American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2022.
This work is written by (a) US Government employee(s) and is in the public domain in the US.
Vol. 191, No. 9
https://doi.org/10.1093/aje/kwac087
Advance Access publication:
May 4, 2022

Practice of Epidemiology

Data-Adaptive Selection of the Propensity Score Truncation Level for Inverse-Probability–Weighted and Targeted Maximum Likelihood Estimators of Marginal Point Treatment Effects

Susan Gruber*, Rachael V. Phillips, Hana Lee, and Mark J. van der Laan

* Correspondence to Dr. Susan Gruber, Putnam Data Sciences, LLC, 85 Putnam Avenue, Suite 2, Cambridge, MA 02139 (e-mail: sgruber@putnamds.com).

Initially submitted October 17, 2021; accepted for publication April 29, 2022.

Inverse probability weighting (IPW) and targeted maximum likelihood estimation (TMLE) are methodologies that can adjust for confounding and selection bias and are often used for causal inference. Both estimators rely on the positivity assumption that within strata of confounders there is a positive probability of receiving treatment at all levels under consideration. Practical applications of IPW require finite inverse probability (IP) weights. TMLE requires that propensity scores (PS) be bounded away from 0 and 1. Although truncation can improve variance and finite sample bias, this artificial distortion of the IP weights and PS distribution introduces asymptotic bias. As sample size grows, truncation-induced bias eventually swamps variance, rendering nominal confidence interval coverage and hypothesis tests invalid. We present a simple truncation strategy based on the sample size, n, that sets the upper bound on IP weights at $\sqrt{n}\ln n/5$. For TMLE, the lower bound on the PS should be set to $5/(\sqrt{n}\ln n/5)$. Our strategy was designed to optimize the mean squared error of the parameter estimate. It naturally extends to data structures with missing outcomes. Simulation studies and a data analysis demonstrate our strategy's ability to minimize both bias and mean squared error in comparison with other common strategies, including the popular but flawed quantile-based heuristic.

bounds; inverse probability of treatment weighting; propensity score truncation

Abbreviations: ATE, additive treatment effect; CI, confidence interval; CTMLE, collaborative targeted maximum likelihood estimation; IP, inverse probability; IPTCW, inverse-probability-of-treatment-and-censoring weighting; IPTW, inverse-probability-of-treatment weighting; MSE, mean squared error; OR, odds ratio; PS, propensity score; SE, standard error; TML, targeted maximum likelihood; TMLE, targeted maximum likelihood estimation.

The inverse-probability-of-treatment–weighted estimator provides unbiased parameter estimates when the propensity scores (PS) used to construct the inverse probability (IP) weights are correctly specified (1). Causal identifiability rests on a positivity assumption that $0 < \text{PS} < 1$. In other words, within every observed stratum of confounders, there must be variation in the assigned treatment. In finite samples, inverse-probability-of-treatment–weighting (IPTW) parameter estimates can be unstable when PS values are close to 0 or 1, as this produces very large IP weights (2, 3). How close the PS can get to 0 or 1 before estimation becomes problematic depends primarily on sample size, and on the underlying distribution of the data. Placing an upper bound on the IP weights has been proposed as a way to improve the bias-variance trade-off of the IPTW estimator (4). This is equivalent to requiring that the PS be bounded away from 0 and 1 for targeted maximum likelihood (TML) estimators (5). Truncation can improve variance and finite sample bias. However, because it artificially distorts the IP weights and the PS distribution, truncation introduces large-sample (asymptotic) bias in the estimated treatment effect. As sample size grows, truncation-induced bias eventually swamps variance, rendering nominal confidence interval (CI) coverage and hypothesis tests invalid.

Popular heuristics for weight truncation rely on a pre-specified quantile of the weights (e.g., 95th or 99th) or specification of a fixed upper bound (e.g., 20, 40) (4, 5). Data-adaptive truncation strategies that aim to optimize the

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

mean squared error (MSE) of the parameter estimate have been proposed for some IPTW estimators and TML estimators (6–8), yet the simpler heuristics are used in practice (9). It is easily demonstrated through simulation that none of these heuristics is best across all data distributions. In this paper, we propose a truncation strategy that tailors the bound to the sample size. It is designed to minimize MSE while preserving valid inference. We compare its performance with other common strategies in simulation studies, and in the analysis of data from a retrospective observational cohort study of the association between total dose of ritodrine hydrochloride and pulmonary edema in twin pregnancy in Japan (10).

To gain insight into how the choice of bound can affect bias and variance, we first consider the stabilized IPTW estimator of the marginal additive treatment effect (ATE) (1),

$$\psi^{\text{IPTW}} = \frac{1}{\sum_{i=1}^{n} A_i\, g\,(1|W_i)} \sum_{i=1}^{n}\left[\frac{\overline{A}A_i}{g\,(1|W_i)}Y_i\right] - \frac{1}{\sum_{i=1}^{n}(1-A_i)\,(1-g\,(1|W_i))} \times \sum_{i=1}^{n}\left[\frac{(1-\overline{A})(1-A_i)}{1-g\,(1|W_i)}Y_i\right], \quad (1)$$

where n is the number of independent and identically distributed observations, Y is the outcome, A is a binary treatment assignment indicator, $\overline{A}$ is the proportion treated, W is a vector of baseline covariates, and $g(1|W) = P(A = 1|W)$ is the PS. Because $g(1|W)$ and $g(0|W) = 1 - g(1|W)$ are in the denominators, the PS must be bounded away from 0 for treated subjects and away from 1 for untreated subjects. A treated subject with a PS near 0 makes a large contribution to the first term on the right-hand side of equation 1. A similarly influential contribution is made to the second term by an untreated subject with a PS near 1. Thus, we see that instability in the IPTW is due to the interplay between large weights and inherent variability or heterogeneity in the outcome. For this reason, the same choice of truncation level that works well in one setting might work poorly in another.We can extend this reasoning directly to data structures in which some outcomes are missing and inverse-probability-of-treatment-and-censoring weights are required. These weights are the inverse of the product of the PS and the conditional probability that the outcome is observed, G_Δ, with appropriate stabilizing constants in the numerator. The inverse-probability-of-treatment-and-censoring weighting (IPTCW) estimator of the ATE is given by

$$\psi^{\text{IPTCW}} = \frac{1}{\sum_{i=1}^{n} A_i\, g\,(1|W_i)} \sum_{i=1}^{n}\left[\frac{\bar{\Delta}\Delta_i}{G_\Delta\,(A_i,W_i)} \times \frac{\overline{A}A_i}{g\,(1|W_i)}Y_i\right] - \frac{1}{\sum_{i=1}^{n}(1-A_i)\,(1-g\,(1|W_i))} \times \sum_{i=1}^{n}\left[\frac{\bar{\Delta}\Delta_i}{G_\Delta(A_i,W_i)} \times \frac{(1-\overline{A})(1-A_i)}{1-g\,(1|W_i)}Y_i\right], \quad (2)$$

where Δ is a binary indicator ($\Delta = 0$ indicates the outcome is missing), $\bar{\Delta}$ is the proportion of nonmissing outcomes, and $G_\Delta(A,W) = P(\Delta = 1 \mid A,W)$. The product of terms in the denominator must be bounded away from 0. When no outcomes are missing, $\bar{\Delta}\Delta_i/G_\Delta(A_i,W_i) = 1$ for all subjects, and equation 2 reduces to equation 1.

