Web Supplement for: Improving Precision and Power in Randomized Trials for COVID-19 Treatments Using Covariate Adjustment, for Binary, Ordinal, and Time-to-Event Outcomes

David Benkeser[1]

Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, U.S.A.

and

Iván Díaz[1]

Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine

and

Alex Luedtke[1]

Department of Statistics, University of Washington

Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center

and

Jodi Segal

Department of Medicine, School of Medicine, Johns Hopkins University

and

Daniel Scharfstein and Michael Rosenblum[2]

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University

SUMMARY: The Web Supplement is organized as follows. Appendix A introduces the estimands and estimators for ordinal outcomes. Appendix B introduces the estimands and assumptions on censoring that we make for time-to-event outcomes. Appendix C presents additional simulation studies, including for non-hospitalized COVID-19 patients. Appendix C.1 presents the data-generating distributions for non-hospitalized COVID-19 patients. Appendix C.2 presents the results of simulation studies for the case that the outcome is binary. Appendix C.3 presents additional simulation results for ordinal outcomes, namely the results for Wald-style inference and for the non-hospitalized population. Appendix C.4 presents additional simulation results for time-to-event outcomes, namely when a restricted set of covariates (age and sex) were used for adjustment and for the difference of survival probabilities in the

,

hospitalized population. Appendix D describes the availability of code that reproduces our simulation experiments and that implements our estimator and confidence intervals.

[1] Contributed equally to this manuscript
[2] Correspondence to M. Rosenblum

A. Estimands and estimators when the outcome is ordinal

A.1 *Estimands*

Let (A, Y) and $(\widetilde{A}, \widetilde{Y})$ denote independent treatment-outcome pairs, and let $u(\cdot)$ be a pre-specified, real-valued transformation of an outcome. The three estimands are defined as follows:

$$\text{DIM: } E[u(Y)|A=1] - E[u(Y)|A=0],$$

$$\text{MW: } P\left(\widetilde{Y} > Y \middle| \widetilde{A}=1, A=0\right) + \frac{1}{2}P\left(\widetilde{Y} = Y \middle| \widetilde{A}=1, A=0\right),$$

$$\text{LOR: } \underset{\beta\in\mathbb{R}}{\arg\min} \textstyle\sum_{j=1}^{K-1} \left\{\text{logit}\, P(Y \leqslant j|A=1) - \text{logit}\, P(Y \leqslant j|A=0) - \beta\right\}^2.$$

All three estimands are smooth summaries of the cumulative distribution functions $F_a(\cdot) := P(Y \leqslant \cdot\,|A=a)$ for $a \in \{0,1\}$. To see that this is the case, let $f_a(j) := F_a(j) - F_a(j-1)$, $a \in \{0,1\}$, denote the corresponding probability mass functions and note that the estimands can be equivalently expressed as follows:

$$\text{DIM: } \sum_{j=1}^{K} u(j)\{f_1(j) - f_0(j)\},$$

$$\text{MW: } \sum_{j=1}^{K} \left\{F_0(j-1) + \frac{1}{2}f_0(j)\right\} f_1(j)$$

$$\text{LOR: } \underset{\beta\in\mathbb{R}}{\arg\min} \textstyle\sum_{j=1}^{K-1} \left\{\text{logit}\, F_1(j) - \text{logit}\, F_0(j) - \beta\right\}^2.$$

A.2 *Covariate adjusted estimator*

Consider a setting in which we observe n independent copies of (X, A, Y), where X represents a d-dimensional vector of baseline covariates, A represents treatment, and Y represents outcome. We assume that $A \perp\!\!\!\perp X$. We use the subscript i to denote data specific to individual i. We now derive an estimator for the CDF that is closely related to an estimator presented in Scharfstein et al. (1999) and to targeted minimum loss-based estimators (van der Laan and Rubin, 2006; van der Laan and Rose, 2011).

For $\alpha := (\alpha(j))_{j=1}^{K-1} \in \mathbb{R}^{K-1}$ and $\beta \in \mathbb{R}^d$, define the following $\mathbb{R}^{K-1} \times \mathbb{R}^d$ function:

$$\text{logit}\, m_{\alpha,\beta}(j,x) = \alpha(j) + \beta^\top x.$$

We will consider the treatment-stratified proportional odds working model for $P\{Y \leqslant j|A = a, X = x\}$ in which there exist (α_0, β_0) and (α_1, β_1) such that $P\{Y \leqslant j|A = a, X = x\} = m_{\alpha_a,\beta_a}(j,x)$ for all j, x, a. Importantly, we do not rely on this model being correct.

