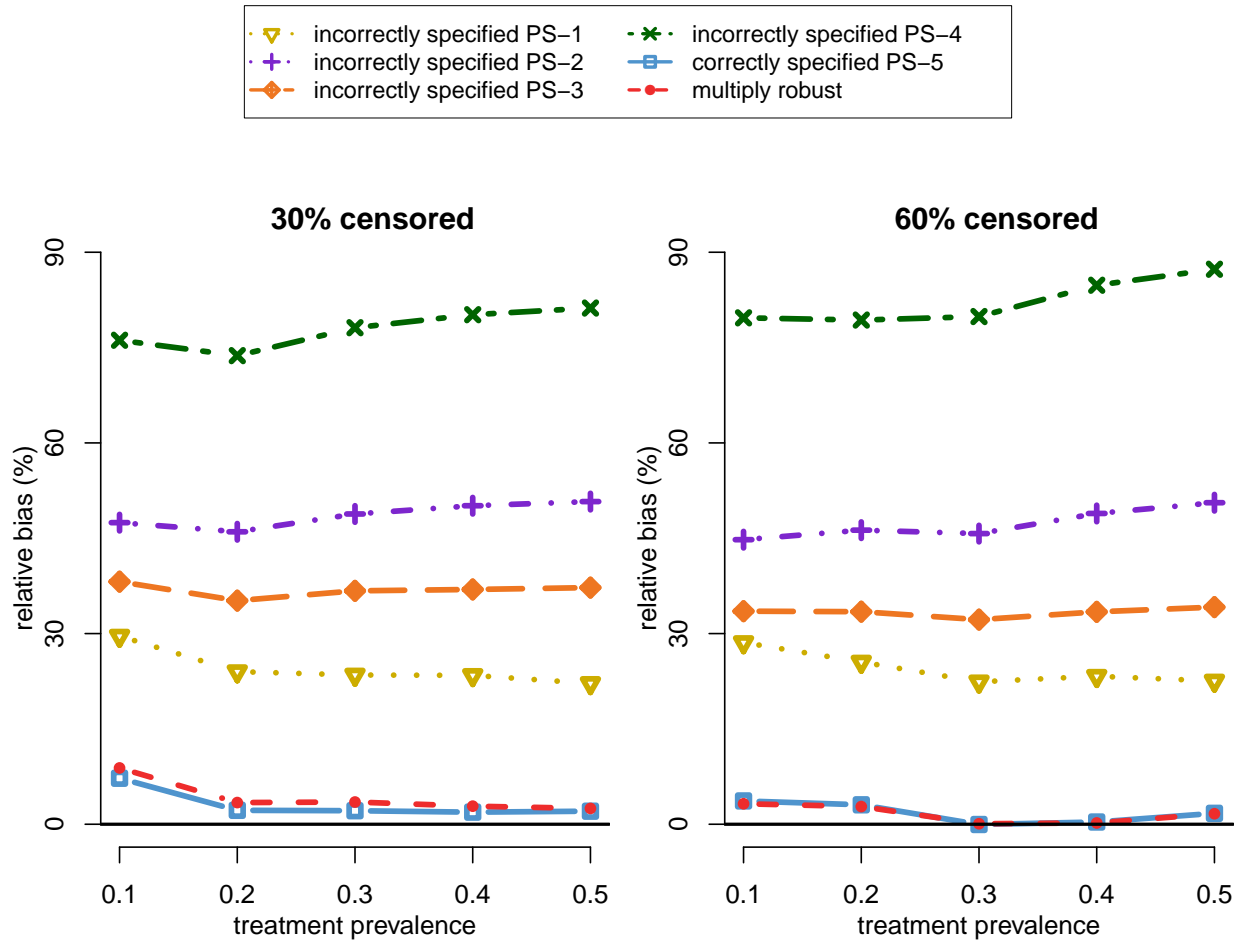


Figure S1: Empirical relative bias in percent using conventional weights with $n = 500$. incorrectly specified PS-1 to PS-4: IPW Cox estimators from four incorrectly specified propensity score models (11)-(14), respectively; correctly specified PS-5: IPW Cox estimator from a correctly specified propensity score model (15); multiply robust: the proposed multiply robust estimator using multiple models (11)-(15).