TML estimators of point treatment effects model the outcome regression, the PS, $g(A|W)$, and the probability that the outcome is observed, $G_\Delta(A,W)$. The latter two are known as the G components of the likelihood (11). The product of the G components must be bounded away from 0 (5). When there are no missing outcomes, $G_\Delta(A,W) = 1$ for all observations. Software implementations of TML estimators enforce user-specified fixed bounds (12–14). For a TML estimator of the ATE, Ju et al. (8) proposed a data-adaptive approach based on a collaborative TML estimator that used cross-validation to select the best bound from a set of candidate values.

Bias and variance both contribute to MSE. For this reason, an optimal truncation strategy would balance both, attempting to minimize MSE while also constraining bias to preserve proper inference. When estimates are normally distributed and CIs are constructed using an accurate estimate of the standard error (SE), coverage is a function of the ratio of the bias to the SE. The coverage of a nominal 95% CI can be calculated as $\Phi(1.96 + r) - \Phi(-1.96 + r)$, where Φ is the cumulative distribution function of the standard normal distribution and r is the ratio of bias to SE. For example, when bias is one-fifth of the SE ($r = 1/5$), CI coverage is 94.5%, but when bias equals the SE ($r = 1$), coverage falls to 83%. The ratio, r, grows with sample size, and CI coverage approaches 0.

A biased estimator will have larger MSE than an unbiased estimator in large samples. However, in finite samples, an estimator that accepts a little bias in return for a large reduction in variance may have smaller MSE. Generally speaking, for TML estimators and IPW estimators, increasing a truncation bound, b, increases bias while simultaneously decreasing variance. Our proposed sample-size–based truncation strategy allows b to grow towards 0 as sample size, n, increases. In finite samples, this strategy strives to minimize the MSE while preserving valid inference. We outline the underlying rationale below. A more theoretical justification is presented in Web Appendix 1 (available at https://doi.org/10.1093/aje/kwac087).

Simulation studies and an applied data analysis demonstrate performance relative to strategies that rely on bounding at fixed quantiles or fixed values. We also compare performance with the collaborative TMLE (CTMLE)-based strategy, generalized to additional parameters (the ATE, relative risk, and odds ratio (OR)), and to data with missingness in the outcome (Web Appendix 2). In our experience, data-adaptive strategies can be more unstable in response to fluctuations in the data distribution than a deterministic procedure when there is limited information in the data. Sensitivity to tuning parameter specifications also contributes to instability. In contrast, the sample-size–based truncation strategy is robust when estimating treatment effects with sparse and nonsparse data.

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

A SAMPLE-SIZE–BASED TRUNCATION STRATEGY

To devise a strategy that minimizes MSE = (bias^2 + variance), we need to understand the impact of the choice of bound, b, on bias and variance and how that changes as a function of sample size, n. As sample size grows towards infinity, variance converges to some limit (the variance of the estimator's influence curve, divided by n). However, if an estimator is biased, the bias remains, no matter how large n grows. Asymptotically, a biased estimator will have larger MSE than an unbiased estimator. However, in finite samples, an estimator that accepts a little bias in return for a large reduction in variance will have smaller MSE. Thus, the goal is to understand how fast b should approach 0 as sample size grows, in order to optimize the trade-off.

For simplicity, consider the effect of truncation on the bias and variance of an (unstabilized) IPTW estimator of the marginal mean outcome under treatment, μ_1, when no outcomes are missing, $O = (W, A, Y)$. The IPTW estimator is given by $\hat{\mu}_1 = 1/n \sum_{i=1}^{n} Y_1 \times A_i/g(1, W_i)$. Truncating the PS at b means setting values smaller than b to b itself. Thus, bounding the PS at b has no effect on the weights for observations where $g(1, W) \geq b$. When this is true for all observations, truncation changes none of the weights, and no bias is introduced. Eventually, as b increases, there will be some observations for which $g(1, W) < b$. Truncating these values introduces asymptotic bias while reducing variance and its associated finite-sample bias. This can be expressed in terms of the cumulative distribution function of the PS and the choice of b. For example, consider the number of observations affected by truncation at level b, and the difference between the truncated and nontruncated values when $n = 100$ versus $n = 1,000,000$. An observation with an IP weight of 100 is quite influential in the smaller data set and only mildly influential in the larger data set. Furthermore, truncating a PS of 0.009 at $b = 0.01$ changes the IP weight from 111 to 100, while truncating a PS of 0.001 at $b = 0.01$ changes the IP weight from 1,000 to 100, a much larger distortion.

In order to control the MSE of the estimate at a $1/n$ rate, we need to balance the reduction in bias that occurs as b is allowed to shrink towards 0, with the corresponding increase in variance. We analyzed the bias and variance of the truncated IPTW estimator as a function of b, under some reasonable statistical assumptions (e.g., the parameter of interest is practically identifiable from data; thus, $\sqrt{n}$-estimation is feasible (see Web Appendix 1)).

Our analysis shows that MSE can be optimized (in rate) by setting $b = 1/(\sqrt{n} \log_x n)$. The $\log_x n$ term in the denominator shrinks the bias to be less than the required $1/\sqrt{n}$, making the bias:SE ratio negligible in finite samples. Thus, in any scenario where $\sqrt{n}$ estimation is possible, our proposed truncation level achieves the optimal rate for MSE, while making bias negligible relative to variance, so that Wald-type inference is preserved.

Through a change of base, b can also be written in terms of the natural log as $c/(\sqrt{n} \ln n)$. The value of the constant, c, in the numerator is unimportant asymptotically. However, at small sample sizes it can have a pronounced effect. In simulation studies, we found good practical performance by setting c equal to 5 (Web Appendix 3, Web Table 1).

Although the analysis focuses on estimating the marginal mean outcome under treatment, this same reasoning applies to estimating the marginal mean outcome under exposure to the comparator or control treatment and to evaluating parameters that are functions of these 2 marginal means (e.g., the ATE, relative risk, or OR). Results for stabilized weights are the same, because multiplication by constants has no effect on rates of convergence. The analysis also extends to data structures with missing outcomes, where the bound is with respect to the product of the G components. When there is missingness in the outcome, n in our formula should be set to the number of observed outcomes rather than the total number of observations. The sample-size–based strategy will also apply to more general IPW estimators (e.g., those used for longitudinal data analyses), but the tuning of the constant will require separate simulations tailored towards the specific IPW estimator considered.

SIMULATION STUDIES

We conducted 2 Monte Carlo simulation studies to demonstrate the performance of TML estimators and IPW estimators that rely on different strategies for selecting a truncation level. Correctly specified models for the PS and the missingness mechanism, the G components of the likelihood, were used for all analyses. Recall that targeted maximum likelihood estimation (TMLE) is a 2-stage procedure (11). In stage 1, the outcome regression is modeled, $\overline{Q}_n^0 = E(Y|A, W)$. If this initial model, $\overline{Q}_n^0$, is misspecified, then if we rely on it to estimate the parameter, there will be residual bias. Stage 2 of TMLE aims to remove this residual bias, by using information in G to update the initial model. The updated model is denoted by $\overline{Q}_n^*$. We expect the impact of the choice of truncation strategy to vary, depending on the magnitude of the residual bias. Therefore, TMLE results are presented under correct (cor) and misspecified (mis) stage 1 outcome regression models (Q_{cor} and Q_{mis}, respectively). Q_{mis} is the unadjusted regression of the outcome, Y, on treatment, A. The IPW estimators used stabilized IP weights.