In addition to the above working model, we consider a treatment-assignment propensity score working model. It is used to define inverse-probability weights that are used when fitting the aforementioned proportional odds working models. Let $\hat{\pi}(a|x)$ be an estimate of $P(A = a|X = x)$, e.g., using a logistic regression model. In the clinical trial setting that we considered in our simulation studies, we used a logistic regression model with just an intercept, i.e., we ignored baseline variables. This is equivalent to using no weights (i.e., all weights equal to a constant) when fitting the proportional odds models. At the end of this subsection, we describe alternative approaches for estimating $P(A = a|X = x)$ and the implications of doing so.

Suppose that, for $a \in \{0, 1\}$, $\hat{\alpha}(a)$ and $\hat{\beta}(a)$ are chosen to minimize the following weighted empirical risk in (α, β):

$$-\sum_{j=1}^{K-1}\sum_{i=1}^{n} \frac{I\{A_i = a\}}{\hat{\pi}(A_i|X_i)} \log\left(m_{\alpha,\beta}(j,X_i)^{I\{Y_i \leqslant j\}}[1 - m_{\alpha,\beta}(j,X_i)]^{I\{Y_i > j\}}\right). \quad (1)$$

Each of these a-specific optimizations can be solved by running a weighted logistic regression on a repeated measures dataset of size $n \times (K-1)$. Alternatively, they can be fit using software for a proportional odds model that allows for weights. For both levels of the treatment a, it can be shown that $\hat{\alpha}(a)_1 \leqslant \hat{\alpha}(a)_2 \leqslant \ldots \leqslant \hat{\alpha}(a)_{K-1}$, and so, for any covariate value x, $m_{\hat{\alpha}(a),\hat{\beta}(a)}(\cdot, x)$ is a monotone nondecreasing function. Moreover, if our treatment-stratified proportional odds working model is correct, then $\hat{\alpha}(a)$ and $\hat{\beta}(a)$ are consistent and asymptotically normal estimators of the true underlying parameters.

Our covariate adjusted estimate of the CDF $\psi_a(j) := P(Y \leqslant j|A = a)$ is given by

$$\hat{\psi}_a(j) := \frac{1}{n}\sum_{i=1}^{n} m_{\hat{\alpha}(a),\hat{\beta}(a)}(j, X_i). \tag{2}$$

Because $m_{\hat{\alpha}(a),\hat{\beta}(a)}(\cdot, X_i)$ is monotone nondecreasing for all $i = 1, \ldots, n$, $\hat{\psi}_a(\cdot)$ is also monotone nondecreasing. The above estimator also satisfies the known constraint that $\hat{\psi}_a(j) \in [0, 1]$.

It can also be shown that $\hat{\psi}_a(j)$ is (i) doubly robust and (ii) efficient if both the treatment mechanism ($P(A = a|X)$) working model and the stratified proportional odds working model are correctly specified. To show this, we now establish that $\hat{\psi}_a(j)$ is in fact an augmented inverse probability weighted estimator. First, note that minimizing (1) to find $\hat{\alpha}_a$ and $\hat{\beta}_a$ implies that the following first-order condition is satisfied for all $j \in \{1, \ldots, K-1\}$:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{I\{A_i = a\}}{\hat{\pi}(A_i|X_i)} \left[I\{Y_i \leqslant j\} - m_{\hat{\alpha}_a,\hat{\beta}_a}(j, X_i)\right] = 0.$$

Next, note that adding this to the right-hand side of (2) shows that

$$\begin{aligned}\hat{\psi}_a(j) &= \frac{1}{n}\sum_{i=1}^{n} \frac{I\{A_i = a\}}{\hat{\pi}(A_i|X_i)} \left[I\{Y_i \leqslant j\} - m_{\hat{\alpha}_a,\hat{\beta}_a}(j, X_i)\right] + \frac{1}{n}\sum_{i=1}^{n} m_{\hat{\alpha}(a),\hat{\beta}(a)}(j, X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{I\{A_i = a\}}{\hat{\pi}(A_i|X_i)} I\{Y_i \leqslant j\} + \frac{1}{n}\sum_{i=1}^{n} m_{\hat{\alpha}(a),\hat{\beta}(a)}(j, X_i) \left[1 - \frac{I\{A_i = a\}}{\hat{\pi}(A_i|X_i)}\right].\end{aligned}$$

The above shows that $\hat{\psi}_a(j)$ is an augmented inverse probability weighted estimator (see Section 7 of Robins et al., 1994) for $\hat{\psi}_a(j)$, with the estimate of the outcome regression $x \mapsto P\{Y \leqslant j|A = a, X = x\}$ given by $x \mapsto m_{\hat{\alpha}(a),\hat{\beta}(a)}(j, x)$.