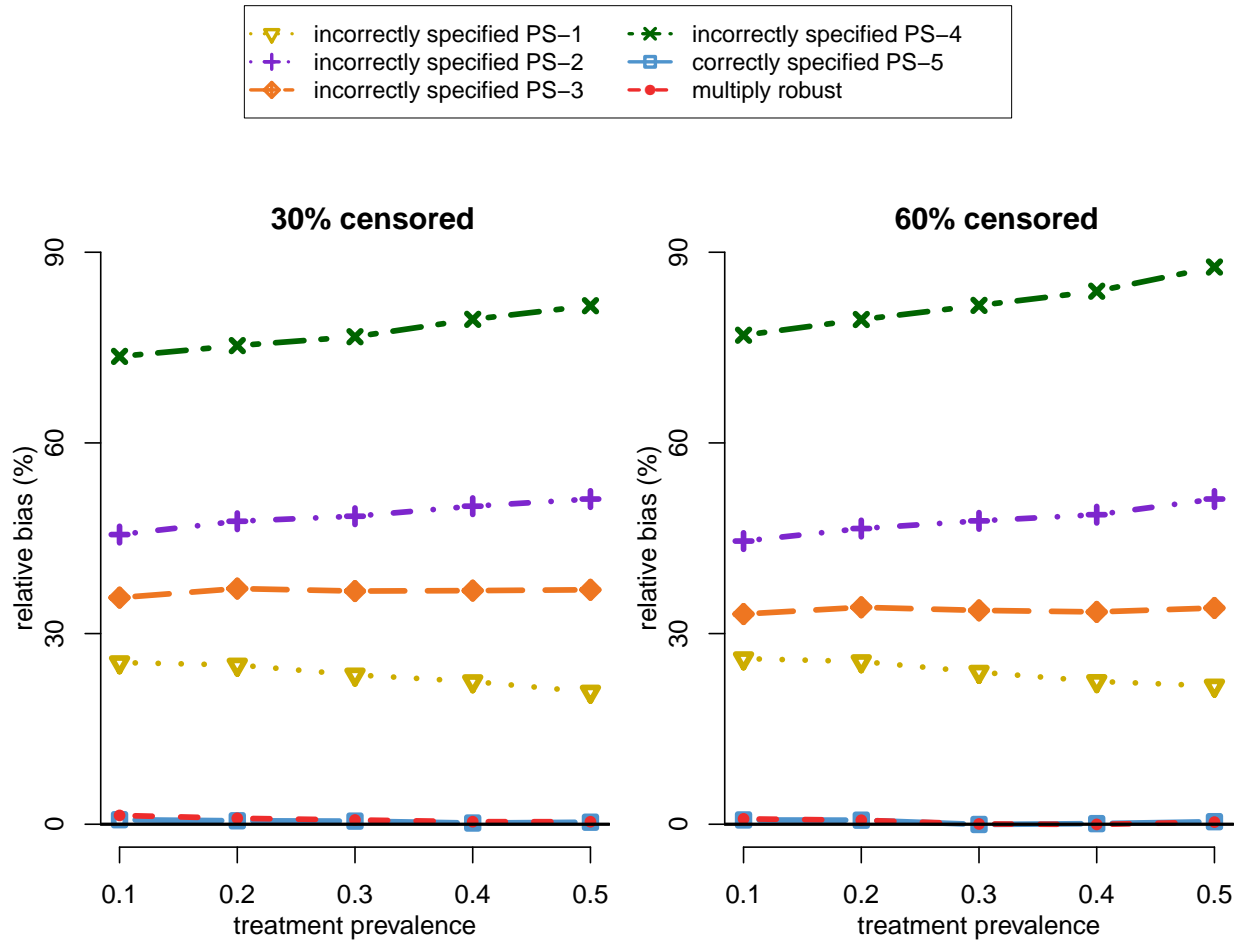


Figure S2: Empirical relative bias in percent using conventional weights with $n = 5000$. incorrectly specified PS-1 to PS-4: IPW Cox estimators from four incorrectly specified propensity score models (11)-(14), respectively; correctly specified PS-5: IPW Cox estimator from a correctly specified propensity score model (15); multiply robust: the proposed multiply robust estimator using multiple models (11)-(15).