For each simulation study, we drew 500 data sets from the underlying data-generating process at 3 different sample sizes, $n = 100$, $n = 1,000$, and $n = 10,000$. In each of the 1,500 data sets, approximately 15% of outcomes were set to missing. The target parameter in simulation study 1, where the outcome was continuous, was the ATE. The target parameter in simulation study 2, where the outcome was binary, was the OR. For all estimators, the PS and missingness probabilities were estimated using correctly specified logistic regression models. All analyses were carried out in the R programming environment, version 4.0-2 (R Foundation for Statistical Computing, Vienna, Austria) (15).

Nine truncation strategies were evaluated. In addition to our proposed sample-size–based lower bound of $5/(\sqrt{n} \ln n)$, 2 were quantile-based, setting the lower bound on the PS at either the 95% quantile or the 99% quantile.

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

Five strategies specified a fixed bound: 0.1, 0.05, 0.025, 0.01, or 10^{-6} (equivalent to capping IP weights at 10, 20, 40, 100, or 10^6, respectively). We also applied the CTMLE-based strategy (Web Appendix 2), with the following tuning parameter settings: squared error loss when the outcome was continuous and negative log likelihood loss when the outcome was binary; penalty equal to the estimated variance of the ATE parameter; 20 cross-validation folds; and 10 candidate truncation levels (0.1, 0.09, . . . , 0.01, and 0.005). In addition, to avoid dividing by 0, the minimum lower bound (lb_{min}) on the distance between $g(A|W)$ and 0 was allowed to grow towards 0 with sample size $\text{lb}_{\text{min}} = 6/\ln n^3$.

Simulation study 1: continuous outcome

Simulation. The data consisted of observations $O = (W, A, \Delta, \Delta Y)$. W is a vector of 3 independent covariates, W_1, W_2, and W_3. W_1 and W_2 are continuous confounders generated from $N(1,1)$ and $N(0,1)$, respectively. W_3 is binary, generated from Bernoulli(0.4). A is the binary indicator of treatment, with 53% treated overall—that is, $P(A = 1 \mid W) = \text{expit}(-1.1 + 0.8W_1 + 0.9W_2 + 1.2W_3)$. Outcome Y is continuous, $Y = 10 + 3\text{A} + 2W_1 - 0.5W_2 + W_3 + 0.5\text{A}W_1 + \varepsilon$, with $\varepsilon \sim N(0,1)$. Δ is an indicator of whether the outcome was observed ($\Delta = 1$) or missing ($\Delta = 0$), with 15% missing on average: $P(\Delta = 1 \mid A, W) = \text{expit}(0.9 + 0.6A + 0.8W_1)$.

The true value of the ATE was $\psi_0^{\text{ATE}} = 3.5$. The true PS ranged from 0.01 to 0.99. Conditional probabilities of remaining uncensored were between 15% and 99.9%. Among observations where the outcome was observed, the untruncated product of the G components ranged from 0.01 to 0.99, and stabilized IP weights ranged from 0.40 to 42.

Results. For each combination of estimator and truncation strategy, we calculated the mean bias, variance, MSE, and ratio of bias to SE over the 1,500 Monte Carlo iterations, combining all 3 sample sizes (Table 1). Our sample-size–based truncation strategy minimized the MSE for all 3 estimators, IPTCW, TMLE-Q_{cor}, and TMLE-Q_{mis}. Truncation at a fixed level of 0.1 also worked well but was more biased and had larger MSE when coupled with IPTCW and TMLE-Q_{mis}. As anticipated, all strategies had similar performance when coupled with TMLE-Q_{cor}. When the initial model for Q is correct, there will be little to no residual bias to handle in stage 2 of the procedure. Thus, the magnitude of the update will be close to 0, and the impact of truncation will be minimal. For this reason, we were not surprised to see more sensitivity to the choice of truncation bound for a TML estimator under Q_{mis}.

Results stratified by sample size illustrate that overall performance was not dominated by poor performance at one particular sample size (Table 2). Although the sample-size–based strategy does not always have the smallest MSE, across all estimators and sample sizes its bias-variance trade-off is consistently among the best, typically with small bias and a favorable ratio of bias to SE.

Box plots of the bias in estimates obtained using IPTCW, TMLE-Q_{cor}, and TMLE-Q_{mis} demonstrate that the sample-size–based strategy performs well at all sample sizes for all estimators (Figure 1). At the smallest sample size ($n = 100$), truncating at the 99th or 95th quantile was successful, but at larger samples sizes these strategies' bias is evident. The choice of bound had the least impact for the TMLE-Q_{cor} estimator at all sample sizes.

Simulation study 2: binary outcome

Simulation. The data consisted of observations $O = (W, A, \Delta, \Delta Y)$. W is a vector of 5 independent covariates, W_1, W_2, W_3, W_4, W_5, that are all confounders, $W_1 \sim N(0,1)$, $W_2 \sim N(W_1, 1)$, $W_3 \sim$ Bernoulli(0.4), $W_4 \sim$ Bernoulli(expit$(0.2W_1 - W_3)$), and $W_5 = \log(U(1,100))$. A is the binary indicator of treatment, with 35% treated overall: $P(A = 1 \mid W) = \text{expit}(-2 + 0.6W_1 + 0.48W_2 - 0.6W_3 + 0.24W_4 + 0.36W_5)$. Outcome Y was binary, with a marginal event rate of 32%: $P(Y = 1 \mid A, W) = \text{expit}(-1.5 + 0.8\text{A} + 0.03W_1 + 0.2W_2 - 0.4W_3 + 0.3W_4 + 0.1W_5)$. Δ is an indicator of whether the outcome was observed, with 15% missing on average: $P(\Delta = 1 \mid A, W) = \text{expit}(0.9 + 0.6\text{A} + 0.2W_2 + 0.2W_5)$. The true parameter value is $\psi_0^{\text{OR}} = 2.10$.

Near violations of the positivity assumption were more extreme than in simulation study 1. The true treatment assignment probabilities typically ranged from 0.0008 to 0.99. True probabilities of remaining uncensored ranged from approximately 53% to 98%. Among observations for which the outcome was observed, the product of the untruncated components ranged from 0.005 to 0.97, and stabilized IP weights ranged from 0.3 to 54.

Results. For each combination of estimator and truncation strategy, we calculated the mean bias, variance, MSE, and bias:SE ratio of the 1,500 Monte Carlo iterations, combining all 3 sample sizes (Table 3). Our sample-size–based truncation strategy minimized bias for the IPTCW estimator and was among the top 3 minimizers of the MSE. Our strategy minimized MSE for both TML estimators. Truncation at the 99th quantile worked well for IPTCW, while truncation at a fixed level of 0.1 worked well for TMLE. Because the near violation of the positivity assumption was more severe than in simulation study 1, there were differences in the MSE among the different truncation strategies even for TMLE-Q_{cor}. Bias was similar for all truncation strategies, while variance was minimized by our sample-size–based strategy.