We next discuss our estimation of the treatment probability $P(A = a|X = x)$. Though this quantity can always be estimated by the empirical treatment probability in our randomized trial setting, there are generally advantages to estimating this quantity within a richer model. For example, a logistic regression of treatment on covariates (main effects only) could be used — in a randomized trial setting, this model is correctly specified provided that it includes an intercept term. The advantage of estimating known treatment probabilities via correctly specified parametric models has been discussed elsewhere – see, for example, Williamson et al. (2014) or, for a general treatment, Section 2.3.7 in van der Laan et al. (2003).

We recommend handling missing ordinal outcomes using doubly robust methods whose validity relies on the outcomes being missing at random conditional on the covariates and treatment assignment. To implement this approach, one can apply the methods described above, but with study arm recoded as 0 to indicate that a patient was both randomized to study arm 0 (control) and had their outcome measured, 1 to indicate that a patient was both randomized to study arm 1 (treatment) and had their outcome measured, and -1 to indicate that the outcome is missing. When study arm is recoded in this way and the outcome is not missing completely at random but is missing at random conditional on covariates, it is important that the model used for $\hat{\pi}$ described above conditions on the baseline covariates, since this recoded treatment is not fully randomized.

B. Estimands and censoring assumptions for time-to-event outcomes

Let T be a time-to-event outcome, C be a right-censoring time, A be a treatment indicator, and X be a collection of baseline covariates. Let τ be an investigator-specified truncation time that will be used to define the RMST, and let t^* be an investigator-specified time at which a comparison between the arm-specific survival probabilities is of interest.

The three estimands are defined as

$$\text{RMST: } E[\min\{T,\tau\}|A=1] - E[\min\{T,\tau\}|A=0],$$

$$\text{RD: } P\left(T \leqslant t^*|A=1\right) - P\left(T \leqslant t^*|A=0\right),$$

$$\text{RR: } \frac{P\left(T \leqslant t^*|A=1\right)}{P\left(T \leqslant t^*|A=0\right)}.$$

Unadjusted methods assume that

$$C \perp\!\!\!\perp T|A. \tag{3}$$

The adjusted methods discussed in the main text assume that

$$C \perp\!\!\!\perp T|A, X, \tag{4}$$

which may be more plausible than (3).

C. Additional simulation studies

C.1 *Data generating distributions for non-hospitalized, COVID-19 patients*

We also conducted simulations to mimic a population of non-hospitalized individuals who test positive for COVID-19 and where the primary outcome is ordinal (1=death, 2=hospitalized and survived, 3=not hospitalized and survived) and the baseline covariate is age category. We set the control arm probabilities of being in each age group and of hospitalization and death as in Table 1, which was extracted from CDC COVID-19 Response Team (2020) analogous to how this was done in Section 4.2.2 for the hospitalized population; the treatment arm distribution was constructed similarly as in Section 4.2.2.

Analogous to the hospitalized population data generating distributions, we assumed that a treatment would have no effect on the probability of death but would decrease the odds of hospital admission (hospitalization) by the same relative amount in each age category. For ordinal outcome scenarios with smaller sample sizes, there were sometimes data sets that had no participants in the lowest or highest outcome category in at least one study arm. For these data sets, the log-odds ratio estimators are undefined. As such, we omitted these sample sizes from our evaluations.

[Table 1 about here.]

C.2 *Additional simulation studies for binary outcomes*

We repeated the simulation studies in hospitalized and non-hospitalized patients for ordinal outcomes, but collapsing the death and ICU admission outcomes (hospitalized setting) and the death and hospitalized outcomes (non-hospitalized setting) to make a binary composite outcome. The binary outcome in the non-hospitalized population is defined as death or hospitalization ($Y = 0$) or survived and no hospitalization ($Y = 1$). The binary outcome for

the hospitalized population is as defined in Section 4.2.1 of the main paper. We compared covariate-adjusted vs. unadjusted estimates of the risk difference of the binary outcome in terms of mean squared error, bias, and variance. We also compared the probability of rejecting the null hypothesis of 0 risk difference using a test based on our covariate-adjusted estimator versus a traditional Chi-squared test. Results are shown in Tables 2-4. When considering the same population and estimand, the only difference between tables that use BCa nonparametric bootstrap-based inference versus tables that use Wald-style inference is in the P(Reject H_0) column.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

C.3 *Additional simulation studies for ordinal outcomes*

We first present simulation results when using Wald-style inference for the population of hospitalized patients, in Tables 5-7 for the three ordinal estimands. Results were largely similar to those that used BCa nonparametric bootstrap-based inference as presented in the main text.