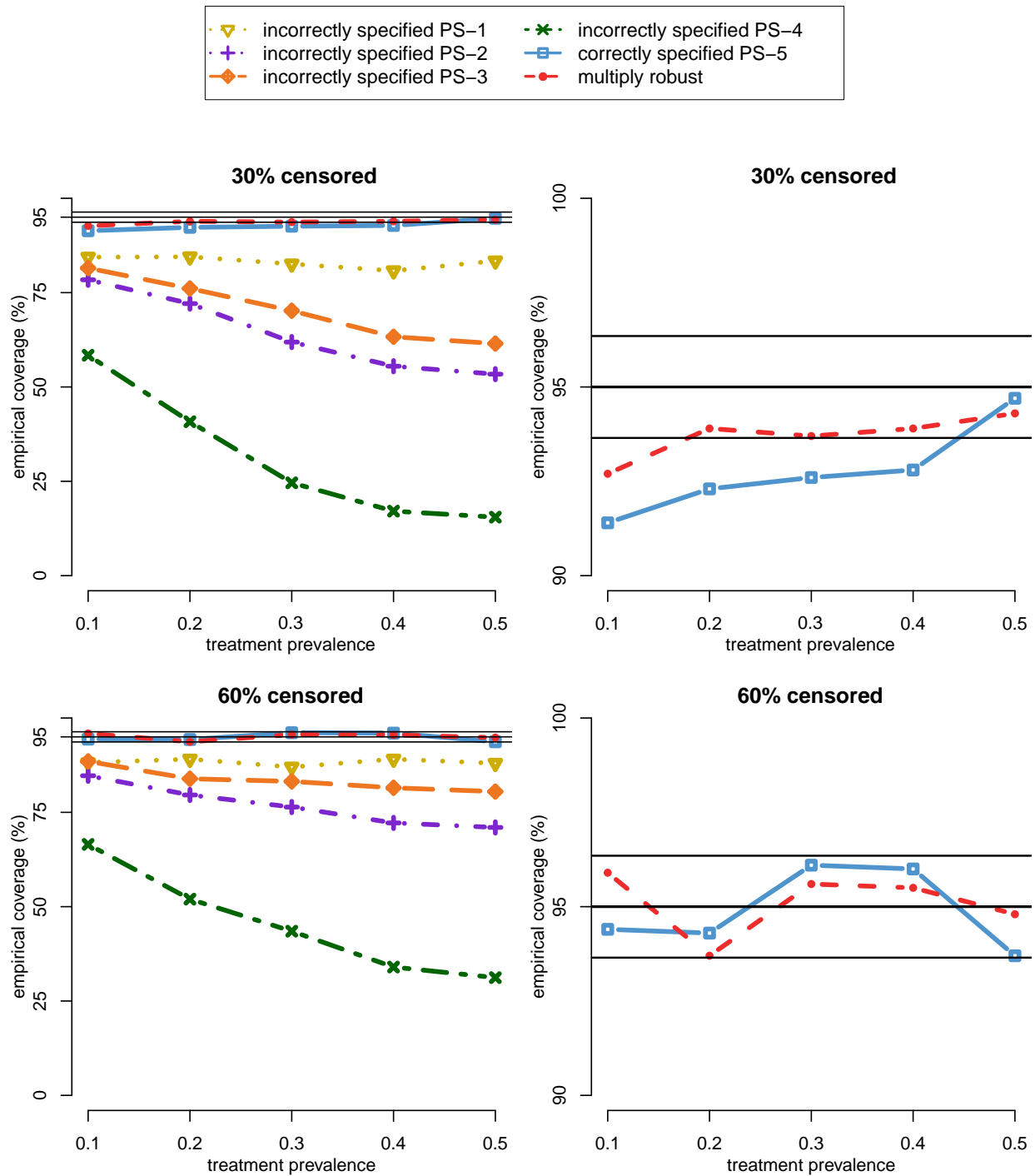


Figure S3: Empirical coverage in percent using conventional weights with $n = 500$. The right panel shows a zoom-in version of the left panel.

incorrectly specified PS-1 to PS-4: IPW Cox estimators from four incorrectly specified propensity score models (11)-(14), respectively; correctly specified PS-5: IPW Cox estimator from a correctly specified propensity score model (15); multiply robust: the proposed multiply robust estimator using multiple models (11)-(15).