Results stratified by sample size again showed that the overall results were not dominated by poor performance at one particular sample size (Table 4). Although the sample-size–based strategy does not always have the smallest MSE, its bias-variance trade-off was consistently among the best across all estimators and sample sizes, usually with small bias, and a favorable ratio of bias to SE. For the IPTCW estimator, our strategy minimized bias but did not always minimize variance or MSE. It did minimize MSE for TMLE-Q_{mis}, and for TMLE-Q_{cor} when n equaled 100. At the larger sample sizes it was not the best, but performance differences were minor. This is illustrated in the box plots of the bias in the OR estimates for each estimator and truncation strategy (Figure 2).

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

Table 1. Simulation Study 1 Bias, Variance, Mean Squared Error, and Ratio of Bias to Standard Error of Additive Treatment Effect Estimates, Obtained From 1,500 Monte Carlo Simulations, Combining Results From All Sample Sizes (n = 100, n = 1,000, and n = 10,000)

Estimator and Truncation Strategy	Bias	Var	MSE	Bias/SE
IPTCW				
99th quantile	0.157	0.073	0.098[a]	0.581
95th quantile	0.371	0.063	0.200	1.478
10^{-6}	0.015	0.122	0.122	0.044
0.01	0.015	0.122	0.122	0.044
0.025	0.017	0.120	0.120	0.048
0.05	0.023	0.111	0.111	0.069
0.1	0.049	0.095	0.097[a]	0.159
CTMLE	0.025	0.102	0.102	0.078
$5/(\sqrt{n} \ln n)$	0.032	0.093	0.094[a]	0.104
TMLE-Q_{cor}				
99th quantile	−0.008	0.030	0.030	0.049
95th quantile	−0.009	0.028	0.028	0.052
10^{-6}	−0.008	0.030	0.030	0.047
0.01	−0.008	0.030	0.030	0.047
0.025	−0.008	0.030	0.030	0.046
0.05	−0.008	0.029	0.029	0.049
0.1	−0.009	0.027	0.027[a]	0.058
CTMLE	−0.009	0.027	0.027[a]	0.057
$5/(\sqrt{n} \ln n)$	−0.009	0.027	0.027[a]	0.054
TMLE-Q_{mis}				
99th quantile	0.067	0.089	0.093[a]	0.227
95th quantile	0.187	0.073	0.108	0.691
10^{-6}	0.015	0.122	0.122	0.044
0.01	0.016	0.121	0.121	0.046
0.025	0.026	0.109	0.110	0.078
0.05	0.060	0.09	0.093[a]	0.201
0.1	0.149	0.07	0.091[a]	0.567
CTMLE	0.144	0.07	0.094	0.536
$5/(\sqrt{n} \ln n)$	0.075	0.08	0.084[a]	0.269

Abbreviations: cor, correctly specified; CTMLE, collaborative targeted maximum likelihood estimation; IPCTW, inverse-probability-of-treatment-and-censoring weighting; mis, misspecified; MSE, mean squared error; SE, standard error; TMLE, targeted maximum likelihood estimation; Var, variance.

[a] One of 3 lowest MSE values for each estimator.

DATA ANALYSIS

Shinohara et al. (10) conducted a retrospective observational cohort study of the association between total dose of ritodrine hydrochloride and pulmonary edema in twin pregnancy in Japan. Ritodrine was approved by the Food and Drug Administration in the 1970s. Though subsequently withdrawn from the US market, it continues to be available in other parts of the world (16). In Japan, ritodrine is a recommended first-line therapy for halting preterm labor. Ritodrine has previously been shown to increase risk of pulmonary edema in pregnant women (14). Shinohara et al. wanted to establish this result in the subpopulation of women pregnant with twins, who are at high risk of preterm labor.

The data set containing observations on 225 women was downloaded from an online data repository (17). We conducted an analysis to estimate the OR associated with exposure to ritodrine hydrochloride at any level versus no exposure. Potential confounders included in the adjustment set were age, height, weight, body mass index (weight (kg)/height (m)^2), and binary indicators of obesity (body mass index ≥25), first pregnancy, single placenta, assistive

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

Table 2. Simulation Study 1 Bias, Variance, Mean Squared Error, and Ratio of Bias to Standard Error of Additive Treatment Effect Estimates, Obtained From 500 Monte Carlo Simulations at 3 Sample Sizes (n = 100, n = 1,000, and n = 10,000)

Estimator and Truncation Strategy	Sample Size											
	n = 100				n = 1,000				n = 10,000			
	Bias	Var	MSE	Bias/SE	Bias	Var	MSE	Bias/SE	Bias	Var	MSE	Bias/SE
IPTCW												
99th quantile	0.158	0.200	0.224[a]	0.35	0.160	0.018	0.044	1.18	0.153	0.002	0.025	3.83
95th quantile	0.370	0.171	0.308	0.90	0.376	0.017	0.158	2.92	0.365	0.001	0.135	9.90
10^{-6}	0.031	0.329	0.329	0.05	0.012	0.034	0.034	0.06	0.004	0.004	0.004[a]	0.06
0.01	0.031	0.329	0.329	0.05	0.012	0.034	0.034	0.06	0.004	0.004	0.004[a]	0.06
0.025	0.032	0.324	0.324	0.06	0.012	0.032	0.033[a]	0.07	0.005	0.003	0.003[a]	0.09
0.05	0.040	0.300	0.301	0.07	0.017	0.030	0.030[a]	0.10	0.012	0.003	0.003[a]	0.22
0.1	0.065	0.258	0.261[a]	0.13	0.045	0.024	0.026[a]	0.29	0.036	0.002	0.004[a]	0.76
CTMLE	0.056	0.268	0.271	0.11	0.015	0.032	0.033[a]	0.08	0.004	0.004	0.004[a]	0.06
$5/(\sqrt{n}\ln n)$	0.079	0.240	0.246[a]	0.16	0.012	0.032	0.033[a]	0.07	0.004	0.004	0.004[a]	0.06
TMLE-Q_{cor}												
99th quantile	−0.025	0.081	0.081	0.09	−0.001	0.007	0.007	0.01	0.001	0.001	0.001	0.03
95th quantile	−0.026	0.077	0.078	0.09	−0.001	0.007	0.007	0.02	0.001	0.001	0.001	0.03
10^{-6}	−0.025	0.082	0.082	0.09	0.000	0.007	0.007	0.01	0.001	0.001	0.001	0.04
0.01	−0.025	0.082	0.082	0.09	0.000	0.007	0.007	0.01	0.001	0.001	0.001	0.04
0.025	−0.025	0.081	0.081	0.09	0.000	0.007	0.007	0.00	0.001	0.001	0.001	0.04
0.05	−0.025	0.078	0.079	0.09	−0.001	0.007	0.007	0.01	0.001	0.001	0.001	0.03
0.1	−0.028	0.074	0.074[a]	0.10	−0.001	0.007	0.007	0.01	0.001	0.001	0.001	0.03
CTMLE	−0.028	0.074	0.075[a]	0.10	−0.001	0.007	0.007	0.01	0.001	0.001	0.001	0.03
$5/(\sqrt{n}\ln n)$	−0.028	0.073	0.073[a]	0.10	0.000	0.007	0.007	0.00	0.001	0.001	0.001	0.04
TMLE-Q_{mis}												
99th quantile	0.089	0.241	0.249	0.18	0.061	0.022	0.026[a]	0.41	0.053	0.002	0.005	1.14
95th quantile	0.198	0.201	0.239	0.44	0.184	0.017	0.051	1.41	0.178	0.001	0.033	4.67
10^{-6}	0.031	0.329	0.329	0.05	0.012	0.034	0.034	0.06	0.004	0.004	0.004	0.06
0.01	0.031	0.326	0.327	0.06	0.012	0.033	0.033	0.07	0.005	0.003	0.003[a]	0.08
0.025	0.042	0.296	0.298	0.08	0.020	0.029	0.029[a]	0.12	0.015	0.003	0.003[a]	0.28
0.05	0.077	0.244	0.250	0.16	0.056	0.023	0.026[a]	0.37	0.047	0.002	0.004	1.02
0.1	0.159	0.189	0.214[a]	0.37	0.150	0.018	0.040	1.12	0.139	0.002	0.021	3.57
CTMLE	0.145	0.199	0.220[a]	0.33	0.149	0.018	0.040	1.12	0.139	0.002	0.021	3.57
$5/(\sqrt{n}\ln n)$	0.201	0.179	0.219[a]	0.48	0.021	0.029	0.029[a]	0.12	0.004	0.003	0.003[a]	0.07