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

We also noted considerable numerical instabilities in implementations of the proportional odds model included in the `MASS` package (function `polr`), which led to our using the more stable implementation in the `ordinal` package (function `clm`) throughout. The latter function is the default in the `drord` package.

Tables 8-13 present simulation results for ordinal outcomes for the non-hospitalized population described in Section C.1 of the Supplementary Materials.

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

C.4 *Additional simulation studies for time-to-event outcomes*

We present results for the difference in restricted mean survival times (RMST) at 14 days estimand in the hospitalized population, when the adjusted estimator uses only age and sex (Table 14). Results are also presented for the difference of survival probabilities (RD) at 7 days estimand in the hospitalized population (when the adjusted estimator uses all six baseline variables from Section 4.2.3) in Table 15.

[Table 14 about here.]

[Table 15 about here.]

D. Code availability

D.1 *Simulation code*

All code needed to reproduce the simulations for ordinal and binary data is available on GitHub (`https://github.com/mrosenblum/COVID-19-RCT-STAT-TOOLS`). The code for the survival simulations is also included in that repository. However, because the simulation is based on private data from Weill Cornell Medicine, the results of the simulation reported

in the manuscript are not reproducible based on the available code. We provide a simulated dataset (not based on real data) with the same structure of the real dataset. This dataset can be used to run the simulation code.

D.2 *R packages*

The `drord` package (available at `https://github.com/benkeser/drord`) implements the proposed estimators for ordinal outcomes and can also be used for analyzing binary outcomes. The package vignette (`https://benkeser.github.io/drord/articles/using_drord.html`) describes implementation of the estimators and all available options in the package. In particular, the package includes: bootstrap-based and closed-form inference for all estimands described here-in, as well as for the treatment-specific PMFs and CDFs; a fully nonparametric covariate-adjusted estimator that uses stratification to estimate the covariate-conditional PMF and estimators; and a plotting method for visualizing covariate-adjusted estimates of the treatment-specific PMFs and CDFs that includes pointwise confidence intervals and simultaneous confidence bands.

The `survtmlerct` package, available at `https://github.com/idiazst/survtmlerct`, implements the targeted minimum loss based estimator for the RMST of Díaz et al. (2019). The package also implements an analogous estimator for the risk difference RD, as well as unadjusted counterparts for both the RMST and the RD. Standard errors are computed using the influence function of the estimators, and Wald-type confidence intervals are implemented. The functions in the package can incorporate any user-provided, preliminary estimates of the outcome and hazard functions, including parametric and data-adaptive estimates that use model selection. The help command applied to the specific functions of the package gives examples of the estimators.

References

CDC COVID-19 Response Team (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm . *MMWR Morb Mortal Wkly Rep 69*(12), 343–346.

Díaz, I., E. Colantuoni, D. F. Hanley, and M. Rosenblum (2019). Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards. *Lifetime Data Analysis 25*(3), 439–468.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association 89*(427), 846–866.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association 94*(448), 1096–1120.

van der Laan, M. J., M. Laan, and J. M. Robins (2003). *Unified methods for censored longitudinal data and causality*. Springer Science & Business Media.

van der Laan, M. J. and S. Rose (2011). *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media.

van der Laan, M. J. and D. Rubin (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics 2*(1).

Williamson, E. J., A. Forbes, and I. R. White (2014). Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Statistics in Medicine 33*(5), 721–737.

Table 1

***Non-hospitalized**, COVID-19 positive population: Age and conditional outcome distributions based on data from (CDC COVID-19 Response Team, 2020) that we use for defining the control arm distribution in the ordinal outcome simulation studies for the **non-hospitalized** population. "Hosp." abbreviates "hospitalized"; "surv." abbreviates "survived".*

Age	P(age)	P(death | age)	P(hosp. & surv. | age)	P(not hosp. & surv. | age)
0–19	0.05	0.00	0.02	0.98
20–44	0.29	0.00	0.18	0.82
45–54	0.18	0.01	0.25	0.74
55–64	0.18	0.02	0.25	0.73
65–74	0.17	0.04	0.36	0.60
75–84	0.09	0.07	0.45	0.48
$\geqslant$ 85	0.06	0.19	0.51	0.30