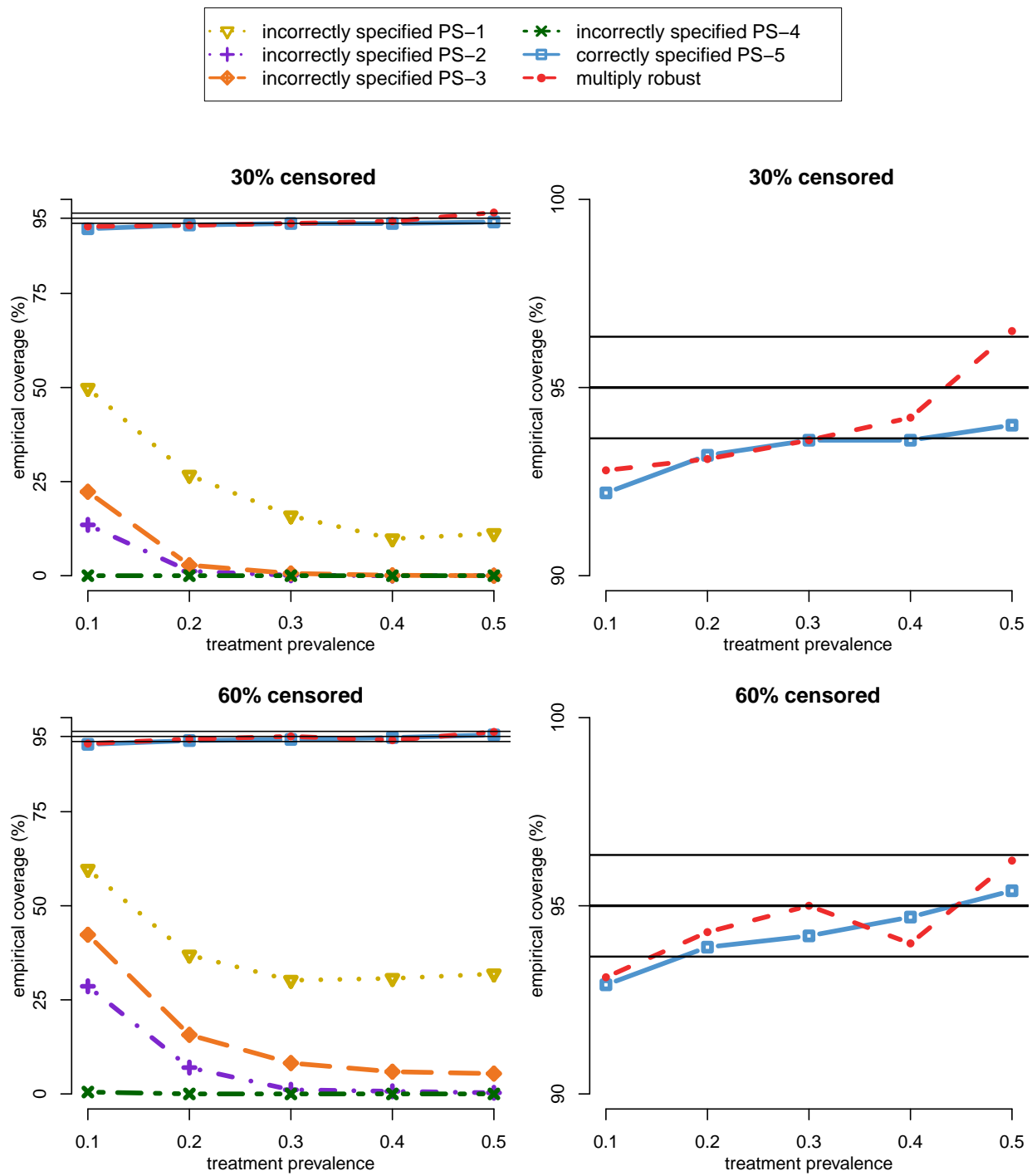


Figure S4: Empirical coverage in percent using conventional weights with $n = 5000$. The right panel shows a zoom-in version of the left panel.

incorrectly specified PS-1 to PS-4: IPW Cox estimators from four incorrectly specified propensity score models (11)-(14), respectively; correctly specified PS-5: IPW Cox estimator from a correctly specified propensity score model (15); multiply robust: the proposed multiply robust estimator using multiple models (11)-(15).