Abbreviations: cor, correctly specified; CTMLE, collaborative targeted maximum likelihood estimation; IPCTW, inverse-probability-of-treatment-and-censoring weighting; mis, misspecified; MSE, mean squared error; SE, standard error; TMLE, targeted maximum likelihood estimation; Var, variance.

[a] One of 3 lowest MSE values for each estimator.

reproductive technology, magnesium administration, and corticosteroid use. No outcomes were missing.

We coupled each previously considered truncation strategy with IPTW and TMLE to estimate the OR. At this sample size, our proposed lower bound on 1/PS for treated and 1/(1 − PS) for untreated is given by $5/(\sqrt{225}\ln 225) = 0.062$. The PS was estimated using ensemble super learning (18). Candidate algorithms were logistic regression, the least absolute shrinkage and selection operator (LASSO), and Bayesian additive regression trees (19, 20). The number of cross-validation folds was set to $V = 20$. Super learning was also used to model the outcome regression for TMLE, using the default library specification in the *tmle* package, specifying $V = 20$. Ninety-five percent CIs were constructed on the log OR scale. Robust SEs were evaluated for IPTW estimators (21, 22). Influence-curve–based estimates of the SE for TML estimators were internally calculated by the *tmle* function.

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

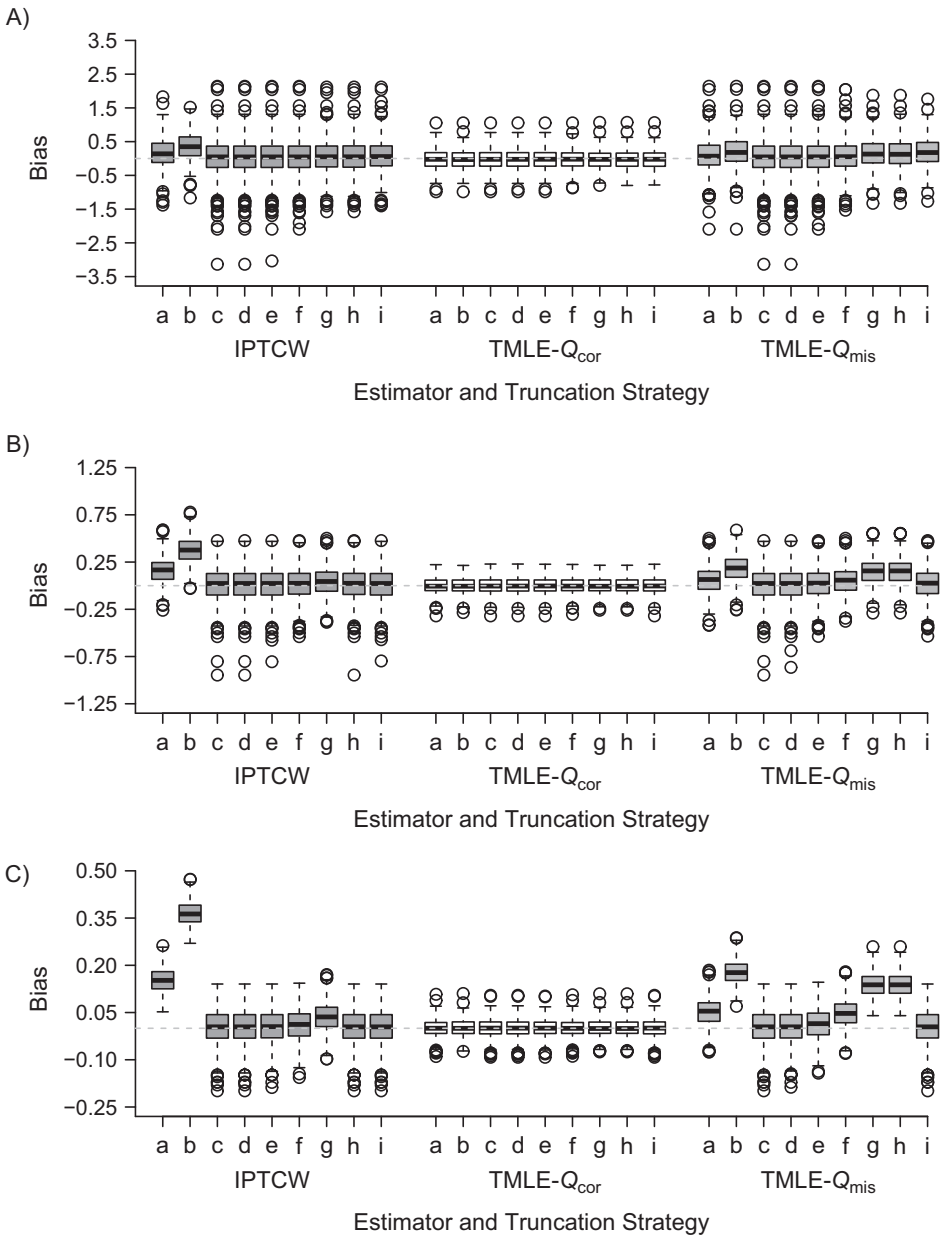

Figure 1. Box plots of bias in the estimated odds ratio for 9 truncation strategies for inverse-probability-of-treatment-and-censoring weighting (IPTCW), targeted maximum likelihood estimation (TMLE) with initial *Q* correctly specified (cor), and TMLE with initial *Q* misspecified (mis) at 3 sample sizes (simulation study 1). A) $n = 100$; B) $n = 1{,}000$; C) $n = 10{,}000$. Truncation strategies are a) 99th quantile, b) 95th quantile, c) 10^{-6}, d) 0.01, e) 0.025, f) 0.05, g) 0.1, h) collaborative targeted maximum likelihood estimation, and i) $5/(\sqrt{n} \ln n)$. The edges of the boxes represent the 25th and 75th percentiles, the horizontal line inside the box represents the median, the whiskers represent 1.5 times the interquartile range, and outliers are shown as circles.