Table 2

Results for the binary outcome and risk difference (RD) estimand in the hospitalized population. *Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.043	0.010	0.003	0.010	1.000
100	Adjusted	0	0.056	0.009	0.004	0.009	0.844
100	Unadjusted	-0.269	0.719	0.009	0.003	0.009	1.000
100	Adjusted	-0.269	0.847	0.008	0.004	0.008	0.859
200	Unadjusted	0	0.031	0.005	0.003	0.005	1.000
200	Adjusted	0	0.041	0.004	0.004	0.004	0.885
200	Unadjusted	-0.199	0.768	0.005	0.003	0.005	1.000
200	Adjusted	-0.199	0.846	0.004	0.004	0.004	0.880
500	Unadjusted	0	0.047	0.002	0.001	0.002	1.000
500	Adjusted	0	0.051	0.002	0.000	0.002	0.878
500	Unadjusted	-0.124	0.770	0.002	0.000	0.002	1.000
500	Adjusted	-0.124	0.837	0.002	0.000	0.002	0.899
1000	Unadjusted	0	0.041	0.001	0.000	0.001	1.000
1000	Adjusted	0	0.042	0.001	0.000	0.001	0.860
1000	Unadjusted	-0.090	0.796	0.001	0.000	0.001	1.000
1000	Adjusted	-0.090	0.861	0.001	0.000	0.001	0.890

Table 3

Results for the binary outcome and risk difference (RD) estimand in the non-hospitalized population. *BCa bootstrap is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.029	0.008	0.003	0.008	1.000
100	Adjusted	0	0.060	0.008	0.001	0.008	0.918
100	Unadjusted	-0.209	0.652	0.005	0.007	0.005	1.000
100	Adjusted	-0.209	0.811	0.005	0.005	0.005	0.941
200	Unadjusted	0	0.043	0.004	0.000	0.004	1.000
200	Adjusted	0	0.059	0.004	0.001	0.004	0.885
200	Unadjusted	-0.161	0.747	0.003	-0.002	0.003	1.000
200	Adjusted	-0.161	0.840	0.003	-0.001	0.003	0.883
500	Unadjusted	0	0.042	0.002	0.000	0.002	1.000
500	Adjusted	0	0.057	0.002	0.000	0.002	0.887
500	Unadjusted	-0.112	0.811	0.001	0.000	0.001	1.000
500	Adjusted	-0.112	0.887	0.001	0.000	0.001	0.879
1000	Unadjusted	0	0.047	0.001	0.000	0.001	1.000
1000	Adjusted	0	0.056	0.001	0.000	0.001	0.920
1000	Unadjusted	-0.073	0.712	0.001	0.001	0.001	1.000
1000	Adjusted	-0.073	0.791	0.001	0.001	0.001	0.930

Table 4

Results for the binary outcome and risk difference (RD) estimand in the non-hospitalized population. *Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.029	0.008	0.003	0.008	1.000
100	Adjusted	0	0.063	0.008	0.001	0.008	0.918
100	Unadjusted	-0.209	0.652	0.005	0.007	0.005	1.000
100	Adjusted	-0.209	0.805	0.005	0.005	0.005	0.941
200	Unadjusted	0	0.043	0.004	0.000	0.004	1.000
200	Adjusted	0	0.058	0.004	0.001	0.004	0.885
200	Unadjusted	-0.161	0.747	0.003	-0.002	0.003	1.000
200	Adjusted	-0.161	0.842	0.003	-0.001	0.003	0.883
500	Unadjusted	0	0.042	0.002	0.000	0.002	1.000
500	Adjusted	0	0.058	0.002	0.000	0.002	0.887
500	Unadjusted	-0.112	0.811	0.001	0.000	0.001	1.000
500	Adjusted	-0.112	0.888	0.001	0.000	0.001	0.879
1000	Unadjusted	0	0.047	0.001	0.000	0.001	1.000
1000	Adjusted	0	0.055	0.001	0.000	0.001	0.920
1000	Unadjusted	-0.073	0.712	0.001	0.001	0.001	1.000
1000	Adjusted	-0.073	0.793	0.001	0.001	0.001	0.930

Table 5

Results for the ordinal outcome and difference in means (DIM) estimand in the hospitalized population. *Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.067	0.023	-0.005	0.023	1.000
100	Adjusted	0	0.065	0.019	-0.007	0.019	0.822
100	Unadjusted	0.303	0.503	0.022	-0.007	0.022	1.000
100	Adjusted	0.303	0.592	0.019	-0.004	0.019	0.845
200	Unadjusted	0	0.042	0.010	-0.002	0.010	1.000
200	Adjusted	0	0.047	0.009	-0.003	0.009	0.862
200	Unadjusted	0.303	0.792	0.012	-0.003	0.012	1.000
200	Adjusted	0.303	0.858	0.010	0.000	0.010	0.872
500	Unadjusted	0	0.060	0.005	-0.001	0.005	1.000
500	Adjusted	0	0.057	0.004	0.000	0.004	0.837
500	Unadjusted	0.195	0.816	0.005	0.000	0.005	1.000
500	Adjusted	0.195	0.869	0.004	0.001	0.004	0.891
1000	Unadjusted	0	0.045	0.002	0.000	0.002	1.000
1000	Adjusted	0	0.044	0.002	0.000	0.002	0.849
1000	Unadjusted	0.136	0.826	0.002	0.000	0.002	1.000
1000	Adjusted	0.136	0.885	0.002	0.000	0.002	0.889