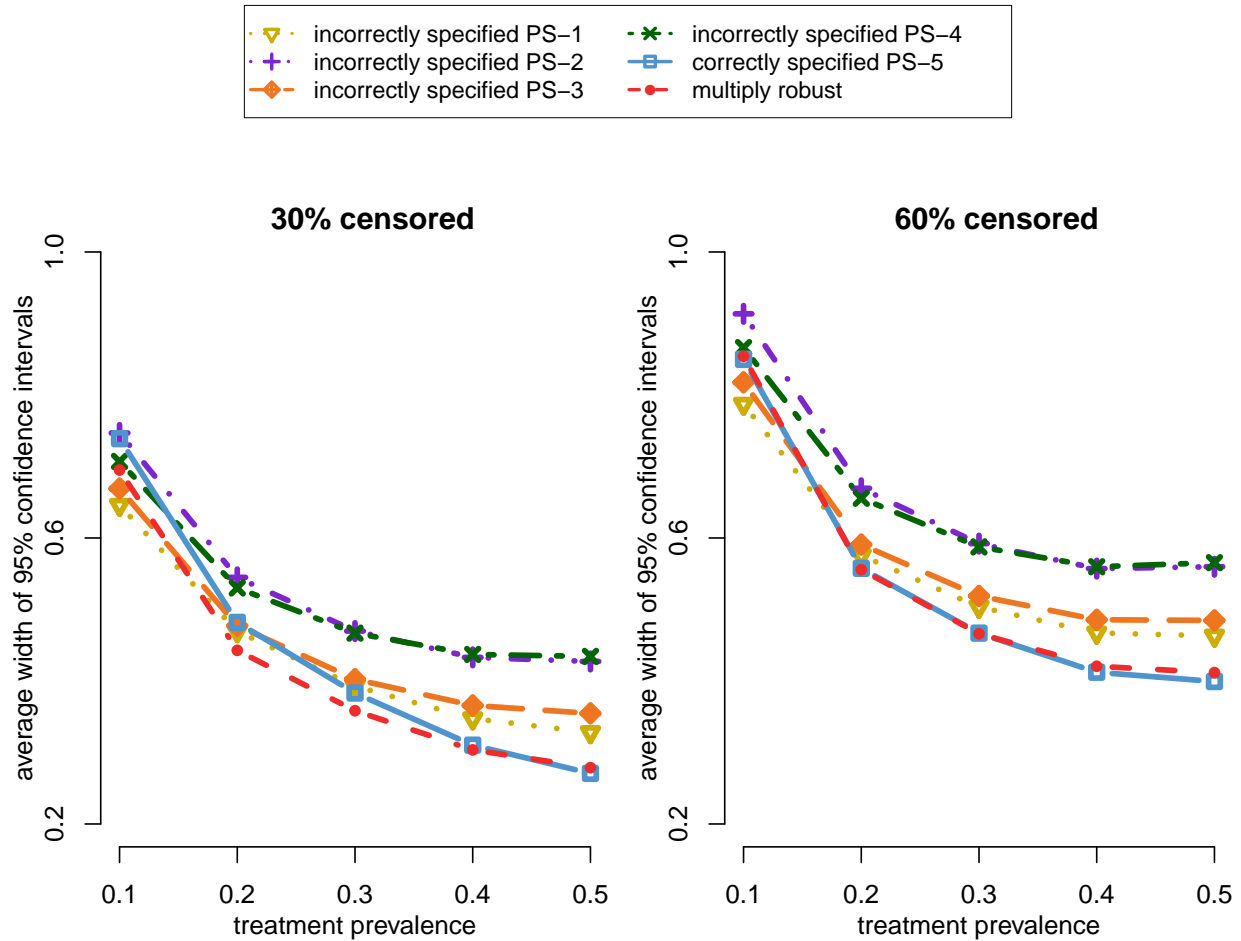


Figure S5: Average widths of 95% confidence intervals using conventional weights with $n = 500$.

incorrectly specified PS-1 to PS-4: IPW Cox estimators from four incorrectly specified propensity score models (11)-(14), respectively; correctly specified PS-5: IPW Cox estimator from a correctly specified propensity score model (15); multiply robust: the proposed multiply robust estimator using multiple models (11)-(15).