With no truncation, PS estimates ranged from 0.11 to 0.75, and stabilized IP weights were between 0.49 and 3.36 (mean = 0.97). Truncation at each of the fixed truncation levels considered had no impact on the estimated ORs, SEs, or CIs, because the maximum weight was less than the minimum lower bound considered (Table 5). This is because no PS values or IP weights exceeded any of the fixed thresholds. Both quantile-based truncation strategies moved OR estimates towards the null. These estimates had slightly smaller SEs than those associated with the other strategies but were presumably more biased.

DISCUSSION

Theory teaches that our sample-size–based truncation strategy for IP-weighted estimators and TML estimators will

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

Table 3. Simulation Study 2 Bias, Variance, Mean Squared Error, and Ratio of Bias to Standard Error of Additive Treatment Effect Estimates, Obtained From 1,500 Monte Carlo Simulations, Combining Results From All Sample Sizes (n = 100, n = 1,000, n = 10,000)

Estimator and Truncation Strategy	Bias	Var	MSE	Bias/SE
IPTCW				
99th quantile	0.43	2.19	2.37[a]	0.29
95th quantile	0.66	1.82	2.25[a]	0.49
10^{-6}	0.33	3.17	3.28	0.19
0.01	0.33	3.17	3.28	0.19
0.025	0.33	3.17	3.28	0.19
0.05	0.33	3.01	3.12	0.19
0.1	0.34	2.68	2.80	0.21
CTMLE	0.33	2.69	2.80	0.20
$5/(\sqrt{n} \ln n)$	0.32	2.55	2.66[a]	0.20
TMLE-Q_{cor}				
99th quantile	0.26	2.83	2.89	0.16
95th quantile	0.26	2.67	2.73	0.16
10^{-6}	0.28	3.61	3.68	0.15
0.01	0.28	3.23	3.30	0.15
0.025	0.26	2.63	2.70	0.16
0.05	0.25	2.15	2.21	0.17
0.1	0.23	1.79	1.85[a]	0.18
CTMLE	0.23	2.07	2.13[a]	0.16
$5/(\sqrt{n} \ln n)$	0.23	1.70	1.76[a]	0.18
TMLE-Q_{mis}				
99th quantile	0.33	2.97	3.08	0.19
95th quantile	0.37	2.85	2.99	0.22
10^{-6}	0.33	3.17	3.28	0.19
0.01	0.33	3.08	3.19	0.19
0.025	0.33	2.82	2.93	0.20
0.05	0.36	2.43	2.56	0.23
0.1	0.45	2.06	2.26[a]	0.31
CTMLE	0.34	2.25	2.36[a]	0.23
$5/(\sqrt{n} \ln n)$	0.36	2.04	2.16[a]	0.25

Abbreviations: cor, correctly specified; CTMLE, collaborative targeted maximum likelihood estimation; IPCTW, inverse-probability-of-treatment-and-censoring weighting; mis, misspecified; MSE, mean squared error; SE, standard error; TMLE, targeted maximum likelihood estimation; Var, variance.

[a] One of 3 lowest MSE values for each estimator.

make the appropriate bias-variance trade-off asymptotically. In finite samples, our strategy balances minimizing the MSE with preserving valid inference. Although truncating IP weights at a prespecified quantile is common practice, our simulation studies and data analysis illustrated that relying on a sample-size–based truncation strategy can improve performance. This was most evident in the real-world data analysis. Even though PS values were bounded well away from 0 and 1, and the maximum IP weight was less than 4, the quantile-based strategy insists on altering a certain percentage of the values. Although we cannot know the truth, this indicates that quantile-based truncation can alter study findings even when there are no extremely influential observations in the data.

Although in any single analysis truncation at a fixed threshold or quantile can outperform the sample-size–based approach, no alternative strategy was superior at all sample sizes, across all data distributions. The fixed bound of 0.01 often performed well in many scenarios, but it was typically more biased than our sample-size–based strategy. The

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

Table 4. Simulation Study 2 Bias, Variance, Mean Squared Error, and Ratio of Bias to Standard Error of Additive Treatment Effect Estimates, Obtained From 500 Monte Carlo Simulations at 3 Sample Sizes (n = 100, n = 1,000, n = 10,000)

Estimator and Truncation Strategy	Sample Size											
	n = 100				n = 1,000				n = 10,000			
	Bias	Var	MSE	Bias/SE	Bias	Var	MSE	Bias/SE	Bias	Var	MSE	Bias/SE
IPTCW												
99th quantile	0.92	6.03	6.87[a]	0.38	0.204	0.154	0.195	0.52	0.166	0.017	0.045	1.27
95th quantile	1.14	4.92	6.22[a]	0.51	0.435	0.163	0.352	1.08	0.399	0.019	0.178	2.89
10^{-6}	0.91	8.83	9.64	0.31	0.063	0.171	0.175	0.15	0.017	0.018	0.018[a]	0.12
0.01	0.91	8.83	9.64	0.31	0.063	0.171	0.175	0.15	0.017	0.018	0.018[a]	0.12
0.025	0.91	8.82	9.64	0.31	0.063	0.168	0.172	0.15	0.018	0.018	0.018[a]	0.13
0.05	0.90	8.37	9.16	0.31	0.065	0.163	0.167[a]	0.16	0.023	0.018	0.018[a]	0.17
0.1	0.89	7.43	8.21	0.33	0.082	0.157	0.163[a]	0.21	0.042	0.017	0.019	0.32
CTMLE	0.89	7.43	8.21	0.33	0.080	0.159	0.165[a]	0.20	0.017	0.018	0.018[a]	0.13
$5/(\sqrt{n} \ln n)$	0.89	7.00	7.78[a]	0.34	0.063	0.168	0.172	0.15	0.017	0.018	0.018[a]	0.12
TMLE-Q_{cor}												
99th quantile	0.72	8.00	8.51	0.26	0.047	0.150	0.152	0.12	0.014	0.017	0.017	0.11
95th quantile	0.72	7.53	8.04	0.26	0.046	0.140	0.141[a]	0.12	0.014	0.016	0.016[a]	0.11
10^{-6}	0.78	10.28	10.87	0.24	0.053	0.164	0.167	0.13	0.014	0.017	0.017	0.11
0.01	0.76	9.17	9.74	0.25	0.050	0.156	0.158	0.13	0.014	0.017	0.017	0.11
0.025	0.72	7.42	7.92	0.26	0.046	0.146	0.148	0.12	0.014	0.017	0.017	0.11
0.05	0.68	6.02	6.47	0.28	0.046	0.139	0.141[a]	0.12	0.014	0.016	0.016[a]	0.11
0.1	0.64	4.99	5.39[a]	0.29	0.047	0.131	0.133[a]	0.13	0.015	0.015	0.015[a]	0.12
CTMLE	0.66	5.80	6.22[a]	0.27	0.028	0.140	0.140[a]	0.07	0.009	0.016	0.016[a]	0.07
$5/(\sqrt{n} \ln n)$	0.63	4.71	5.10[a]	0.29	0.046	0.146	0.148	0.12	0.014	0.017	0.017	0.11
TMLE-Q_{mis}												
99th quantile	0.89	8.28	9.05	0.31	0.069	0.162	0.166[a]	0.17	0.029	0.017	0.018[a]	0.22
95th quantile	0.92	7.94	8.78	0.33	0.116	0.153	0.166[a]	0.30	0.079	0.017	0.023	0.61
10^{-6}	0.91	8.83	9.64	0.31	0.063	0.171	0.175	0.15	0.017	0.018	0.018[a]	0.12
0.01	0.90	8.58	9.38	0.31	0.064	0.167	0.171	0.16	0.019	0.018	0.018[a]	0.14
0.025	0.89	7.83	8.61	0.32	0.075	0.159	0.164[a]	0.19	0.033	0.017	0.018[a]	0.25
0.05	0.89	6.70	7.49	0.34	0.115	0.153	0.166[a]	0.29	0.073	0.017	0.022	0.57
0.1	0.95	5.65	6.54[a]	0.40	0.220	0.152	0.200	0.56	0.177	0.017	0.048	1.36
CTMLE	0.90	6.10	6.90[a]	0.37	0.081	0.163	0.169	0.20	0.042	0.019	0.021	0.30
$5/(\sqrt{n} \ln n)$	0.98	5.36	6.31[a]	0.42	0.076	0.158	0.164[a]	0.19	0.017	0.018	0.018[a]	0.13