Table 6

Results for ordinal outcome and Mann Whitney (MW) estimand in the hospitalized population.
Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0.500	0.071	0.003	-0.002	0.003	1.000
100	Adjusted	0.500	0.062	0.002	-0.003	0.002	0.822
100	Unadjusted	0.627	0.607	0.002	-0.002	0.002	1.000
100	Adjusted	0.627	0.696	0.002	-0.002	0.002	0.852
200	Unadjusted	0.500	0.048	0.001	-0.001	0.001	1.000
200	Adjusted	0.500	0.047	0.001	-0.001	0.001	0.864
200	Unadjusted	0.627	0.917	0.001	-0.001	0.001	1.000
200	Adjusted	0.627	0.959	0.001	0.000	0.001	0.878
500	Unadjusted	0.500	0.060	0.001	0.000	0.001	1.000
500	Adjusted	0.500	0.054	0.000	0.000	0.000	0.843
500	Unadjusted	0.582	0.926	0.001	0.000	0.001	1.000
500	Adjusted	0.582	0.950	0.000	0.000	0.000	0.905
1000	Unadjusted	0.500	0.044	0.000	0.000	0.000	1.000
1000	Adjusted	0.500	0.047	0.000	0.000	0.000	0.844
1000	Unadjusted	0.557	0.915	0.000	0.000	0.000	1.000
1000	Adjusted	0.557	0.940	0.000	0.000	0.000	0.890

Table 7

Results for the ordinal outcome and log-odds ratio (LOR) estimand in the hospitalized population. *Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.053	0.185	0.018	0.185	1.000
100	Adjusted	0	0.057	0.153	0.021	0.152	0.824
100	Unadjusted	-0.686	0.306	0.231	0.006	0.231	1.000
100	Adjusted	-0.686	0.372	0.196	0.001	0.196	0.848
200	Unadjusted	0	0.042	0.080	0.004	0.081	1.000
200	Adjusted	0	0.044	0.069	0.007	0.069	0.854
200	Unadjusted	-0.686	0.562	0.111	0.000	0.111	1.000
200	Adjusted	-0.686	0.633	0.096	-0.003	0.096	0.863
500	Unadjusted	0	0.060	0.035	0.002	0.035	1.000
500	Adjusted	0	0.065	0.029	0.000	0.029	0.826
500	Unadjusted	-0.408	0.574	0.038	-0.001	0.038	1.000
500	Adjusted	-0.408	0.640	0.033	-0.002	0.033	0.869
1000	Unadjusted	0	0.041	0.015	0.000	0.015	1.000
1000	Adjusted	0	0.047	0.013	0.000	0.013	0.851
1000	Unadjusted	-0.278	0.577	0.016	0.000	0.016	1.000
1000	Adjusted	-0.278	0.641	0.014	0.002	0.014	0.878

Table 8

Results for the ordinal outcome and difference in means (DIM) estimand in the non-hospitalized population. *BCa bootstrap is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.057	0.011	-0.002	0.011	1.000
100	Adjusted	0	0.061	0.010	-0.001	0.010	0.947
100	Unadjusted	0.193	0.504	0.009	-0.006	0.009	1.000
100	Adjusted	0.193	0.535	0.008	-0.004	0.008	0.950
200	Unadjusted	0	0.062	0.006	0.000	0.006	1.000
200	Adjusted	0	0.063	0.005	-0.001	0.005	0.892
200	Unadjusted	0.193	0.816	0.005	0.003	0.005	1.000
200	Adjusted	0.193	0.844	0.004	0.003	0.004	0.915
500	Unadjusted	0	0.055	0.002	0.000	0.002	1.000
500	Adjusted	0	0.048	0.002	0.000	0.002	0.896
500	Unadjusted	0.125	0.791	0.002	0.000	0.002	1.000
500	Adjusted	0.125	0.838	0.002	0.000	0.002	0.894
1000	Unadjusted	0	0.054	0.001	0.000	0.001	1.000
1000	Adjusted	0	0.060	0.001	0.000	0.001	0.924
1000	Unadjusted	0.092	0.806	0.001	-0.001	0.001	1.000
1000	Adjusted	0.092	0.832	0.001	0.000	0.001	0.948