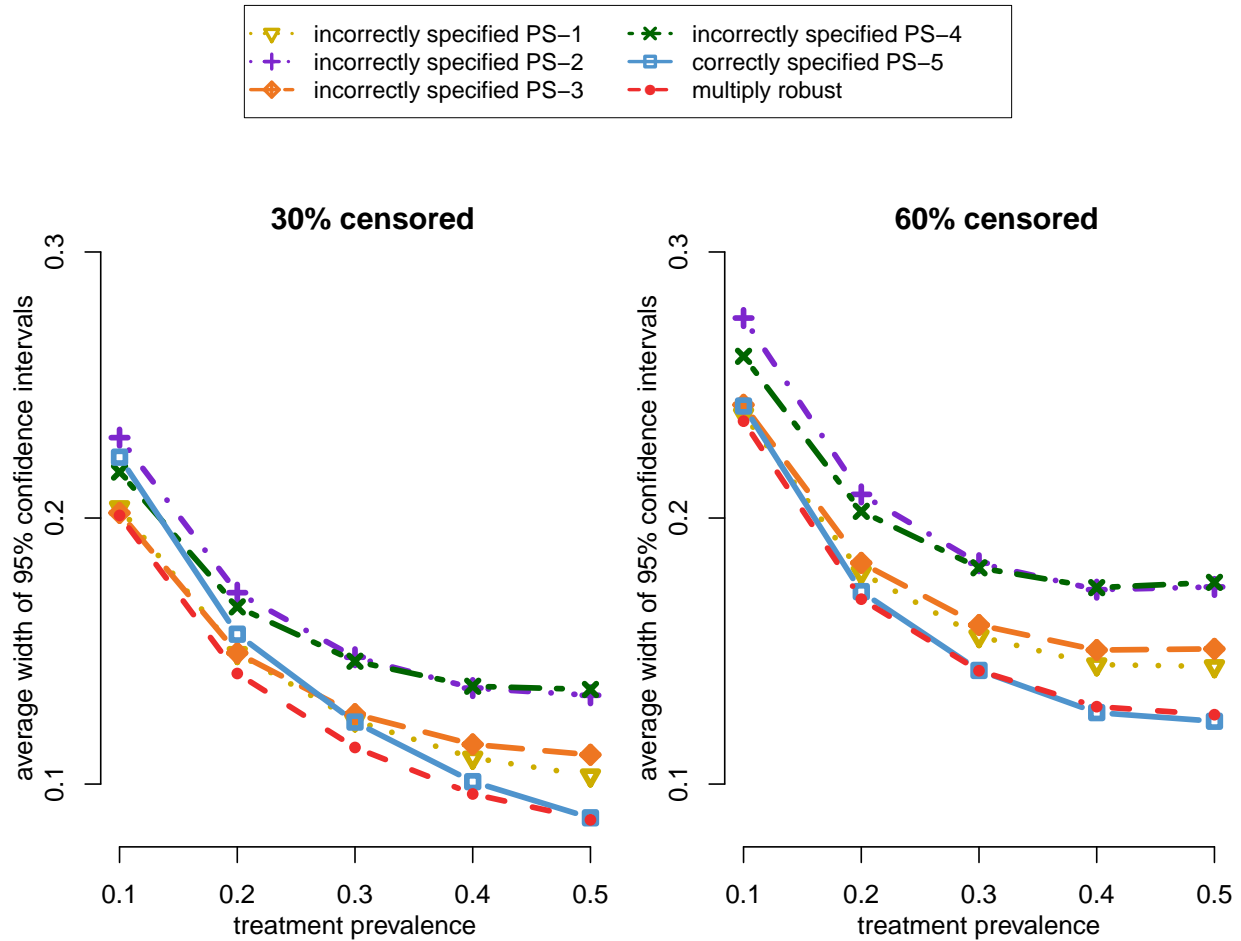


Figure S6: Average widths of 95% confidence intervals using conventional weights with $n = 5000$.

incorrectly specified PS-1 to PS-4: IPW Cox estimators from four incorrectly specified propensity score models (11)-(14), respectively; correctly specified PS-5: IPW Cox estimator from a correctly specified propensity score model (15); multiply robust: the proposed multiply robust estimator using multiple models (11)-(15).