Abbreviations: cor, correctly specified; CTMLE, collaborative targeted maximum likelihood estimation; IPCTW, inverse-probability-of-treatment-and-censoring weighting; mis, misspecified; MSE, mean squared error; SE, standard error; TMLE, targeted maximum likelihood estimation; Var, variance.

[a] One of 3 lowest MSE values for each estimator.

CTMLE-based strategy also exhibited performance superior to the quantile-based truncation strategies and many of the fixed bounds, but it had larger variance and MSE than the sample-size–based strategy. This may be because strategies that "listen to the data" overreact to small perturbations in the distribution when the data are sparse. This manifests as increased variance, as illustrated in the CTMLE-based strategy's results. For TMLE, when the stage 1 outcome regression model was correctly specified, the choice of truncation strategy had little impact. The impact was more pronounced when the outcome regression was misspecified, and targeting in stage 2 of the TMLE procedure was needed to reduce residual bias.

The default truncation strategy in the *tmle* package (version 1.5.0-1) in the Comprehensive R Archive Network (CRAN) is the sample-size–based lower bound presented

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

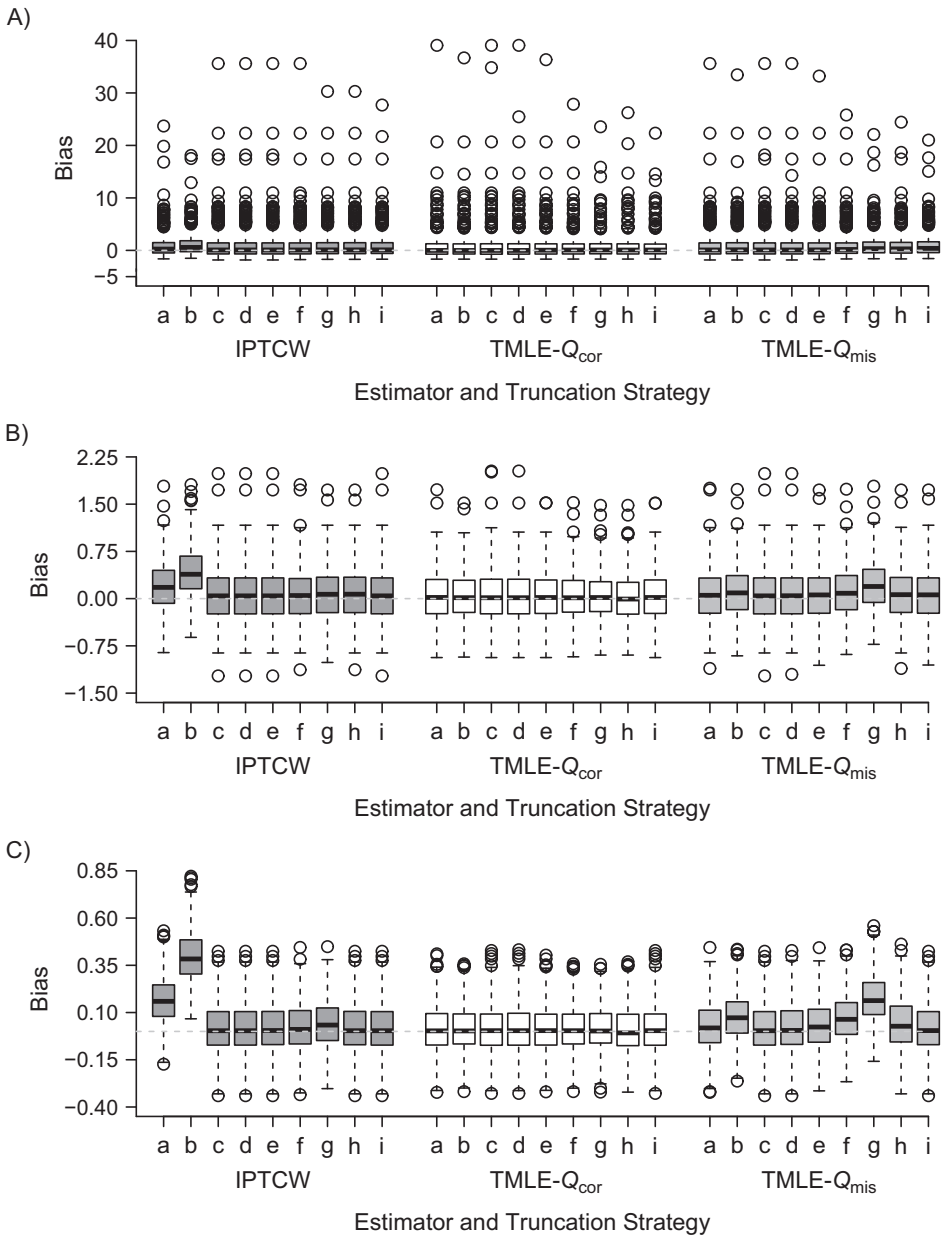

Figure 2. Box plots of bias in the estimated odds ratio for 9 truncation strategies for inverse-probability-of-treatment-and-censoring weighting (IPTCW), targeted maximum likelihood estimation (TMLE) with initial *Q* correctly specified (cor), and TMLE with initial *Q* misspecified (mis) at 3 sample sizes (simulation study 2). A) $n = 100$; B) $n = 1{,}000$; C) $n = 10{,}000$. Truncation strategies are a) 99th quantile, b) 95th quantile, c) 10^{-6}, d) 0.01, e) 0.025, f) 0.05, g) 0.1, h) collaborative targeted maximum likelihood estimation, and i) $5/(\sqrt{n}\ \ln n)$. The edges of the boxes represent the 25th and 75th percentiles, the horizontal line inside the box represents the median, the whiskers represent 1.5 times the interquartile range, and outliers are shown as circles.

in this paper, $5/[\sqrt{n}\ \ln(n)]$ (12). Other TMLE-based R packages in CRAN for analyses of longitudinal and time-to-event data accept fixed user-specified upper and lower bounds, with suggested default lower bounds of 0.001 (survtmle) and 0.01 (ltmle) (13, 14). We hypothesize that for these TML estimators, a time-varying bound on the *G* components of the likelihood based on the number of observations contributing to the targeting step at each time *t* would have good performance. Our future work will involve investigating data-adaptive estimation of the constant in the numerator of the PS bound, *c*, instead of relying on the heuristic $c = 5$.