Table 9

Results for the ordinal outcome and difference in means (DIM) estimand in the non-hospitalized population. *Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.053	0.011	-0.002	0.011	1.000
100	Adjusted	0	0.058	0.010	-0.001	0.010	0.947
100	Unadjusted	0.193	0.495	0.009	-0.006	0.009	1.000
100	Adjusted	0.193	0.562	0.008	-0.004	0.008	0.950
200	Unadjusted	0	0.062	0.006	0.000	0.006	1.000
200	Adjusted	0	0.065	0.005	-0.001	0.005	0.892
200	Unadjusted	0.193	0.831	0.005	0.003	0.005	1.000
200	Adjusted	0.193	0.863	0.004	0.003	0.004	0.915
500	Unadjusted	0	0.050	0.002	0.000	0.002	1.000
500	Adjusted	0	0.054	0.002	0.000	0.002	0.896
500	Unadjusted	0.125	0.790	0.002	0.000	0.002	1.000
500	Adjusted	0.125	0.846	0.002	0.000	0.002	0.894
1000	Unadjusted	0	0.052	0.001	0.000	0.001	1.000
1000	Adjusted	0	0.061	0.001	0.000	0.001	0.924
1000	Unadjusted	0.092	0.808	0.001	-0.001	0.001	1.000
1000	Adjusted	0.092	0.845	0.001	0.000	0.001	0.948

Table 10

Results for ordinal outcome and Mann Whitney (MW) estimand in the non-hospitalized population.
BCa bootstrap is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0.500	0.051	0.002	-0.001	0.002	1.000
100	Adjusted	0.500	0.050	0.002	-0.001	0.002	0.939
100	Unadjusted	0.594	0.642	0.001	-0.003	0.001	1.000
100	Adjusted	0.594	0.682	0.001	-0.002	0.001	0.945
200	Unadjusted	0.500	0.062	0.001	0.000	0.001	1.000
200	Adjusted	0.500	0.058	0.001	-0.001	0.001	0.893
200	Unadjusted	0.594	0.930	0.001	0.001	0.001	1.000
200	Adjusted	0.594	0.943	0.001	0.001	0.001	0.915
500	Unadjusted	0.500	0.052	0.000	0.000	0.000	1.000
500	Adjusted	0.500	0.051	0.000	0.000	0.000	0.898
500	Unadjusted	0.561	0.883	0.000	0.000	0.000	1.000
500	Adjusted	0.561	0.922	0.000	0.000	0.000	0.893
1000	Unadjusted	0.500	0.055	0.000	0.000	0.000	1.000
1000	Adjusted	0.500	0.062	0.000	0.000	0.000	0.929
1000	Unadjusted	0.544	0.898	0.000	0.000	0.000	1.000
1000	Adjusted	0.544	0.919	0.000	0.000	0.000	0.954

Table 11

Results for ordinal outcome and Mann Whitney (MW) estimand in the non-hospitalized population. *Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0.500	0.057	0.002	-0.001	0.002	1.000
100	Adjusted	0.500	0.069	0.002	-0.001	0.002	0.939
100	Unadjusted	0.594	0.538	0.001	-0.003	0.001	1.000
100	Adjusted	0.594	0.612	0.001	-0.002	0.001	0.945
200	Unadjusted	0.500	0.068	0.001	0.000	0.001	1.000
200	Adjusted	0.500	0.069	0.001	-0.001	0.001	0.893
200	Unadjusted	0.594	0.901	0.001	0.001	0.001	1.000
200	Adjusted	0.594	0.924	0.001	0.001	0.001	0.915
500	Unadjusted	0.500	0.050	0.000	0.000	0.000	1.000
500	Adjusted	0.500	0.057	0.000	0.000	0.000	0.898
500	Unadjusted	0.561	0.861	0.000	0.000	0.000	1.000
500	Adjusted	0.561	0.910	0.000	0.000	0.000	0.893
1000	Unadjusted	0.500	0.054	0.000	0.000	0.000	1.000
1000	Adjusted	0.500	0.060	0.000	0.000	0.000	0.929
1000	Unadjusted	0.544	0.885	0.000	0.000	0.000	1.000
1000	Adjusted	0.544	0.911	0.000	0.000	0.000	0.954

Table 12

Results for the ordinal outcome and log-odds ratio (LOR) estimand in the non-hospitalized population.
BCa bootstrap is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
500	Unadjusted	0	0.029	0.108	0.006	0.108	1.000
500	Adjusted	0	0.035	0.101	0.007	0.101	0.938
500	Unadjusted	-0.354	0.139	0.115	0.003	0.115	1.000
500	Adjusted	-0.354	0.141	0.108	0.006	0.108	0.934
1000	Unadjusted	0	0.036	0.052	0.001	0.052	1.000
1000	Adjusted	0	0.040	0.049	0.001	0.049	0.939
1000	Unadjusted	-0.246	0.172	0.055	0.002	0.055	1.000
1000	Adjusted	-0.246	0.174	0.052	0.003	0.052	0.945

Table 13

Results for the ordinal outcome and log-odds ratio (LOR) estimand in the non-hospitalized population.