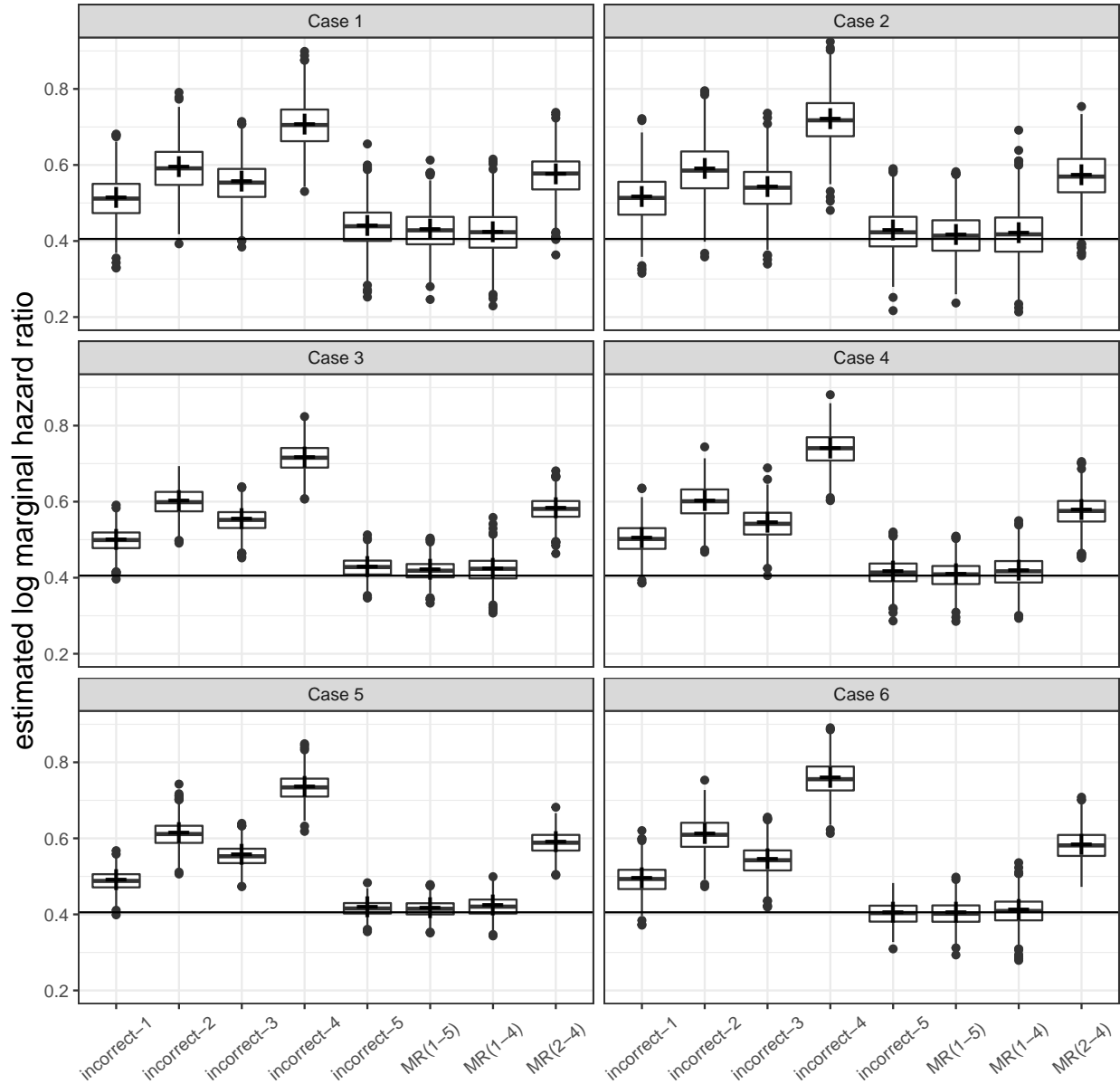


Figure S7: Simulation results when all postulated propensity score models are wrong using conventional weights with $n = 5000$.

incorrect 1-5: IPW Cox estimators from five incorrectly specified propensity score models (11)-(14) and (16), respectively; MR(a-b): the proposed multiply robust estimator using incorrect models a to b;

Cases 1, 3, and 5: 30% censoring with treatment prevalence 10%, 30%, and 50%; Cases 2, 4, and 6: 60% censoring with treatment prevalence 10%, 30%, and 50%;

“+”: average of estimates across 1000 simulation runs. The horizontal solid line indicates the true log marginal hazard ratio.