Ensuring that IP weights are finite is required for any IPW estimator, and ensuring that the product of the *G* components

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

Table 5. Estimates[a] of the Association Between Pulmonary Edema and Any Exposure to Ritodrine Hydrocholoride Versus None

Truncation Strategy	IPTW			TMLE		
	OR	SE	95% CI	OR	SE	95% CI
99th quantile	5.18	0.45	2.15, 12.46	5.58	0.45	2.30, 13.53
95th quantile	4.91	0.44	2.07, 11.63	5.30	0.44	2.23, 12.61
10^{-6}	5.53	0.46	2.26, 13.54	5.55	0.45	2.29, 13.48
0.01	5.53	0.46	2.26, 13.54	5.55	0.45	2.29, 13.48
0.025	5.53	0.46	2.26, 13.54	5.55	0.45	2.29, 13.48
0.05	5.53	0.46	2.26, 13.54	5.55	0.45	2.29, 13.48
0.1	5.53	0.46	2.26, 13.54	5.55	0.45	2.29, 13.48
CTMLE	5.53	0.46	2.26, 13.54	5.55	0.45	2.29, 13.48
$5/(\sqrt{n}\ln n)$	5.53	0.46	2.26, 13.54	5.55	0.45	2.29, 13.48

Abbreviations: CI, confidence interval; CTMLE, collaborative targeted maximum likelihood estimation; OR, odds ratio; SE, standard error.
[a] OR, SE, and 95% CI associated with each truncation strategy.

of the likelihood is nonzero is required for TMLE. Thus, some bound must be specified when these estimators are used in practice. However, although the choice of bound can help, it is not a panacea when there is a sparsity of information in the data for estimating the target parameter. A thoughtful analyst will examine PS diagnostics and conduct sensitivity analyses to assess the robustness of the study finding to departures from causal assumptions (9, 23). When IP weights are extreme (i.e., the PS is near 0 or 1), the analyst might also consider targeting a less ambitious causal parameter (e.g., the effect of a realistic treatment rule) or a stochastic intervention (24).

ACKNOWLEDGMENTS

Author affiliations: Putnam Data Sciences, LLC, Cambridge, Massachusetts, United States (Susan Gruber); Department of Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, California, United States (Rachael V. Phillips, Mark J. van der Laan); and Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, US Department of Health and Human Services, Silver Spring, Maryland, United States (Hana Lee).

This work was funded by the Food and Drug Administration (contract 75F40119C10155).

The data set used in this study is available for public download from the Dryad data repository (https://doi.org/10.5061/dryad.1v8v6).

We thank Shinohara et al. (10) for making the data from their study publicly available.

The content of this article reflects the views of the authors and does not necessarily represent the official views of, or an endorsement by, the Food and Drug Administration, the US Department of Health and Human Services, or the US government.

Conflict of interest: none declared.

REFERENCES

1. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11(5):561–570.
2. Kang JDY, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci.* 2007;22(4):523–539.
3. Freedman DA, Berk RA. Weighting regressions by propensity scores. *Eval Rev.* 2008;32(4):392–409.
4. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168(6):656–664.
5. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res.* 2012;21(1):31–54.
6. Bembom O, van der Laan MJ. Data-adaptive selection of the truncation level for inverse-probability-of-treatment-weighted estimators. (U.C. Berkeley Division of Biostatistics Working Paper Series, no. 230). Berkeley, CA: University of California, Berkeley; 2008. https://biostats.bepress.com/ucbbiostat/paper230/. Accessed March 21, 2022.
7. Xiao Y, Moodie EE, Abrahamowicz M. Comparison of approaches to weight truncation for marginal structural Cox models. *Epidemiol Methods.* 2013;2(1):1–20.
8. Ju C, Schwab J, van der Laan MJ. On adaptive propensity score truncation in causal inference. *Stat Methods Med Res.* 2019;28(6):1741–1760.
9. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023

effects in observational studies. *Stat Med.* 2015;34(28):3661–3679.
10. Shinohara S, Sunami R, Uchida Y, et al. Association between total dose of ritodrine hydrochloride and pulmonary oedema in twin pregnancy: a retrospective cohort study in Japan. *BMJ Open.* 2017;7(12):e018118.
11. van der Laan MJ, Rose S. *Targeted Learning: Prediction and Causal Inference for Observational and Experimental Data*. New York, NY: Springer Publishing Company; 2011.
12. Gruber S, van der Laan MJ. tmle: an R package for targeted maximum likelihood estimation. *J Stat Softw.* 2012;51(13):1–35.
13. Schwab J, Lendle S, Petersen M, et al. ltmle: longitudinal targeted maximum likelihood estimation. https://cran.r-project.org/web/packages/ltmle/index.html. Published 2016. Accessed March 21, 2022.
14. Benkeser D, Hejazi N. survtmle: compute targeted minimum loss-based estimates in right-censored survival settings. (Version 1.1.1). Published 2019. https://CRAN.R-project.org/package=survtmle. Accessed March 21, 2022.
15. R Core Team. *R: A Language and Environment for Statistical Computing*. (Release 4.0-2). Vienna, Austria: R Foundation for Statistical Computing; 2020.
16. Elliott JP, Morrison JC. The evidence regarding maintenance tocolysis. *Obstet Gynecol Int.* 2013;2013:708023.
17. Shinohara S, Sunami R, Uchida Y, et al. Association between total dose of ritodrine hydrochloride and pulmonary edema in twin pregnancy: a retrospective cohort study in Japan. *BMJ Open.* 2017;7:e018118.
18. Polley E. SuperLearner: super learner prediction. (Version 2.0-19). http://CRAN.R-project.org/package=SuperLearner. Published 2016. Accessed March 21, 2022.
19. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, NY: Springer Publishing Company; 2002.
20. Dorie V. dbarts: Discrete Bayesian Additive Regression Trees Sampler. (Version 0.9-17). https://CRAN.R-project.org/package=dbarts. Published 2020. Accessed March 21, 2022.
21. Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. *J Stat Softw.* 2004;11(10):1–17.
22. Zeileis A. Object-oriented computation of sandwich estimators. *J Stat Softw.* 2006;16(9):1–16.
23. Díaz I, van der Laan MJ. Sensitivity analysis for causal inference under unmeasured confounding and measurement error problems. *Int J Biostat.* 2013;9(2):149–160.
24. van der Laan MJ, Petersen ML. Causal effect models for realistic individualized treatment and intention to treat rules. *Int J Biostat.* 2007;3(1):Article 3.

Downloaded from https://academic.oup.com/aje/article/191/9/1640/6580570 by guest on 01 October 2023