Wald-style inference is used for confidence intervals and hypothesis testing. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
500	Unadjusted	0	0.031	0.108	0.006	0.108	1.000
500	Adjusted	0	0.040	0.101	0.007	0.101	0.938
500	Unadjusted	-0.354	0.184	0.115	0.003	0.115	1.000
500	Adjusted	-0.354	0.208	0.108	0.006	0.108	0.934
1000	Unadjusted	0	0.045	0.052	0.001	0.052	1.000
1000	Adjusted	0	0.063	0.049	0.001	0.049	0.939
1000	Unadjusted	-0.246	0.180	0.055	0.002	0.055	1.000
1000	Adjusted	-0.246	0.210	0.052	0.003	0.052	0.945

Table 14

Results for difference in restricted mean survival times (RMST) at 14 days estimand in hospitalized population, when the adjusted estimator uses only age and sex. *Confidence intervals and hypothesis tests are Wald-style. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

n	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.011	0.781	0.018	0.780	1.000
100	Adjusted	0	0.015	0.771	0.014	0.771	0.987
100	Unadjusted	1.06	0.085	0.570	-0.265	0.500	1.000
100	Adjusted	1.06	0.098	0.570	-0.263	0.501	1.000
200	Unadjusted	0	0.048	0.481	-0.013	0.481	1.000
200	Adjusted	0	0.050	0.476	-0.014	0.476	0.989
200	Unadjusted	1.06	0.326	0.328	-0.145	0.307	1.000
200	Adjusted	1.06	0.337	0.326	-0.141	0.306	0.995
500	Unadjusted	0	0.050	0.201	-0.003	0.201	1.000
500	Adjusted	0	0.049	0.196	-0.003	0.196	0.975
500	Unadjusted	1.06	0.729	0.151	-0.070	0.146	1.000
500	Adjusted	1.06	0.742	0.147	-0.069	0.143	0.978
1000	Unadjusted	0	0.048	0.100	0.001	0.100	1.000
1000	Adjusted	0	0.047	0.096	0.001	0.096	0.963
1000	Unadjusted	1.06	0.959	0.079	-0.060	0.076	1.000
1000	Adjusted	1.06	0.963	0.077	-0.060	0.073	0.972

Table 15

Results for difference of survival probabilities (RD) at 7 days estimand in hospitalized population when the adjusted estimator uses all six baseline variables from Section 4.2.3. *Confidence intervals and hypothesis tests are Wald-style. "Effect" denotes the true estimand value; "MSE" denotes mean squared error; "Rel. Eff." denotes relative efficiency which we approximate as the ratio of the MSE of the estimator under consideration to the MSE of the unadjusted estimator. In each block of four rows, the first two rows involve no treatment effect and the last two rows involve a benefit from treatment.*

Sample Size	Estimator Type	Effect	P(reject H_0)	MSE	Bias	Variance	Rel. Eff.
100	Unadjusted	0	0.052	0.007	0.001	0.008	1.000
100	Adjusted	0	0.065	0.007	0.001	0.007	0.935
100	Unadjusted	0.087	0.185	0.007	-0.002	0.007	1.000
100	Adjusted	0.087	0.209	0.006	-0.001	0.006	0.973
200	Unadjusted	0	0.050	0.004	-0.001	0.004	1.000
200	Adjusted	0	0.058	0.003	-0.001	0.003	0.869
200	Unadjusted	0.087	0.316	0.003	-0.003	0.003	1.000
200	Adjusted	0.087	0.357	0.003	-0.003	0.003	0.904
500	Unadjusted	0	0.053	0.002	0.001	0.002	1.000
500	Adjusted	0	0.052	0.001	0.001	0.001	0.838
500	Unadjusted	0.087	0.648	0.001	-0.002	0.001	1.000
500	Adjusted	0.087	0.717	0.001	-0.002	0.001	0.858
1000	Unadjusted	0	0.052	0.001	0.001	0.001	1.000
1000	Adjusted	0	0.051	0.001	0.001	0.001	0.833
1000	Unadjusted	0.087	0.918	0.001	-0.002	0.001	1.000
1000	Adjusted	0.087	0.947	0.001	-0.002	0.001	0.851